UNIVERSIDADE CATÓLICA PORTUGUESA

DETERMINING THE IMMUNE RESPONSE IN HUMAN IMMUNODEFICIENCY VIRUS INFECTION

HIV-1 diversity as a tool for epidemic monitoring

Thesis presented to the Catholic University Portugal to attain the degree of Doctor of Philosophy in Health Sciences – Medicine

by

Alexandre Manuel Câmara de Carvalho

Institute of Health Sciences

October 2014

UNIVERSIDADE CATÓLICA PORTUGUESA

DETERMINING THE IMMUNE RESPONSE IN HUMAN IMMUNODEFICIENCY VIRUS INFECTION

HIV-1 diversity as a tool for epidemic monitoring

Thesis presented to the Catholic University Portugal to attain the

PhD degree in Health Sciences –Medicine

By Alexandre Manuel Câmara de Carvalho

under supervision of

Professor Henrique Lecour

Professor Jorge Pedrosa

Institute of Health Sciences

October 2014

ABSTRACT

The Human Immunodeficiency Virus type 1 (HIV-1) is characterized by extensive genetic diversity at the population level but also within a single infected individual. The swift capacity of the virus to generate extensive diversity within the human host played a central role in the origin of the disease and is also key for the current global proportions of the HIV-1 pandemic. The epidemic started in Africa with multiple zoonotic transmissions of simian immunodeficiency virus (SIV) to humans. This was followed by a period of diversification and adaptation to the human population that, enhanced by the high rates of mutation and recombination of the virus, allowed the emergence of a virus capable of efficient sexual transmission among humans. The spread of the human adapted virus is estimated to have initiated from late 1950s to the early 1960s from Africa to the rest of the world. The predominance of the subtype B HIV-1 virus in Western Europe suggests that this was the first subtype to be introduced in this region. The subtype diversity pattern of HIV-1 in Portugal resembles the ones found in Central Africa being far more complex than the viral diversity patterns observed in the rest of Western Europe highlighting the relevance of in detailed studies of the Portuguese HIV-1 epidemics.

In this work we have characterized the local HIV-1 epidemic of the Portuguese city of Braga in the years from 2000 to 2012. We found that the most frequent HIV-1 subtypes were G and B and by combining epidemiological and phylogenetic analysis we were able to uncover local transmission clusters of non-B and non-G subtypes among locals in association with sexual transmission networks that initiated transmission in the early 2000s. This corroborates Portugal as an early point of introduction of non-B HIV-1 subtypes in Western Europe. Having performed this characterization at the level of this local population we then focused on analyzing the duration of infection at the level of the infected patient. For this purpose we have optimized a methodology to differentiate recent from chronic infections. It was based on the study of ambiguous nucleotide calls obtained from routine HIV-1 genotyping. We found that the analysis of these ambiguities, as an expression of intra-host HIV-1 diversity, allowed the inference of the duration of infection in this study population. Subsequently, we questioned if high HIV-1 subtype diversity found in this region correlated with higher rates of transmission of drug resistance mutations. We found that the level of transmitted drug resistance in this population was similar to other European regions and independent

predictors of transmitted drug resistance (TDR) could not be identified supporting the recommendation of universal viral sequencing at patient admission.

This study performed in a country that is unique in Western Europe in what regards to HIV-1 diversity supported Portugal as one of the early entry-points of non-B HIV-1 subtypes in Western Europe and also reinforced the need for more efficacious local control measures targeting sexual transmission routes. We believe this study is of general importance especially in a time when several reports suggest that the prevalence of non-B subtypes in Western Europe is increasing. The knowledge herein generated also contributed for the development of method to discriminate recent from non-recent HIV-1 infections, a step of crucial importance to validate prevention strategies. Importantly, it was also shown that the higher HIV-1 subtype diversity found in this study population does not correlate with an increase in the rate of transmission of drug resistance when compared to the rest of the Western Europe.

RESUMO

O Vírus da Imunodeficiência Humana Tipo 1 (VIH-1) é caracterizado por uma extensa diversidade genética não só a nível da população, mas também a nível individual, em cada hospedeiro. A rapidez do vírus para gerar grande diversidade dentro do hospedeiro humano desempenhou um papel central na génese da doença e é também essencial para as proporções globais atuais da pandemia do VIH-1. A epidemia começou em África, com várias transmissões zoonóticas de vírus da imunodeficiência símia (SIV) para seres humanos. Isto foi seguido por um período de diversificação e adaptação na população humana que, amplificada pelas altas taxas de mutação e de recombinação do vírus, permitiu o surgimento de uma nova espécie de vírus capaz de transmissão sexual eficiente entre os seres humanos. O início da propagação deste vírus já adaptado ao ser humano é estimada a partir do final dos anos 1950 ao início dos anos 1960, da África Central para o resto do mundo. A predominância do subtipo B do VIH-1 na Europa Ocidental sugere que este foi o primeiro subtipo a ser introduzido nesta região. O padrão de diversidade dos subtipos do VIH-1 em Portugal assemelha-se ao encontrado na África Central, sendo muito mais complexo do que os padrões de diversidade vírica observados no resto da Europa Ocidental. Este facto justifica o relevo que estudos detalhados sobre o VIH-1 em Portugal possam ter para a compreensão das epidemias de VIH-1.

Neste trabalho foi caracterizada a epidemia local por VIH-1 na cidade portuguesa de Braga, entre os anos 2000 e 2012. Descobrimos que os subtipos VIH-1 mais frequentes foram G e B. Pela combinação de análise epidemiológica e filogenética pudemos demonstrar grupos de transmissão locais de subtipos não-B e não-G entre os residentes em associação com redes de transmissão sexual que iniciaram a transmissão no início da década de 2000. Isto indicia o papel de Portugal como um ponto de início da introdução de subtipos não-B do VIH-1 na Europa Ocidental. Tendo realizado esta caracterização a nível da população local, o trabalho concentrou-se na análise da duração da infeção ao nível individual. Para este efeito, aperfeiçoou-se uma metodologia para diferenciar infeção recente de infeção crónica. Baseados no estudo de posições nucleotídicas ambíguas obtidas a partir de genotipagem rotineira doVIH-1,descobrimos que a análise dessas ambiguidades, como uma expressão da diversidade intra-hospedeiro do VIH-1, permite inferir a duração da infeção nesta população

em estudo. Posteriormente questionamos se a elevada diversidade do VIH-1 encontrada nesta região se poderia correlacionar com maiores taxas de transmissão de mutações de resistência aos antiretrovíricos. Descobrimos que o nível de resistência à terapêutica transmitido nesta população foi semelhante a outras regiões europeias. Não foram identificados preditores independentes de resistência transmissível aos antiretrovíricos, suportando a recomendação de sequenciamento viral universal no momento do contacto do doente com os serviços de saúde.

Este estudo realizado num país que é único na Europa Ocidental no que diz respeito à diversidade do VIH-1,validoua noção de Portugal como um dos pontos de entrada iniciais de subtipos não-B do VIH-1 na Europa Ocidental e também reforçou a necessidade de medidas locais de controlo mais eficazes, que visem modos de transmissão sexual. Acreditamos que este estudo é relevante, especialmente num tempo em que vários artigos sugerem que a prevalência de subtipos não-B na Europa Ocidental está a aumentar. O conhecimento aqui gerado também contribuiu para o desenvolvimento de um método para discriminar infeções recentes pelo HIV-1 das não-recentes, um passo de importância crucial para validar as estratégias de prevenção. Relevantemente, também foi demonstrado que à maior diversidade doVIH-1 encontrada na população em estudo, não correspondeu um aumento na taxa de transmissão de resistência à terapêutica, quando comparada com o resto da Europa Ocidental.

AGRADECIMENTOS

Aos meus dois orientadores, Professor Henrique Lecour, Professor Jorge Pedrosa e ainda ao Professor Gil Castro, por todo o apoio, por me obrigarem a pensar e a avançar e por nunca me deixarem sentir desacompanhado.

Às minhas duas colegas de labuta, Cristina Ângela e Teresa Pimentel, por me desobstruírem o caminho sempre que foi preciso.

A dois parceiros essenciais, Fernando Branca e Nuno Osório, pelo apoio técnico-científico e não só…

Às pessoas que trabalham nos Laboratórios de Biologia Molecular e de Imunologia do Hospital de Braga, particularmente o Dr. Mota-Garcia, por nunca negarem o que eu lhes pedia.

Mas aquilo que é verdadeiramente essencial: aos meus quatro filhos, que tantas vezes sacrificados no tempo que lhes poderia dedicar, sempre me incentivaram a não parar.

E à Cândida, porque mesmo não sendo eu um grande homem sei que tenho ao meu lado uma grande Mulher.

Este trabalho foi parcialmente financiado pela Fundação GlaxoSmithKline das Ciências de Saúde por intermédio de uma Bolsa de Investigação para apoio a Doutoramento na área de infeção VIH/SIDA.

Table of Contents

ABSTRACT V
RESUMO VII
AGRADECIMENTOS IX
ABBREVIATIONS 15
GENERAL OBJECTIVES AND OUTLINE OF THE THESIS 17
CHAPTER1: GENERAL INTRODUCTION 19
History of AIDS pandemic 21
HIV-1 origin 25
BIOLOGY AND PHYLOGENY OF HIV 33
What is HIV? 33
HIV-1 structure 33
HIV-1 life cycle 35
EVOLUTION TO DIVERSITY 41
HIV-1adaptation to a new host 41
APOBEC3G 45
Tetherin 46
SAMHD1 47
Genetic diversity 49
Intra-host immune selection 54
Recombination 64
Geographic diversity 69
Subtype B: start of an epidemic in a high-income setting 73
India: a straight C connection 74
Maximum founder effect: the case of former Soviet Union 75

Europe in present times: melting pot increases HIV-1 genetic diversity 76
Current status 76
Portugal's unique pattern of diversity 78
Assessing HIV-1 diversity implications 84
Transmission and disease progression 84
Diagnosis and disease management 86
Response and resistance to antiretroviral therapy 87
Vaccine development 89
CHAPTER 2: ANALYSIS OF A LOCAL HIV-1 EPIDEMIC IN PORTUGAL HIGHLIGHTS ESTABLISHED TRANSMISSION OF NON-B AND -G SUBTYPES 93
Abstract 96
Introduction 97
Methods 99
Study population 99
Sequencing of viral samples 99
Phylogenetic analysis 100
Estimation of evolutionary dates 101
Statistical analysis 101
Ethics 101
Results 102
High HIV-1 subtype diversity 102
Increasing incidence of infection with A1 and F1 HIV-1 subtypes 102
Evidence for local transmission clusters 105
Established local transmission of non-prevalent HIV-1 subtypes 106
Discussion 109
CHAPTER 3: VIRAL DIVERSITY IN STANDARD HIV-1 SEQUENCES: CONTRIBUTION OF AMBIGUITIES TO DIFFERENTIATE RECENT FROM CHRONIC INFECTION 113

Abstract 115
Introduction 116
Methods 118
Ethical Considerations 119
Results 120
Characterization of study participants 120
Diversity evaluation 121
Assessment of age of infection 122
Evaluation of algorithm performance 123
Discussion 125
Conclusion 127
CHAPTER 4: DRUG RESISTANCE, TRANSMISSION CLUSTERS AND POLYMORPHISMS IN A LOCAL HIV-1 EPIDEMIC 129
Abstract 131
Introduction 133
Methods 135
Study population 135
Characterization of disease stage 135
RNA extraction, amplification and sequencing 135
Drug resistance mutations (DRM) 136
RNA extraction, amplification and sequencing 136
Phylogenetic analysis and identification of clusters 137
Evaluation of cluster dynamics 137
Statistical analysis 138
Ethical considerations 138
Results 139
Study population 139

Levels and trends of overall HIV-1 drug resistance 139
Transmitted drug resistance 141
Trends of TDR rates 142
Transmission clusters and clustering 143
Transmission clusters and HIV-1 resistance 145
Cluster dynamics 146
Reverse transcriptase A98S polymorphism and protease E35_T insertion 147
Discussion 149
CHAPTER 5: GENERAL DISCUSSION AND CONCLUSIONS 155
Conclusions 164
REFERENCES 167

ABBREVIATIONS

Abbreviation	Meaning
AIDS	Acquired immunodeficiency syndrome
APOBEC3G	Apolipoprotein B mRNA-editing enzyme, catalytic polypeptide-like editing complex 3
ART	Antiretroviral therapy
ARV	Antiretroviral
AZT	Azidothymidine
BCN	Broadly cross-neutralizing
BEAST	Bayesian Evolutionary Analysis Sampling Trees
BLAST	Basic Local Alignment Sequence Tool
BST-2	Bone marrow stromal antigen 2
CDC	Centers for Disease Control
CI	Confidence interval
CMV	Cytomegalovirus
CRF	Circulating recombinant form
CTL	Cytotoxic T-cell
DNA	Deoxyribonucleic acid
DRC	Democratic Republic of the Congo
DRM	Drug resistance mutation
FDA	Food and Drug Administration
FIV	Feline immunodeficiency virus
HCV	Hepatitis C virus
HIV	Human immunodeficiency virus
HLA	Human histocompatibility leukocyte antigen
HPD	Highest posterior density
HTLV	Human T-lymphotropic virus
IDU	Intravenous drug users
IQR	Interquartile range
LANL	Los Alamos National Laboratory
LAV	Lymphadenopathy-associated virus
LTR	Long terminal repeat
MCMC	Markov Chain Monte Carlo
MHC	Major histocompatibility complex

ML	Maximum likelihood
MSM	Men who have sex with men
NNRTI	Non nucleoside reverse transcriptase inhibitor
NRTI	Nucleoside reverse transcriptase inhibitor
OR	Odds ratio
PAS	Proportion of ambiguous sites
PCR	Polymerase chain reaction
PIC	Pre-integration complex
PIC	Protease inhibitor
PNG	Potential N-linked glycosylation
RNA	Ribonucleic acid
SAMHD1	Sterile alpha motif and histidine-aspartate domain-containing protein 1
SE	Standard error
SIV	Simian immunodeficiency virus
TDR	Transmitted drug resistance
tMRCA	Time of the most recent common ancestor
URF	Unique recombinant form
USA	United States of America

GENERAL OBJECTIVES AND OUTLINE OF THE THESIS

The human immunodeficiency virus (HIV) appeared as a human pathogen recently, but it is unquestionably a successful one, in great part due to an extraordinary ability to elude human immune defenses. High evolution rates that led to high viral diversity are key features for the success of HIV as a pathogen. This genetic diversity comes with a price to HIV, though: it can be used against it, in predicting its progression and in combating its propagation. Actually, studies on the origin of HIV and its evolution within individuals and between populations, leading to present day diversity, are essential to understand HIV pathogenesis and the emergence acquired immunodeficiency syndrome (AIDS).

Although there are hundreds or even thousands of HIV variants circulating worldwide, the vast majority of HIV infections are caused by type 1,group M[1], hereinafter referred as HIV-1. HIV-1 genetic diversity has important implications in screening, diagnostic testing, disease monitoring and treatment outcomes and may also affect viral transmissibility and pathogenicity.

To explain how HIV-1 achieved its present degree of diversity and how this diversity can be used to monitor the progression of the epidemic at a local or regional level, the thesis is organized in five chapters.

The major aim of the first chapter is to describe the path forHIV-1 diversity, from a local zoonosis to a global pandemic composed of regional/local epidemics, and how this process shaped the evolution of the pandemic. HIV-1 infection is a well-suited example of evolutionary dynamics, occurring simultaneously on spatial and temporal scales. Determining the factors that conducted the evolution of the pandemic is far beyond historical interest: it can also be the key to control this disease.

In the second chapter of this work it is proposed to (a) detect and depict transmission networks and (b) to take advantage of the information provided by molecular epidemiologic characterization of HIV-1 infection phylodynamics at a local level in order to better design preventive measures to curtail further spread of local or regional epidemics.

In chapter 3 it was intended to design an algorithm applicable to the determination of incidence. Infections were categorized as recent or not recent by evaluating HIV-1 viral diversity and its association with the length of infection in HIV-1 infected patients attended in Hospital de Braga with at least one HIV-1 *pol* gene sequencing available. This was accomplished by examining the relationship between the proportion of ambiguous sites in initial sequencing and time of infection, in association with related variables, such as time since HIV-1 infection diagnosis, CD4+ cell count and AIDS status.

In chapter 4, drug resistance surveillance, by means of transmitted drug resistance analysis and identification of polymorphisms was performed, aiming at better management of antiretroviral therapy

In chapter 5, a general discussion of the thesis is presented, where the main results are highlighted and placed in the context with the relevant literature. Based on data presented in this thesis and on the current literature, study of HIV-1 diversity is applied to epidemic monitoring. Lastly, the main conclusions and future perspectives resulting from this work are addressed, and the potential clinical applicability of the reported findings is proposed.

CHAPTER1: GENERAL INTRODUCTION

History of AIDS pandemic

The disease currently known as acquired immunodeficiency syndrome (AIDS) was first described in June 1981 in a paper published in *Morbidity and Mortality Weekly Report* (MMWR) named "*Pneumocystis* pneumonia — Los Angeles"[2].

Definitely, the authors couldn't predict then all the devastation they were unveiling or imagine the epidemics future scale or even its global dissemination. The World Health Organization estimates there are more than 34 million people infected with HIV and about 2.5 million new infections acquired in only 2012[3]. It is calculated that around 0.8% of all the people worldwide aged 15 to 49 years is infected with HIV, although the prevalence is quite variable geographically, from 0.1% in Middle East and North Africa to 4.7% in Sub-Saharan Africa[3].

In the begging of the AIDS pandemic, patients suffering from this mysterious syndrome were reported in crescent numbers since its initial description. In a few months, the basics of disease transmission were identified: even though all the initial patients were homosexual men, soon other risk factors were identified: Haitians[4], hemophiliacs[5] and intravenous drug users[6]. For the first time contaminated blood was hypothesized to be propagating the disease.

The diagnosis of several Haitians with Kaposi's sarcoma and other AIDS-related conditions, with no history of homosexual practices or intravenous drugs use, raised the hypothesis that AIDS had come from Haiti, and that Haitians were responsible for the AIDS epidemic in the United States. These claims, which were often founded on dubious evidence, based on a culture of blame and prejudice that surrounded HIV and AIDS in the early years, fuelled racism in the United States and many Haitians suffered from it. It was then politically difficult to present epidemiological findings in a neutral and objective way. Ironically, later it was discovered that Haiti is the local of oldest-known HIV/AIDS epidemic outside Africa and that probably exported HIV to America mainland[7].

In 1982, four events changed the way this syndrome was seen. First, it was named Acquired Immunodeficiency Syndrome (AIDS)[6].

Secondly, the first report of a possible heterosexual transmission, in five women[8], turned a illness that has thought to affect only marginal and well-defined groups into an universal

disease with potential to become a pandemic, also making the "risk behavior" concept somewhat obsolete.

Thirdly, in November were published the first guidelines on health and laboratory personnel precautions[9], comparing the spread of this new disease with the B hepatitis mode of transmission.

And, finally, it was demonstrated the role of contaminated blood as a source of this new syndrome, when a child was diagnosed after a blood transfusion[10]. Soon after, four affected children were described as the first vertical transmission occurrences[11].

When 1982 ended, the disease had a name – AIDS - and all the major routes of transmission were identified: mother-to-child, sexual and blood borne. However, without a laboratorial diagnostic test, blood banks had difficulty in preserving the safety of blood supply, leading to at least 15 000 hemophiliacs infected in the United States alone, between 1981 and 1984[12].

Similarly, intravenous drug users were at high risk due to exchange of blood during needle sharing. The first needle exchange program was set up in Amsterdam in 1984, trying to decrease the circulation of contaminated injection equipment, thereby reducing the spread of blood-borne pathogens in the community[13].

So far, all the known patients lived in United States of America, Haiti or Western Europe. In March 19, 1983, the first five African patients are described[14]. The authors recognized the potentially devastating future of this disease in Africa, forwarded by a rise in opportunistic infections.

All the data at this time pointed to an infectious disease but the infectious agent responsible for AIDS was still unknown. The first possible cause of AIDS to be suggested was the cytomegalic virus (CMV), a *herpesviridae* family member. This assumption was based on the coincidence of those affected by this new immune deficiency with high rates of CMV[2,15] infection, coupled with the fact that CMV displays some immunosuppressive potential. However, this theory could not fully explain all the known cases and was therefore quickly abandoned. Two related substances were alternatively suggested to be the cause of AIDS, amyl nitrite and isobutyl nitrite, used at the time as sexual stimulants and admittedly immunosuppressant agents[16]. However, soon cases emerged in people who had never had contact with any of the drugs. Another hypotheses put forth was that repeated exposure to exogenous sperm running with alloantigen could trigger a chronic immune stimulation,

flowing in a condition similar to graft-versus-host disease and, consequently, in opportunistic infections[17]. Outside the scientific community, this new disease has also been considered as a sort of punishment, punishing "less desirable" lifestyles (homosexuality, intravenous drug use) [18]. At this time a viral cause for AIDS was just one among many other hypotheses. Those who worked with hepatitis B, considered quite plausible such hypothesis, since AIDS affected similar groups of people. Also who was familiar with retroviruses, like feline leukemia virus, which causes a generalized immunodeficiency in felines, had the notion that this syndrome in humans was comparable and could therefore be also caused by a retrovirus. Doubts about the viral etiology of AIDS remained until the effective isolation and identification of HIV.

In May 1983, researchers from Luc Montagnier's laboratory at the Pasteur Institute in Paris reported to have isolated a new retrovirus from lymphoid ganglions that they hypothesized to be the cause of AIDS[19]. The virus was later named lymphadenopathy-associated virus (LAV) and a sample was sent to the United States Centers for Disease Control, which was later passed to the National Cancer Institute. In May 1984, a team from the National Cancer Institute led by Robert Gallo confirmed the discovery of the virus, renaming it human T lymphotropic virus type III (HTLV-III)[20]. In January 1985, a number of more-detailed reports were published concerning LAV and HTLV-III, and by March it was clear that the virus were the same, were from the same source, and were the etiological agent of AIDS.

Later in 1985, the FDA approved the first commercial test for the detection of the recently discovered virus [21], which had a significant impact on the safety of blood banks, one of the biggest concerns at the time, and on people with high-risk behaviors.

Finally, in 1986, the Centers for Disease Control and Prevention (CDC) reported the first AIDS case definition that, for being largely consensual, has been universally adopted as a diagnostic criterion[22]. This definition recognizes 21 opportunistic conditions, infectious or neoplastic, diagnosing AIDS, when associated with a positive serologic test. This classification would suffer only one update, in 1993[23], to include pulmonary tuberculosis, recurrent pneumonia, invasive cervical carcinoma and the CD4 cell count of less than 200/mm^3 in the list of AIDS defining conditions. This last criterion was responsible for an increase of almost 200% in cases in the United States of America, when comparing 1993 with 1992.

Meanwhile, another virus had been identified in September 1985[24],the HTLV-IV. It was at this point that the name of the virus was changed, assuming the identity by which is known until today –human immunodeficiency virus –HIV, further divided in types 1 and 2[25]. The HIV-1 is responsible for the global pandemic, while the HIV-2 is restricted to West Africa and a few countries historically related, like Portugal and France[26]. The epidemic quickly evolved from a series of small outbreaks in the so-called risk groups in the United States and Western Europe to a global and catastrophic threat to public health worldwide. Great efforts have been then undertaken, ranging from molecular biology to public health, leading to relevant advances in epidemiology, etiology, diagnosis and patient management.

Two initiatives, apparently humble, have shown positive results in the developed world: sex education with a particular focus on the use of condoms[27] and investment in prevention among intravenous drug users, particularly through needle exchange programs[28]. It was possible to limit the horizontal spread in places where such programs were implemented firmly and without prejudices. However, there are still social, political, religious, cultural and even personal barriers that prevent the more widespread use of these two measures.

In 1993, AIDS became the leading cause of death in the United States of America in the age group 25-44 years[29]. But this epidemic will literally devastate the sub-Saharan region of Africa, where 69% of all the people infected with HIV live nowadays, where 1 in every 20 adults is infected and where 71% of the approximately 7000 new daily infections occur – 2.5 million new infections globally in 2012[3]. The impact on the productive layers of the population is reflected in the economic and social disruption and political destabilization of entire countries where HIV prevalence is high – Zimbabwe, Swaziland, Lesotho, Botswana and Namibia, citing some examples[30]. Here the epidemic evolves in two different ways: horizontally, by sexual transmission between adults; and, to a lesser extent, vertically, in which infected mothers give birth to infected children[31].

One last word regarding specific therapy: in March 1987 the first antiretroviral drug was approved, AZT[32]. Another chapter started here, transforming this disease from an invariably fatal condition into a chronic pathology, requiring daily medication, periodic blood testing and routine medical appointments. The introduction of combined antiretroviral therapy of high efficacy, in 1996 in developed countries and since 2004 in the rest of the world, greatly contributed to further consolidate this change[32]. This therapeutic regimen is considered one of the most cost-effective measures in healthcare, allowing to save 14 million years-life in

developing countries. In 2012, for the first time, a majority of people eligible for treatment (54%) was receiving antiretroviral therapy.

However, in spite of all our present knowledge, there are still significant failures in prevention and control, which translate to the following paradox: despite the growing available financial and scientific resources, global attention and a variety of dedicated political agendas, HIV continues to infect and kill, especially in poorer nations. Effective vaccines are not expected at short term and productive public discussion about sex and drugs, after all the main transmission routes, is missing on the agenda of most international public powers. In contrast, other transmission paths, such as vertical and transfusional, have been fought with success, at least in the developed world. Still, it is estimated that about 1.7 million people may have died with AIDS-related causes during the year 2011[33] and probably more than 25 million since the beginning of this epidemic.

HIV-1 origin

AIDS emerged as result of a combination of human behavior, social conditions and zoonotic transmissions of a virus, probably in the first decade of the 20th century in a region known then as French Equatorial Africa.

As strains of HIV-1 were sampled from around the world, it became apparent that they exhibit extremely high genetic heterogeneity and that analysis of the evolution of this diversity could be critical in revealing insights into the enigmatic origin of the virus. What were the reasons for its sudden emergence, epidemic spread and unique pathogenicity? Furthermore, the key to understand the origin of HIV was the discovery that closely related it whit other virus — the simian immunodeficiency virus (SIVs) —present in a wide variety of African primates. In fact, both HIV and SIVs are classified as lentivirus, a retrovirus family, and SIVs have been identified in more than 40African primate species[34].The nomenclature in use appends the initials of each non-human primate after “SIV”. For example, SIVcpz is a chimpanzee-infecting virus and SIVsmm infects a monkey called sooty mangabey. A high genetic diversity is observed among the different SIVs, but generally each primate species is infected with a species-specific virus, which forms monophyletic lineages in phylogenetic trees. Among these monophyletic lineages, there are examples of co-evolution between virus

and their primate hosts, but there are also many examples of cross-species transmissions followed or not by recombination between distant SIVs[35]. Notably, SIVs don´t seem to be pathogenic to its natural hosts, although there are few studies conducted in wild populations[36]. These two findings suggest an ancient host-virus relationship, although not a host dependent evolution. In fact, it seems unlikely a co-evolution pattern given that monkeys clades are millions of years old and the most recent common ancestors of SIVs are far younger[37].Preferential host switching and subsequent diversification accounts for resemblance between host and pathogen phylogenies[35].

Based on phylogenetic trees of the primate lentivirus, the evolutionary history of HIV-1 and HIV-2 could be reconstructed. The enigma of HIV-2 origin was solved first: it is intimately related to SIVsmm[38], a virus infecting with high prevalence a mangabey monkey (*Cercocebus atys*) living in West Africa, where HIV-2 probably emerged.

The close genetic relationship found between HIV-1 and SIVcpz, the chimpanzee infecting SIV, allowed hypothesizing the later as a precursor of the former[39]. Initially, the rarity of isolates of SIVcpz among chimpanzees raised doubts about whether chimpanzees represented a true SIV reservoir[39]. A non-invasive technology of detecting SIVcpz antibodies in urine and fecal samples allowed drawing a precise picture of SIVcpz distribution[40]. Chimpanzees are classified into two species, the common chimpanzee (*Pan troglodytes*) and the bonobo (*Pan paniscus*). Common chimpanzees have traditionally been further subdivided into four geographically differentiated subspecies: western (*P. t. verus*); Nigeria-Cameroonian (*P. t. ellioti*, formerly termed *P. t. vellerosus*); central (*P. t. troglodytes*); and eastern (*P. t. schweinfurthii*). In only two of these four subspecies SIVcpz was found: central (*P. t. troglodytes*) and eastern (*P. t. schweinfurthii*) chimpanzees[40]. SIVcpz isolated in *P. t. troglodytes* has a particular close relationship to the HIV-1 group M, the major cause of AIDS pandemic[39]. Central chimpanzee (*P. t. troglodytes*) lives in the African region(Congo river basin) that presents the greatest genetic diversity of HIV[41],a fact that could be expected if this was the place where HIV-1 first emerged. SIVcpz infection is common among chimpanzees in that region[42,43]and so it was established that chimpanzees of that particular subspecies are the natural reservoir for SIVcpz and the source of HIV-1 group M, answering the "where" question: HIV-1 group M ancestor came from southeastern rain forests of Cameroon (modern East Province) near the Sangha River (figure 1).This region is presumably where the virus was first transmitted from chimpanzees to humans[40]. However, reviews of the epidemiological evidence of early HIV-1 infection in stored blood samples,

and of old cases of AIDS in Central Africa, seems to point to a HIV-1 group M early human center not in Cameroon, but rather farther south in the Democratic Republic of Congo, more probably in its capital city, Kinshasa[41].Leopoldville, as it was known before independence, was not only the largest city in the region, but also a likely destination for a virus escaping from southeast Cameroon. Indeed, in the early 1900s, the main routes of transportation out of that remote forest region were rivers; those surrounding this area flow south, ultimately draining into the Congo River, and leading to Leopoldville (figure 2).

The absence of SIVcpz infections in two of the four subspecies suggested that chimpanzees had acquired this virus recently. Nonetheless, SIV evolution is complex and recombination does occur between different variants. Probably SIVcpz is a recombinant virus, originated by a dual infection with another SIVs, the descendants of which are now found in red-capped mangabeys (*Cercocebus torquatus*, SIVrcm) and greater spot-nosed monkeys(*Cercocebus*

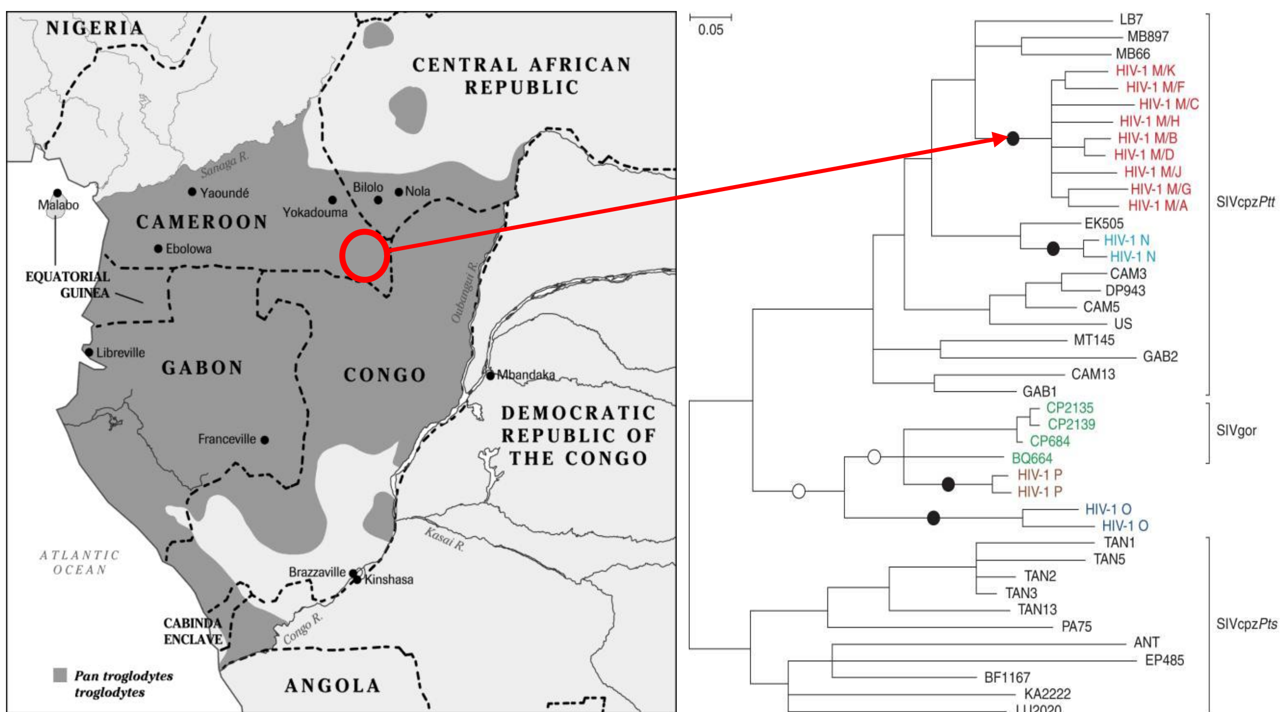

Figure 1: Habitat of Central chimpanzee (*Pan troglodytes troglodytes*) and HIV-1 phylogenetic tree. Red circle marks the region in southeast Cameroon where SIVcpz strains closely related to HIV-1 group M are found. The phylogenetic relationships of representative SIVcpz, HIV-1, and SIVgor strains are shown. SIVcpz and SIVgor sequences are shown in black and green, respectively. The four groups of HIV-1, each of which represents an independent cross-species transmission, are shown in different colors. Black circles indicate the four branches where cross-species transmission-to-humans has occurred. White circles indicate two possible alternative branches on which chimpanzee-to-gorilla transmission occurred. The phylogenetic tree was estimated using maximum likelihood methods. The scale bar represents 0.05 nucleotide substitutions per site. Based on reference 1.

nictitans, SIVgsn)[44]. Chimpanzees share the habitat and hunt and kill these smaller monkeys, opening chances for cross-species infection during predation[45].

Noticeably, every HIV-1groupcorresponds to a different SIV introduction, as we can perceive because HIV lineages are interspersed among the SIVcpz branches[1].Indeed, the SIVcpz strains from *P. t. troglodytes* form a monophyletic cluster with all HIV-1 strains. This mixing of HIV and SIV lineages, patent in figure 1, points to an independent origin of groups M, N, O and P. The vagaries in sampling make impossible to determine exactly when and how those cross-species transfers have occurred, but at least four jumps were needed to originate the four HIV-1 groups (marked with a black circle in figure 1). For example, HIV-1 group N seems to be a recombinant between a SIVcpz strain and a virus related to the ancestor of group M, with this recombination occurring before the establishment in humans. The recently recognized HIV-1 group P clusters significantly with SIVgor strains[46]and the most likely explanation for its emergence is gorilla-to-human transmission of SIVgor. However, a chimpanzee origin for the HIV-1 group O and SIVgor lineage has been proposed, so the possibility that SIVcpz also gave rise to HIV-1 group P cannot be ruled out, either indirectly by transmission to gorillas and then to humans or directly by transmission to humans and also to gorillas. Regarding group M, there are presently 9 subtypes identified (A-D, F-H, J, K) and dozens of recombinant forms that diverged from the same ancestral (figure 1). In 2008, Worobey *et al* found HIV-1 sequences preserved in a Congolese woman lymph node biopsy done in 1960 and called it DRC60. For HIV, as for all rapidly evolving virus, it is possible to calibrate their evolutionary rates by comparing virus isolated at different time points and then generalizing this molecular clock to estimate dates of divergence of the different clusters in the evolutionary trees.ComparingDRC60 with viral sequences from a blood-plasma sample originally obtained in 1959, also from Kinshasa (ZR59)[47], a high divergence was found (approximately 12%).These viral sequences were classified in different subtypes: ZR59 was a subtype D and DRC60 was a subtype A. This divergence demonstrated that at least 55 years ago group M HIV-1 strains had already undergone substantial diversification. Using relaxed molecular clock analyses incorporating DRC60 and ZR59,the authors could date the M group's most recent common ancestor near the beginning of the 20th century, between 1884 and 1924[48].This is the most convincing answer we have to the “when” question until now.

Initially, SIVcpz was thought to be harmless for its natural host, since sooty mangabeys have no sign of immunodeficiency despite SIVsmm high viral loads. Several studies proved other way, providing compelling evidence that SIVcpz was pathogenic to its natural host, causing CD4+ T-cell depletion and histopathological findings similar to end-stage AIDS[42,43]. This natural history is also consistent with recent acquisition.

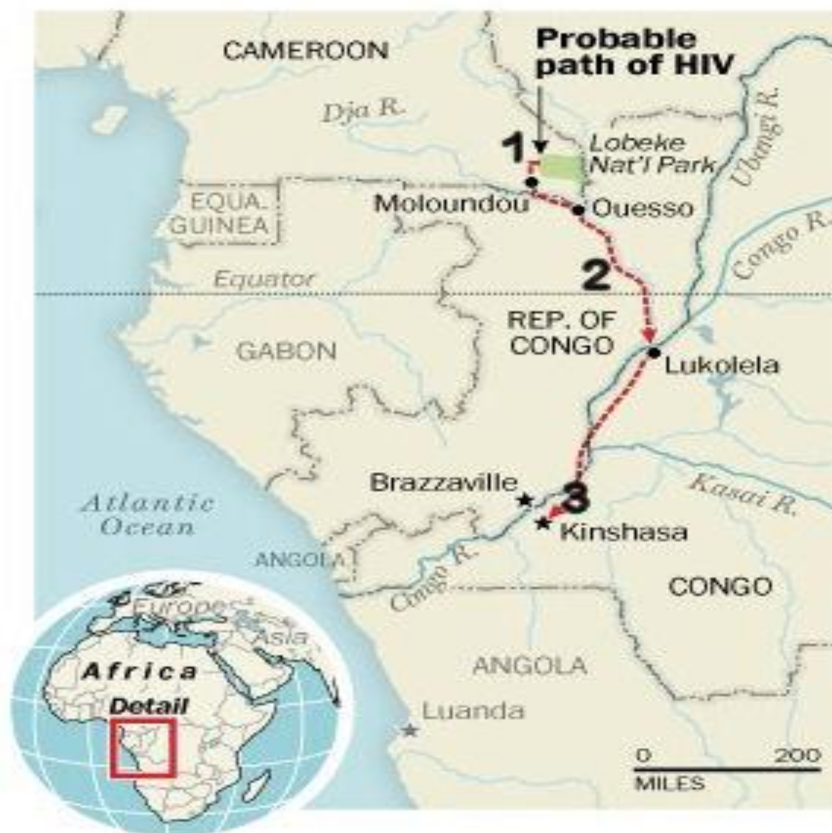

Figure 2: A plausible mechanism for acquisition and early propagation of HIV-1.

1 – Humans butcher chimpanzees infected with SIV; **2** – The virus is carried by people travelling along the rivers… **3** - …to Kinshasa, where the epidemic begins.

Probably, SIV transmission to humans has taken place at low frequency for hundreds of years. But there's no evidence of endemic HIV-1infections in late 19th or early 20th centuries. Why is there an AIDS epidemic now? Some strains could be nonpathogenic; others were maybe non-transmissible between hosts; or infected individuals got sick and died without spreading the virus, due to low population density or lack of risky sexual behavior.

Several factors were needed to start a pandemic disease, such as AIDS (see figure 2):

a) Increased exposure risk

b) Increased probability of transmission

c) Increased probability of virus adaptation to humans

Increased exposure risk – Colonial authorities used intensive and forced labor in Central Africa to build railroads and extract rubber between late 1890s and 1930s. Local and non local workers were concentrated in camps where there was a disproportional man to women ratio, nearing 12:1, and not enough food supplies[49]. This dislocation and concentration in wild areas, like forests, probably led to significant increases in hunting for food. Fire guns were more accessible, especially after World War I. This facilitated killing of large animals, like chimpanzees.

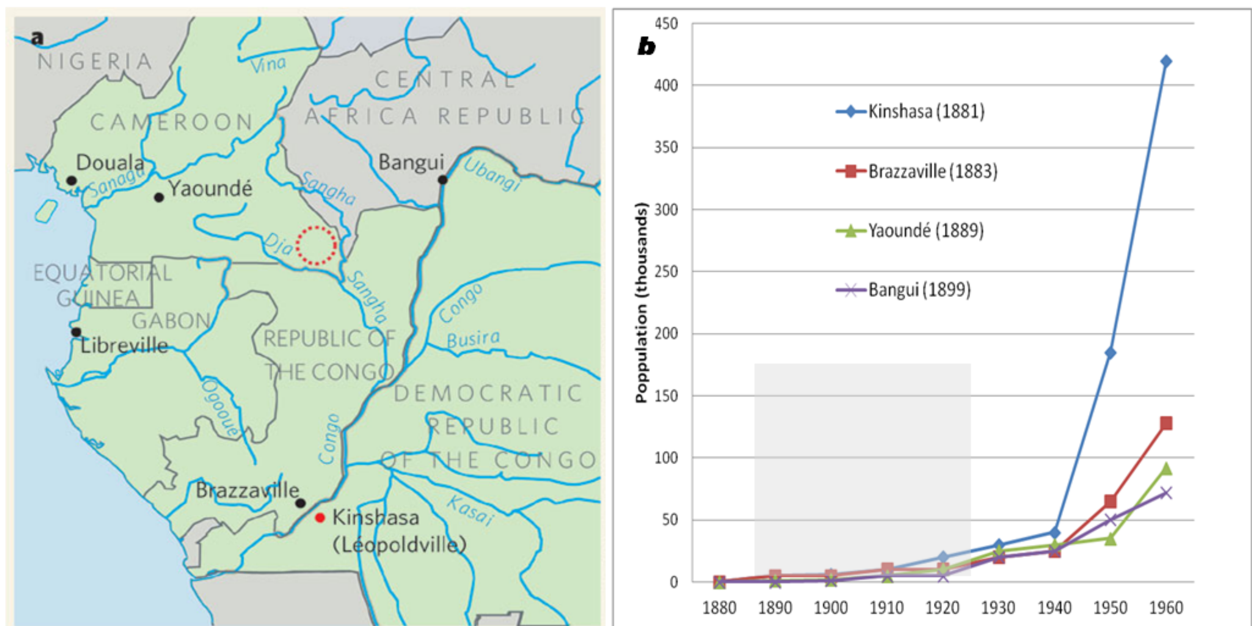

Figure 3: Demographic changes in the cradle of HIV-1

a, Map of west-central Africa showing major rivers (major communication routes in that area) and cities with explosive population growth in the twentieth century. SIVcpz strains most closely related to the viruses of HIV-1 group M have been found in chimpanzees in southeast Cameroon (red circle). **b**, Demographic evolution of the major cities near the epicenter of the HIV-1 epidemic. Noteworthy, the absence of a single site with a population exceeding 10,000 until after 1910 and the massive growth after 1940. In the grey area the estimate time for HIV-1 emergence (1884-1924). The founding date of each major city in the region is listed after its name. Data extracted from reference 49.

Increased probability of transmission – The distorted sex ratio in labor camps certainly was a risk to sexual promiscuity as most women were placed in those camps for "recreational" purposes. Since the end of 19th century, riverboat traffic along the Congo River and its effluents became regular and intense, connecting areas where people had little or no contact

with each other previously[49].After 1940s, railroads helped establishing massive migrations to major cities, like Kinshasa (former Leopoldville) and Brazzaville, leading to its rapid expansion (figure 3). Such a quick "urbanization" created favorable conditions to disease establishment and propagation.

Last, we must remember that there were massive sanitary campaigns happening at that time, carried out with very limited resources: in 1916 more than 80000 people in the Nola region (in Central African Republic, near Cameroonian border, by Sangha River) were treated with intravenous drugs for trypanossomiasis with only six syringes[50].

It was reported by Pépin *et al* how intravenous treatment of endemic infections (trypanossomiasis in southeast Central African Republic and malaria in southern Cameroon) led to an increased level of transmission of two blood borne virus, hepatitis C virus and HTLV-1, in the first half of 20th century[51].

As mentioned before, those two regions are coincident with *Pan troglodytes troglodytes* habitats, the chimpanzee that harbored SIVcpz, believed to be the HIV-1 precursor, so authors hypothesize that those iatrogenic exposures may have proportionate an opportunity for HIV to evolve from a few isolated cases to a larger reservoir of infection. In thatpaper[51], the authors describe an increased mortality in the elder groups in association with these therapeutic interventions.

Although impossible to confirm, the excess mortality was probably due to HIV, acting as a quicker killer than HTLV-1 and HCV, two virus compatible with prolonged survival. Parenteric therapy for endemic infections more than 50 years ago might have contributed to ignite HIV pandemic.

Increased viral adaptation – For decades, cross-species events must have occurred, resulting in epidemiological dead ends: a hunter kills an ape, infects himself manipulating the carcass, eventually infects his wife and both die of AIDS in their village. Establishing transmissibility is crucial for a successful pandemic. SIV must therefore to have changed into a new species, adapted to its new host. This topic will be further developed ahead.

In conclusion, HIV probably crossed the species barrier in the bloody process of butchering chimps for food. The spread of the virus in the Belgian and French colonies of central Africa in the mid-20th century was probably accelerated by rapid, male dominated urbanization and the concentration of workers in camps with attendant prostitution. Health campaigns where

needles were repeatedly reused may also have contributed. The links between francophone Central Africa and Haiti provided HIV with a staging post to the western hemisphere; genetic analysis suggests that the virus was imported from Haiti to the United States of America in the late 1960s[7].

The sexual liberation of the 1960s and 1970s combined with homophobia concentrated large groups of sexually active gay men in small enclaves of tolerance in cities like New York and San Francisco. Anal sex is inherently more dangerous for passing on virus than vaginal sex because it is more likely to cause small tears and lesions. That, together with a high turnover of partners, provided perfect conditions for the rapid spread of HIV. Because these men were young, otherwise healthy and largely well educated and white, their illness attracted attention in a way that earlier cases in Africa and the Caribbean had not. AIDS came out of the shadows.

BIOLOGY AND PHYLOGENY OF HIV

What is HIV?

As mentioned above, HIV-1 is a member of the genus *Lentivirus*, part of the family *Retroviridae.* This virus is the causal agent of the Acquired Immune Deficiency Syndrome. There are two phylogenetically distinct types of HIV, known as HIV-1 and HIV-2. HIV-1 was isolated in 1983 in the United States of America from homosexual men afflicted with rare opportunistic infections[19]; HIV-2, which has a slightly different genomic structure, was isolated from West African AIDS patients in 1986 [24]. HIV-1 is further divided into four groups: M (for main or major) is responsible for the global pandemic; N (for non-M non-O) is confined to a dozen patients infected in Cameroon; O (for outlier) infected less than hundred thousand people and is restricted to Cameroon and Gabon; and P, discovered in 2009, has infected two patients in Cameroon so far[46].On other hand, HIV-2, that appears to be less pathogenic[52,53], is mainly restricted to West Africa[26] and countries with historical links to that area, like Portugal and France.

Lentiviruses are characteristically responsible for long-duration illnesses with a long incubation period, also features of HIV-1 infection[54].The long infection period of HIV within the host, in the absence of long term treatment, is followed by almost universal progression to AIDS, reaching a mortality rate of near 100%. The long infection period in conjunction with population demographics has led to the HIV-1s unique epidemiology, as described previously.

HIV-1 structure

HIV-1 is different in structure from other retroviruses. It is spherical in shape, roughly 80 – 100 nm in diameter and possesses two usually identical copies of a single stranded RNA genome of approximately 9 kilobases in length. The HIV-1 genome, represented in figure 4, is primarily a coding RNA and contains nine open reading frames that produce 15 proteins[55].

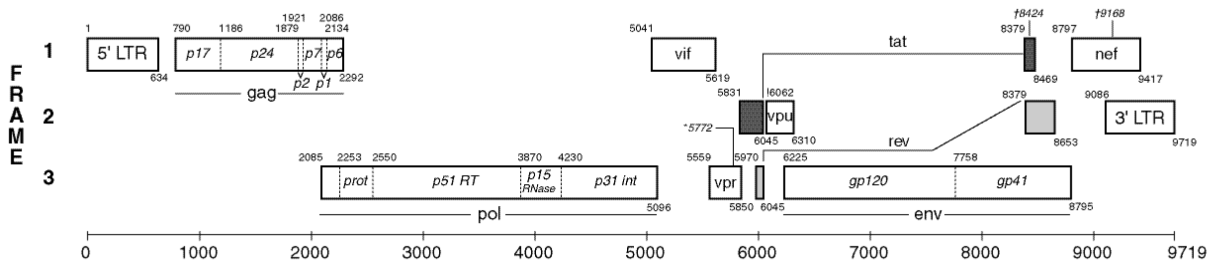

Figure 4: HIV-1 genomic landmarks. Los Alamos National Laboratory reference strain (HXB2), showing HIV-1 genomic structure consisting of nine genes (*gag*, *pol*, *env*, *tat*, *rev*, *nef*, *vif*, *vpr* and *vpu*). Figure and description of gene functions based on http://www.hiv.lanl.gov.

GAG: The genomic region encoding the capsid proteins (group specific antigens). The *gag* polyprotein precursor is proteolytically processed by the viral protease to generate the matrix (MA or p17), capsid (CA or p24), nucleocapsid (NC or p7) and p6 proteins. *Gag* associates with the plasma membrane, where virus assembly takes place.

POL: This genomic region encodes three *pol* proteins, protease (PR), reverse transcriptase (RT) and integrase (IN), providing essential enzymatic functions for HIV life cycle. These enzymes are produced as a *gag*-*pol* precursor polyprotein, which is processed by the viral protease.

ENV: Viral glycoproteins produced as a precursor (gp160), which later is processed resulting in two envelope proteins, surface or gp120, which mediates virus attachment and entry into host cells via interaction with host cell receptors, and transmembrane or gp41, responsible for membrane fusion. Protein gp120 contains the binding site for the CD4 receptor.

TAT: Trans-activator of transcription (HIV gene expression). One of two essential viral regulatory factors (the other is rev) for HIV-1 gene expression. Mainly, it acts by activating RNA transcription initiation and preventing its premature termination.

REV: Regulator of expression of virion proteins. The second regulatory factor for HIV expression, rev acts by promoting the nuclear export, stabilization and utilization of the viral mRNAs. Rev is considered the most functionally conserved regulatory protein of lentiviruses.

VIF: Viral infectivity factor, promotes the infectivity but not the production of viral particles. In the absence of *vif*, the produced viral particles are unable to freely infect cells, although cell-to-cell transmission is not affected. It also prevents the action of the cellular APOBEC-

3G, an host factor that inhibits infection by deaminating DNA:RNA heteroduplexes in the cytoplasm.

VPR: Viral protein R, which enhances post cell entry infectivity.

VPU: Viral protein U, which is found in HIV-1 and SIVcpz but not in HIV-2 and several other SIVs. *Vpx*, found in HIV-2 is a *vpr* homolog, apparently resulting of a gene duplication event, possibly by recombination. Both have similar functions, suggesting redundancy. They degrade CD4 in the endoplasmic reticulum thus allowing virion release from the plasma membrane of HIV infected cells.

NEF: negative factor, which down regulates CD4, the primary viral receptor, and MHC class I expression, increasing viral infectivity. *Nef* genes are essential for efficient viral spread and disease progression, maintaining high viral loads.

HIV-1 life cycle

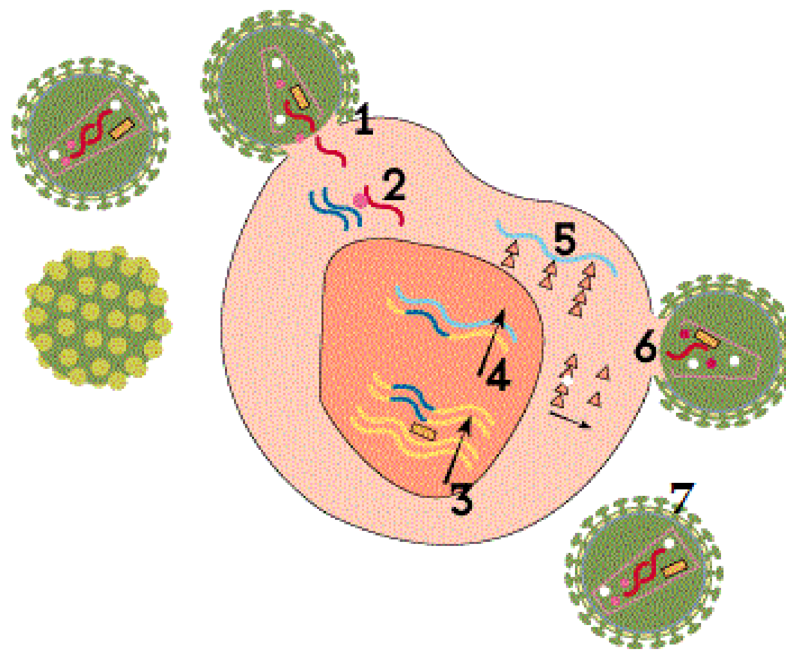

Figure 5: HIV-1 life cycle. HIV-1 is represented in green, RNA in red, pro-viral DNA in blue and host cell DNA in yellow. **1** –Binding to the target cell; **2** – Reverse transcription of HIV RNA; **3** – Integration into the host genome; **4** – Transcription of viral proteins; **5** – Assembly of a new viral particle; **6** – Release of immature virions; **7** – Maturation of the virion

In order to replicate, virus must infect a host cell. A property called tropism determines which cell type a given virus will infect. HIV-1 searches for cells that have CD4 surface receptors, because this particular protein enables the virus to bind to the host cell. T-lymphocytes and cells of monocyte macrophage lineage both express the CD4 receptor and can therefore be infected by HIV-1. As described earlier, HIV-1 genome consists in nine genes, encoding for 15 proteins. Each of these proteins plays a role in HIV-1 life's cycle, which includes binding, reverse transcription, integration, transcription, assembly, release and maturation[55].

CELL BINDING AND ENTRY: HIV is an enveloped virus. In this envelope, the viral protein type I transmembrane glycoprotein (gp), coded by the *env* gene, is expressed. HIV-1*env* is produced in a host cell as gp160, later cleaved to gp120 (extracellular component) and gp41 (transmembrane component). The first step of HIV-1 binding occurs when extracellular gp120 binds to its primary cell surface receptor CD4. Binding of gp120 to CD4 results in a conformational change in *env*, enabling binding of gp120 to a cellular co-receptor. While a number of cellular co-receptors for HIV-1 binding have been described, the most important are CCR5 and CXCR4.Both co-receptors are expressed in memory CD4+ T cells[56],but CCR5 can be expressed also in macrophages and monocytes[57]. Individual virus strains can display increased affinity for both or either receptors, showing R5 tropism, X4 tropism or dual tropism. Initial infection typically occurs via an R5-tropic virus, which evolves over the course of an untreated infection into an X4-tropic virus, often with a dual-tropic intermediate[58]. There is evidence that X4-tropic virus may be more pathogenic that R5-tropic virus[59].

Engagement of gp120 with the cellular co-receptor induces additional conformational changes with resultant interaction of the gp41 component of *env* with the plasma membrane, fusion of the viral and host membranes, and insertion of the viral core into the cytoplasm[60]. Binding is the first step for HIV-1 infection and *env* is the only viral protein expressed in the HIV-1 envelope, so it seemed attractive targeting both in vaccine development. However, HIV-1displays an immune evasion strategy consisting in hypervariability and extensive glycosylation of gp120, turning all vaccination efforts useless to date.

HIV-1 REVERSE TRANSCRIPTION: Once fusion of the viral and cellular membranes takes place, the viral core enters the cytoplasm of the target cell. The HIV-1 viral core is a multimeric structure composed of capsid (CA or p24, a component of *gag*) proteins. Contained within the viral core are two copies of the viral single plus strand RNA genome, as

well as viral proteins (reverse transcriptase, *nef*, *vpr*, nucleocapsid, integrase), and primer tRNA. Following uncoating and degradation of the viral core, the viral genome enters the nucleus for integration[61].

Reverse transcription is a multistep process entailing the transition of the single-stranded viral RNA into double-stranded proviral DNA, which is then integrated into the host chromosome. This is accomplished by HIV-1 reverse transcriptase (a component of *pol*), ultimately a DNA polymerase (simultaneously RNA- and DNA-dependent) and a ribonuclease, combining DNA polymerization and RNA cleavage. Outstandingly, HIV-1reverse transcriptase lacks the functional "proof-reading" activity that is present in other DNA polymerases, contributing to the high mutation rate of the virus[62].However, the dimeric nature of the retroviral HIV RNA genome is the true responsible for its high genetic variability by means of forced and non-forced copy-choice recombinations during reverse transcription[63]. Later, this recombination as a form of primitive sexual reproduction will be detailed.

HIV-1 DNA INTEGRATION: Transcription of RNA to DNA and subsequent integration into the host cell genome are distinctive features of retroviruses. HIV-1, as other retroviruses, benefit of integration in terms of protection, as unintegrated genetic material is quickly degraded.

HIV-1 integration begins as the newly synthesized viral DNA binds HIV-1 integrase in the LTR region in the cytosol of the host cell. Combined with additional host and viral proteins, a pre-integration complex (PIC) is formed[64]. This complex translocates to the host cell nucleus. Once inside the nucleus, the PIC binds to the host cell DNA. Next, the 3' overhang of the viral DNA is inserted into the 5' target host DNA. Evidence exists that HIV-1 DNA preferentially integrates into actively transcribed regions of the host cell DNA[65]. At this point, there still remain unpaired flanking regions and gaps between the viral DNA and the host genome. Since the cell does not recognize this newly inserted DNA as foreign, host DNA repair elements will respond and repair the ends, leading to fully integrated HIV-1 DNA[66].

After successful integration, the provirus can either be transcribed, leading to new progeny virions, which is discussed in the next section, or can lie dormant without viral protein gene expression, in a state of latency. HIV-1 latency is a major barrier in accomplishing a cure of the infection. Once integrated, if the virus lays dormant, it will evade the host's immune response and is unaffected by current therapies that target active viral proteins[67]. Current

research suggests that the half-life of a latently infected, resting memory T cell, can be as long as four years[68]. Transcription of provirus upon reactivation allows replenishment with HIV virions, making nearly impossible to eradicate HIV from the system without first eradicating the latently infected cells.

TRANSCRIPTION AND REGULATION OF VIRAL PROTEINS: After integration, HIV-1 transcription is controlled in part by the HIV-1 protein tat. This protein binds to the 5' LTR of the integrated viral DNA and ultimately leads to HIV's transcription elongation. Some of the HIV-1 genes overlap and require splicing, which may result in splice variants. Splicing variants that contain sequences for proteins involved in virion formation, including structural proteins and enzymes, containing a rev-responsive element that allows HIV-1 rev to bind the mRNA and transport it out of the nucleus to be translated[69].

ASSEMBLY OF HIV-1 VIRIONS: The next step of the HIV-1 life cycle is packaging and creating the virus (assembly). *Gag* and *gag*/*pol* proteins are of major importance for assembly to take place. *Gag* is responsible for recruiting HIV-1 proteins to the cell membrane where assembly takes place. Matrix, capsid, nucleocapsid, and p6 are encoded by *gag*. Before being trafficked to the membrane, the newly synthesized *gag* binds the genomic HIV RNA that has been exported from the nucleus via its nucleocapsid region and to a less selective region of matrix. Once at the plasma membrane, matrix anchors the polyprotein to the plasma membrane. *Env* is trafficked separately from *gag*. The p6 and capsid domains are responsible for protein: protein interactions necessary to assemble the virion. Capsid and its interactions with matrix, nucleocapsid and p6 align and concentrate *gag* proteins in preparation for budding[70].

BUDDING OF HIV-1 VIRIONS: Budding consists in a virion leaving the infected cell with components of the plasma membrane encircling it. By crossing the plasma membrane, HIV-1 obtains its lipid envelope. It is a complex process, requiring recruitment of host factors and opposition to mechanisms developed to combat the release of immature virions[71]. HIV-1 *vpu* as an important role in antagonizing such mechanisms[72].

MATURATION OF HIV-1 VIRIONS: The final step of HIV-1 life cycle is called maturation and corresponds to the process of structural changes that allows the virus to infect new host cells. The HIV-1 protease is the key to the assembly of a mature infectious virion. When translated, the viral proteins are translated as polyproteins. The HIV-1 protease is responsible for cleaving the polyproteins so they can structurally form the virion. After cleavage of *gag*,

the viral RNA becomes condensed and stabilized[73]. The structural components are assembled and an HIV-1 virion is formed. This is considered a turning point, because now the virion has transformed from a vesicle budded from the host cell to a mature virus that can infect, replicate, and survive in new cells.

EVOLUTION TO DIVERSITY

HIV-1adaptation to a new host

When a virus enters a new host species for the first time, several outcomes may occur, ranging from incidental infection to epidemic spread. As referred above, HIV-1 has originated from cross-species transmission of another lentivirus, SIVcpz, probably in southeast Cameroon, *circa* 1900. There are five arguments in favor of this hypothesis, all of them mentioned in the previous section:

1. Similarities between both genomes
2. Phylogenetic trees showing close relations
3. SIV's prevalence in their natural hosts
4. Geographical coincidence
5. Plausible mechanisms of transmission

In 2000, a study was published providing proof that HIV-1 group N related virus circulate in wild chimpanzees, confirming inter-species transmission[74].

However, we also know that SIVs infects its natural hosts for hundreds or thousands of years without causing disease or immunodeficiency even in the presence of high viral loads, contrasting with its name[36,75]. Nevertheless, when a SIVsmm was accidentally transmitted in captivity from a sooty mangabey to an Asiatic macaque, the later developed a fatal immunodeficiency, resembling AIDS[76]. This occurrence was experimentally reproduced later, originating SIVmac, which allowed rhesus monkeys to serve as an AIDS animal model[77]. We must remember that no SIV was ever identified in non human primates outside Africa.

SIVcpz, who shares a common ancestral with HIV-1 groups M and N, is in turn a chimera between older SIVs[44]. This make it a phylogenetically recent virus, eventually explaining why it can produce CD4+ cell depletion in chimpanzees and curtail their life expectancy[42], although not as dramatically as in HIV or SIVmac infections.

In 1992 and 1994 two cases of accidental transmission of SIVsmm to laboratory technicians resulted in a spontaneous clearance of the virus[78] or in an asymptomatic infection[79].

However, it was recently demonstrated that SIVcpz is able to replicate in human lymphoid tissue, as long as a viral matrix protein undergo an adaptation[80]. All these facts allow concluding that SIV is not the cause of AIDS and that AIDS is not a true zoonosis.

Indeed, zoonosis (transfer of a pathogen from non-human animals to humans) and subsequent spread of the pathogen between humans, requires the following conditions:

1. Ahuman population
2. A nearby population of a host animal
3. An infectious pathogen in the host animal that can spread from animal to human
4. Interaction between the species to transmit enough of the pathogen to humans to establish a human foothold, which could have taken millions of individual exposures
5. Ability of the pathogen to spread from human to human (perhaps acquired by mutation)
6. Some process allowing the pathogen to disperse widely, preventing the infection from "burning out" by either killing off its human hosts or establishing immunity in a local population of humans.

Criteria 1 to 4 are clearly fulfilled by HIV-1. Criteria 5 and 6 imply some kind of adaptation to the new host species. Classifying HIV infection as a zoonosis is not just semantics. If AIDS was a true zoonosis, it meant that it was naturally acquired directly from an animal source. SIVcpz is accepted as the original source of HIV-1 but not as the original source of AIDS. The importance of classifying HIV infection as being or not a zoonosis resides in understanding the processes by which animal virus can cause epidemics or pandemics. In another words, understanding the adaptative process (es) in a new host that will eventually lead to a new emerging disease.

There are several arguments disfavoring AIDS as a classic zoonosis, being rather a zoonotic transmission of a virus ultimately causing a human infectious disease:

1. A large number of exposures to different SIVs are calculated to have happened in Central and West Africa for centuries[81]. In spite of that, only 12 cross-species transmissions are documented, giving birth to four HIV-1 and eight HIV-2 groups[82]. Even in the presence of massive exposure, SIV infections in humans are

exceedinglyrare[83]. In fact, only four of these crossover events resulted in epidemic strains. They are the HIV-1 group M, the major group of virus of the pandemic, group O, which is responsible for around5% of cases in Cameroon[84] and groups A and B of HIV-2, which are the epidemic forms of HIV-2[85,86]. All other HIV-1 and 2 groups are outstandingly rare.

2. SIV infections in accidental (non natural) primate hosts are weakly or non-pathogenic and usually spontaneously cleared[87,88], with the previously noted exceptions of SIVmac. It seems obvious that cross-species transmission of a lentivirus is not the only requirement for the selection of a pathogenic virus in the new host and that some mechanism of virus adaptation to the new host is mandatory.
3. The SIVs infections in their natural host are generally asymptomatic in spite of high viral loads over long periods of time[36,75]. This finding reinforces the assumption that a change in the pathological potential of the virus is needed for SIV to become pathogenic in anew primate host.
4. The epidemic only emerged in the second half of the 20th century in spite of several previous decades of exposure, which suggests the intervention of some human factors, like deforestation, urbanization and travel, providing the necessary epidemiologic conditions for a pandemic to rise.

Direct contacts with infected animals were clearly not enough to start the pandemic. Those early contacts certainly assumed several forms, like accidental blood borne acquisition, by hunting or manipulating carcasses or even bites of pet monkeys. Once breaking the species barrier, a sum of biologic, demographic and epidemiologic factors favored the infection progression among human populations. In this section, the biologic factors will be discussed.

After exposure and direct contact, the second step for a pathogen to be able to infect his new host is compatibility. That means the existence of appropriate receptors on host cells. The genetic affinity between humans and chimpanzees may enlighten the cross-species jump of HIV-1, as the tropism of SIVcpz must have been fundamental for human cells permissivity to infection. As explained before, there must have been at least 12 cross-species transmissions, originating two HIV types further divided into groups (Figure 6). These are quite different in terms of spread and pathogenicity. Due to their different origins, HIV-1 and HIV-2 are phylogenetically and genomically distinct, probably explaining their differences. A question

remains: why is HIV-1 Group M the pandemic virus? This could be due to a combination of sociological factors and intrinsic viral properties.

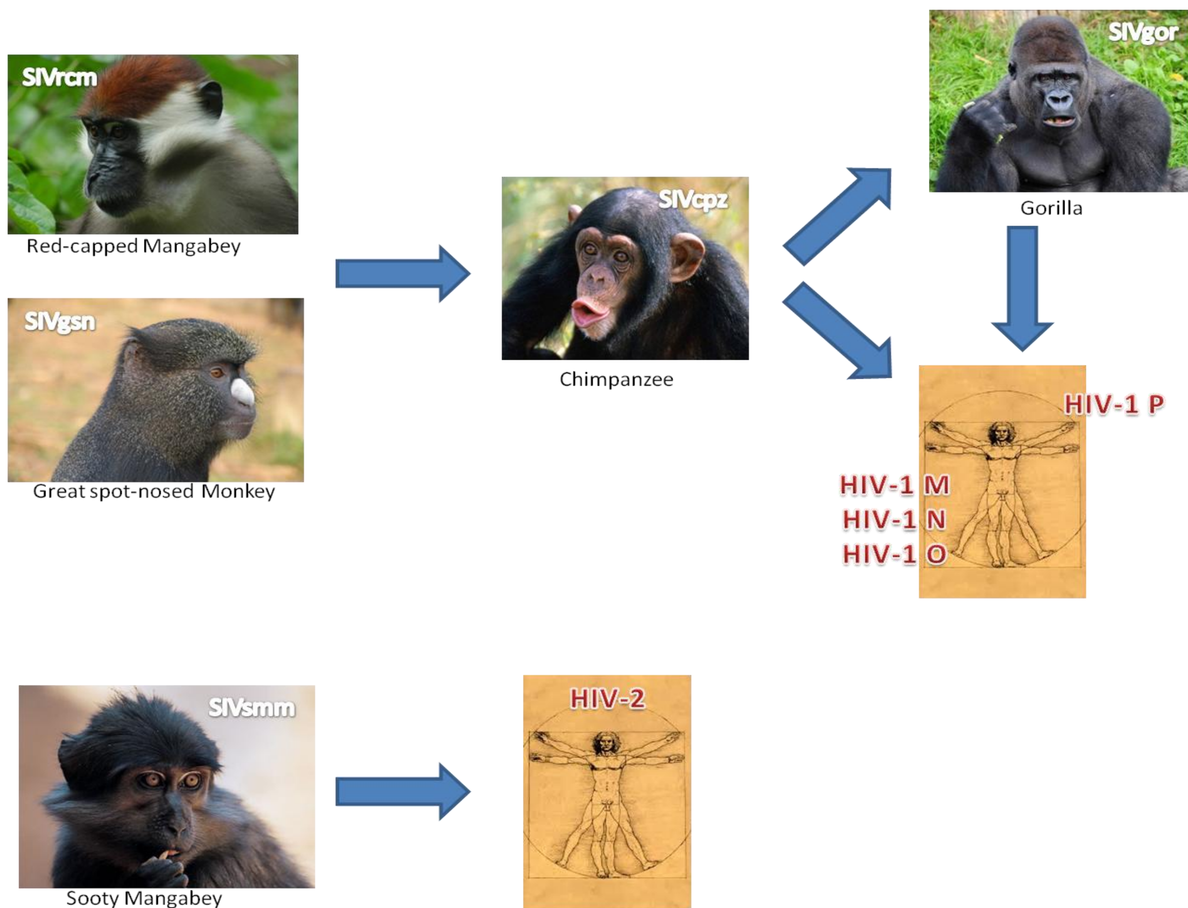

Figure 6: **HIV evolutionary history.** The precursor of HIV-1, SIVcpz is of hybrid origin, resulting from a chimera between SIVrcm and SIVgsn. HIV-1 is the result of three independent cross-species transmission of SIVcpz from chimpanzees and one transmission of SIVgor from gorilla (itself the descendant of transmission from SIVcpz); and HIV-2 arose from eight independent cross-species transmission of SIVsmm from sooty mangabeys to humans. Blue arrows mark the necessary adaptations to a new host.

As seen in the previous section, molecular clock studies estimated the foundation of HIV-1 group M at a time (early 20th century) that coincides with high population growth in Africa and migration to major cities around the zone of cross-species transmission. This demographic particularity provided a sufficient host-population size and transmission networks to facilitate the establishment of HIV-1.Contrasting, the emergence of the most recent common ancestor (tMRCA) of HIV-1 Group O is about 1920 (1890 – 1940) and HIV-1 Group N tMRCA is about 1963 (1948 – 1977)[89], when demographic and socio-economical factors were less favorable to start a pandemic.

Nevertheless, there is also evidence of intrinsic properties of HIV-1 group M strains that might explain its more effective viral replication. Specifically, HIV-1 virus isolates from group M have a higher replicative capacity and display more effective viral countermeasures that antagonize host restriction factors[90]. Restriction factors are effectors of the innate immune response against viral pathogens that inhibit viral replication by operating as molecular barriers to essential steps of the viral life cycle. There are several host restriction factors againstHIV-1 and other retroviruses that have been identified to date. They are germline-encoded, not involving somatic learning and immunologic memory. Indeed, many of these factors are proteins highly induced upon interferon stimulation as part of switch to an "antiviral state" of the cell[91]. Some of these restriction factors assume relevance in defending host cells from retroviral infection. It is believed that HIV-1 group M adaptation to human hosts and resultant successful infection resides in countermeasures (represented by accessory genes) to restriction factors. Some of these restriction factors drove this adaptation and ultimately gave birth to HIV-1: APOBEC3G, tetherin and SAMHD1.

APOBEC3G

APOBEC3G (apolipoprotein B mRNA-editing enzyme, catalytic polypeptide-like editing complex 3) was identified because of its antagonism by HIV-1 *vif*[92]. APOBEC3G is a host cellular cytidine deaminase enzyme, primarily expressed in the natural targets of HIV infection: CD4+ T lymphocytes, macrophages and dendritic cells[93]. APOBEC3G can inhibit viral propagation mainly through deamination activity, mutating dC in dU in nascent viral DNA strands generated by reverse transcription, leading to the destruction of some of these strands. For those who evade destruction, the G to A hypermutations will lead to premature termination codons or mutated viral proteins[94].APOBEC3G also displays some inhibitory deaminase independent activity, by binding viral RNA, by interfering with the DNA strand transfer acrobatics of reverse transcription, by physically blocking reverse transcriptase and by obstructing integration into the host cell genome[95]. APOBEC3G is packaged into virions in infected virus-producing cells and it has been shown that it is largely this virion-packaged fraction, rather than the pool of cytoplasmic APOBEC3G, that is most active on the viral genome in newly infected cells[96]. APOBEC3G restriction of HIV-1 is antagonized by *vif*[97]. If HIV accessory gene *vif* is expressed, APOBEC3G averts packaging into nascent virions and lose its antiviral activity. In the target cell, *vif* directly binds to APOBEC3G, targeting it for ubiquitin-dependent proteasomal degradation. However, there is species-specificity between APOBEC3G-*vif* interactions. For example, human APOBEC3G can be targeted by HIV-1 *vif*

but not by SIVagm *vif*; reciprocally, African green monkey's APOBEC3G can be targeted by SIVagm *vif* but not by HIV-1 *vif*[96]. A single amino acid change on APOBEC3G can toggle the species-specific switch of *vif* sensitivity[98]. This means that host evasion from *vif* antagonism requires a single adaptive change and consequently imposes a species-specific barrier. Thus, HIV-1 *vif* acts as an adaptor protein that targets the host cellular machinery to remove APOBEC3G. As a result, APOBEC3G is not packaged into virion particles and virus keeps its infectivity. Of relevance, suboptimal blockade of APOBEC3G function by partially defective *vif* mutants will lead to an increase of G-to-A viral mutations that can facilitate the emergence of some antiretroviral resistance mutations, even in the absence of drug exposure[99].

Tetherin

Tetherin, also known as CD317 or BST-2 (bone marrow stromal antigen 2) is a host restriction factor capable of inhibiting the release of a broad range of virus from the plasma membrane of infected cells. Tetherin restricts virus release by physical linkage between the nascent virion and the host cell plasma membrane[100]. Tetherin is expressed by all major cellular targets of HIV-1 infection, including CD4+ T lymphocytes. Its tetramer structure is formed by two parallel dimers with pronounced flexibility. This intrinsic protein topology determines its function rather than the coding amino acid[100]. As a result, tetherin also inhibits the release of other virus, such as filovirus (Ebola and Marburg virus), arenavirus (Lassa virus), and other retrovirus, including gammaretrovirus (murine leukemia virus) and spumaretrovirus (foamy virus)[101].

Retroviruses have developed at least three strategies to counteract tetherin. First, the *vpu* protein from HIV-1 abrogates the retention phenotype in cells that express tetherin and has been observed to be co-localized with tetherin. *Vpu* displays species specificity and counteracts human tetherin by targeting it for degradation. Notably, *vpu* is encoded by a unique lineage of primate lentivirus, including HIV-1, SIVcpz and SIVgsn/mus/mon. HIV-1 *vpu* directly binds to tetherin leading to the ubiquitination of tetherin and to its lysosomal degradation[102]. In contrast, SIVcpz, the immediate precursor of HIV-1, whose *vpu* shares a common ancestry with HIV-1 *vpu*, uses *nef* instead to counteract chimpanzee tetherin. Human tetherin, however, is resistant to *nef* and thus poses a significant barrier to zoonotic transmission of SIVcpz to humans[103]. Outstandingly, to counteract tetherin, HIV-1 *vpu* evolved, substituting *nef*. Here, HIV-1 adaptation derives from an adaptative escape from a past antagonist and, unlike expected, HIV-1 neofunctionalized a different protein instead of

adapt the original antagonist[100]. This evolution of *vpu* to antagonize tetherin may have freed HIV-1 *nef* to evolve and contribute to its pathogenicity. This switch, being exclusive of HIV-1 group M, suggests that it could be essential to the establishment of pandemicity.

Second, lentiviruses from the SIVsmm/SIVmac lineage also have evolved the ability to counteract tetherin through the *nef* protein. The *nef* proteins from SIVsmm and SIVmac display species specificity in their ability to counteract their hosts, sooty mangabey and rhesus macaque respectively, and closely related tetherins. *Nef* binds to the cytoplasmic tail domain of tetherin leading to internalization of cell-surface tetherin and intracellular retention, counteracting its anti-viral activity[103].

Third, as HIV-2 does not encode *vpu* to overcome the action of tetherin, it uses an alternative way through *env*. HIV-2 *env* antagonism of tetherin promotes the cell surface downregulation of tetherin, similar to *vpu*. However, *env* does not induce protein degradation, but leads to the intracellular sequestration of tetherin[104].

As referred above, at least three SIV/HIV gene products (*vpu*, *nef* and *env*) have the potential to counteract primate tetherin proteins, often in a species-specific manner. Human tetherin developed a deletion resistance to *nef*[105]. During its adaptation process to humans, HIV-1 group M obviated this hassle by switching from *nef* to *vpu* to antagonize human tetherin[106]. In contrast, *vpu* proteins from non-pandemic group O and group P virus do not antagonize tetherin[106,107] and those of the group N virus gained some modest antitetherin activity but do not degrade CD4[106]. The finding that pandemic HIV-1 group M, but not non-pandemic groups O, N and P virus, efficiently antagonize the human tetherin supports that pandemic HIV-1 group M strains are better adapted to humans. The inability to antagonize human tetherin may potentially explain the limited spread of these later groups in the human population.

Several other virus can counteract tetherin actions, by means of downregulation or sequestration. These independent acquisitions of a viral tetherin antagonist by multiple virus further emphasize the role of tetherin and the strong selective pressure imposed by its potent and large restriction spectrum.

SAMHD1

SAMHD1 (sterile alpha motif (SAM) and histidine-aspartate (HD) domain-containing protein 1) is a deoxynucleoside triphosphate triphosphohydrolase, a cellular enzyme, responsible for

blocking replication of HIV-1 in dendritic cells, macrophages and monocytes, where it is highly expressed and functional[108]. By converting nucleotide triphosphates to a nucleoside and triphosphate, SAMHD1 depletes the pool of nucleotides available to a reverse transcriptase for viral cDNA synthesis and thus prevents viral replication.

SAMHD1 was identified when the accessory protein *vpx* was shown to be decisive for the capability of primate lentiviruses to efficiently infect monocytes and dendritic cells[109]. Only HIV-2 lineage and another lineage of SIVs, represented by SIVrcm, encode the auxiliary protein *vpx* that potently overcomes the block to viral replication constituted by SAMHD1, by promoting its degradation by the proteasome machinery. *Vpx* is absent from SIV strains related to HIV-1 as well as from HIV-1 itself. So, in contrast to the other restriction factors, HIV-1 has no means to counteract SAMHD1[110], which was dispensable for efficient HIV-1 spread. This raises the hypothesis of a possible advantage for HIV-1 and the related SIV strains not to counteract SAMHD1.

Evolutionary analysis by Laguette *et al* showed that SAMHD1 experienced strong positive selection episodes during primate evolution[111]. SAMHD1 proteins of apes, monkeys, and lemurs were all active against HIV-1, whereas *vpx* degraded and antagonized SAMHD1 in a species-specific manner. Laguette and colleagues questioned whether the presence of *vpx* represents an advantage favoring cross-species transmission and observed that *vpx* appears to be dispensable for persistence and spread in humans.

Lim *et al* also noted that only two out of eight primate lentivirus lineages encode *vpx*, whereas its paralog, *vpr*, is conserved across all existing primate lentiviruses[112]. By functional analysis, these authors found that multiple *vpx* proteins shared the ability to degrade SAMHD1, but that this ability was often host specific. Additionally, some *vpr* proteins from virus lacking *vpx* could also potently degrade SAMHD1. Evolutionary analysis showed that the ability to degrade SAMHD1 resulted from neofunctionalization of *vpr* that preceded the acquisition of *vpx* in primate lentiviruses. It was concluded that *vpr* gained a new function to degrade SAMHD1 once during viral evolution, thereby initiating an evolutionary “arms race” with SAMHD1. However, the authors also noted that many lentiviral lineages, including the precursors of HIV-1 and HIV-1 itself, never acquired this function[113].

Apparently, no advantage for HIV-1 seems to be inferred from not antagonizing SAMHD-1.A cell intrinsic sensor for HIV-1 exists in dendritic cells and mediates an antiviral immune response. If there is no dendritic cells infection, this sensor is not typically engaged [114]. The

virulence of HIV-1 may be related to the evasion of this mechanism. Accordingly to this hypothesis, HIV-1 did not evolve an anti-SAMHD1 counteracting protein to avoid the cellular detection of viral proteins and consequent immune activation.

Still, the higher pathogenicity of HIV-1 group M relative to other HIV-1 groups and to HIV-2 may be explained by other unique antagonistic functions (or more effective countermeasures) that collectively allow HIV-1 group M to achieve sufficient replicative potential in target cells.

The innate immune defense proteins battle viral infection and their viral counterparts fight back, hindering their function mainly by degradation, internalization and sequestration. This leads to a never-ending evolutionary race between the host and the pathogen. When SIV took a foothold in humans, this two sided arms race leaded to adaptation, overcome of species barrier and ultimately to the emergence of a new lentivirus, HIV.

Genetic diversity

Genetic diversity proved essential for HIV-1 success as a human pathogen. It explains viral persistence in face of immune selection, hindering virus identification and facilitating immune escape, the emergence of drug resistance and the absence of a vaccine after all these years. This diversity is patent as sequence variability, particularly within the *env* variable (V) regions[115].

Korber *et al* described the amazing diversity of HIV quasispecies in a single infected individual, comparing it to the annual variation of influenza[116]. It was found that a phylogenetic tree of influenza virus sequences sampled world-wide in 1996 showed much less diversity than a sampling of subtype B HIV-1 envelope sequences from a single city, Amsterdam, in 1990–1991. This extent of diversity can be explained by a set of factors, namely high viral turnover, high mutation rates, frequent recombination, host immune selection and genetic and phenotypic limitations to variation. A quasispecies can be defined as a cloud or swarm of genetically diverse variants that are linked through mutations that interact cooperatively on a functional level but collectively contributing to the characteristics of a population[117].

Nucleotide sequence variation between HIV-1 groups may reach 30%[118]. Diversity within HIV-1 group M is large, even if compared with other rapidly evolving virus, like influenza[116], thus it is catalogued in nine subtypes (A-D, F-H, J and K)[119]. Subtypes A and F have been further divided into sub-subtypes A1-A3 and F1-F2, respectively. Subtypes E and I don't exist, as their recombinant nature was revealed by complete genome sequencing and were later reclassified as CRF01_AE and CRF04_cpx respectively[120]. This classification suffers from arbitrariness in terms of definition, however it provides a common language for referring to related lineages and captures a fundamental feature of the virus: the gene and protein sequences within a HIV-1 subtype are more closely related to one another than to the genes and proteins from other subtypes. The associations and groupings of subtypes can be statistically validated through phylogenetic analysis[121]. CRF stands for Circulating Recombinant Forms and describe viral genomes that contain clearly delineated sections derived from different subtypes or other recombinants, that share a common ancestor, and that are the basis of multiple infections – in at least three different patients epidemiologically unrelated (http://www.hiv.lanl.gov/content/sequence/HelpDocs/subtypes-more.html). CRFs are thus epidemic strains, which, like subtypes, are of global importance. There are currently 61 defined CRFs (according to Los Alamos HIV Sequence Database, accessed on February 14th, 2014), from first generation recombination or from second generation recombination, the later occurring when recombination takes place between more than one first generation CRF or between CRFs and different subtypes. Some of the CRFs cause more infections than "pure" subtypes, like CRF01_AE in Asia or CRF02_AG, found throughout Western and Central Africa[122].More extensive sampling in regions of sub-Saharan Africa with great viral diversity has resulted in ever greater indications of the potential complexity of HIV-1 diversity. In regions where multiple subtypes are co-circulating with a high prevalence, intersubtype recombination iscommon[123] and recombination between recombinants has also been reported[124]. The large number of novel recombinants suggests that multiple infections of HIV in the same individual are not uncommon. Several new isolated examples of strains that do not clearly fit into any defined subtype or known circulating recombinants have been described and are called unique recombinant forms (URFs)[123]. The subtypes themselves are growing more diverse with time. By definition, a subtype is a cluster in which virus sequences have approximately equal genetic distances between them (10-30%), depending on the genes compared. The mean genetic divergence of subtype A, B, C and D genomes isolated in 1999 (9.8%, 10.0%, 7.9% and 8.5%, respectively) each equal that of the whole group M epidemic in 1985 (8.8%). The genetic divergence of the entire group M has

increased to 14.9% in the same period (LANL HIV Sequence database, http://hiv-web.lanl.gov).

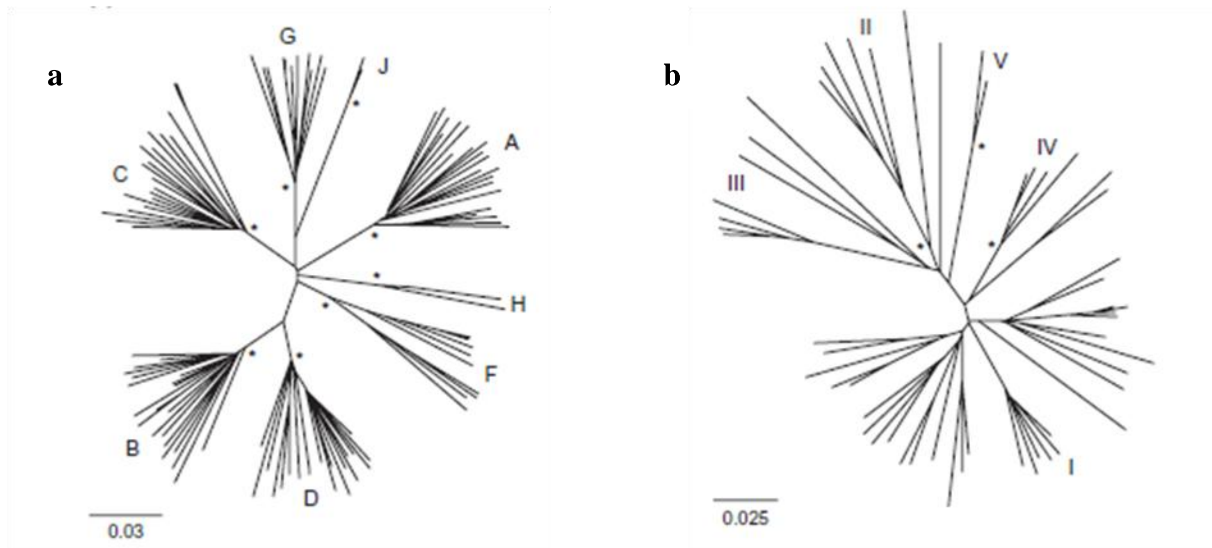

Figure 7: HIV-1 Phylogenetic trees constructed using global group M (a) and Group O (b) envelope gp160 sequences. The •indicates bootstrap support greater than 90%. The scale bar corresponds to nucleotide substitutions per site. Bold lettering corresponds to the Group M subtype designations from the LANL HIV Sequence Database and bold numbers represent the Group O clusters that correspond to previously proposed clusters. Extracted from http://www.hiv.lanl.gov.

The representative tree (figure 7) for the group M envelope region displays the characteristic starburst structure that apparently defines this group's phylogenetic diversification: an organization indistinct and strongly supported clades in phylogenetic trees. This phylogenetic substructure is characterized by a double-star phylogeny, i.e., a tendency for long branch lengths within subtypes that coalesce near the ancestral node of the subtype and long pre-subtype branches that coalesce near the root of the entire tree[125].As a result, strains within any given subtype are always more closely related to each other than they are to strains belonging to a different subtype. The lack of equidistant group O clusters or lack of a starburst structure is also observed in figure 7. It would seem that the group O phylogeny reflects an epidemiology that is dependent on host transmissions on a highly localized (endemic) scale[125]. As a result, the branches leading to the group O clusters tend to be deeper within the trees resulting in shorter branch lengths leading to the individual clusters. In fact, group O has

been geographically restricted to Cameroon where it is responsible for about 1.4%of all HIV-1 infections[84].

This great diversity is seeded by the lack of a proof-reading mechanism in RNA viral reverse transcriptase, and the consequential high error rate (1.5–2.4×10^{-5}mutations/bp/cycle)[126]. Mutations can occur at three stages of retroviral replication:(a) when viral RNA is transcribed from the provirus by host DNA-dependent RNA polymerase II (*pol* II); (b) when the single-stranded viral RNA genome is converted into double-stranded DNA (dsDNA) by viral reverse transcriptase (RT); or (c) when the provirus is copied by the host DNA-dependent DNA polymerase during replication of the infected cell.HIV-1 generates, on average, one error genome per replication cycle[126]. Potentially, each provirus is anew mutant strain, unique at least in one base site. Mutations accumulate over successive replication cycles, leading to a myriad of closely related but not identical virus in every infected individual. Blood and lymphoid tissue from a HIV-1 infected adult contains 10^{11} CD4+ lymphocytes, of which between 10^{9} and 10^{10} can be showed to harbor viral DNA[127]. On account of 3×10^{7} HIV-1 infected patients worldwide, there may be as many as 3×10^{17} HIV-1 genetically unique strain variants in circulation. This vast reservoir of genetic variants may increase the potential for successful adaptation of the HIV-1strains. This fact is the first step for forming quasispecies, which are a hallmark of HIV-1 infection. Quasispecies represent a genetically complex population of virus from an initially limited number of infectious particles, eventually just one.

Besides replication errors, we must also account on the rapid turnover of HIV-1 in vivo. The estimated average total HIV-1 production is 10.3×10^{9} virions per day. The duration of the HIV-1 life cycle in vivo is 1.2 days on average, and the average HIV-1 generation time - defined as the time from release of a virion until it infects another cell and causes the release of a new generation of viral particles - is 2.6 days[128].This formidable numbers allow us to imagine the huge diversity a single virion can incite in a single host.

Two more important factors driving HIV-1 diversity within-host are the host selective immune pressures[129], and the recombination events during replication[130]. Ultimately, HIV-1 evolves during the course of an infection and adapts to its host. These features will be discussed next in more detail, knowing that, although impossible to identify and characterize by direct analytical methods the founder virus at or near the moment of transmission, it is this early virus and sequences evolving from it that will lead to productive clinical infection.

Around the pandemic, HIV-1 is transmitted mainly by sexual contact across mucosal surfaces, but also by maternal-infant exposure and by percutaneous inoculation. For reasons that are still incompletely understood, CCR5-tropic virus (R5 virus) are preferentially transmitted by all routes[131]. Transmission is followed by an orderly appearance of viral and host markers of infection in the blood plasma. In the acute phase of infection, HIV-1 replicates exponentially and diversifies randomly, allowing for an unambiguous molecular identification of transmitted/founder viral genomes and a precise characterization of the population bottleneck to virus transmission. This concept of genetic bottleneck in HIV-1 transmission was suggested in 2004, when Derdeyn and coworkers found that the virus that establishes HIV infection was monophyletic and highly homogeneous despite high diversity in donor[132]. Interestingly, these authors also found that the transmitted virus tended to have shorter V1-V4 regions, which meant it had fewer glycosylation sites. A likely functional consequence of having fewer glycosylation sites is a greater exposure of the CD4 binding domain, which can benefit binding to the target cell but often results in an augmented susceptibility to antibody neutralization. The concept that sexual transmission of HIV-1 resulting in productive clinical infection arises from a single virus, was reinforced by Haaland and coworkers, highlighting the extreme bottleneck and inherent inefficiency in HIV-1transmission[133]. Nevertheless, this severe genetic bottleneck can be mitigated by the presence of inflammatory genital infections in the at risk partner, suggesting that this restriction on genetic diversity is imposed in large part by the mucosal barrier. In fact, in epidemiologically unlinked transmissions, when individuals became HIV infected by somebody other than their spouse, the frequency with which two or more virus establish infection actually occurs in up to 24%, in association with the presence of either a chronic ulcerative disease or an inflammatory genital infection[134]. Considering several published studies indicating that HIV-1 subtype could be a determinant of transmissibility and disease progression[135,136], it is conceivable that the presence of multiple subtypes, CRFs and URFs could impact in transmission bottleneck. However, Nofemela and collaborators found that the frequency of single variant infections in a genetically complex HIV epidemic was similar to the one found in cohorts of genetically restricted subtype B or C epidemics, suggesting that multiple circulating subtypes and unique recombinant forms do not have a significant impact on the transmission bottleneck[137].

Summarizing, it remains to be determined if HIV-1 transmission is largely a stochastic process whereby any reasonably fit R5 virus can be transmitted or if it is a deterministic one,

with features of transmitted/founder virus that facilitate their transmission in a biologically meaningful way. Clarification of this aspect is of extreme importance for vaccine development.

Intra-host immune selection

There are no doubts about the plasticity and mutability of HIV-1 genome. HIV-1 genetic variability at the nucleotide level can reach up to 5% within an infected individual[129], explaining HIV-1 diversity and divergence. In a data set of HIV-1sequences sampled longitudinally from an infected patient, diversity is the average pair wise genetic (nucleotide or amino acid) distance within the sequences sampled at a given time point, while divergence is their average genetic distance from the most recent common ancestor (MRCA), *i.e*. the root of the viral genealogy. Surprisingly, HIV-1 evolves considerably faster within infected individuals than it does at the population level, with positive selection dominating in the former. Intra-host HIV-1 phylogenies have a strong temporal structure[138], reflecting the successive fixation of advantageous mutations and the extinction of unfavorable ones, in a vivid example of Darwinian natural selection (see figure 8).

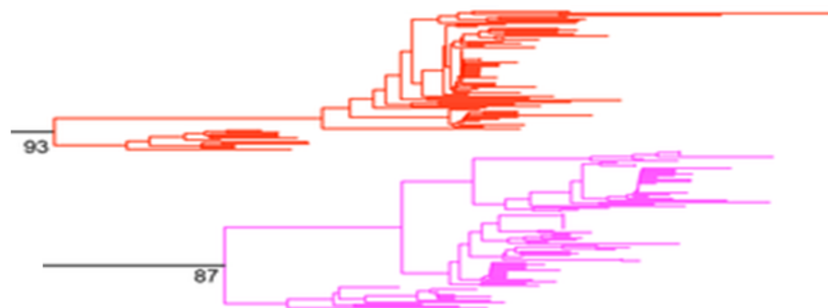

Figure 8: Patterns of intra-host evolution of HIV-1. For two mother-to-child vertically transmitted HIV-1, a maximum likelihood rooted tree was constructed. Each color represents a different pair. The scale corresponds to 5% of divergence between sequences. The bootstrap values are indicated.
In each case, intra-host HIV evolution is characterized by continual immune-driven selection, such that there is a successive selective replacement of strains through time, so that multiple lineages are able to coexist at any time point.
A major bottleneck is anticipated if the virus is transmitted to a new host, as the infecting virion will represent a single lineage. Data extracted from reference 138.

In contrast, the spatial and temporal diffusion of HIV-1 is neutral, not depending on positive selection but rather reflecting chance. Genetic diversity is strongly reduced by a bottleneck linked to transmission, leading to a homogenization of the virus during primary infection, a question that will be further detailed ahead.

Behavioral aspects of HIV-1 transmission are an important factor in determining inter-host variation. All depends on individuals, inserted in transmission chains, so a particular viral strain success at population level is based on factors extrinsic to HIV-1. Even if a virus has extraordinary fitness, conferred by highly advantageous escape mutations, by finding himself infecting an individual with low rates of partner exchange, it will fail successful establishment in that human population.

Lythgoe and Fraser proposed that the mismatch between viral diversity in a patient and the homogeneous viral infection he actually transmitted could be explained by a store and retrieve theory, consisting in preferential transmission of ancestral virus stored in long-lived latent CD4+ T lymphocytes[139]. It is known from long time that these cells effectively store virus, creating a stable archive[140]. The existence of a mechanism allowing for the preferential transmission and/or establishment of ancestral viral sequences in new hosts is therefore highly probable. There is indeed evidence that ancestral virus are, at least sometimes, preferentially transmitted[141]to new hosts. Investigating the virus in transmitting couples, has shown that virus circulating in newly infected heterosexual recipients tend to be more closely related to donor ancestral sequences than contemporary sequences circulating within the donor at the time ofinfection[142]. Moreover, the HIV-1 sequence a person acquires through heterosexual transmission tends to be similar to the sequence that will be transmitted later[141]. The question is what kind of mechanism can support these findings. Compartmentalization is ruled out, because among intravenous drug users the rate of divergence is slower than in sexual transmission[143]. Ancestral virus are likely to have an advantage in breaking the mucosal barrier, invading T cells in the intestinal mucosa, or replicating faster in the first days of infection. That is because they will be similar to the original successfully transmitted strain, without escape mutations and free from these mutations fitness cost.

Joint viral-host genome analysis is estimated to reflect genetic signals of escape and explain intra-host diversity and evolution. However, there are constraints to viral escape, which reflect RNA and protein structural requirements that may translate into loss of fitness. Although the HIV-1 genome is considered to be highly variable, 77% of amino acid positions

are conserved, whereas10% of the genome is under positive selection[144].This class of sites defines critical residues in host-pathogen interaction, whether resulting from cytotoxic T-cell (CTL) or other host-selective pressures. Nevertheless, it is remarkable how the virus tolerates such a high degree of diversity while maintaining fitness, even considering that far more genomes come to dead ends than those that generate viable descendants.

Complexity and multiplicity of HIV-1 intra-host evolution depends on multiple factors, although HIV-1mode of viral transmission can be considered a major determinant[134,145,146].Transmission of a single viral variant occurs in about 76–78%of cases of heterosexual transmission[134,147], in about 60% of cases of HIV-1 infected men who have sex with men (MSM)[146], and only in about 40% of intravenous drug users (IDU) [145]. Conversely, transmission of multiple viral variants of HIV-1 gradually increases from about 20% during heterosexual transmission to about 40% in MSM, and to 60% in IDU. As it is estimated that about 76-78% of heterosexual transmissions are initiated by a single virion and the remaining by two to five virus, the typical HIV infection is accompanied by a severe genetic bottleneck[133], as noticed above. Convincing evidence that the mucosal barrier plays an important role in reducing multiplicity of transmitted HIV-1 is demonstrated in a paper by Bar *et al*[145].

As mentioned before, HIV-1 infection evolves in a patient over time to produce a quasispecies complex of viral genomes, originated from a founder virus. During the acute phase of infection, the viral population rapidly expands and viral loads reach extremely high levels (often in excess of 1 million RNA copies per ml), chiefly at expenditure of T-lymphocytes of the mucosal associated lymphoid tissue[148]. Infected individuals often become ill with an acute syndrome resembling infectious mononucleosis. However, limited viral evolution precedes this viremia peak, meaning that in the first two to six days of infection the extent of viral diversity is probably minute, opening an opportunity to possible vaccine usefulness.

About2 weeks later, plasma RNA levels drop sharply and the acute symptoms subside. This drop in viral load may be due in part to depletion of the most susceptible target cells, as well as to subsequent innate, cell-mediated and humoral adaptive immune responses to viral antigens. The individual now enters into a period of clinical latency that can last for several years. This period is characterized by a steady state level of viral replication, the set point, which reflects a balance between viral replication and clearance by the immune system. The

viral load following acute infection is a relatively good predictor of diseaseprognosis[149], the higher, the worst. The determinants of viral load set-point in newly infected individuals include viral genetic factors and host genetic factors, representing the immune defense capacity of the recipient. Concomitant with ongoing viral replication is the progressive depletion of the levels of circulatingCD4+ T cells, eventually leading to the development of AIDS. In the absence of treatment, the virus and the immune system settle into an "evolutionary arms race". Ensuing, genome diversification broadening occurs, leading to a progressive increase in both viral divergence and diversity. Eventually, the immune system collapses and, as progression to AIDS begins, viral divergence stabilizes and viral diversitydeclines[150]. This fact has been justified as a consequence of CD4+ T cell depletion, which likely results in less effective selection pressure on the virus, as well as significant decrease in target cells capable of sustaining viral replication.

Several factors contribute for the described diversification. Mutation was previously pointed out, driven by infidelity of replication. Immune responses, conditioning escape phenomena, will be discussed now.

In chronic infection, HIV-1 is continually being selected by the host immune response, both through the activity of neutralizing antibodies and through cytotoxic T-lymphocyte responses. This immunological pressure forces the virus to mutate and generate escape mutants. At least in regions where antiretroviral therapy is currently available, a third force determines HIV-1 evolution: drug pressure.

To escape antibodies, HIV-1 changes the shape of its envelope proteins[151]. To escape cell-mediated immunity, HIV-1 switches amino acids in epitopes[152], which are short snippets of viral protein that are displayed on the surface of HIV-infected host cells to alert killer T cells. These escape mutations are identifiable because genes undergoing positive selection leave a classic genetic signature—non-synonymous mutations (mutations that change the amino acid) occur more frequently than synonymous mutations (mutations that preserve the amino acid).

A technique called single-genome amplification, interpreted in the context of a model of random virus evolution, allowed an unprecedented look at the phylogenetic structure of HIV-1 in the early hours post-infection, making possible an unambiguous molecular identification of actual transmitted/founder virus, responsible for establishing productive clinical HIV-1 infection in humans[134]. This approach showed that, during the acute phase of replication, the

viral population expands rapidly under little selective pressure, displaying a star likephylogeny[134]. Such a precise molecular identification of transmitted/founder viral genomes and their evolving progeny enabled an assessment of the earliest adaptive immune responses that shape and constrain the early replicating HIV-1 quasispecies.

The neutralizing antibody response against HIV-1 develops after resolution of acute infection, about two to five weeks after transmission[153].Neutralizing antibodies typically play a key role in controlling viral infections and contribute to the protective effect of many successful vaccines but, in the case of HIV-1 infection, this delayed reaction does not lead to clearance from the infected host. However, antibodies represent a formidable force behind immune selection, as discussed in more detail further ahead in this section.

Among the earliest selective forces acting on viral evolution, are the HIV-specific cytotoxic-T-lymphocyte (CTL) responses. Positive selection is first observed in epitopes encoding CD8+ cytotoxic T-cell escape variants and this often occurs before seroconversion[154], associated with a dramatic increase in CTL immune-adaptive mutations, particularly in highly variable proteins such as *env* and *nef*[155].This findings support the hypothesis that effective CTL responses are essential to the resolution of the acute phase of viremia[156]. However, there is a delicate balance between cytotoxic T lymphocyte escape and viral fitness, as the virus populations explore multiple adaptive pathways. Obviously, there are limits to the plasticity of individual virus, even in highly variable proteins. Studies of HIV-1 drug resistance are crucial to understand the complex interplay between the immune response and the sequence evolution of HIV-1, and to understand why sustained immune control of HIV-1 is so difficult to attain. It is well known that HIV-1 can rapidly develop drug resistance mutations when single sites in the viral genome are under intense selective pressure, such as in patients undergoing mono- and dual-drug therapy[157]. Although viral escape from antiretroviral therapy through the development of drug resistance mutations may be immediately advantageous to the virus in the presence of the drug, it also may result in a reduction of viral replicative fitness. Mutation M184V of reverse transcriptase is a classic example, offering full resistance to lamivudine and emtricitabine, on expenses of reduced replication capacity and, in the absence of the antiretroviral drug, of less fitness than wild-type virus, presumably because these mutations reduce the overall activity and processivity of HIV-1 reverse transcriptase[158]. This concept of drug-induced selection pressure can also be applied to selective pressures instructed by virus-specific $CD8^+$ T cell responses. To illustrate the complex relationship between immune-mediated selection pressure, the

emergence of viral escape mutations within targeted $CD8^+$ T cell epitopes, and the impact of these mutations on viral replicative fitness, is the example of the human histocompatibility leukocyte antigen (HLA)-B57. This HLA class I allele has been consistently associated with protection from HIV-1 disease progression[159]. Individuals expressing HLA-B57 mount a strong $CD8^+$ T cell response against a highly conserved epitope within *Gag*, called TW10, very early in acute HIV-1 infection[160]. The development of this TW10-specific $CD8^+$ T cell response is associated with the reduction of viral load by a thousand times[161]. The virus eventually evades this dominant TW10-specific $CD8^+$ T cell response by selecting for escape variants within the epitope[162]. But despite immune escape, viral replication remains well controlled in these individuals, and large numbers of individuals expressing HLA-B57 have long-term non progressive HIV-1 infection[163]. The underlying mechanism, responsible for efficient control of virus replication despite viral escape from $CD8^+$ T cell–mediated immune pressure, appears to be related to the reduced replicative fitness of virus containing escape mutations in the TW10 epitope[164]. Undeniably, the rapid in vivo reversion of these mutations back to wild-type after transmission into a new HLA-B57 negative host, and the direct impact of these mutations on viral replication in vitro, confirm the deleterious impact of escape mutations in TW10 on viral replicative fitness[162]. These studies also show that the virus tries to minimize the impact of these mutations by developing secondary compensatory mutations that can partially restore the replication defects[165]. Furthermore, a population study of HIV-1 subtypes B and C infected individuals demonstrated an inverse correlation between the proportion of mutations within $CD8^+$ T cell epitopes and viral load[166]. The replicative capacity conferred by the transmitted *gag* correlates with set point viral loads in newly infected individuals, as well as with the viral load of the transmitting source. Transmitted virus with high replicative capacity will cause more rapid CD4+ decline over the first three years, independent of viral load. This suggests that the trajectory of pathogenesis may be affected very early in infection, before adaptive immunity can respond[167].

Taken together, these studies propose a model in which the virus is either controlled by potent virus-specific T cell responses or evades antiviral immune pressures through sequence variations that reduce its fitness.

In the absence of treatment, HIV-1 intra-host genealogies tend to show an evolutionary temporal structure, where sequences from the same sampling time tend to cluster together and are the direct ancestors of sequences from the following time point, as seen in figure 8, where this observation is corroborated in sequences from pediatric patients infected vertically[138].

Some studies suggested that genetic drift, as recognized in influenza A inter-host evolution, was the shaping force of intra-host HIV evolution[168]. Divergently, studies based on sequence comparisons of synonymous and nonsynonymous substitutions, taking into account phylogenetic relationships between those sequences, showed that both purifying and positive selection play a substantial role in the continuous emergence of new variants and in the ability of the virus to evade the immuneresponse[169,170]. Comparing HIV-1 sequences isolated early at infection with those of later virus demonstrated an increase of fitness during chronic infection[171] and greater efficacy in escaping from both CD8+ T-cell responses and neutralizing antibodies[172]. Other factors can affect ladder-like temporal evolution of HIV-1 within its host. The type of class I HLA allele, as mentioned before, can slow or accelerate disease progression, and the recruitment of archived genomes from viral reservoirs increase the chances of recombination. Comparison between temporally different sequences in a subject submitted to different conditions can provide important insights to evolutionary shaping of HIV phylogeny.

Antibodies have the potential to block HIV-1 replication in at least three ways: neutralizing antibodies bind cell-free virus, preventing infection of target cells; infected cells are targeted and killed by antibody dependent cytotoxicity; and antibody dependent cell-mediated virus inhibition leads to reduced virus production, by antiviral cytokines and phagocytosis enhancement. Also antibodies exert immune pressure on the virus that will lead to escape. It is not clear which of these mechanisms will be most effective in containing HIV-1, because the relative contribution of cell-free versus cell to cell spread in HIV-1 transmission and pathogenesis is not entirely defined. HIV-1 can evade the antiviral action of antibodies due to the characteristics of its envelope spike, a target for neutralizing antibodies. This spike consists of a trimer of heterodimers, formed by two glycoproteins, gp120 and gp41[173]. The first one, gp120, is a highly glycosylated protein, organized into variable (V1-V5) and conserved (C1-C5) regions[174]. The receptor site for the CD4 molecule resides in a conserved region that is of difficult access for antibodies. The coreceptor site, both in R5 and in X4 virus, is largely inaccessible, unless CD4 binds and trigger conformational changes. On the other hand, gp41, which is well conserved, is also hidden from antibody recognition[175].

Sequence variability of HIV-1 *env*, which is a major target for neutralizing antibody responses, is concentrated in the variable loops (V1-V5). Escape from antibody responses is easily attained, as variable loops mutations do not interfere with viral fitness[176]. Every time an efficient neutralizing antibody response develops, an escape variant is selected[177,178]. Once

a new response is mounted to this variant, a new escape variant will emerge, in an endless repetition of the process. There is a striking contrast between the widespread diversity found on these epitopes versus the very low levels of diversity found in more conserved regions[179].

It has long been accepted that a key factor to protection against HIV-1 rests with anti-envelope antibodies that directly neutralize the virus. These neutralizing antibodies are extensively studied being the main focus for identifying protective HIV-1 immunogens[180]. Early in HIV-1 infection (within the first week of detectable viremia), antigen-antibody complexes are detected[181]. From here, there is a well defined sequence of antibody appearance: first, anti-gp41 (in a few days), then anti-gp-120 (in a few weeks). Both are binding antibodies, incapable of lowering viremia[181] and apparently not exerting any selective immune pressure on HIV-1 envelope[134]. Truly neutralizing antibodies, against the original infecting strain, take months to emerge, but are not able to neutralize virus from other patients[182]. Through substitution, insertions and deletions, immune escape will occur, originating less sensitive virus to antibody neutralization. This loss of sensitivity is time and evolution dependent, as contemporaneous virus are more resistant to autologous neutralization then earlier ones[183]. Another way HIV-1 evade antibody neutralization is through an evolving "glycan shield", in which shifting glycan hide epitopes, preventing access of neutralizing antibodies[178], explaining complete replacement of neutralization-sensitive virus by successive populations of resistant virus without well-defined escape mutations in the *env* region. Extensive glycosylation also acts as a shield for other vulnerable sites, particularly in V1 V2 regions[184].

As mentioned above, neutralizing antibodies tend to be highly specific, targeting variable envelope regions of HIV-1[185]. As escape mutants emerge, eventually more conserved regions will be exposed and this will lead to the induction of broadly cross-neutralizing (BCN) antibodies in some patients[186]. Why this happens in some individuals is unclear, but is probably related to the duration of infection, as years of persistent viral stimulation are needed to generate BCN antibodies[187]. Broader neutralizing antibodies are detected 3–4 years after infection, but being generated in only 20-30% of individuals, showing no efficacy *per se* in the control of HIV replication[188].

The definition of “broadly” is hard to standardize, being generally based in the capacity to neutralize heterologous virus of multiple subtypes. The relative breadth of BCN antibodies activity is probably the result of multiple antibody specificities, each one targeting a few viral

variants[189]. Theoretically, because BCN antibodies target more conserved sites, escape would be more difficult. This is supported by the occurrence of heterologous neutralization, seen more frequently among plasma with broad-cross neutralizing activity[190]. However, when one of this potent antibodies, targeting a highly conserved region of HIV-1 envelope glycoproteins, was purified, it was found that cloned virus from the same patient were resistant to it[191].

It can be therefore concluded that HIV-1 escape mechanisms exists to any given antibody. It is the polyclonal nature of each individual antibody responses that justifies the apparent absence of escape, as multiple virus variants are neutralized by different specific antibodies, targeting different regions of HIV-1 envelope glycoprotein. These antibodies are molded and maturated over time, undergoing multiple rounds of affinity/escape cycles. These antibodies present high levels of conformational changes and hypermutation[192]. However, it is not fully understood how the immune system can generate broadly-cross neutralizing antibodies.

After all the immune pressure and related constraint mechanisms, the virus that gets transmitted has been selected for survival in one immunogenetic environment. What will happen to that viral quasispecies when it enters a new host, how rapidly it escapes, and what ensue to those escape mutations that were selected in the previously chronically infected host are relevant questions. It is possible that these escape mutations imply deficiencies in the fitness of the transmitted virus, before it has a chance to revert them, therefore affecting subsequent viral load and infection progression.

In a study of 114 discordant couple transmission pairs, Goepfertand and colleagues found that multiple mutations in the *nef* gene did not seem to impair the transmission of the virus. In contrast, there was a progressive effect of escape mutations in the highly conserved HIV-1 *gag* region, particularly in p24, a key structural component of the virus. When five or more escape mutations were present, the viral load in the newly infected partner was significantly lower[193]. This suggests that the virus is placed at a disadvantage by having to escape the immune response in the infecting partner. Those partners whose immune systems are most effective at targeting *gag* may actually transfer virus that are least fit, and this may have a positive long-term effect for the newly infected partner because.

Even in a genetically homogeneous susceptible population, with shared similar HLA-I alleles, some advantageous mutations, such as those conferring CTL escape, might not appear until late in infection[194]. If these late-escape mutants do not arise until after most individuals

have transmitted the virus, natural selection will be less effective at the population level. As a consequence, HIV-1 strains might not readily adapt to the HLA haplotype distributions of the respective local populations, because some CTL-escape mutants have little opportunity for further transmission.

Moreover, although certain CTL-escape mutants can be transmitted through the population[195], it is possible that CTL-escape mutations that are passed to individuals with the 'wrong' HLA background will probably be deleterious and removed by purifying selection at the population level. Indeed, the fact that repeated individual adaptation continues to occur, indicates that the HIV population as a whole is not adapted to the host HLA distribution[196]. For a CTL restricted response to be establish in a population, it is necessary a similarity in HLA-I haplotypes in the individuals integrating that same population.

In summary, the inter-host HIV evolutionary process will not select for virus with enhanced transmissibility. Indeed, the intra-host evolutionary rate is higher and faster than between-hosts[139]. The virus probes an extraordinary variety of potential escape routes. Immune escape is very fast, with dozens of escape routes being explored and fixing advantageous epitopes, in a delicate and complex balance between viral fitness and immune escape. Furthermore, as the star burst radiation of HIV-1 subtypes in human populations can eventually become blurred by the emergence of recombinant circulating strains, recombination may also play an important role in the evolution ofHIV-1 in an individual, a topic discussed in the following section. At the population level, there is evidence that HIV-1 envelope proteins are evolving to become more resistant to neutralization over time[197], turning the virus less susceptible to immune control. On the other hand, these mutations accumulation may also lead to reduced virulence, as a consequence of least HIV-1 fitness.

Recombination

Retroviruses are unique because they are the only known group of virus that contains two copies of RNA within the virion. RNA dimerization facilitates viral diversification by promoting the utilization of both RNA copies during reverse transcription[198]. The HIV reverse transcriptase can use both copies of the co-packaged viral genome in a process named retroviral recombination. Recombination occurs much more frequently than mutation(2 to 20 events/genome/replicative cycle)[199], and is a major determinant of viral diversification. This phenomenon, considered a primitive form of sexual reproduction by mixing two viral genomes in the creation of viral offspring which will carry genetic information from both "parents", can assume one of two forms: template switching or dual infection.

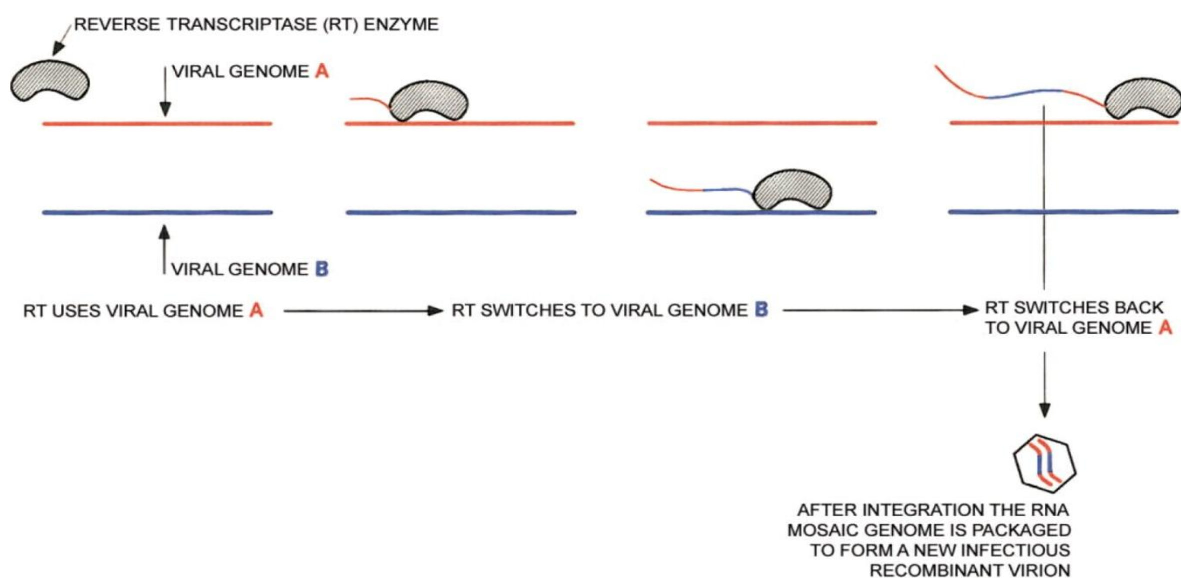

Figure 9: **Simplified scheme of template switching during reverse transcription of HIV RNA** - during HIV recombination, the viral reverse transcriptase (RT) enzyme can pass from one virus template to another, creating a new infectious virion that is a mosaic, or chimera, of the parental virions. RNA degradation is not represented. Adapted from reference 200.

Template switching consists in a switch between co-packaged RNA templates within internal regions of the genome during reverse transcription, leading to formation of recombinant DNAmolecules[200](figure 9).Template switching is an intrinsic part of the retroviral life cycle and has at least two explaining models[201]. Both rely on low binding affinity of reverse transcriptase. Retroviral RNA typically contains several small ruptures, although contained in a stable complex. When reverse transcriptase finds one of this ruptures, it jumps to the homologous sequence of the neighbor chain to keep synthesizing negative DNA strand. This chain transfer can occur whenever there is a rupture, resulting in a hybrid DNA with genetic

information from both RNA strains. Other consensual mechanism for template switching is explained by a dynamic copy choice model: during DNA synthesis, reverse transcriptase, a dual-function enzyme with a DNA polymerase domain and an RNase H domain, switches its template from one copy of the repeated sequence to the other, thereby deleting one of the copies plus any intervening sequences. Both direct repeat deletion and recombination operate by the same molecular mechanism that causes template switching, in which a steady state between the rates of polymerization and RNA degradation determines the frequency of reverse transcriptase template switching. Either way, recombination has a powerful effect on the diversification of HIV-1 by mixing mutations within the viral quasispecies.

Dual infection, defined as infection of the same individual by two or more different HIV-1 strains, is another source of genetic diversity through recombination. Dual infection can be further distinguished in co-infection, when there is an infection with different strains of HIV-1 simultaneously or at least so temporally close that an immune response to the first strain could not be mounted (before seroconversion), and superinfection, a sequential passage of virus during multiple transmission events (after seroconversion)[202].

There are evidences of dual infections with other lentiviruses, like FIV[203] and SIV[204], suggesting that recombination within retroviral genomes is a common occurrence.

Recombination due to dual infection is probably more frequent than reported. Intuitively, the existence of CRFs seems linked to dual infection and subsequent intersubtype recombination. We know that about 20% of HIV-1 infections around the world are caused by CRFs and URFs[205] and that geographical regions with multiple circulating subtypes have a significant prevalence of intersubtype recombinant infections[206]. Knowing that AIDS evolves slowly and is a persistent infection, there are high probabilities for dual infection, just depending on patients' behaviors.

When two genetically distinct strains of HIV-1 infect the same cell, recombination will certainly occur. Globally, dual infection is modeling the pandemic through CRFs. There is a gradual replacement and phasing-out of the initial predominant HIV-1 strains (the *pure* subtypes) by the increasingly-epidemiological important CRFs. Substitution of subtype B and CRF01_AE for BC recombinants in Southern China[207], and for AE/B recombinants in Malaysia[208]and subtype B and F for BF recombinants in Brazil[209,210] are examples of this evolution.

At the individual level, these events can alter virulence or pathogenicity, confer resistance to antiretroviral therapy and influence disease progression. Recombination emerges as a powerful evolutionary force, distinct from mutation, which can only be accountable for slow and steady changes, as high mutations rates will inevitably lead to degenerated viral copies. For an organism with a very high mutation rate, such as HIV-1, an efficient recombination mechanism provides theoretical advantages in attaining beneficial genetic diversity. Mutations allow for a rapid exploration of all nucleotide sequences conceivable. Once an organism is presented with a sequence encoding to peak fitness, every subsequent mutation will be unfavorable. Furthermore, unfavorable mutations accumulate more rapidly than restorative back-mutations. Genetic recombination can regenerate this lost fitness.

Besides creating and maintaining genetic diversity, recombination has several other advantages from HIV point of view[211](figure 10). It can also repair viral genomes, both genetically and physically.

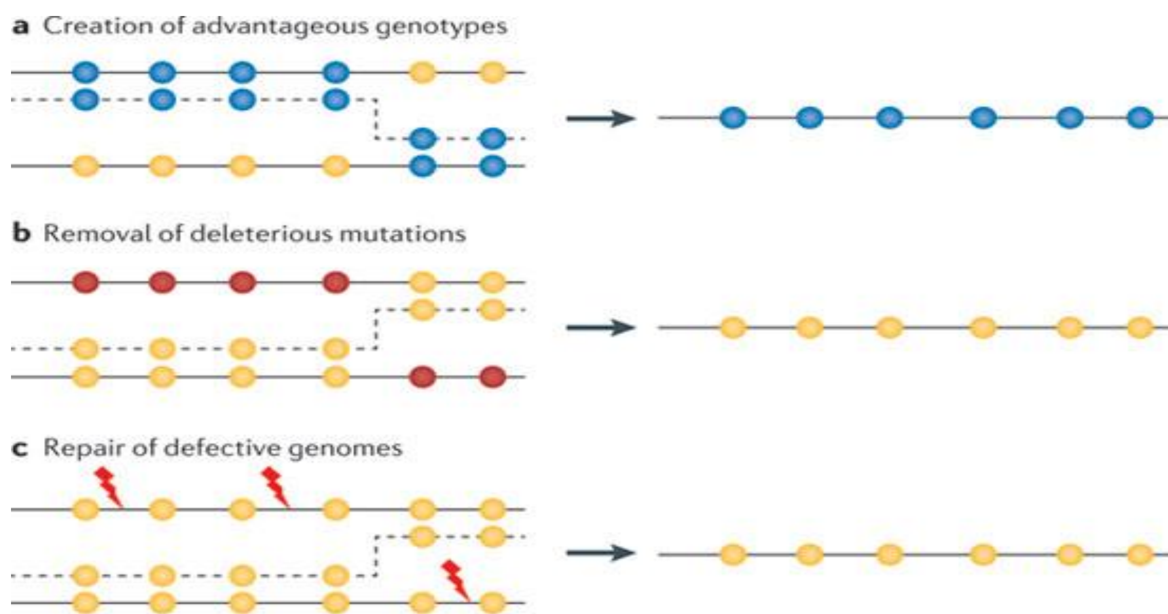

Figure 10: **Positive effects of recombination on HIV-1 genome**. **a** - Recombination can create advantageous combinations of mutations (blue circles) that increase the rate of adaptive evolution compared with mutation alone, or it can disassociate advantageous and deleterious mutations, allowing the former to spread. **b** - Recombination can remove deleterious mutations (red circles) and restore the wild-type genotype, which can lead to a selective advantage for recombination if deleterious mutations occur frequently enough. **c** - Recombination can also generate a functional genome from damaged parental molecules. Genetic damage, such as strand breaks or oxidative base modifications are represented by red lightning symbols. Yellow circles indicate wild-type loci. Adapted from reference 211.

During infection, HIV-1 is exposed to huge selective pressures by the host immune system. In order to survive, the virus must continually adapt to evade a rapidly changing immune response. HIV-1 has a much faster pace of change, but it is recombination that steer adaptation. Between HIV-1 and its human host, an evolutionary arms race develops as adaptation and counter-adaptation to each other weaponry leads to evolution through natural selection. Recombination also creates new diversity by mixing pre-existing mutations within a population, generating complex combinations of mutations that would otherwise have to arise sequentially in an asexual mode of reproduction, in a much slower and unpredictable pace[212]. Recombination allows different areas of HIV-1 genome to evolve separately by breaking linkages between mutations. For example, if a beneficial mutation is physically linked to a deleterious one, recombination ensures the first can be maintained and the second can be discarded by breaking their association. Another way recombination can favor HIV-1 evolution is by eliminating competition among beneficial mutations. In an asexual mode of reproduction, two beneficial mutations will arise as individual mutations in separate lineages. Before fixation in a population they will compete with each other. In a sexual mode of reproduction, both mutations can be recombined into the same lineage.

In the absence of recombination, organisms tend to accumulate deleterious mutations at each replication cycle, as new mutations are more probable than reversions and new mutations are more likely to have harmful than advantageous effects. As a result, with each replication cycle, the accumulation of mutations decreases the fitness of the virus until eventually it will disappear. Recombination can bypass this by recreating mutation free individuals from a population of mutants[213].

Recombination is not limited by sequence similarity. It has been demonstrated to occur inter-group (between HIV-1 group M and O)[214], and inter- and intra-strain within HIV-1 group M[215].Recombination has become a common occurrence among different HIV-1 strains, and the intersubtype recombination is the most frequently observed, although intra-subtype recombination is also possible[216].

However, the impact, at the level of individual patients, of HIV-1 recombination is not entirely defined. The shuffling of polymorphisms found in distinct viral quasispecies certainly play a role in generating viral diversity[129]. Recombination is much more efficient and rapid than mutation in making HIV-1 evolve. The ability to maintain extensive diversity may be extremely important for viral pathogenesis, because it ensures the availability of viral

quasispecies able to escape changes in the selective pressures exerted by the immune response or by antiretroviral therapy. For HIV-1, recombination means increasing its potential evolutionary success.

In addition to generating diversity, viral recombination could also be useful in preserving already existing diversity. In the course of infection in an individual patient, diversity can be threatened by evolutionary bottlenecks. Although reflective of an initial selective advantage, the emergence of a genetically homogeneous viral population could prove disadvantageous, because the descendants might later become susceptible to elimination by an immune response focused against shared antigenic determinants. An evolutionary bottleneck resulting from the emergence of a unique viral species with high resistance to antiretroviral agents creates such a risk. If, however, during or subsequent to emergence, such strains could recombine at high rates with preexisting strains, much viral diversity could be maintained in regions outside those responsible for the bottleneck.

Geographic diversity

Over the past 30 years, HIV-1 infection/AIDS has evolved into an increasingly heterogeneous pandemic composed of multiple localized epidemics. Since its emergence, as a result of the high error rate of reverse transcriptase, recombination and selective pressure exerted by human immune system, HIV-1 group M has diverged into clades or subtypes as well as numerous circulating and unique recombinant forms. As referred previously, HIV-1 is native from Central Africa, most probably from Democratic Republic of the Congo. The DRC is the most diverse set of HIV-1 M group sequences currently known [41]. Analysis of this set of Central African virus suggest a period of slow expansion early in the epidemic, with more rapid expansion in recent decades, consistent with the time estimates that suggest that the virus was present in the human population for many decades prior to AIDS being detected and defined. An early ''starburst'' expansion of HIV-1 variants was hypothesized, leading to the different subtypes, corresponding to a period of rapid expansion in Central Africa. Accordingly to this model, the establishment of multiple epidemics around the world, started with a phase of regional expansion, with posterior emergence of geographically specific subtypes, in part due to specific modes and routes of transmission.

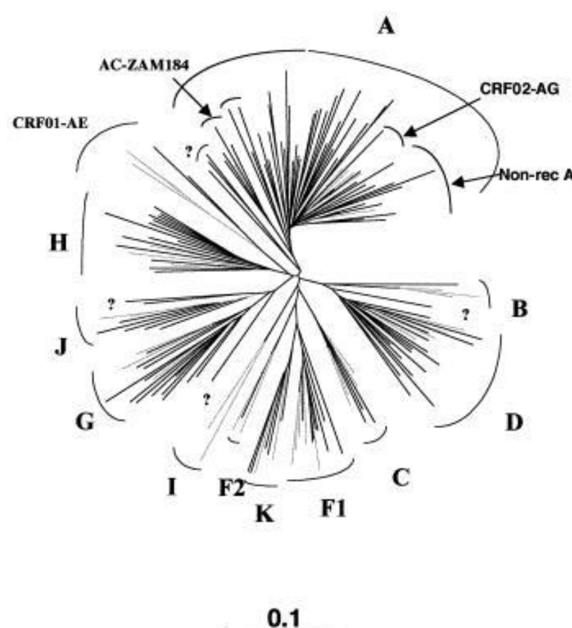

Figure 11: HIV-1 diversity in Democratic Republic of Congo. Phylogenetic tree based on the *env* V3-V5 region of 197 new HIV-1 isolates, from Kinshasa (DRC), showing a very high degree of divergence within each subtype. Reference strains (lighter in the picture) were from strains representing each known subtype. Data extracted from reference 41.

In figure 11, we perceive that group M strains from DRC have a less organized substructure, with nodes near pre-subtype branches, and more intra-subtype divergence when comparing to the global group M phylogenetic tree (see figure 7a).

This difference between trees allows inference of chance exportations of strains from DRC to new susceptible, geographically distant, populations. Intensive viral collection from West and Central Africa has now uncovered strains that fall between the previously described subtypes[41]. This indicates that these regions of Africa were the source of the strains that ignited successful epidemics in other locations, in Africa and beyond, and that the subtype structure of the HIV-1 tree can reflect, to a large extent, sampling bias. For example, most HIV-1 strains isolated in North America and Europe fall into subtype B, and their relative similarity reflects their recent common origin from a founder in, or from, Africa. This is called a founder effect, consistent with the global phylogenetic tree of HIV-1 group M (see figure 7a), depicting long branch lengths within subtypes that coalesce near the ancestral node of the subtype and long pre-subtype branches that coalesce near the root of the entire tree: an exported virus finds a new cluster in a region distinct from its ancestral and start to diversificate from that moment onwards. The global expansion of relatively few viral subtypes is indicative of clustering at a global level.

The commencement of radial evolution of group M viruses in multiple subtypes is likely due to adaptation and expansions in the first human hosts, in Central Africa. On the other hand, the host genetic background, host restrictive factors, transmission bottlenecks, social/behavioral and environmental limitations, founder effects and other viral factors could have contributed to variable geographic spread through the human population, throughout time (figure 12). That contributes to explain why specific HIV subtypes tend to be linked to particular geographic regions. This uneven distribution is in favor of the founder effect theory: a single introduction in a susceptible population followed by a rapid spread. The founder effect associated with accidental exportation of a given strain from the region of the initial epidemic, followed by subsequent local epidemics in previously non-infected regions, gave rise to the current global subtype's distribution[129]. This is a stochastic or pure chance explanation. In current times, the impact of this founder event tends to be minimized due to multiple introductions of different subtypes and subsequent co-circulation in transmission groups, with higher probabilities of dual infections and recombination. Some variables that may have also shaped global subtype distribution include the influence of genetic patrimony of human populations, the relative fitness of each subtype and some kind of preferential

transmission mode according to subtype specificities. Evolutionary epidemiology contributes to the understanding of the establishment of an HIV-1 strain in a susceptible population: strains may differ by virulence, transmission or recovery rates. Hosts will differ in their individual susceptibility, their response to infection and in their capability of transmission (primarily risk behavior, in HIV infection). A model which comprises all these variables is astonishing difficult to obtain, so studies use simplified models, generally not taking into account within-host heterogeneity and compartmentalization, population density and mobility and still unknown viral parameters. Notwithstanding, with current understanding of phylogeny, some emblematic geographic patterns could be elucidated. First, a global picture will be described.

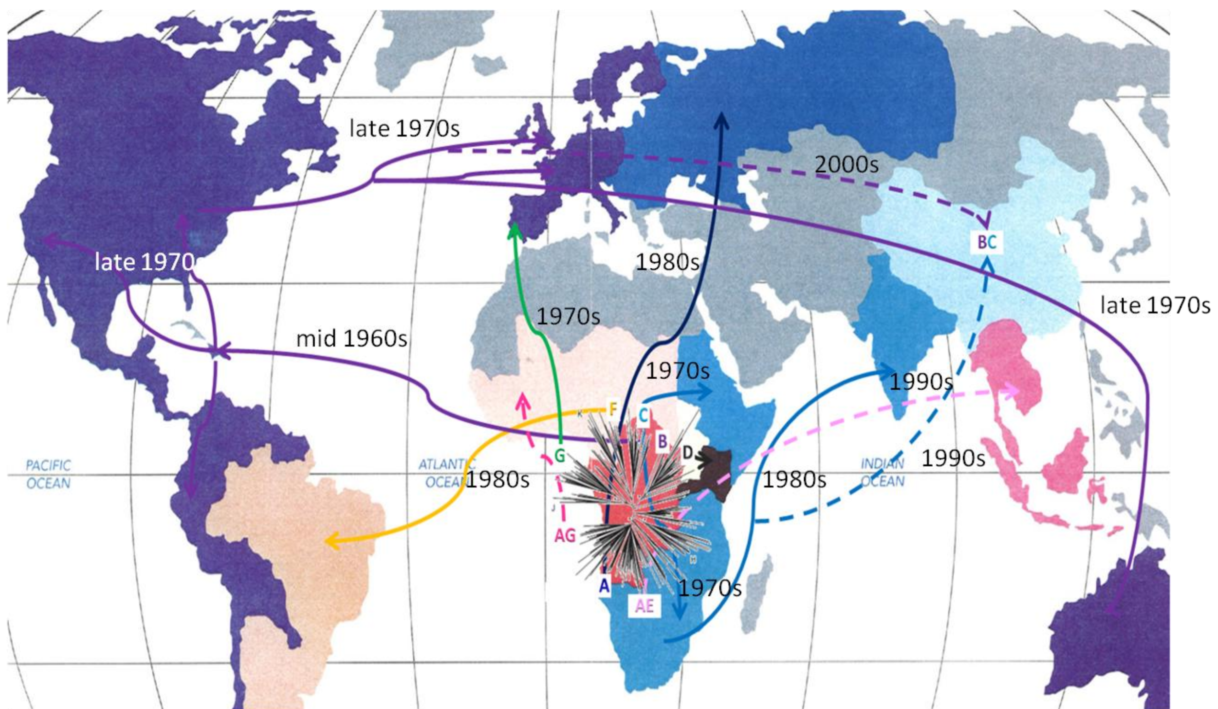

Figure 12: Potential routes and timings of migration of HIV-1 group M subtypes A, B, C, D, G and F. In dashed arrows potential migration routes for CRF01_AE and CRF02_AG. Recombinant BC in China is represented as resultant from spread of B and C subtype. Based on reference 217.

As mentioned before, HIV-1 strains differ enormously in terms of global prevalence. Six strains account for the majority of HIV-1 infections: subtypes A, B, C, D and two circulating recombinant forms, CRF01_AE and CRF02_AG. In 2007, global proportions of HIV-1 subtypes and recombinants have shown that subtype C is the most successful of HIV-1 group M lineages, accounting for more than 50% of world's infections followed by subtypes A (12%), subtype B (10%), subtype G (5%) and subtype D (3%). Subtypes F, H, J and K

overall account for less than 1% of all HIV-1 infections. CRF01_AE and CRF02_AG together are responsible for 10% of infections while CRF03_AB is responsible for 0.1% of global infections with the other recombinants contributing to the remaining 8% of all HIV-1 infections[217].

At a regional level, several trends were noticed early in the pandemic. For example, intravenous drug use in Southeast Asia in the mid-1980s and in Eastern Europe and Russia during the early 1990s led to the rapid spread of CRF01_AE and of subtype A, respectively[218,219]. A similar expansion of subtype B HIV-1 transmission occurred among MSM in North America and Europe in the early 1980s. However, HIV- 1 subtype C (the

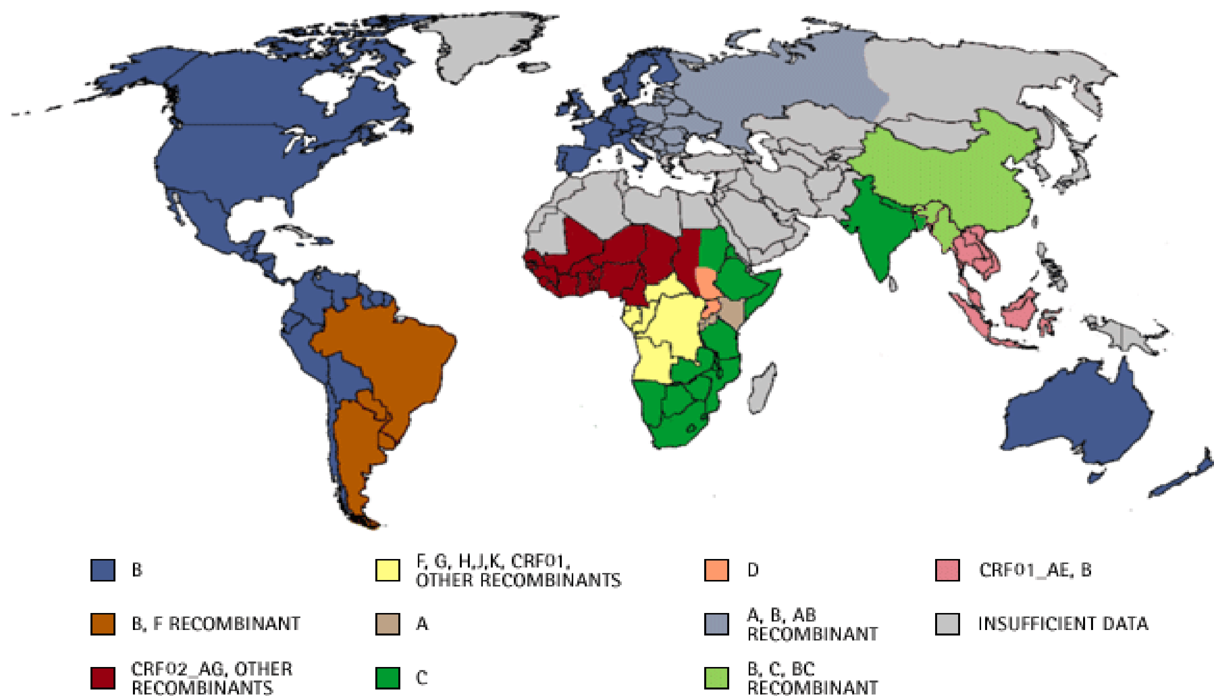

Figure 13: Patterns of HIV-1 subtype's distribution. In the Americas, Australia and Western Europe, subtype B predominates everywhere but eastern South America, where there is a substantial proportion of BF recombinants in addition to subtype B. In Eastern Europe, subtypes A, B, and AB recombinant strains dominate the epidemic. Three different patterns have been observed in Asia: subtype C, a mixture of B, C, and BC recombinants, and a mixture of subtype B and CRF01_AE.Africa shows the greatest diversity. Subtype C dominates the South and East, except for significant foci of subtypes A and D, as shown. West and West Central Africa harbor mainly CRF02_AG, alongside a complex array of other recombinants each present at a low frequency. The most complex epidemic is in Central Africa, where rare subtypes and a wide variety of recombinant forms circulate without any discernible predominant strain. This map demarcates boundaries more distinctly than they exist in reality. Gray representation of Northern Africa, the Middle East, and Central Asia is essentially due to lack of data on HIV-1 subtypes. Reprinted with permission from International AIDS Vaccine Initiative Report.

dominant subtype in the world) appears to have slowly emerged globally over the past 10 to 15 years as a consequence of multiple introductions[218]. In recent years, a substantial increase of recombinant forms has been observed as a consequence of the increased genetic complexity of the global epidemic[217,220-224]. However, the prevalence of recombinant forms (estimated to be around the 20%[217]) is still underestimated. Indeed, genetic complexity is not always detected, and this is mainly due to the subtyping of only one genetic region and not of the full genome. Consequently, specimens previously considered "pure" variants may be classified as recombinants when additional viral genes are analyzed.

Condensing the available information about HIV-1 spatial distribution, ten different epidemic patterns can be identified, as indicated by different colors in figure 13. Interestingly, the greatest diversity of subtypes and recombinants is present in DRC, Central African Republic, Gabon, Angola and Chad, where only about 5% of the world's infected individuals live[205]. Thus, a general observation is that a higher diversity of subtypes is associated with relatively slower epidemics whilst explosive epidemics generally have only one prevalent subtype, a consequence of the founder effect. In Africa, starting in DRC, HIV-1 divergence decreases to west and to south and prevalence increases as we move towards south.

Four different examples of geographically localized epidemics may allow to understand how HIV-1 become global but diverse.

Subtype B: start of an epidemic in a high-income setting

HIV-1 group M subtype B is the most geographically widespread HIV variant, ranging from Americas to Europe, Asia and Australia[218], even being a rare subtype in Central Africa, where it was presumably born. This was the first subtype discovered, because it was the cause of the first AIDS cases, reported in the United States in 1981.

Soon after AIDS identification, evidence of a high prevalence among Haitian immigrants in the United States[4] stimulated conjecture that Haiti may have been the origin of the newly identified syndrome, a claim that was not confirmed, as previously mentioned. However, Haitian HIV-1 sequences proved to be of major importance as they tend to occupy basal positions on the subtype B phylogeny[118]. It is now known that HIV-1 M subtype B started to diverge in human beings around 1950[118] and relaxed molecular clock methods allowed to mark HIV-1 subtype B entry in Haiti in 1966 and in United States soon after, in 1969[7]. Apparently, the return of one of the many Haitians who worked in the newly independent Congo gave the virus a chance to migrate to America in the 1960s[225]. Upon arrival, the virus

spread in Haiti, seeded a clade in Trinidad and Tobago[7] and ultimately made its way into the United States via emigration (eventually also to other countries in the Caribbean and in South America), where an ancestral virus of all subtype B infections across the world crossed from the Haitian community to the non-Haitian population. This means that a single introduction could begin a remarkable epidemic, although there is no evidence that this ancestor possessed any selective advantage over other strains; this event may simply reflect chance colonization instead of competitive selection. Why there is no evidence of more Haiti-to-USA successful outbreaks, despite presumably frequent movement of the virus because of its rising prevalence and the thriving sex industry linking those two countries, it is a mystery. In 1978, there was an estimated prevalence of 5% of HIV infection among MSM in New York and in S. Francisco[226], revealing several years of successful spread of the virus in those communities. It is plausible that HIV-1 was slowly spreading in the heterosexual population for an extensive period before entering the highest-risk homosexual population, where it found best conditions to spread explosively enough to finally be noticed: the sexual liberation of the 1960s and 1970s combined with large groups of sexually active gay men with a high turnover of partners in tolerant cities like New York and San Francisco. Recently, employing a high number ofvirus sequences from South America countries, Junqueira and colleagues add further clues to the history of HIV-1 subtype B in Americas, stating that part of the epidemic in South America derived directly from the Caribbean epidemic[227]. These authors propose an epidemiologic link between Latin America and the United States, based in waves of human migrations, which could have also contributed to the spread of subtype B in North America.

India: a straight C connection

At the beginning of 1986, India had no reported cases of HIV infections. Later in that year, the first patients were diagnosed among sex workers[228]. It was noted that contact with foreign visitors was almost constant in those first patients. Steadily, the number of patients started to rise, first in Indians returning from African countries, like Uganda and Zambia, later with a pattern of spreading through rail tracks and particularly through roads and highways. Truck drivers were detected to be a major high risk group because of promiscuous sexual behavior, engaging in unsafe sex along major Indian roads[229]. Most of the sex workers were married and entertained up to 6 customers per day[230]. Thus, the epidemic shifted from high risk truck drivers to low risk house wives. HIV-1 spread was fueled in those initial times by blood borne infection, as blood products used in transfusions were not consistently screened until

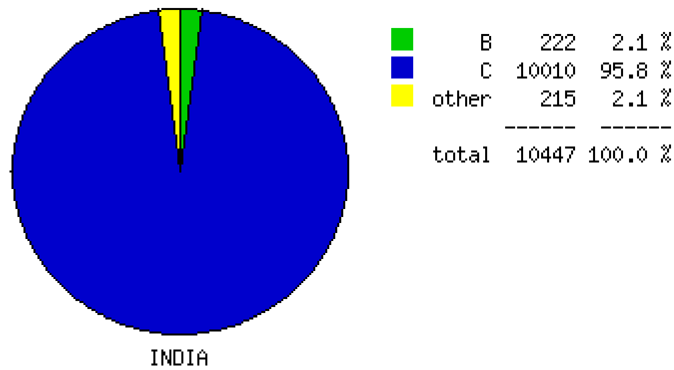

Figure 14: HIV-1 subtypes distribution in India in 2014. Extracted from http://www.hiv.lanl.gov.

1997[231]. When efforts were made to ascertain virus subtypes prevalent in India, results showed that 82% of the patients harbored a strain closely related to a subtype C South African strain[232]. A link between South Africa and India could be established on account of methaqualone. Methaqualone was, in 1980s and 1990s, a recreational drug, widely used in South Africa, where it was known as mandrax. In India, many drug manufacturer companies shifted to mandrax production, much more profitable than acetaminophen, for example. Huge quantities of mandrax were estimated to be smuggled to South Africa. It is possible that, on return route, HIV-1 was brought back to coastal areas of India and spread to the today's overwhelming subtype C dominance (figure 16).

Maximum founder effect: the case of former Soviet Union

In early 1990s, Soviet Union seemed relatively preserved from HIV-1 pandemic. There were reports of a nosocomial outbreak among children[233], sporadic MSM infected with subtype B virus[234] and several patients infected with diverse clades as a consequence of heterosexual contacts with people of African origin[233]. In 1995, a dramatic change occurred: starting in the Ukrainian city of Odessa, successive outbreaks of HIV-1 infections among IDUs spread widely throughout all former Soviet Union countries[235]. This epidemic, recognized in 2004 as the fastest-growing in the world, originated 250000 new infections among IDUs in 2002 alone[236]. The epidemic was largely dominated by a subtype A variant of monophyletic origin, distinct from African clades[237]. Nowadays, a prevalence of 27% of HIV-1 infections among IDUs is estimated in Eastern Europe, one of the highest in the world[33]. The breakup of the Soviet Union with subsequent economic decline, combined with a novel travel freedom, had devastating consequences regarding HIV-1 transmission. The virus found itself among a very large population of IDUs adopting unsafe practices. The unique genetic features of this subtype A virus allowed easy tracking, from Ukraine to St. Petersburg and Estonia and later to Moscow and Siberia as well as almost all former Soviet Union republics[238]. Unfortunately, this clade origin was never elucidated.

Europe in present times: melting pot increases HIV-1 genetic diversity

The epidemic spread in Western Europe began in early 1980s, as an extension of the epidemic among MSM in the United States. This is epidemiologically inferred, as most AIDS cases were diagnosed in homosexual men with recent travels to the United States[239]. As expected, almost all patients were infected with subtype B viruses[240], like in the United States and with phylogenetic trees supporting multiple independent introductions in Europe[241]. Eventually, this subtype B epidemic spread to injection drug users. In addition, African clades were introduced in some countries, either by African immigrants or by individuals who had travelled to Africa and returned. These patients were infected almost exclusively by heterosexual contact[242,243]. Non-B subtypes remained largely confined to people linked to Sub-Saharan Africa and so subtype B dominates HIV-1 epidemic in Western Europe. In some countries, particularly those who have large immigrant populations or strong relations with Sub-Saharan Africa, like France, United Kingdom or Portugal, increases in non-B infections were noted in the last 7-8 years[244]. Part of these infections is attributable to cases epidemiologically linked to Africa, but some clades circulate among local populations. In Portugal, subtype G is endemic, both in heterosexual and IDUs populations[245]. In Switzerland, CRF11_cpx infects half of the IDUs[246]. In Finland, CRF01_AE, a variant of southeast Asian origin, caused an outbreak among IDUs[247]. In Greece, a recently introduced subtype A virus, related to former Soviet Union subtype, is rapidly substituting endemic subtype B[248]. This last example from Greece is paramount of other reports stating that clades circulating at low prevalence became prominent in a few years. It seems like chance introductions into unexposed populations are provoking shifts in clades distribution. All these findings indicate a trend toward an increasing genetic complexity in Europe.

Current status

From all over the world, there are reports of fast growing localized epidemics, usually associated to a single behavior risk: MSM in China[249,250], IDUs in India[251], MSM in Thailand[252]. In all of them, a mechanism is common: a new virus strain is introduced via group risk behavior, usually IDU or MSM. The new HIV-1 strains rapidly emerge and recombine with the "native" ones. When this happens in heavily populated regions, like India or China, epidemics rapidly shift from MSM or IDU to the rest of the population, increasing the spread potential.

In a previous section, intra-host evolution was discussed. But, in an infectious disease, the implications of the evolutionary processes go far beyond the single host level and tend to

affect an entire population. Actually, bottlenecks that occur at transmission may "reset" viral evolution between hosts, with the result being that within-host selection may have little effect on the long-term evolution of HIV-1at the population level. However, it was demonstrated that HIV-1 has already reached the optimal virulence to maximize transmission between hosts. That could only be achieved if there is a heritable variation in virulence[253,254].

Historically, many more viruses undoubtedly emerged from Africa; however, these viruses did not establish themselves within transmission networks, or were of lower fitness which limited their dispersal. Founder effects can probably account for most of the current dominant epidemics whereby a single chance introduction of an ancestral virus resulted in major and successful spread. This is clearly illustrated by the success of subtype B in North America and Western Europe and by the rapid expansion of subtype C in India or subtype A in Eastern Europe.

Although the distribution of HIV-1 subtypes is relatively more or less localized, there is a tendency toward progressive dispersion of all subtypes in different geographical areas and toward new recombinant subtypes. As a consequence, clustering will become eventually weaker and subtype classification will lose importance, at the same time as recombinant forms will become more prevalent and ubiquitous.

By the end of 2012, 38 000 to 62 000 adults and children were estimated to be living with HIV/AIDS in Portugal, an increase from the 26000 to 45000 people believed to be infected in 2001[33].

These figures show that the prevalence of HIV/AIDS in Portugal is one of the highest in Europe (around 0.6%). Concerning transmission routes, after an initial period dominated by homosexual transmission of HIV-1, a shift towards transmission through heterosexual contacts and intravenous drug use occurred. Today, as depicted in figure 15, with a decrease of new infections in IDUs, heterosexual contact is now the main route of HIV-1 transmission in Portugal[255].

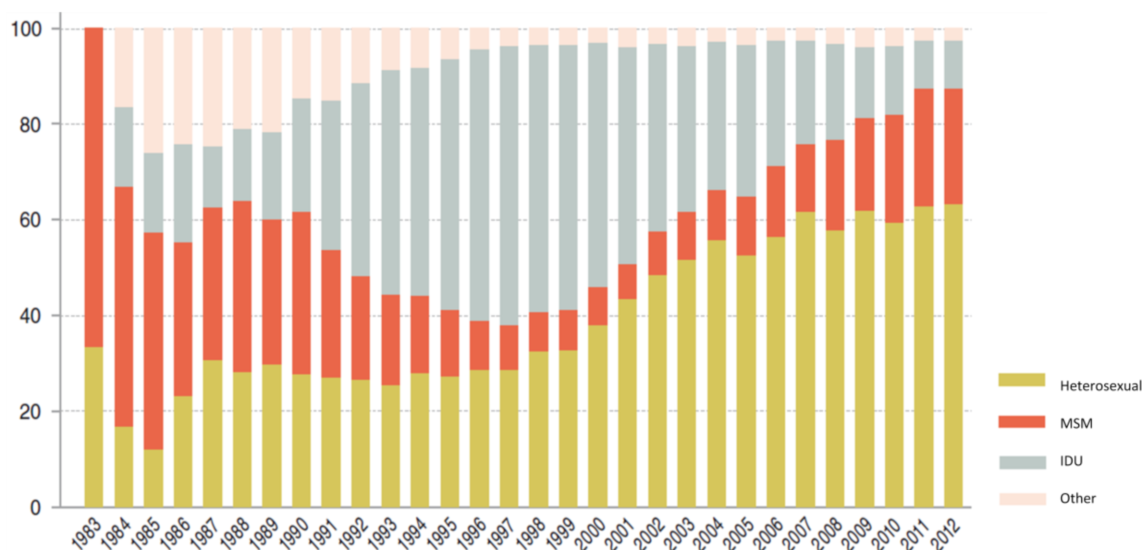

Figure 15: Total Portuguese HIV infections (1983-2013). Percentage distribution, according to transmission route and year of diagnosis (adapted from reference 255)

In the European Union, the highest proportion of new HIV diagnoses in 2012 was reported among MSM (40%, 11,877 cases), followed by heterosexual transmission (34%, 9,944 cases). The latter includes 12% (3,474 cases) of heterosexually-acquired cases originating from sub-Saharan African countries with generalized epidemics. People who inject drugs accounted for 6% (1,785 cases) of all HIV infections[256].

Portugal has a quite unique HIV-1 subtype distribution. The current HIV-1 epidemic in Portugal is caused by multiple subtypes, with predominance of subtypes B and G (together they are responsible for almost three quarters of infections – see figure 16). The high

prevalence of these two subtypes has promoted the appearance of several unique B/G recombinants, and one of them is believed to be native from Portugal, CRF14_BG[257].This CRF emerged in the early 1990s, spread little after to Galicia, Spain, and then to the rest of Europe, carried mainly by intravenous drug users. Its prevalence is decreasing though, probably because of difficult transmission, consequence of its predisposition to assume X4 tropism, and a tendency to recombine, as suggested by CRF14_BG like subgenomic fragments existence in newly diagnosed individuals[258].

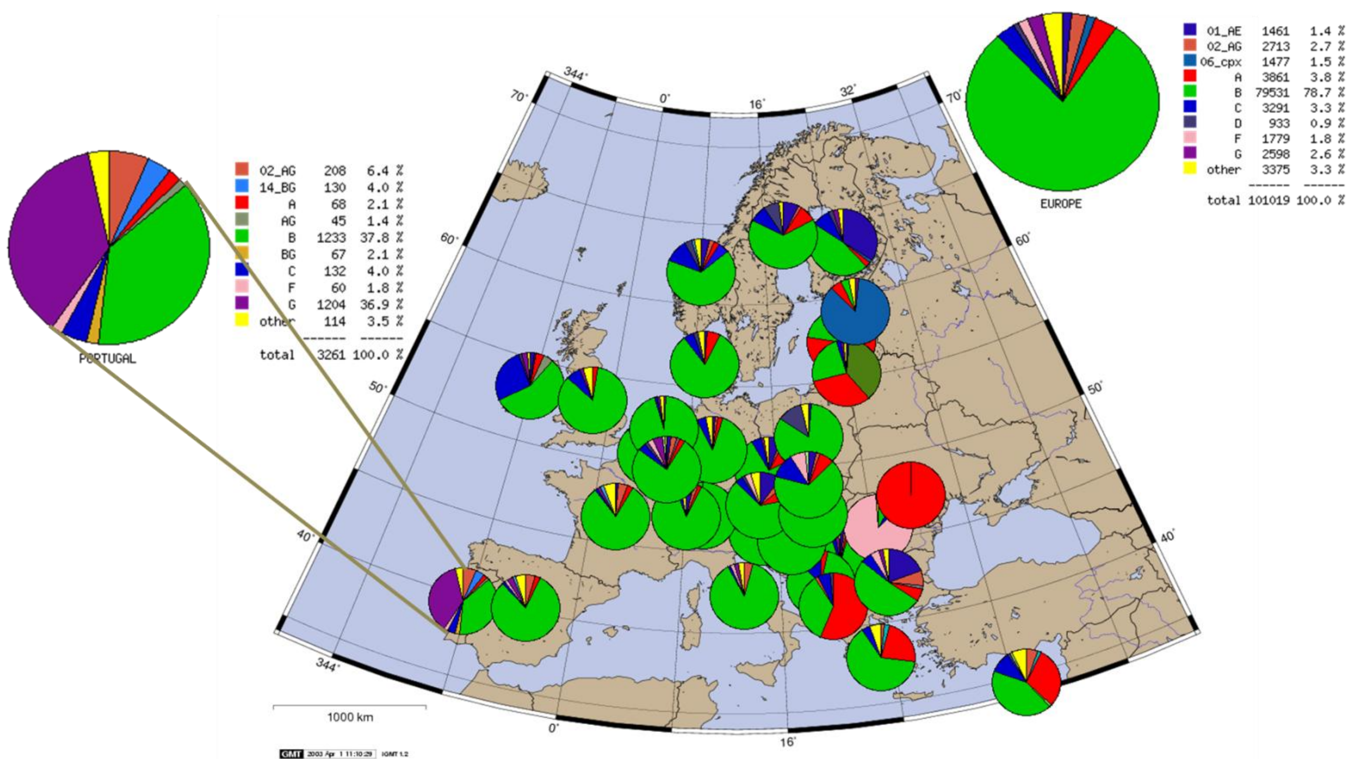

Figure 16: HIV-1 subtypes distribution in Europe in 2014, highlighting Portugal. Extracted from http://www.hiv.lanl.gov.

Classically, in Western Europe, subtype B is largely prevalent. However, several studies are proving a disturbing evolution. In Madrid, an usual destination and transit point for immigrants from Africa and South America, 71.4% of HIV-1 infected immigrants were carriers of non-B strains[259]. In Denmark, in a study designed to characterize the phylogeny of new infections, only 12% were of subtype B and almost all non-B subtypes were detected in people traced to countries with high-prevalence for that specific strain [260]. In Belgium, a total rise in non-B subtypes from 0% in 1983 to 47% in 2001 was detected among new infections,

and being a patient of African origin had an odds ratio of 6.93[242]. Differences in transmission modes seem to occur between B and non-B subtypes in Western Europe. In fact, in Finland and in the United Kingdom it was shown that most non-B subtype's infections were transmitted heterosexually in direct or indirect contact with African, Caribbean or Asiatic individuals, contrasting with subtype B infections, transmitted mainly among MSM from native origin[261,262].

This emerging broad HIV-1 diversity in Europe is mainly caused by population movements, such as migration and travelling. This evolution is amplified by sexual contacts with individuals from countries where those variants are highly prevalent, principally from Africa and Asia. In Portugal, that effect must have occurred early in the epidemic, as demonstrated by the singular genetic multiplicity of HIV-1 strains. Given the close interactions between Portugal and its former colonies in Africa, studies were performed to evaluate the genetic diversity in those countries. However, a characteristic that differentiates the HIV-1 epidemics in these countries and Portugal is the main viral transmission routes: in Mozambique, Angola, Guinea-Bissau and Cape Verde HIV-1 is almost exclusively transmitted through heterosexual contact[263], whereas in Portugal transmission by intravenous drug use accounts for 39.9% of the HIV infections[255].

In 2000, HIV-1 infection was relatively new in Guinea-Bissau. Indeed, most of HIV infections were caused by HIV-2, endemic in that region. Almost all patients infected with HIV-1 carried some variety of subtype A recombination[264] (figure 17). Such a genetically restricted epidemic can be explained by events of introduction, accidentally trafficked into Guinea-Bissau. Due to unstable economical and political conditions persisting for the last decades, population movements led to recombination and dispersion of other HIV-1 subtypes, with the current distribution represented in figure 19, noting CRF02_AG, the main responsible for regional epidemic in West Africa, as the predominant circulating strain.

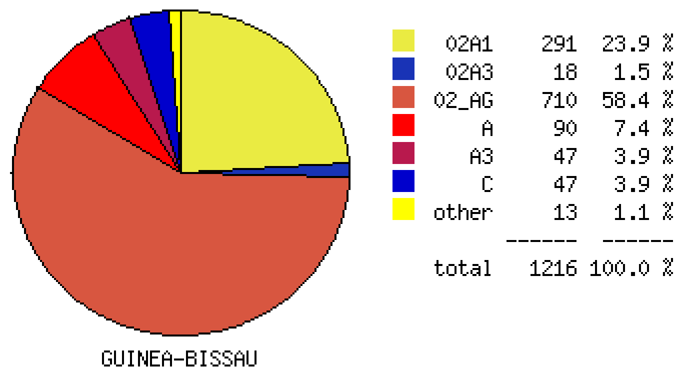

Figure 17: HIV-1 subtypes distribution in Guinea-Bissau in 2014. Extracted from http://www.hiv.lanl.gov.

Cape Verde is an archipelago, offshore West African coast. Being a small insular country, it was rather isolated and partially protected from mainland viruses until early 1980s. Since then, and due to its strategic location as a hub between Africa, Europe and America, international tourism and workers migrations allowed the introduction of different HIV strains. According to the study by Oliveira et al, in 2012 HIV-1 subtype G was the most prevailing[265]. Consistently, G strains are highly divergent, as a consequence of multiple introductions. The authors also found that those G variants were imported mostly from Angola and Portugal, where highly divergent subtype G strains prevail. Of particular interest, 30% of the viruses were recombinant, with 63% of them carrying at least one untypable genomic fragment[265]. A more recent and comprehensive study maintained subtype G as the more prevalent in Cape Verde, but report that CRF02_AG is present in 30.6% of patients and that the intersubtype recombinant viruses comprise 46.1% of all HIV-1 samples analysed[266]. Although present in lower frequency, subtype F1 is prevalent in one of the nine Cape Verde islands, possibly originating from a single introduction. Absence of pure subtype A is remarkable, as it is one of the most prevalent subtypes in Guinea-Bissau. Like in Guinea-Bissau, country with great affinities with Cape Verde, HIV-2 represents an important proportion of infections (around 25%)[263] by HIV.

People displacements to and from Cape Verde gave HIV-1 an opportunity to enter the country and to diversify. Portugal is one of the main destinies for Capeverdeans emigrants and is a source of seasonal workers, businessmen and tourists. It is quite plausible that HIV-1 moves along in both ways with people flow.

Mozambique is a Southeast African country, a region where HIV-1 infection prevalence is the highest worldwide, ranging from 5.1% in Tanzania to 26.5% in Swaziland[3]. Mozambique itself has the world 8th higher prevalence (11.1% in 2012)[33]. Notably, in the entire region, subtype C is dominant and so it happens in Mozambique, with Mozambican C sequences widely spread across multiple clusters containing C sequences from neighboring countries, moving from west to east. This is strong evidence against a specific Mozambican C subtype[267].

HIV-1 epidemic in Mozambique probably started in early 1980s (first patient diagnosed in 1986) and since then it has spread at a very high rate, especially among women[268]. From a prevalence of 1.9% in 1993, the Mozambican epidemic exploded to 11.5% in 2007 and stabilized since then[33]. This epidemic has some potential to evolve quickly also in genetic complexity, due to recent identification in Maputo, the capital city, of clusters of CRF37_cpx, originally found in Cameroon, and of sub-subtype A3, originally found in Senegal[267]. In the same study, Bártolo and colleagues revealed subtype G strains closely related to Portuguese ones, however with lack of statistical support to conclude in which way they circulated[267].

This explosive growth in Mozambique has the usual culprit, people displacement, mainly caused by a long civil war, leading to a movement of population towards major cities.

Of all former Portuguese African colonies, Angola is the closest to DRC, presumed the original region of AIDS emergence. This geographic proximity explains the extremely high genetic diversity among circulating HIV-1 strains[269]. According to the study by Bártolo and colleagues, only 53% of viruses were pure subtypes, subtype A and its sub-subtypes

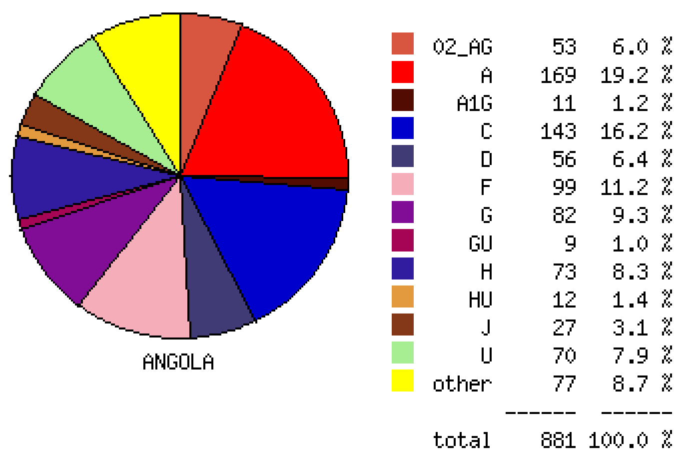

Figure 18: HIV-1 subtypes distribution in Angola in 2014. Extracted from http://www.hiv.lanl.gov.

predominating over the remaining. A significant proportion of subtype C was also reported by the same authors (figure 18).

The Angolan recombinants are complex and account for 47% of total infections. Moreover about a third of these are second generation recombinants[270]. This typically happens in old epidemics, as there is a direct association between HIV-1 genetic diversity and infection time[41]. The independence war urged population mobility between Angola and neighboring countries between 1961 and 1974, making likely that an important number of Angolans became infected with diverse HIV-1 subtypes before eventually returned to their own country.

Spreading of HIV Angolan strains to Portugal is also believed to have started as early as 1961, due to dislocations of soldiers caused by the same independence war. Later, in 1975, nearly 320 000 residents moved to Portugal in few months, soon after independence of Angola, according to National Statistics Institute (INE). Two thirds were under 40 years old. Such a formidable inflow of sexually active people certainly contributed to the introduction of non-B strains of HIV-1 in Portugal. Nowadays, HIV-1 prevalence in Angola (3.7% of adult population) is low when compared with Southern Africa yet similar to DRC. The prevalence was however higher in 1989, reaching 6.1%[271].

The social and economic connections of Portugal with former colonies is certainly the reason for the unique pattern of HIV-1 subtype distribution in Portugal, namely as compared to other European countries.

In summary, the profile of HIV epidemic is changing in Europe, with a growing incidence of non-B subtypes infections. The unique scenery of HIV-1 infection in Portugal offers a wealth of epidemiological lessons, almost as a foretaste of what to anticipate in a near future for the rest of Western Europe.

Assessing HIV-1 diversity implications

Transmission and disease progression

As mentioned above, broad associations between HIV-1 subtypes and transmission modes have long been reported. For instance, MSM are associated with subtype B in United States and Europe[7,241]. As there is no evidence that subtype B is poorly transmitted via other routes, probably the explanation resides in an important role played by founder effects and social transmission networks. In the same way, the apparent predilection for heterosexual transmission by non-B subtypes in Europe and North America may be linked to immigrants from Africa and Asia[272]. We know that in South Africa and in India the heterosexual epidemics is almost exclusively caused by subtype C[273]. Subtype C virus appears to have a stronger presence in female genital mucosa than other subtypes, which may facilitate heterosexual transmission[274]. In Uganda, Kiwanuka and collaborators concluded that subtype A viruses have a significantly higher rate of heterosexual transmission than subtype D viruses while studying differences in heterosexual HIV-1 transmission among HIV-discordant couples[136]. HIV-1 subtype distribution among these couples was 73.9% (198) subtype D, 11.6% (31) subtype A, and 14.5% (39) recombinant viruses. In this study by Kiwanuka, only 92 HIV-1 transmissions were investigated, it represents a small sample. This is an example of the great limitation of these and similar studies: how to compare transmission rates in populations with high genetic diversity, where true subtype advantages would be evident if they exist. Regarding mother to infant transmission, conflicting results were published. In Kenya, this kind of transmission was more common among mothers infected with subtype D compared with subtype A[275], but another study in Tanzania showed preferential transmission of subtype C compared to subtype A or D[276]. Other authors found no association between subtype and rates of mother to infant transmission[277]. Many maternal factors can contribute to this variability, like age, immunological status or viral load. The role of viral determinants remains to be cleared, as we know that viral diversity in the mother is higher than that present in the newborn, suggesting some selection of maternal virus[278].

The earliest events in HIV transmission have received particular attention, including studies at the molecular level. Although a genotypic signature of early-transmitting viruses has proved difficult to identify, two key features have emerged. First, in studies of both

heterosexual and mother to child transmission, early-transmitting gp120 has been found to be shorter in length, and encode fewer potential N-linked glycosylation sites (PNG) than typical chronically replicating isolates. These features have thus far been only found in the context of infection with HIV-1 subtypes A and C[132,279]. Second, length shortening has been observed in the V1/V2 region, as well as in the V4 and flanking regions of gp120[279]. These two characteristics apparently provide early-transmitting isolates with increased transmission fitness. The extent of this fitness-advantage is poorly understood, however it is known that V1/V2, along with V4 and flanking regions, is frequently an early target of autologous neutralizing antibodies[142]. The viral evolution inside a host drives escape to neutralization through amino acid substitutions, insertions/deletions and also by adding/shifting glycosylation sites. In this way, the majority of the viral population drifts away from the genotypic features that distinguish early-transmitting isolates. If variability compromises transmissibility fitness, transmitted variants have less diversity and divergence and are more closely related to the ancestral sequences, which represent a minority subset of HIV-1. This favor for transmission explains the bottleneck in HIV-1 inter-host spread. Nevertheless, it remains to be determined whether there is a true association between HIV-1 diversity and transmission or whether the differences found by some authors are associated with other factors that can influence transmission, like behavioral or epidemiological features. Considering the current body of knowledge, it is still not possible to define which virus would be best fit for transmission; we only have trends and models.

A long time question yet to be answered is whether clade specificities or differences can influence rates of disease progression. An oversimplified view of HIV-1 infection progression considers that CD4+ cell counts are the distance between a train and an obstacle ahead, and that viral loads are the train's speed: the lower the first or the higher the second, sooner the crash (AIDS or death) will happen. Surely disease progression is far more complex. There have been several prospective, observational studies of the course of HIV-1 infection in cohorts infected with various HIV-1 subtypes, trying to correlate genetic diversity with disease progression. Subtype D was associated with the most rapid disease progression relative to other subtypes[280]. A more recent study supported this finding by associating subtype D with a significantly faster decline of CD4+ cell counts, as compared with subtype A ($p<0.001$)[281].A similar result emerged from a study conducted in an ethnically diversified population in London, with subtype D infection associating with a faster decline in CD4+ cell counts as compared whit subtypes B ($p=0.02$), A ($p=0.004$) or C ($p=0.01$)[282]. A

relevant question is if there are any biological bases for these differences. A clue was highlighted when Kaleebu and colleagues found that the emergence of X4 variants was more common in HIV-1 subtype D compared with subtype A[283]. Moreover, Huang et al. showed that subtype D may be dual tropic, with tropism for both R5 and X4 coreceptors, more frequently than other subtypes[284]. It is well known that X4 variants are associated with increased CD4+ cell depletion and faster disease progression[285].

Finally, in face of some conflicting or incongruent results, it must be remembered that progression of disease, such as the one caused by HIV-1, depends on many confounder factors, such as nutrition, co-morbidities, genetic factors of the patients and access to medical care, which can turn very hard to control.

Diagnosis and disease management

The vast majority of the serological and molecular assays for the diagnosis of HIV-1 infection and for patient management are based on subtype B, the commonest of HIV-1 subtypes in the United States of America and in Western Europe. These tests should be able to detect all genetic forms of HIV-1. However, as discussed before, HIV-1 underwent very extensive genetic and antigenic evolution, along with global redistribution of subtypes and recombinants. With the fourth generation assays for antibody detection, which are able to detect all known HIV-1 group M subtypes as well as HIV-2 positive samples with 100% sensitivity and >98% specificity[286], the large proportion of false negatives presently involve HIV-1 group O[287]. Although highly divergent, HIV-1 group N virus are detected by commercial immunoassays[288]. More problematic are failures in detecting HIV-1 infections by rapid tests, widely used in Africa for diagnosis because they are cheap, simple and instrument-free. Minor antigenic differences in subtypes D, F, H and CRF02_AG can compromise sensitivity of these tests, going as low as 94.1%[289-291]. A false-negative HIV-test in a high prevalence area is a definite way to perpetuate transmission. Recent seroconversion and, paradoxically, immune exhaustion in long term infections, might be associated with low HIV antibodies levels, also diminishing the sensitivity of serological assays.

Quantifying HIV-1 RNA levels is essential to monitor disease progression, detect primary or perinatal infections as well as the response to antiretroviral therapy. As these assays rely on

HIV-1 sequence-specific primers or probes, independently of the technology (reverse transcriptase polymerase chain reaction, branched-chain DNA signal amplification, real-time polymerase chain reaction or isothermal nucleic acid sequence-based amplification), if reliable quantification is compromised, the results of such assay can have undesirable consequences, as viral load quantification is an important parameter to estimate adequate therapeutic response. Natural polymorphisms and genetic variation might create unrecognized variants. Several comparative studies demonstrated that the sensitivity and specificity of viral load assays varies depending on HIV-1 group or subtype, especially in non-B subtypes, complex recombinants and groups O, N and P[292-294]. The solution for this limitation was to design a test that targets a highly conserved region across all subtypes and CRF. The *pol* integrase region of the HIV-1 genome is subject to less variability than other regions[295], so it is preferred as a target for amplification in the most reliable tests, covering the highest number of HIV-1 genetic forms.

Response and resistance to antiretroviral therapy

Genetic differences between HIV-1 clades can lead to altered susceptibility to antiretroviral drugs (ARV). A classical example is given by HIV-1 group O and HIV-2, both exhibiting high-level innate resistance to non nucleoside reverse transcriptase inhibitors (NNRTI) [296,297], due to natural polymorphisms. Such as HIV-1 diagnostic tools, ARV were developed in the Western world, so they are based in subtype B. Susceptibility of non-B subtypes to ARV is further obscured because genotypic and phenotypic resistance testing are also originally based in subtype B. There are several evidences of an apparent discrepancy in drug resistance among subtypes. A striking one is a statistically significant difference in the response to nevirapine (a NNRTI) in single dosage to prevent mother to child transmission, in which subtype C shows more resistance than subtypes D or A[298]. On other hand, several studies failed to demonstrate major differences in the response to ARV therapy according to HIV-1 subtype. In France, a cohort of 416 adult patients (24% of whom carried a non-B subtype) showed that HIV-1 subtype did not affect clinical progression, viral load or CD4+ cell count regardless of ARV scheme used[299]. In London, in a subset of patients of African origin infected with non-B subtypes, the use of either protease inhibitors (PI) or NNRTI based therapy did not altered response to treatment[300]. When comparing patients infected with

subtypes A, C, D and CRF02_AG with patients infected with subtype B virus, Geretti and collaborators found no differences in achieving viral load suppression, when treated with a NNRTI or PI based regimen[301]. In the absence of any ARV pressure, reverse transcriptase and protease sequences are naturally polymorphic (between 30 and 40%), when comparing B to non-B subtypes[302]. These genotypic variations do not confer diminished susceptibility, so it seems reasonable to presume, with our current understanding, that different group M subtypes have similar susceptibilities to currently used ARV[303]. Nevertheless, a large global collaborative study identified 55 subtype B drug-resistance mutations and all were found in at least one non-B subtype[304]. On reverse, of 67 resistance mutations found in at least one non-B subtype, only 61 were also seen in subtype B isolates, indicating the occurrence of novel mutations in non-B subtypes[304], justifying a fastidious vigilance. In top of that, there are subtle variations in different subtype's genomes able to influence the emergence of resistance when drug exposure is present. A single nucleotide substitution from the wild-type codon found in subtype C can generate the mutation V106M, which is associated with NNRTI resistance, while at least two substitutions are needed for the wild-type subtype B codon[305]. This indicates that subtype C may have a lower genetic barrier to NNRTI resistance than subtype B. This mutation, V106M, is in fact associated with subtype C infected patients failing therapy, as it is frequently found after treatment with efavirenz or nevirapine[306]. Subtype C also appears to acquire K65R faster than subtype B[307]. This mutation is associated with tenofovir resistance and its presence in higher rates in subtype C, suggesting that these viruses may have a particular predisposition toward acquiring this mutation[308]. Apparently, this subtype has an intrinsic difficulty in synthesizing *pol* sequences that leads to template pausing at codon 65, facilitating acquisition of K65R under selective drug pressure[309].

In the protease gene, polymorphisms do not impair drug susceptibility, due to a high genetic barrier. Several concomitant mutations are necessary to full blown resistance occur. However, those polymorphisms may facilitate the genetic pathway of resistance, as soon as the virus generates a major resistant mutation[310] . The minor mutation V11I, associated with darunavir resistance, occurs naturally in CRF37_cpx isolates[311]. The V82I natural polymorphism in subtype G facilitates the emergence of V82M, a mutation associated with resistance to indinavir[312]. The L90M mutation, that confers resistance to nelfinavir, an saquinavir, is rare in subtype F but common in subtype B in patients from Brazil[313]. All these observations suggest differences in drug resistance pathways between HIV-1 subtypes.

However, the evidence gathered so far is insufficient to assess the actual contribution to resistance of innate genetic HIV-1 diversity.

There are many other variables that influence response to therapy, namely adherence, drug regimens or ethnicity, which must be controlled in large prospective studies to assess the efficacy of different drug schemes in patients infected with non-B subtype HIV-1. Anyhow, the effect of extensive recombination, fueled by geographic diversification and conditioned by drug pressure, can link resistance mutations and lead to multi-drug resistance.

Vaccine development

Long demanded, an effective vaccine against HIV infection remains a chimera hard to conquer. Development of a vaccine, able to hinder the HIV-1 pandemic is halted by the extensive genetic diversity of the virus. Several vaccine approaches have been able to augment the immune responses to HIV infection. However, most of these responses, whether cellular or humoral, have largely failed in controlling HIV infection. Indeed, fully functional escape variants are easily selected and overcome the immune system assault.

For most vaccines, success consists in identifying an immunization approach that mimics or enhances protective host immunity. However, in HIV infection there is no known precedent for spontaneous immune responses leading to clearance of HIV or even durable immune protection from re-infection, as events of superinfection can occur[314].

A successful prophylactic vaccine able to prevent chronicity during natural infection depends of the induction of virus-specific neutralizing antibodies[315]. The initial characterization of HIV-1 as a retrovirus with *env* protein-mediated entry brought the hope that an *env*-based vaccination approach would yield neutralizing antibodies, providing protection against either acquisition or progression of infection. However, the development of HIV/AIDS vaccines targeting humoral immunity has encountered unsurpassed obstacles as the vast majority of antibodies elicited by immunization with monomeric *env* proteins were directed at bait epitopes with little or no neutralization activity[316]. Indeed, two large, well-conducted Phase 3 efficacy trials of alum-adjuvanted *env* protein (gp120; VAX003 and VAX004) showed no efficacy against HIV-1 acquisition or post-infection viremia[317,318]. An alternative approach, capable of eliciting potent, durable, broadly neutralizing antibodies, similar to those produced

in some chronically infected individuals, the so called elite neutralizers[319], seemed promising. These broadly neutralizing antibodies, which are produced years after seroconversion, are of little benefit to their bearer though, as viral escape mutants persist in those patients[320], but, if present in appropriate concentration at the time of contagion, could confer protection from HIV-1 infection, providing they could target multiple conserve epitopes. To date, however, despite extensive effort, no immunogen/vaccine approach has been capable of reliably elicit such broadly neutralizing antibodies, and the prospects for such immunogens remain uncertain[321,322]. The native envelope structure is a trimer of gp120/gp40heterodimers which hides crucial parts of the molecule and undergoes conformational changes upon receptor binding. Furthermore, hypervariable loops mask critical receptor binding sites and carbohydrates impede antibody binding[323]. With respect to inducing antibodies against the envelope protein, it is important to remember that for the annual influenza vaccine less than 2% amino acid change in the circulating influenza strain can cause a failure in the cross reactivity of the polyclonal response induced by the vaccine[324]. In comparison, genetic variation of *env* within a HIV-1 group M subtype can be >15% and variation between subtypes can be >30%[324].

Eliciting a cellular immune response against HIV-1 infection has the potential to lower the viral load setpoint and thereby slow disease progression as HIV-specific CD8+ cytotoxic T cells play an important role in controlling acute HIV infection. However, the initial response is narrow and the targeted epitopes rapidly escape[154]. To reduce the chances of viral escape, multiple high conserved epitopes would need to be targeted by a vaccine. A T cell based vaccination strategy has formidable obstacles: populations around the world vary in types and frequencies of HLA alleles so different individuals infected with the same HIV-1 subtype will recognize distinct T cell epitopes; intra-subtype responses are stronger and more frequent than inter-subtype reactivities[325]; and even conserved epitopes may not be presented to T cells, depending on the sequences flanking those epitopes[326].

However, HIV-1 has an “Achilles heel”, an immune vulnerability that gives its host a chance to prevent or control infection. In fact, there is a growing body of evidence showing the vulnerability of HIV-1 to immunity during the early phase of infection, when the genetic bottleneck that occurs during transmission largely reduces HIV diversity. In fact, several studies proved that the genetic complexity of chronic infection is much higher than that of acute/early infection[132,327]. Moreover, the majority of productive infections start with a single virion[133,134,146,147]. These transmitted/founder viruses should be the major target for vaccine

development, as these are the viral strains primarily involved in infection. However, there are two difficulties to put into practice this line of thought: (i) to identify acute infections and (ii) precisely determine HIV-1 infecting virion(s).

Above all, a single vaccine, even protecting against major subtypes and CRFs, wouldn't probably be enough to adequately defend an immunized population. There would be still a risk of these individuals to get infected with emergent CRFs or even URFs that would rise new epidemic waves, by positive selection. Any successful strategy will depend on sound and robust molecular epidemiology data on HIV-1 subtype distribution, making possible the design of sequential and multivalent vaccination regimens.

CHAPTER 2: ANALYSIS OF A LOCAL HIV-1 EPIDEMIC IN PORTUGAL HIGHLIGHTS ESTABLISHED TRANSMISSION OF NON-B AND -G SUBTYPES

ANALYSIS OF A LOCAL HIV-1 EPIDEMIC IN PORTUGAL HIGHLIGHTS ESTABLISHED TRANSMISSION OF NON-B AND -G SUBTYPES

Alexandre Carvalho*†‡, Patrício Costa*†, Vera Triunfante*†, Fernando Branca‡, Fernando Rodrigues*†, Catarina L. Santos*†, Margarida Correia-Neves*†, Margarida Saraiva*†, Henrique Lecour§, António G. Castro*†, Jorge Pedrosa*†, Nuno S. Osório*†

[*] Life and Health Sciences Research Institute (ICVS), School of Health Sciences, University of Minho, Braga, Portugal;

[†] ICVS/3B's PT Government Associate Laboratory, Braga/Guimarães, Portugal;

[‡]Hospital de Braga, Braga, Portugal;

[§]Institute of Health Sciences, Portuguese Catholic University, UCP Porto, Portugal.

Manuscript submitted to Journal of Acquired Immune Deficiency Syndromes

Abstract

Objective: Existing data supports Portugal as the European country with highest HIV-1 subtype diversity. The characteristics of transmission networks may be a key determinant of the HIV epidemic growth and diversity patterns but Portuguese studies on this subject are scant. Therefore we aimed to analyze the phylodynamics of HIV-1 infection in the Portuguese region of Minho.

Methods: Molecular epidemiological analysis was applied to data from 289 HIV-1 infected individuals followed in the reference Hospital of the province of Minho, Portugal, who had their virus sequenced between 2000 and 2012.

Results: Virus of the G (29.1%) and B (27.0%) subtypes were the most frequent, followed by recombinant forms (17.6%), C (14.5%), F1 (7.3%) and A1 (4.2%) subtypes. Multinomial logistic regression revealed that the year of HIV-1 diagnosis was associated with an increasing risk of infection with the A1 and F1 subtypes when compared with B, G, C or recombinant virus. As expected, polyphyletic patterns suggesting multiple and old introductions of subtypes B and G were found. However, transmission clusters of non-B and -G virus among native individuals were also found with the dates of the most recent common ancestor estimated to the early 2000s.

Conclusions: Our study supports high HIV-1 subtype diversity in the Portuguese region of Minho and local transmission of non-B and -G subtypes that started more than one decade ago. The rate of infection with A1 and F1 virus, found in sexually transmitted clusters, is increasing, reinforcing the need for more efficacious control measures targeting sexual transmission routes.

Key Words: HIV-1 subtypes, viral diversity, molecular epidemiology, transmission clusters, viral transmission networks, Portugal

Introduction

Globally, 35.3 million people were estimated to be infected with HIV-1 at the end of 2012 and AIDS remains one of the world's most serious health challenges[3]. Phylogenetically, HIV-1 is divided into four groups M, O, N and P. Most HIV-1 infections globally are caused by M group virus that can be further divided in at least nine subtypes, A-D, F-H, J, K and different circulating and unique recombinant forms[328]. Although clinical evidence is still limited and current antiretroviral regimens appear to have comparable efficacy in all subtypes there is presently evidence showing that particular HIV-1 subtypes may have transmission advantage[136,329-331], higher replicative efficiency or altered drug susceptibility[282,332-335]. The geographic patterns of M group subtypes are continuously changing in response to human population migrations and active transmission networks thus inciting constant vigilance. Although several reports suggest that the prevalence of non-B subtypes is increasing in Western Europe[242,336-340], B subtype remains the most prevalent. Portugal contrasts with the rest of the Western Europe in its distribution of HIV-1 subtypes. In addition to B subtype, Portugal also has a high prevalence of G subtype[341,342]. The high prevalence of B and G subtypes is thought to have promoted the appearance among intravenous drug users (IDU) of different types of B/G recombinant strains, namely CRF14_BG that is estimated to have emerged in Portugal in the early 1990's and then spread to Spain and other European countries[257,343,344]. The association between HIV-1 subtype and risk-behavior patterns has been complex to define mainly due to difficulties in obtaining large numbers of each viral subtype and transmission route in a homogeneous study population[280].

Portugal has one of the highest HIV-1 prevalence in Western Europe and following a decrease in the last decade of HIV-1 infection in IDU, heterosexual contact is nowadays estimated to be the most relevant transmission route in Portugal[255]. The reconstruction of viral transmission networks is a relevant tool to monitor the disease and determine preventive efficacious measures. Several studies have shown that, in addition to patient interview, phylogenetic analysis of genetic sequences from the virus can provide valuable insights to identify events of onward transmission and evaluate the spread of the virus[345-347]. Nonetheless, limited information is available to understand HIV-1 transmission clusters in Portugal and to explain the high HIV-1 diversity in the region.

The aim of this study was to perform molecular epidemiologic characterization of a cohort of 289 patients followed in the reference hospital for the Minho province, Portugal. Specifically, we aimed at identifying local transmission networks and possible relationships with previously described transmission clusters. In line with previous studies in Portugal we found large subtype diversity. In addition, our analysis supports that the transmission of non-predominant subtypes among the local population initiated more than one decade ago, providing valuable insights into the dynamics of infection in this geographic area.

Methods

Study population

289 individuals were selected from the HIV-1 infected patients followed at Hospital de Braga (HB) according to two criteria: (i) availability of plasma sample or plasma-derived viral sequence sampled from 2000 to 2012 and; (ii) absence of previous antiretroviral treatment at the time of sampling. The following information was collected anonymously from the clinical files of each individual: presumed transmission route, gender, age, nationality, presumed country of infection and date of diagnosis. HB is a university affiliated hospital serving as the reference hospital for the 1 093 021 habitants of the northwest Portuguese province of Minho. The prevalence of HIV-1 infection in Minho (0.12%) is lower than the overall prevalence in Portugal (0.31%)[255]. By the end of 2012, a total of 748 HIV-1 patients were being followed at HB, representing 57% of the HIV-1 infected individuals from Minho's region[255]. According to transmission mode, HB population presented significant differences from the data for the Portuguese HIV-1 infected individuals: more IDU and fewer men who have sex with men (MSM). Also, in HB, there were more men and fewer patients over 40 years old (Table 1). The frequency of individuals reporting heterosexual transmission is similar when comparing HB HIV-1 patients with the overall country data (Table 1).

	HIV-1 infections in Hospital de Braga	HIV-1 infections in Portugal†	p-value
Heterosexual contact	332 (44.4%)	18424 (43.3%)	0.539
IDU	328 (43.9%)	15992 (37.6%)	<0.001***
MSM	71 (9.5%)	5845 (13.7%)	<0.001***
Male	573 (76.6%)	31255 (73.4%)	0.039*
>40 years (on diagnosis)	200 (26.7%)	13903 (32.7%)	<0.001***
Total	748	42580	

Table 1: Comparison of the transmission route, gender and age at diagnosis between the HIV-1 infected individuals followed at Hospital de Braga and all the available data on HIV-1 infections in Portugal by the end of 2012. IDU: intravenous drug users; MSM: men who have sex with men.
† Data from reference 1.***p<0.001 *p<0.05 binomial test

Sequencing of viral samples

Viral RNA was extracted using Magna Pure Total Nucleic Acid Isolation Kits (Roche Applied Science). RT-PCR and DNA sequencing were performed with Trugene HIV-1

Genotyping System (Siemens Healthcare Diagnostics). The sequenced regions include part of the coding sequences of *gag* (492 to 501), p6 (44 to 53), *pol* (60 to 402), p2p7p1p6 (129 to 138), Protease (4 to 99) and RT (1 to 127, reported positions are amino acid positions relative to protein start in the HXB2 reference genome, GenBank: K03455.1). The subtyping of the 289 sequences was made using REGA 3.0[348] and non-automatic phylogenetic analysis. Non-automatic bootscan analysis was also done with the program SimPlot to confirm selected subtypes using the F84 nucleotide substitution model and a sliding window of 200-bp, a 40-bp step[232]. Detection of recombination was confirmed using the program RDP[349]. Sequences were uploaded to GenBank and assigned the following accession numbers: KM205831-KM206119.

Phylogenetic analysis

The 289 HIV-1 sequences obtained in this study and 88 sequences from the databases including the M group consensus and a previously defined set of subtype reference sequences[348] including at least two reference sequences from each M group subtype (A1, A2, B, C, D, F1, F2, G, H, J and K) and from 26 CRF (CRF01_AE, CRF02_AG, CRF03_AB, CRF04_CPX, CRF05_DF, CRF06_CPX, CRF10_CD, CRF11_CPX, CRF12_BF, CRF13_CPX, CRF14_BG, CRF18_CPX, CRF19_CPX, CRF20_BG, CRF24_BG, CRF25_CPX, CRF27_CPX, CRF29_BF, CRF31_BC, CRF35_AD, CRF37_CPX, CRF39_BF, CRF40_BF, CRF42_BF, CRF47_BF) were aligned using MUSCLE[350]. The phylogenetic analysis of the 377 sequences was conducted using RAxML 7.0.3 to produce a maximum likelihood tree using 1000 bootstrapping replicates[351]. Analysis was repeated with PhyML[352] computing the aLRT support of all tree branches and by Bayesian analysis using BEAST[353]. The best fitting nucleotide-substitution model for the Bayesian analysis was estimated using jModeltest v2.1.2[354] to be the general time reversible (GTR) model with a proportion of invariant site (I) and gamma distribution of rates (G), selected among 88 different models according to the Akaike Information Criterion (AIC), the Bayesian Information Criterion (BIC), and the Decision Theoretic Framework (DT). An eventual bias introduced by convergent evolution due to the presence of drug resistant mutations was discarded by repeating the analysis after removal of codons associated with drug resistance in the standardized list of mutations for surveillance of transmitted drug resistance established by the World Health Organization[355]. The general topology of the trees and identification of clustering remained unchanged. Clusters of at least three individuals were identified based on a ML bootstrap support > 95%, a Bayesian posterior probability >0.95.

Estimation of evolutionary dates

Estimates of the time of the most recent common ancestor (MRCA) were performed by inferring simultaneously population parameters, substitution parameters, and tree topology using Bayesian Markov Chain Monte Carlo (MCMC) inference as implemented in BEAST version 1.8.0[353]. Three independent runs 160 million replicates were performed under Bayesian Skyline relaxed molecular clock model, using a general time-reversible nucleotide substitution model with heterogeneity among sites modeled with a gamma distribution. Examination of the MCMC samples with Tracer v1.4 indicated convergence and adequate mixing of the Markov chains. After inspection with Tracer, we discarded an appropriate number of steps from each run as burn-in, and combined the resulting MCMC tree samples for subsequent estimation of posteriors. We summarized the MCMC samples using the maximum clade credibility topology, with branch length depicted in years.

Statistical analysis

To identify the main predictors of HIV-1 subtype groups a multinomial logistic regression model was performed. With this procedure we assessed the association between the date of diagnostic and HIV-1 subtype groups controlling for other relevant variables. HIV-1 subtypes with low number of cases were pooled resulting in 5 groups that were used for statistical analysis: G (n=84); B (n=75); C (n=42); other subtypes (A1, n=12; F1, n=21; total n=33) and recombinants (n=42). The least represented subtypes J (n=1) and D (n=3) had no influence on the results and were excluded from the analysis. The independent variables analyzed were date of diagnostic, age, gender and transmission mode. The SPSS package (IBM SPSS Statistics v19) was used to conduct all statistical analysis and results were considered to be significant for $p<0.05$.

Ethics

The project was approved by the Ethics Committee of the Hospital de Braga. Written consent was obtained for all the patients enrolled in the study. Clinical data was codified to ensure confidentiality of the patients.

High HIV-1 subtype diversity

Among the 289 individuals that met the inclusion criteria, 76.8% were male and the average age on diagnosis of the study population was 44.5 years ranging from 18 to 87 years. The most frequently reported route of infection was heterosexual contact (n=161, 55.7%), followed by IDU (n=99, 34.2%) and MSM (n=26, 9.0%) (Table 2). The study population was highly homogeneous, with >90% of the individuals being Portuguese of white ethnicity and presumed to be infected in Portugal. The most frequent subtypes found were G (n=85, 29.4%), B (n=75, 26%) and C (n=42, 14.5%) followed by F1 (n=22, 7.6%) and A1 (n=12, 4.2%) subtypes. Only 1.4% of the studied individuals had infection with other "pure" subtypes (D, n=3 and J, n=1). The most frequent CRF found was CRF14_BG (n=15, 5.2%), followed by CRF02_AG (n=4, 1.4%). Individuals infected with unique recombinant forms (URF) constituted 9.0% of the population (Figure 19A).

Variables		n	%
Gender	Male	222	76.8%
	Female	67	23.2%
Age on diagnosis (years)	≤20	17	5.9%
	21-40	190	65.7%
	41-50	41	14.2%
	>50	41	14.2%
Patient Nationality	Portuguese	260	90.0%
	Other	29	10.0%
Ethnicity	White	274	94.8%
	Black	15	5.2%
Presumed country of infection	Portugal	280	96.9%
	other	9	3.1%
Route of transmission	Heterosexual contact	161	55.7%
	MSM	26	9.0%
	IDU	99	34.2%
	other	3	1.1%
	Total	**289**	**100%**

Table 2. Descriptive statistics on the demographics of the study population

Increasing incidence of infection with A1 and F1 HIV-1 subtypes

The dates of HIV-1 diagnosis in the study population spanned the period from 1987 through 2012. The proportion of infections diagnosed each year with different subtypes was investigated. The less frequent subtypes were pooled to allow statistical analysis (Figure 19B)

and a significant logistic regression model was obtained ($\chi^2_{(20)}$=77.3, p<.001 and Pseudo $R^2_{Nagelkerke}$ value was 0.245). Multivariate analysis revealed that the year of HIV-1 diagnosis

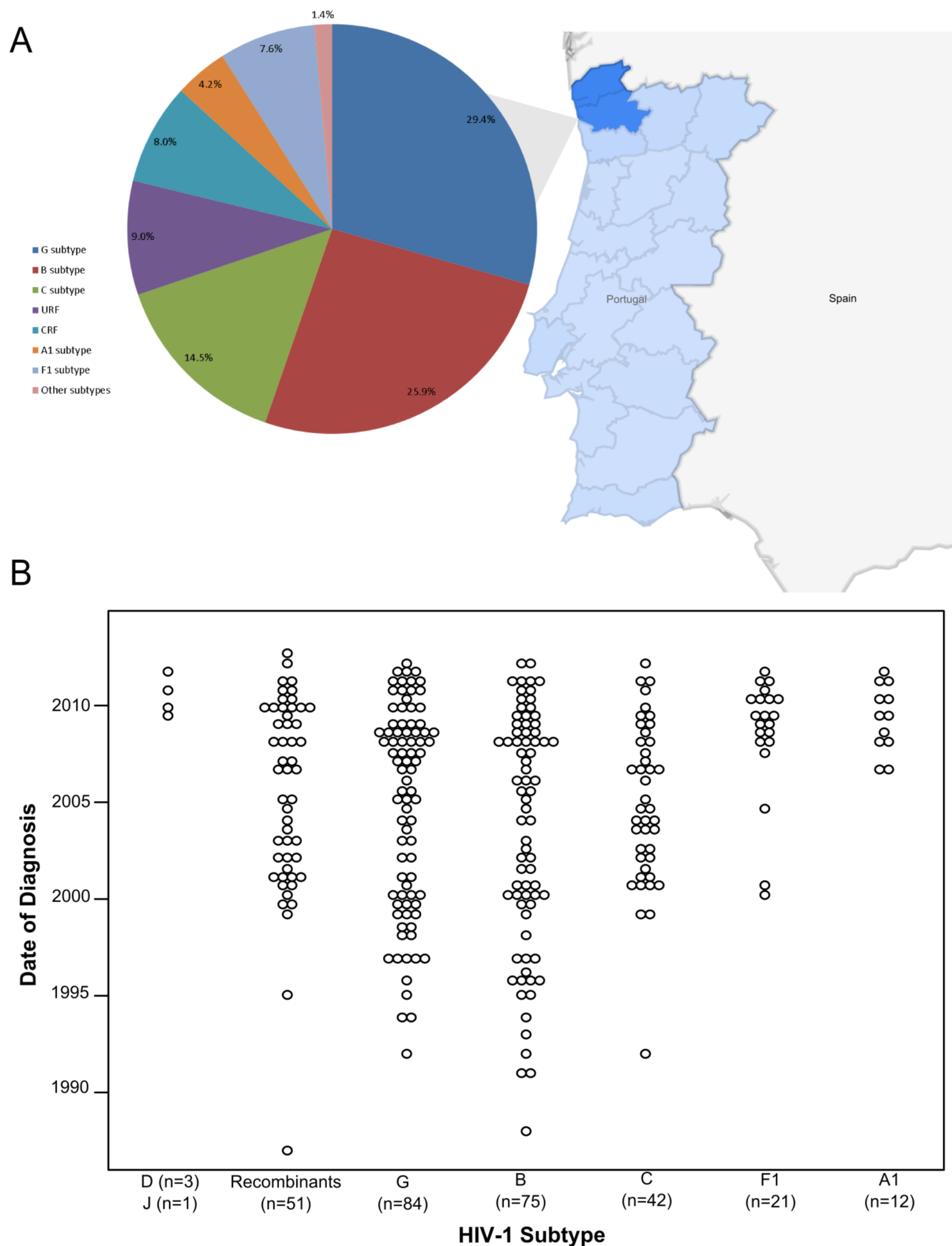

Figure 19: HIV-1 subtype diversity (**A**) and temporal distribution of HIV-1 subtypes (**B**) in the cohort of 289 infected individuals from the Minho region.

was associated with subtype A1 and F1 infection. The odds ratio (OR) for being infected with

each subtype for a 1-year increase in the time period of diagnosis was OR=0.852 (p<.05) for recombinant virus versus A1 and F1 subtypes, OR=0.824 (p<.01) for G subtype versus A1 and F1 subtypes, OR=0.781 (p<0.001) for B subtypes versus A1 and F1 subtypes, OR=0.862 (p<0.05) for B subtypes versus A1 and F1 subtypes (Table 3). Overall these data supports that the rate of infection with A1 and F1 subtypes is increasing in the study population over the years when compared with the risk of infection with C, B, G or recombinant virus.

Subtype		B	SE	Wald	OR	CI 95% OR	
						LB	UB
Recombinants vs. A1/F1 subtypes	Days since date of diagnosis (1 year)	-0.0004	0.000	6.278*	1.000 (0.852)	0.999	1.000
	Age	-0.006	0.019	0.091	0.994	0.959	1.031
	Males	0.298	0.638	0.218	1.347	0.386	4.699
	Heterosexual	-0.808	0.735	1.207	0.446	0.106	1.884
	IDU	-0.346	0.820	0.178	0.708	0.142	3.533
G vs. A1/F1 subtypes	Days since date of diagnosis (1 year)	-0.001	0.000	10.774**	0.999 (0.824)	0.999	1.000
	Age	0.016	0.016	1.1	1.016	0.986	1.048
	Males	-0.847	0.476	3.167	0.429	0.169	1.090
	Heterosexual	1.450	1.161	1.559	4.264	0.438	41.537
	IDU	2.012	1.214	2.747	7.475	0.693	80.672
B vs. A1/F1 subtypes	Days since date of diagnosis (1 year)	-0.001	0.000	17.319***	0.999 (0.781)	0.999	1.000
	Age	-0.016	0.017	0.932	0.984	0.952	1.017
	Males	-0.473	0.512	0.856	0.623	0.229	1.698
	Heterosexual	-1.234	0.663	3.459	0.291	0.079	1.069
	IDU	-1.593	0.771	4.269*	0.203	0.045	0.921
C vs. A1/F1 subtypes	Days since date of diagnosis (1 year)	-0.0004	0.000	5.154*	1.000 (0.862)	0.999	1.000
	Age	0.017	0.019	0.815	1.017	0.980	1.056
	Males	-0.646	0.579	1.242	0.524	0.168	1.632
	Heterosexual	-0.267	0.949	0.079	0.766	0.119	4.919
	IDU	1.216	0.996	1.491	3.374	0.479	23.754

***p<.001; **p<.01; *p<.05. Abbreviations: SE. standard error; OR. Odds ratio; LB. lower bound; UB. upper bound; IDU. intravenous drug users

Table 3. Multinomial logistic regression model relating HIV-1 subtypes with date of diagnostic, age, gender and transmission mode.

Evidence for local transmission clusters

The phylogenetic analysis of the viral sequences allowed the identification of 14 transmission pairs and 17 transmission clusters (Figure 20).

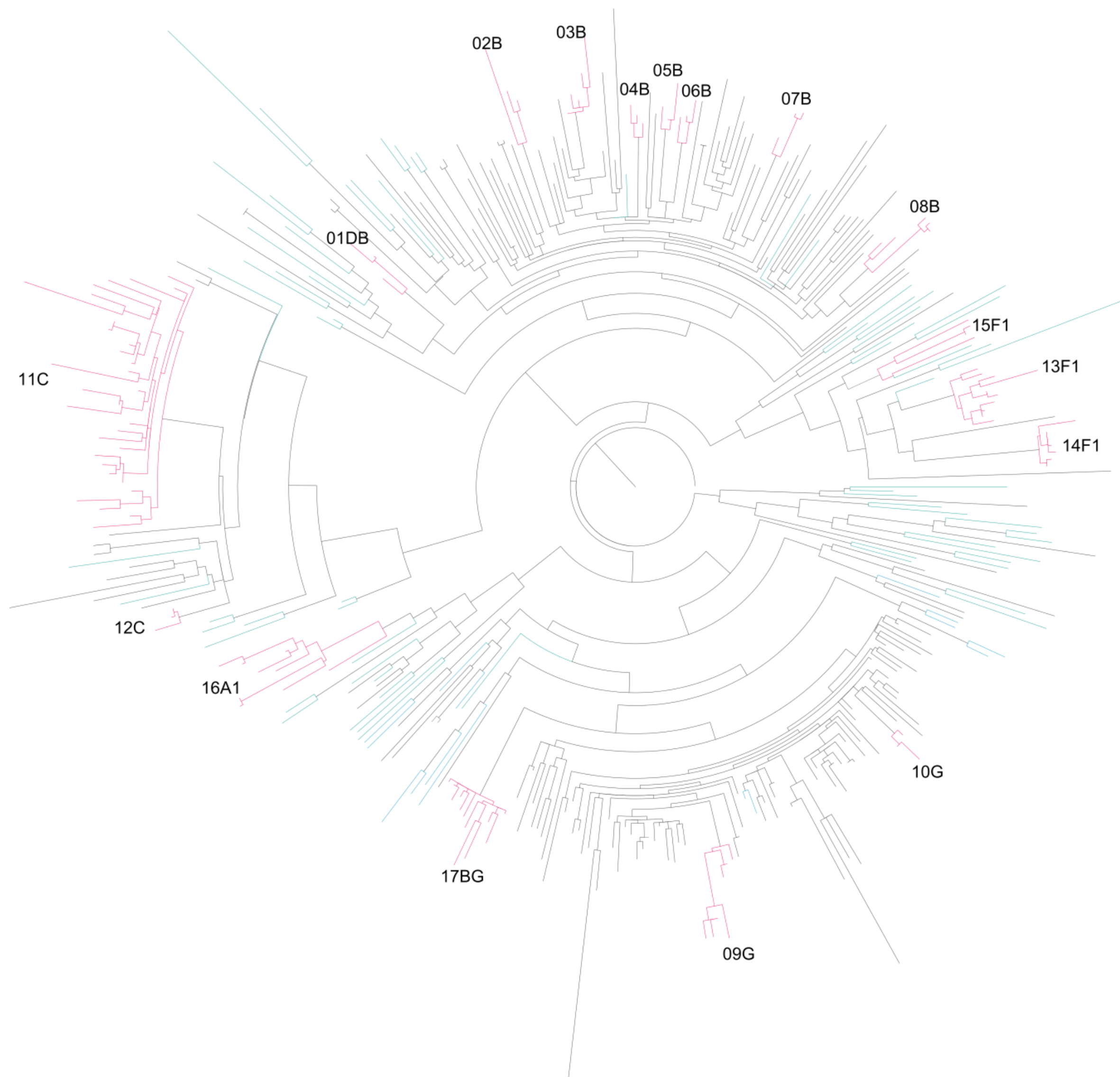

Figure 20: Phylogenetic relationships among the HIV-1 sequences isolated from 289 infected individuals from Minho province, Portugal. Maximum likelihood (ML) phylogenetic analysis with was performed using 289 partial HIV-1 sequences obtained in this study and 89 subtype reference sequences (colored in blue) and rooted using the M-group consensus sequence. Branch lengths were expressed as the number of nucleotide substitutions per site. Transmission clusters were supported by a ML bootstrap support > 95% based on 1000 replicates and were colored in pink.

In terms of HIV-1 subtype distribution, seven clusters included subtype B virus, three included subtype F1, two included subtype G, two included subtype C, one included subtype A1, one included CRF14_BG and one included DB URF (Table 4). The mean number of individuals per cluster was 6.9. Among the clusters with an above the mean number of

individuals (≥7), the largest (11C) was composed almost integrally of IDU (27 out of 31 individuals), 4 of these individuals reported sharing injection material. The second largest (17BG) was also composed of IDU (11 out of 12 individuals) with the non-IDU individual in the cluster reporting sexual contact with one of the other members of the cluster. All F1 clusters (13F1, 14F1 and 15F1) are predominantly sexual in transmission mode (Table 4). 13F1 cluster has one patient out of the ten that is of Brazilian nationality. Furthermore, the phylogenetic analysis shows that the viral sequences from 13F1 and 14F1 clusters share common ancestors with reference sequences collected in Brazil (GenBank: AY173957.1, EU735538.1, EU735540.1). Additionally, BLAST analysis identified 5 sequences highly related to the 14F1 (Figure 21) virus that were isolated in Italy and phylogenetically linked to Brazil[356]. Cluster 16A1 is a sexually transmitted cluster (6 heterosexual and 3 MSM), including one Mozambican infected before immigrating to Portugal.

Established local transmission of non-prevalent HIV-1 subtypes

In order to estimate the evolutionary dates of the reported transmission clusters we performed a Bayesian MCMC analysis. All independent runs converged to almost identical values for all parameters (data not shown). The mean substitution rate was 2.16×10^{-3} (95% highest posterior density [HPD] interval, 1.8505×10^{-3} to 2.4929×10^{-3}) substitutions per site per year. The date of the most recent common ancestor (MRCA) was determined for all the clusters and ranges from 1993-2008 (Table 4). There are no marked differences in the MRCA date among different subtypes. Among the clusters of non-prevalent HIV-1 subtypes 11C, 15F1, 16A1 and 17BG were the ones with older MRCA dates ranging from 1993 to 1999. With the exception of one Mozambican individual from cluster 16A1 these clusters are formed by Portuguese individuals presumed to be infected in Portugal. When removing the Mozambican individual the MRCA date of 16A1 was 1999 (Table 4). These results support the existence of clusters of non-prevalent subtypes that have started transmitting among the local population more than one decade ago. In order to gain insights into the activity of the clusters along the years we have analyzed the date of HIV-1 diagnosis for all cluster-included individuals. For the majority of the clusters the results are suggestive of continuous onward transmission (Table 4).). A possible exception is the cluster 17BG, since it showed the lower mean date of diagnosis (2002) and in the 6 years from 2006 to 2012 no virus belonging to this cluster were found in the study population suggesting a decrease in transmission.

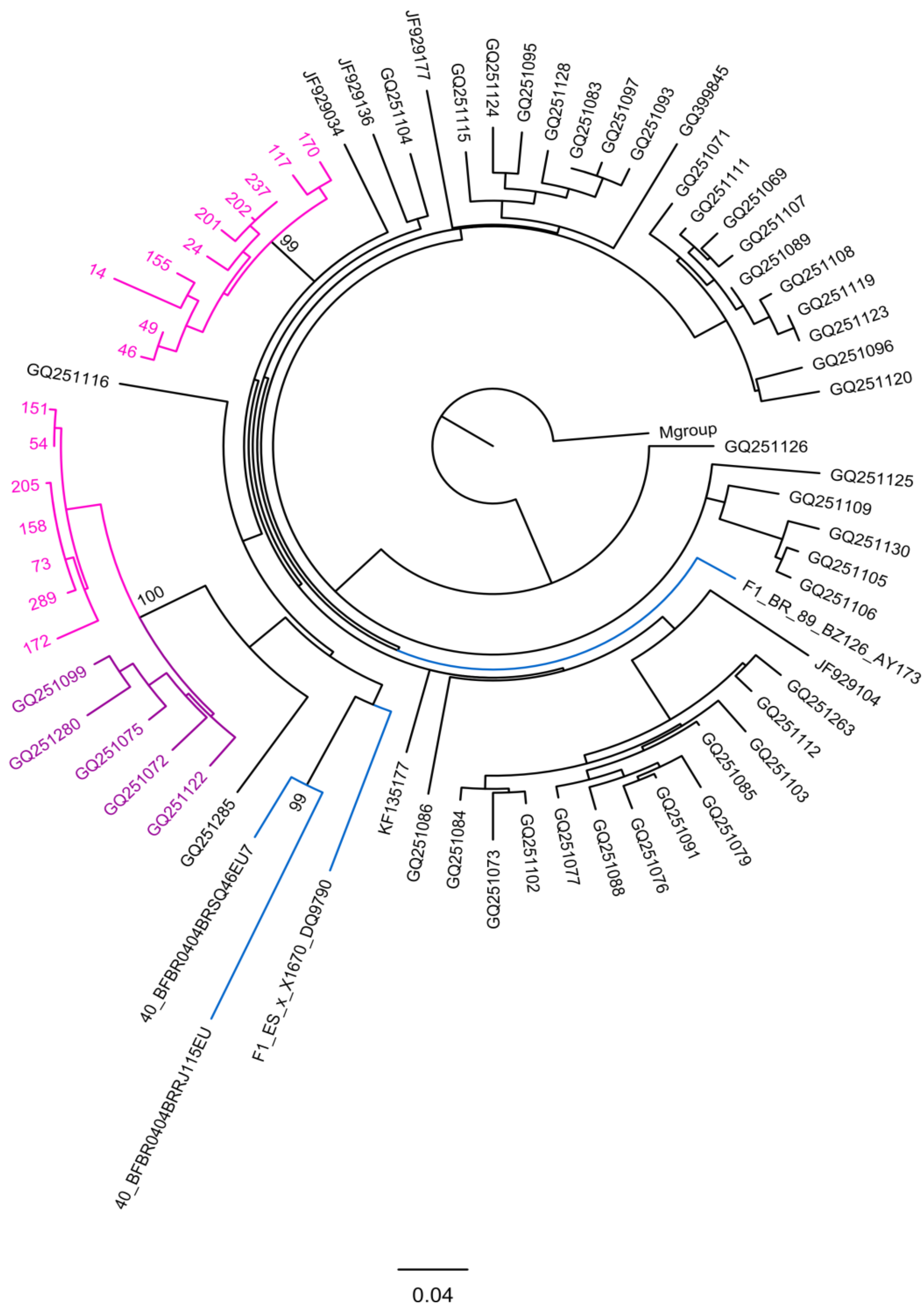

Figure 21: Phylogenetic relationships among the F1 transmission clusters identified in the population of 289 infected individuals from Minho province, Portugal. Maximum likelihood (ML) phylogenetic analysis was performed using sequences obtained in this study (colored in pink), 4 subtype reference sequences (colored in blue) and 50 sequences that have similarity above 95% when compared with the 13F1 and 14F1 sequences, obtained using BLAST search among all public available sequences. Tree was rooted using the M-group consensus sequence.

					year	95% HPD
01DB	5	DB URF	2009**,** 2006-2012	IDU (4), Heterosexual (1)	2002	1998.2-2006.1
02B	3	B	2006**,** 2003-2008	IDU (2), Heterosexual (1)	1999	1994.2 - 2003.3
03B	3	B	2003**,** 2000-2009	Heterosexual (2), MSM (1)	2004	2001.9 - 2006.5
04B	3	B	2007, 2006-2009	Heterosexual (4)	2004	2000.7 - 2006.9
05B	5	B	2006**,** 2003-2008	MSM (3), Heterosexual (2)	2003	2000.0 - 2005.6
06B	3	B	2008**,** 2005-2010	Heterosexual (3)	2002	1998.4 - 2004.8
07B	3	B	2009**,** 2008-2010	Heterosexual (3)	2004	2000.0 - 2006.9
08B	3	B	2007, 2000-2010	Heterosexual (2), MSM (1)	2002	1998.1 - 2004.4
09G	7	G	2006,1999-2009	Heterosexual (5), IDU (2)	2000	1997.1 - 2003.2
10G	3	G	2010,2008-2011	Heterosexual (2), IDU (1)	2008	2004.8 - 2010.4
11C	31	C	2004,1999-2011	IDU (27), Heterosexual (4)	1994	1990.4 - 1998.1
12C	3	C	2006,2001-2009	Heterosexual (3)	2005	2001.6 - 2007.6
13F1	10	F1	2006,2000-2011	Heterosexual (7), IDU (3)	2000	1993.2 - 2003.5
14F1	7	F1	2009,2007-2011	Heterosexual (6), IDU (1)	2005	2002.7 - 2007.2
15F1	4	F1	2010,2009-2010	Heterosexual (4)	1994	1987.5 - 2001.6
16A1	9	A1	2009,2006-2011	Heterosexual (6), MSM (3)	1993 (1999†)	1987.2 - 1998.7 (1994.0-2003.3†)
17BG	12	CRF14_BG	2002,2000-2006	IDU (10), Heterosexual (2)	1999	1994.5 - 2002.8

† time of the most recent common ancestral of the cluster 16A1 excluding one individual known to have been infected before immigrating to Portugal. Abbreviations: MRCA, most recent common ancestral.

Table 4. Characterization of the 17 HIV-1 transmission clusters identified in the study population of the Portuguese region of Minho.

Discussion

Since its origin in Africa approximately 100 years ago, HIV-1 is continuously undergoing genetic diversification that is enhanced by the massive globalization of the human population[257]. Despite the notion that current antiretroviral regimens have comparable efficacy across existing HIV-1 diversity, there is evidence showing that some HIV-1 subtypes may have transmission advantage, higher replicative efficiency or even altered drug susceptibility[282,332-335], raising awareness on the relevance of investigating HIV-1 diversity. Furthermore, for effective targeting of preventive measures it is very relevant to perform persistent monitoring of the HIV-1 pandemic using phylogenetic and epidemiological data analysis as a tool for the reconstruction of viral transmission networks[232,347,348] and to allow the effective targeting of preventive measures.

Existing data shows that Portugal contrasts with the rest of the Western Europe in its distribution of HIV-1 subtypes[257,341]. It is interesting to gain further understanding on the causes underlying this difference also in light of the evidence supporting a recent increase in the infections with non-B HIV-1 subtypes in several Western Europe countries[242,244,339,340]. In this study we have analyzed 289 HIV-1 infected individuals from the Minho province, Portugal. Collectively, the results obtained are consistent with previous studies in Portugal in showing high prevalence of non-B subtypes (73.0%), mainly virus of subtype G (29.4%), followed by the subtype C (14.5%). In our study population, the heterogeneity of HIV-1 subtypes is attributable to Portuguese-born individuals presumed to be infected in the region with only 3.8% of the cases being among immigrants or Portuguese individuals presumed to be infected elsewhere. Contrarily, the rising prevalence of non-B HIV-1 subtypes in Western Europe has been attributed to the growing number of immigrants from Sub-Saharan Africa and South America where these variants are prevalent. As an example, in Spain 27% of the HIV-1 cases diagnosed in 2007 were of non-B subtypes and 90% of these cases were African and South American immigrants[258]. Phylogenetic analysis of our study population indicates that the most prevalent subtypes, B and G, show high inter-individual genetic distances suggesting old and multiple introductions of virus of these subtypes in the region. This observation is in accordance with the fact that B subtype is predominant in Western Europe and was probably introduced in several occasions in the late 1970s and early 1980s[357]. As for subtype G, it is possible that the intense human migrations between Portugal and its former

African colonies in the 1970-80s due mostly to the independence wars[358] contributed to the early introduction others HIV-1 subtypes, namely in the case of the migration connection with Angola where there is a large HIV-1 genetic diversity[269], possibly due to its proximity to Democratic Republic of Congo, the presumed country of HIV-1 origin[41]. The analysis of transmission clusters showed that 39.4% of all sequences grouped in 17 local transmission clusters, with a mean of 6.9 sequences per cluster. The distribution of clustered sequences per subtype showed that almost every F1 sequences (21 out of 22) are incorporated in clusters. In opposition, B and G subtype clusters include only 35.9% and 11.8% of the B and G subtype sequences, which can be considered additional evidence of long time circulation of B and G subtypes among the studied population. Interestingly, we also found clusters of non-B and -G subtypes, namely of the C, F1, A1 and CRF14_BG that have a date of MRCA in the late 90s even when considering only Portuguese born individuals presumed to be infected in the region. This supports that these HIV-1 subtypes have been introduced in Minho more than one decade ago. The two largest transmission clusters, 11C and 17BG, are composed in their large majority (>87%) by IDU reflecting the compartmentalization and closed character of the transmission among individuals from this risk group. The analysis of the date of diagnosis of the individuals in these clusters suggests that its transmission might be decreasing namely in the case of 17BG since in the 6 years ranging from 2006 to 2012 no HIV-1 infection with virus from this cluster were found in the study population. In the last decade, other local epidemics with CRF14_BG have been described among IDU in Spain and Portugal[257,343]. Our results are in line with the data showing a decrease in the prevalence of IDU in among the HIV-1 infected individuals in Portugal[255] suggesting a positive impact of the preventive strategies implemented in the last decade in reducing the transmission of HIV-1 among the IDU population. Importantly, our data support an increased incidence of infection in the study population with F1 and A1 subtypes. The occurrence of events of onward transmission of F1 and A1 HIV-1 subtypes in our study population were strongly linked to sexual transmission. We identified three F1 clusters, mainly formed by individuals who report heterosexual contact as the presumed viral transmission route. The phylogenetic analysis shows that the viral sequences from clusters 13F1 and 14F1 share common ancestors with reference sequences collected in Brazil. The analysis of the public databases allowed the identification of 5 sequences isolated in Italy that belong to the 14F1 cluster, thus suggesting a large geographic range of this transmission network contrarily to what was found in all the other clusters. In common with the 13F1 and 14F1 clusters the previously described Italian cluster was also phylogenetically linked to Brazil[356]suggesting to this country as the origin of

introduction of subtype F1 virus in southern Europe. We also identified one sexually transmitted A1 subtype cluster that was epidemiologically linked to Mozambique since one of the individuals in the cluster is Mozambican and presumed to be infected prior emigrating to Portugal. Despite having MRCA dates similar to 11C and 17BG, F1 and A1 subtype clusters have a more recent mean date of diagnosis and include sequences diagnosed in the last three years suggesting that these clusters might still be engaged in active transmission.

Overall our study of local HIV-1 epidemic in the Portuguese region of Minho supports that these region contrasts with the rest of the Western Europe in its HIV-1 subtype distribution due to established transmission among native individuals of non-B subtypes. Our molecular and epidemiologic analysis highlight increasing incidence and onward transmission of F1 and A1 subtype virus via sexual transmission routes supporting the need for continuous monitoring and strengthening of preventive strategies targeted at these transmission modes.

CHAPTER 3: VIRAL DIVERSITY IN STANDARD HIV-1 SEQUENCES: CONTRIBUTION OF AMBIGUITIES TO DIFFERENTIATE RECENT FROM CHRONIC INFECTION

Chapter 3: VIRAL DIVERSITY IN STANDARD HIV-1 SEQUENCES: CONTRIBUTION OF AMBIGUITIES TO DIFFERENTIATE RECENT FROM CHRONIC INFECTION

Abstract

Introduction: An individual infected with human immunodeficiency virus type 1 (HIV-1) generally harbors a mutant cloud of related variants, usually arising from a single founder virus. Intra-host viral diversity has been reported to increase with time. Ambiguous sites are identified in sequences obtained in the regular follow-up of HIV-1 infected patients and have been proposed as markers of viral diversity. As diversity increases over time, it is possible to estimate age of infection based on these ambiguities.

Objective: To evaluate the contribution of ambiguous sites in HIV-1 sequences routinely obtained, associated with CD4+ cell counts and AIDS stage, as indicators of age of infection in HIV-1 infected persons attended in Hospital de Braga, Portugal.

Methods: For 203 antiretroviral naive patients who had at least one HIV-1 sequence available, the proportion of ambiguous sites (PAS) was calculated and its association with duration of infection was investigated. Evaluation of this association was made using a subset of patients with confirmed recent or chronic infection.

Results: PAS was three times greater in subjects with chronic infection (Md=1.31%) than in those recently infected (Md=0.44%; $p<0.0001$). PAS was negatively correlated with CD4+ cell count ($p<0.0001$) and positively correlated with the existence of an AIDS defining condition ($p=0.034$). An algorithm based in CD4+ cell count, AIDS stage and PAS can discriminate chronic from recent infection (<1 year) with a specificity of 96% and a negative predictive value of 0.89.

Conclusion*:* There is a presently underestimated value for sequence ambiguities as an expression of intra-host viral diversity in estimating time of infection, especially when conjugated with clinical and epidemiological data.

Key-words: HIV, Ambiguous Sites; Viral Diversity; Time of infection; Incidence; CD4+ cell count.

Introduction

Identification of recent HIV infection within populations is of utmost importance as it makes possible the incidence evaluation and state of the HIV epidemic, as well as improving treatment outcomes and controlling HIV transmission. Nevertheless, the *2012 European HIV/AIDS Surveillance Report* found that 50% of new HIV infection diagnoses were of late presenters ($CD4^+$<350/mm^3), including 30% with advanced HIV infection ($CD4^+$<200/mm^3)[359]. Time of infection is one of the most poorly defined parameters in epidemiologic studies, mainly because the gold standard, prospective follow-up of HIV-uninfected people, is expensive, time-consuming and logistically demanding. Determining HIV incidence rates through longitudinal follow-up of uninfected people is also susceptible to several biases, such as the possible loss to follow-up of those at most risk of infection or the repeated testing leading to behavior changes [360,361].

Estimation of the incidence rate through inference from prevalence measurements and mortality data can also be done but it also has its limitations such as the challenge to account for internal and international migration [362]. Yet another method, based in identification of patients with acute retroviral syndrome, is fallible because such syndrome may have not occurred or may not have been recognized.

Furthermore, laboratory methods have been developed that rely on the antibody levels and their avidity as indicators of recent infection. They are founded on the gradual response of the host immune system to the infection, which will result in recent seroconverters to test below a defined level during a post-seroconversion period window [363]. Although their accuracy has been improved with their inclusion in multi-testing algorithms, they still present limitations, such as the misclassification of infections as recent, especially in advanced disease when antibody levels tend to decrease, and the different window periods for different subtypes which creates challenges in their application in populations with multiple viral subtypes [364]. In western and central Europe the most prevalent subtype is subtype B, representing 85.20% of the individuals with HIV-1 infections[359]. Portugal presents itself as special case with one of the lowest proportion of subtype B infections in Europe (39.2%), being the remaining individuals infected with non-B subtypes, mainly subtype G [365].

Up to 80% of HIV infections are monomorphic, initiated by a single virus [134]. But as an RNA virus, HIV has a complex evolution dynamic due to several factors, including high viral

turnover [366], high mutation rates [367], retroviral recombination [368] and selection pressure from the host immune system [172]. In a fast pace, the initial virion starts to diverge and soon the viral population of an infected individual is not represented by one single genome, but rather by a mutant cloud of non-identical but closely related viral variants that continuously goes through genetic variation and competition. As repositories of distinct viral variants, mutant clouds are the source of virus adaptability, acting as a unit of selection [369].

It has been reported that viral diversity increases with the age of infection, first in a linear fashion but then at decreasing rates until a plateau is reached. Eventually, the immune system collapses and, as progression to AIDS begins, viral divergence stabilizes and viral diversity declines [370]. How can this diversity be assessed in a way pertinent to estimating age of infection?

Since 2003, recommendations exist to perform standard genotyping of HIV-1 protease and reverse transcriptase in all newly diagnosed infections, so antiretroviral resistance can be detected prior to therapy initiation [371]. This practice originated abundant genetic material, mainly from *pol* gene, where reverse transcriptase and protease are codified. For economical reasons, these sequences are obtained by bulk sequencing; that is, the sequencing procedure is applied to a diverse sample of the HIV population. If the frequency of the most frequent nucleotide at a given position exceeds a threshold (typically around 80%), bulk sequencing returns the predominant nucleotide at this position. However, if this is not the case, then so called ambiguous nucleotide calls are reported, implying that the patient harbors viral strains with different nucleotides at this locus. Thus the fraction of ambiguous nucleotides is a measure of the degree of polymorphism of the HIV population within a patient. Therefore, the use of these ambiguous sites (also called ambiguities, ambiguous nucleotide calls, degenerate bases or mixtures) as a marker of genetic diversity and hence length of infection, constitutes a simple approach to determine the age of infection and additionally it's inexpensive, as it uses sequences that are already routinely requested and available in clinical practice. Protease and reverse transcriptase sequencing is available in the Hospital de Braga, Portugal, since 2005 and in 2006 became part of the standard patient care for every newly admitted HIV-1-infected patient.

Methods

The population consisted of all the HIV-1 infected and treatment naive patients followed in the Hospital de Braga, an university affiliated hospital with 700 beds, serving as reference hospital to the Northwest region of Minho in Portugal (pop.: 1,093,021), between July of 2005 and April of 2014 with at least one HIV-1 *pol* gene sequencing in that period. Individuals without any HIV sequence or unavailable clinical data were excluded from the study. A total of 203 patients were selected.

Sequencing was done using TRUGENE® *HIV-1* Genotyping Kit (Siemens Medical Solutions Diagnostics, Tarrytown, New York, USA) and the OpenGene™DNA Sequencing System (Siemens Healthcare Diagnostics, Tarrytown, NY). The sequences represent a real time polymerase chain reaction product of part of the protease (codons 4 to 99) and part of the reverse transcriptase (codons 38 to 247) region of HIV-1 *pol* gene. The shortest sequence had 909 base pairs and the longest 918 base pairs.

Sequences were aligned and trimmed with reference HIV-1 *pol* sequences from the Los Alamos National Laboratory using ClustalW available in MEGA5 software [372] and the REGA version 3 HIV-1 Subtyping Tool was used to identify each sequence subtype [348].

Ambiguous sites were considered when an ambiguous base pair code was registered in the sequence instead of a normal nucleotide. The codes considered were defined by the International Union of Biochemistry : R (A or G), Y (C or T), K (G or T), M (A or C), S (G or C), W (A or T), B (C, G or T), D (A, G or T), H (A, C or T), V (A, C or G) and N (A, C, T or G) [373]. PAS was then calculated by dividing the number of ambiguous sites by the total base pair number of each sequence.

Time since HIV infection diagnosis was considered as the period of time between the date of the diagnosis of infection and the date of sampling for the first sequencing available for each patient.

Patients were empirically considered chronically infected (>1 year) if time since HIV diagnosis was superior to 12 months, if $CD4^+$ T-cell T count was inferior to 200 cells/mm^3 by the time of sequencing, as Lodi and collaborators demonstrated this count is related to infections lasting around 8 years [374], or if the patient had AIDS diagnosis at first sequence (S1) obtained (n=118). Otherwise patients were classified as having recent infection. They

were further divided by mode of transmission and by HIV-1 subtype. In 129 patients a recent (n=40) or chronic infection (n=89) status could be well-established based on epidemiological and clinical data and in the presence of an HIV negative test prior to a positive one.

Presumed chronically infected patients were compared with presumed recently infected ones regarding PAS, CD4+ cell count and existence of an AIDS defining condition.

To study the association of PAS with the length of infection, correlation analyses was performed between PAS at S1 and two known indicators of duration of infection: the presence or absence of AIDS criteria and the $CD4^+$ T-cell T count. For this correlation analysis individuals for whom S1 was obtained after eight years of HIV diagnosis were excluded (n=12) as it also has been reported that the increase in PAS is almost linear within 8 years of infection but subsequently it starts to stabilize and even decrease [375]. Afterwards, a sub-analysis was done to investigate the relationship between PAS and $CD4^+$ T-Cell count according to the mode of HIV acquisition and subtype.

On variables with a non-normal distribution non-parametric tests were used: Mann-Whitney U Test to compare independent samples and Spearman's Rank Correlation Coefficient to study correlation between variables.

Data was collected and registered in Microsoft® Excel® 2010 (©2010 Microsoft Corporation) and statistical analysis and graphical presentation was done using GraphPad Prism version 6.04 for Windows, GraphPad Software, La Jolla California USA. All *p* values were considered to be statistically significant if <0.05.

Ethical Considerations

The study was approved by the Health and Science of Life Ethics' Subcommittee of Minho's University and the Administration Board and Health Ethics' Committee of the Hospital of Braga. Written consent was obtained for all the patients enrolled in the study. Clinical data was codified to ensure their confidentiality.

Results

Characterization of study participants

Demographic and clinical parameters of every subject at sequence (S1) sample collection are represented in Tables 5 and 6 and in Figure 22. Most individuals were men (78.8%) of white ethnicity (96.6%), infected in Portugal (89.5%). 77.3% acquired HIV through sexual intercourse and 21.7% through injection drug use (IDU).

Total n=203	n (%)
Gender	
Male	160 (78.8)
Female	43 (21.2)
Ethnicity	
White	196 (96.6)
Other	7 (3.4)
Disease Stage (on sequencing)	
AIDS	85 (41.9)
No AIDS	118 (58.1)
Mode of Transmission	
Heterosexual	131 (64.5)
IDU	44 (21.7)
MSM	26 (12.8)
Others[a]	2 (1.0)
Age on Infection (years)	
≤20	14 (6.9)
21-40	113 (55.7)
41-60	64 (31.5)
>60	12 (5.9)

Table 5: Demographic characterization of study subjects.

[a]- vertical, transfusional

	Median	1st Quartile-3rd Quartile
Nº of ambiguities	8	1.0-15.8
Proportion of ambiguous sites (PAS) (%)	0.87	0.44-1.72
Time Since HIV-1 diagnosis (months)	1.6	0.7-19.2
CD4+ count (cells/mm^3)	264	116-428

Table 6: Characteristics of potential indicators of duration of infection by the time of sequencing (n=203)

An AIDS defining condition was present by the moment of sequencing in 41.9% of the subjects. Considering the moment when HIV infection diagnosis was made, study participants had a median age of 37 years, ranging from 15 to 82 years.

The most common HIV-1 subtype identified at S1 was G (26.6%), followed by B (23.6%) and C (12.3%). CRF02_AG was the most frequent circulating recombination form recognized (3.0%). A unique recombinant form was present in 7.4% of the patients.

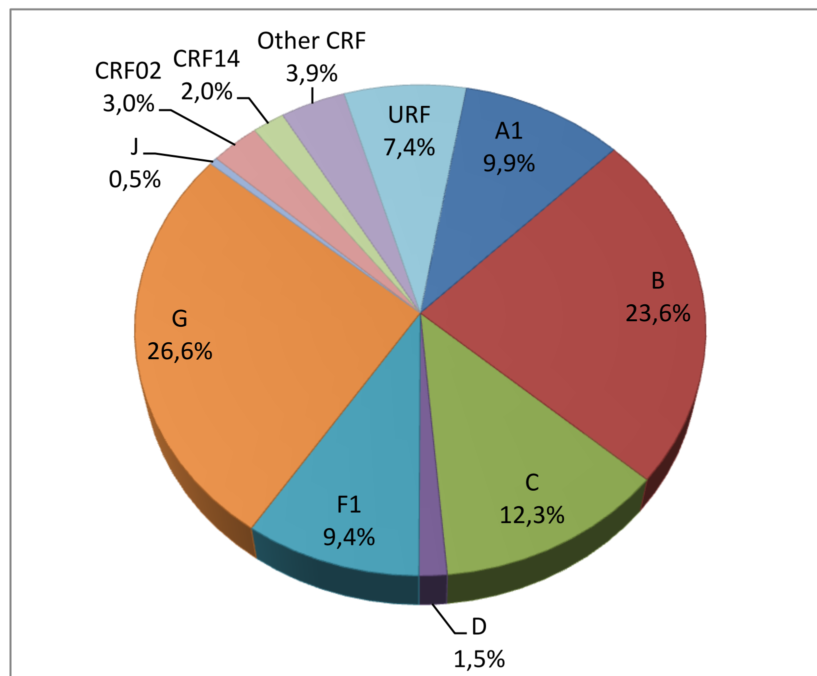

Figure 22: HIV-1 subtype distribution for the 203 sequences available

Diversity evaluation

The median value for PAS at S1 was 0.87% but Mann-Whitney test revealed a significant difference between patients classified as having a potential recent infection (n=85, median=0.44%) and those with chronic infection (n=118, median=1.31%; $p<0.0001$) (Table 7).

	Median	1st Quartile-3rd Quartile	CI (95%)	P value
Presumed recent	0.436	0.109-0.871	0.477-0.915	< 0,0001
Presumed chronic	1.309	0.763-2.097	1.342-1.668	

Table 7: Comparison between PAS of presumed recent and presumed chronic HIV infections

The median time from diagnosis of HIV infection to sequence sampling was 1.6 months, with an interquartile range of 18.6 months.

Assessment of age of infection

To further study the association of PAS at S1 with factors traditionally related to length of infection, a correlation analysis was performed between PAS and three different variables at S1 sampling: existence of an AIDS defining condition, CD4+ cell count and estimated time of infection, here defined as time in months passed between diagnosis and sequencing. As mentioned previously, for this analysis, individuals with time of infection above 96 months were excluded. A negative correlation was demonstrated between the PAS and $CD4^{+}$ T-Cell Count (r_s=-0.418, n=203, p<0.0001) and a positive one between the PAS and the time since diagnosis ($r_{s=}$0.194, n=203, p=0.0056), as shown in figure 23. The existence of an AIDS condition on the moment of sequencing correlated positively with PAS (r_s=0.223, n=85, p=0.034).

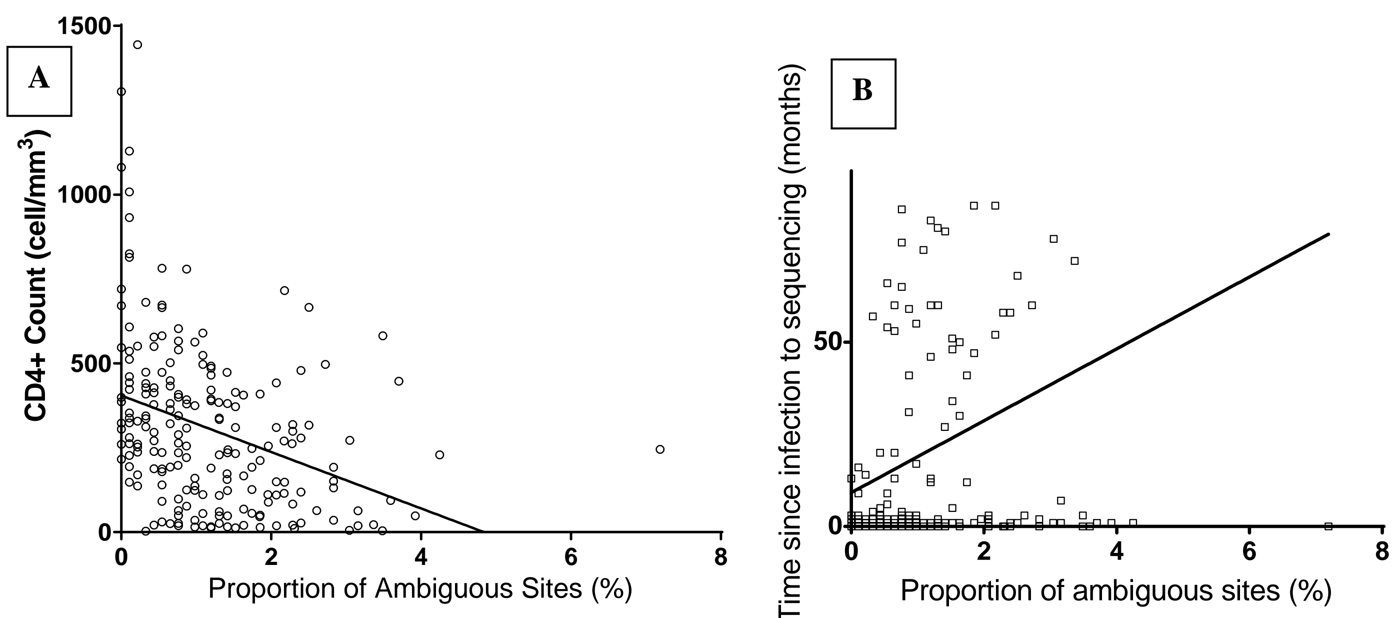

Figure 23: Linear regression of the correlation between (A) CD4+ count, (B) time since infection to sequencing and proportion of ambiguous sites

Trying to establish a relation between PAS, transmission modes and HIV-1 subtypes, a polynomial regression model failed to demonstrate any statistically significant difference (Table 8). To further investigate the relationship between PAS and CD4+ cell count, a correlation analysis was performed using the three most important transmission modes

(representing 99% of total) and the five more prevalent HIV-1 subtypes and URF (altogether accountable for 89.2% of infections). No significant correlations were also found between PAS and CD4+ cell counts regarding transmission mode or HIV-1 subtype.

N	203
R^2	0,291
R^2 adjusted	0,205
SE offit (RMSE)	0,914579

Parameter	Estimate	95% CI	SE	p-value
Transmission: Heterosexual	-0,4786	-1,025 to 0,06729	0,27667	0,0853
Transmission: IDU	-0,1167	-0,7160 to 0,4826	0,30372	0,7012
Transmission: MSM	-0,1525	-0,7611 to 0,4561	0,30844	0,6217
SUBTYPE A1	0,1773	-0,3128 to 0,6674	0,24836	0,4762
SUBTYPE B	0,3295	-0,05789 to 0,7170	0,19634	0,0950
SUBTYPE C	-0,06062	-0,5328 to 0,4115	0,23929	0,8003
SUBTYPE F1	0,1553	-0,3560 to 0,6666	0,25912	0,5498
SUBTYPE G	0,1651	-0,2213 to 0,5514	0,19580	0,4003
URF	0,04621	-0,4853 to 0,5777	0,26934	0,8640

Table 8: Relations between PAS, transmission modes and HIV-1 subtypes analyzed by polynomial regression model. IDU: intravenous drug user; MSM: men who have sex with men; SE: standard error; CI: confidence interval

To evaluate the capacity of PAS to discriminate between recent and chronic infections, a receiver-operator characteristic (ROC) curve was constructed. With a PAS value<0.544 we obtained a specificity of 91% (95% CI=82.4-96.3; likelihood ratio=3.171)

Based on these results, we proposed a three steps algorithm (Figure 24): candidates to be infected for less than 12 months must have a CD4+ cell count superior to 200/mm^3, no AIDS defining condition and a PAS inferior to 0.55.

Evaluation of algorithm performance

In 40 patients, a more accurate recent infection condition could be acknowledged, based in time of seroconversion, as each one have a HIV negative serology followed by a HIV positive test in less than 12 months. Furthermore, in 89 subjects a time of infection longer than 12 months could be established, based in epidemiological and serological data. There were no significant differences between these two sets of patients and the presumed recently or chronically infected ones, in Mann-Whitney test. Table 9 confirms a statistically significant difference between truly recent and truly chronic infections regarding PAS. This

	Median	1st Quartile-3rd Quartile	CI (95%)	P value
True recent infections	0.218	0.109-0.599	0.270-0.724	< 0,0001
True chronic infections	1.412	0.763-2.175	1.350-1.714	

Table 9: Comparison between PAS of truly recent and truly chronic HIV infections

subset of patients was used to assess the performance of the proposed algorithm. In recognizing a recent infection, the proposed algorithm achieved a sensibility of 0.73, a specificity of 0.96, a negative predictive value of 0.89 and a likelihood ratio of 16.

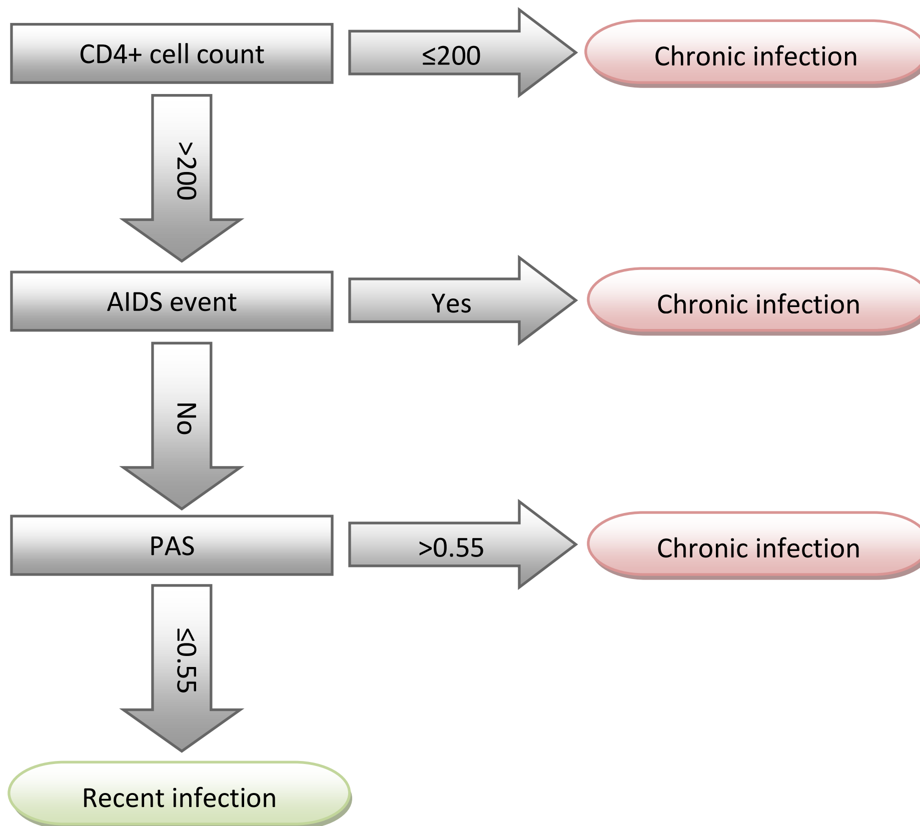

Figure 24: Proposed algorithm to discriminate between chronic (>12 months) and recent (≤12 months) based on CD4+ cell counts, presence of an AIDS defining event and proportion of ambiguous sites

Discussion

Lack of accurate methods to discriminate recent from chronic HIV-1 infections is a major obstacle for measuring HIV-1 incidence, which is essential for monitoring transmission dynamics in order to design, optimize and evaluate intervention programs. One other utility for identification of recent infections will be awareness of transmitted antiretroviral resistance. Knowing that around 80% of HIV-1 infections are founded by a single virus [134] that will steadily diverge in the course of infection, how can we evaluate diversity and how can we apply such evaluation to estimating age of infection? No perfect method has been developed yet. However, ambiguous sites in the resistance detection requested protease/reverse transcriptase sequences, an abundant source of genetic material, are increasingly being recognized as indicators of viral diversity in HIV-1 and, consequently, as a potential marker of age of infection. This study investigated this potential role in 203 HIV-1 infected patients followed in the Hospital de Braga with at least one HIV-1 sequence available. As shown, the median PAS at S1 was almost three times greater in patients with presumed chronic infect than in patients with presumed recent infection, a highly significant difference ($p<0.0001$).

As described, the sample in this study was constituted by multiple subtypes which seem to be consistent with other reports about subtype prevalence in Portugal [376]. The significant correlations between PAS, presumed length of infection and CD4+ cell count were maintained regardless of HIV-1 subtype. Same was true when we studied different transmission modes: PAS correlations with presumed length of infection and CD4+ cell count were independent of route of infection.

These finding are in line with previous studies and support the evidence that PAS, as a surrogate of viral diversity, relates to the duration of HIV-1 infection.

In several studies, the estimated time between seroconversion and reaching a CD4+ cell count inferior to $200/mm^3$ varied from 6.2 [377] to 7.6 [378] or 7.9 years [374]. It seems rational to use this threshold as first step to discriminate recent from chronic infections. Every patient with an AIDS condition was classified as chronically infected, even bearing a CD4+ cell count$>200/mm^3$, as every individual presenting with an AIDS-defining event, regardless of the CD4+ cell count is considered a late presenter [379]. That is the rationale for the second step.

In this study there were available a subset of patients in whom we could differentiate time since infection in two categories: under a year and over a year. On examining its performance with this group of subjects, our logarithm exhibited a low rate of false positives (specificity of 0.96) and a negative predictive power of 0.89, meaning it has a good discriminatory power to classify an infection as non recent. A high rate of false negatives (sensibility=0.73) can be attributed to the fraction of infections in which the number of virus leading to productive clinical infection is multiple, driving to higher diversity in recent infections. In our sample, 27.5% of recent infections (9/40), with a mean estimated time of infection of 7.3 months, were misclassified as chronic because of PAS>0.55% (median =0.986). All but one had a CD4+ cell count above 500/mm^3 so diversity should be corrected for recombination events in infections with multiple variants. In the Swiss Zurich Primary HIV Infection Study, 18% (24/130) of patients genotyped within a month from infection had >0.68% ambiguous bases in *pol*, suggesting infection with several founder virus as well [375].

There is a relative inefficiency of virus transmission by sexual routes. In infection following intravenous inoculation, as in IDU, due to the absence of a mucosal barrier, it is more likely to find a higher frequency of multiple-variant transmission and therefore an earlier and superior diversity [145]. Kouyos and collaborators found more genetic diversity in early infection amongst patients infected by a intravenous route compared to infection acquired sexually [375]. In our sample, PAS of presumed recently infected IDU (n=11; median=0.55) is higher than PAS of presumed recently infected subjects by sexual transmission (n=73; median=0.33), although this is not a significant difference probably due to limitations caused by a small sample.

In Braga (and in Portugal) the coexistence of multiple subtypes has the potential to expand the scope of HIV investigation, traditionally concentrated on subtype B, the one with the greatest prevalence in Western Europe and North America. Attention to the importance of an increasing number of recombination forms and to the diversity of the global pandemic make clear the importance of studies that take into account multiple subtypes.

An advantage of our approach is that only requires biological samples obtained in the standard follow-up of HIV infected patients and does not require follow-up cohorts, although it uses a multiple parameter approach in order to improve accuracy. This means easy implementation and no added costs. It has a robust performance to factors such as viral subtype and transmission route. To be considered an optimal algorithm, accuracy must be

improved mainly by much more detailed analysis of diversity which could be achieved with pyrosequencing or ultra deep sequencing. Anyhow, inherent limitations of this methodology suggest that it will be more useful in contributing to population-level information on HIV-incidence than to assess individual length of infection.

As supplemental major limitations of this study, it must be referred the small sample size considered and the use of an empirical definition as an initial discriminator for chronic or recent infection.

Conclusion

This study demonstrates the existence of a relationship between the proportion of ambiguous sites with the age of infection, the $CD4^+$ T-cell count and the AIDS status in ART naïve patients in a population with multiple subtypes, with a higher PAS being associated with a longer duration of infection. This finding served to propose an easy implementable algorithm to help differentiate HIV-1 sequences as being from recently (≤1 year) or chronically (>1 year) infected patients.

CHAPTER 4: DRUG RESISTANCE, TRANSMISSION CLUSTERS AND POLYMORPHISMS IN A LOCAL HIV-1 EPIDEMIC

Chapter 4: DRUG RESISTANCE, TRANSMISSION CLUSTERS AND POLYMORPHISMS IN A LOCAL HIV-1 EPIDEMIC

Abstract

Objective: To characterize HIV-1 transmitted drug resistance (TDR) and to assess its relations with transmission networks and respective dynamics in a circumscribed geographical area.

Methods: Between 2005 and 2012, 289 HIV-1 positive individuals followed in Braga, Portugal, had their virus sequenced and analyzed using the WHO list of mutations for surveillance of TDR. Non drug exposed sequences were categorized as recent infections, newly diagnosed infections of unknown duration or previously diagnosed infections of unknown duration. Clusters of at least 3 members were identified based on a maximum likelihood bootstrap support > 95% and Bayesian techniques.

Results: 17 sequences revealed TDR, representing a prevalence of 9.4% (95% confidence interval, 5.6% - 14.7%), comprising 3.3% resistant to nucleoside reverse transcriptase inhibitors, 3.9% resistant to non-nucleoside reverse transcriptase inhibitors and 3.3% resistant to protease inhibitors. No dual class resistance was detected but triple class resistance was identified in one patient. No predictor associated with TDR was found. Phylogenetic analyses revealed 17 transmission clusters, involving 114 out of 289 sequences and ranging in size from 3 to 31 members. 11 of these clusters comprised drug-resistant strains, including one cluster in which 2 out of 3 patients were infected with a strain carrying both K101E and M184V mutations. Intravenous drug use and non-B non-G subtypes were correlated with clustering. MSM exposure was positively related and IDU exposure was negatively related with cluster average growth. A98S polymorphism was present in 95.3% of subtype G virus and protease codon 35 insertion E35E_T was found integrating a cluster of subtype C, in a prevalence larger than previously described.

Conclusions: In Braga, level of transmitted drug resistance is similar to other European regions and largely involved in transmission chains. This study failed in demonstrating independent predictors for TDR in transmission clusters, reinforcing the utility of universal sequencing at admission.

Key Words: Transmission clusters; transmitted drug resistance; Cluster dynamics; phylogeny; polymorphisms; HIV subtypes; protease codon 35 insertion

Introduction

Transmitted or primary drug resistance, defined as the existence of drug resistance mutations (DRM) in individuals never before exposed to antiretroviral therapy (ART) is an important concern to public health because it jeopardizes the response to such therapy. In regions where ART is widely available, transmitted drug resistance (TDR) is usually the outcome of virus transmission between ART experienced and ART naïve people. Prevalence of TDR varies geographically and ranges from 5.6% in Sweden[380] to 14.6% in United States of America[381]. In Europe, the multicentric SPREAD study reported a TDR prevalence of 11.1% in men who have sex with men from 2002 through 2007, revealing a significant difference in TDR according to route of transmission[382]. Temporal stage of HIV infection may also influence viral transmission and contribute to propagation of primary drug resistance. Early stages of infection are usually associated with higher viral load and unawareness of HIV status and so may disproportionally contribute to TDR[346]. An important role for transmission networks in dissemination of drug resistance is increasingly recognized[346,383,384]. Understanding the HIV-1 transmission patterns becomes crucial to optimize prevention and control of the epidemic and reconstruction of transmission networks can provide valuable insights in the spread of the virus[385-387]. Until a decade ago, it was a task based almost exclusively in interview data collected from the patients. Since 2003, international guidelines recommend baseline testing for drug resistance in all HIV-1 infected patients[388]. This has led to a substantial increase in the availability of viral sequence data and allowed a new approach to study the HIV epidemiology, phylogenetic analysis[345]. This technique allows the identification of mutual characteristics of clusters, i.e. specific groups of patients in which multiple transmissions of HIV-1 have taken place. Studies using phylogenetics based on the *pol* gene of HIV were performed throughout the world to map local HIV epidemics in correlation with transmission pathway, drug resistance, risk behavior and cluster size. Some focused on the contribution of primary infection to onward transmission[346], while others investigated the transmission of drug resistant virus[389-391] or concentrated on specific populations[347,392]. Generally, these studies are centered on the predominant subtype or the most predominant route of transmission, with little information on the other circulating subtypes or transmission routes. In Braga, the local specificities in subtype diversity and patterns of transmission provided a chance to correlate the presence of transmission networks as established by the genetic relationship of the virus, with information on demographics, transmission mode, CD4

counts and the presence of drug resistant virus. In addition, there are several studies indicating regional differences regarding what fuels HIV-1 local epidemics[393,394], motivating the attainment of a better insight in the dynamics of the infection and transmitted drug resistance in this specific geographical area.

Methods

Study population

Hospital de Braga is a university affiliated hospital serving as reference hospital to Minho region, in the Northwest of Portugal (population: 1093021). It provides care to over 900 patients infected with HIV. Between 2005 and 2012, 289 patients older than 18 years had their virus sequenced and were included in the study. This represents 73.9% of all HIV-1 infected persons admitted in the hospital in that period.

Characterization of disease stage

By reviewing clinical records, previous exposure to antiretroviral therapy was confirmed in 109 patients – the ART experienced group. The drug-naïve cohort of 180 patients were further categorized in recent (≤ 1 year) infections (n=26), newly diagnosed patients (sequenced in the first 12 months after diagnosis) with infection of unknown duration (n=98) and previously diagnosed patients with infection of unknown duration (n=56).

An individual fulfilling the following criteria was considered recently infected: an interval between the last negative HIV serology and sequencing inferior to 12 months, CD4+ count>200 cells/mm^3 and absence of any AIDS defining condition.

RNA extraction, amplification and sequencing

Viral RNA was extracted using Magna Pure Total Nucleic Acid Isolation Kits (Roche Applied Science). RT-PCR and DNA sequencing were performed with Trugene HIV-1 Genotyping System (Siemens Healthcare Diagnostics). The sequenced regions include part of the coding sequences of *gag* (492 to 501), p6 (44 to 53), *pol* (60 to 402), p2p7p1p6 (129 to 138), Protease (4 to 99) and RT (1 to 127, reported positions are amino acid positions relative to protein start in the HXB2 reference genome, GenBank: K03455.1). The subtyping of the 289 sequences was made using REGA 3.0[348] and non-automatic phylogenetic analysis. Non-automatic bootscan analysis was also done with the program SimPlot to confirm selected subtypes using the F84 nucleotide substitution model and a sliding window of 200-bp, a 40-bp step[232]. Detection of recombination was confirmed using the program RDP[349]. Sequences were uploaded to GenBank and assigned the following accession numbers: KM205831-KM206119.

Drug resistance mutations (DRM)

Definitions accepted by World Health Organization for drug resistance surveillance were adopted, according to three criteria: (i) mutations should cause or contribute to drug resistance, (ii) mutations should not occur as polymorphisms in the absence of therapy and (iii) mutations should be identified in most common group M subtypes. The following resistance mutations were scored: to nucleoside reverse transcriptase inhibitors (NRTI): M41L, K65R, D67N/G/ E, T69D/insertion, K70R/E, L74V/I, V75M/T/A/S, F77L, Y115F, F116Y, Q151M, M184V/I, L210W, T215Y/F/I/S/C/ D/V/E, K219Q/EN/R; to non-nucleoside reverse transcriptase inhibitors (NNRTI): L100I, K101E/P, K103N/S, V106A/M, V179F, Y181C/I/V, Y188C/L/H, G190A/S/E, P225H, M230L; and to protease inhibitors (PI): L23I, L24I, D30N, V32I, M46I/L, G48V/M, I50L/V, F53L/Y, I54V/L/M/A/T/ S, G73S/T/C/A, L76V, V82A/T/F/S/C/M/L, N83D, I84V/A/C, N88D/S, L90M. Every mutation had its criteria confirmed on HIV Drug Resistance Database, from Stanford University (available online on hivdb.stanford.edu). Transmitted drug resistance (TDR) was defined as drug resistance in previously untreated persons. As drug resistance rarely occurs without antiretroviral exposure, TDR implies that a virus with DRM was transmitted, either directly or through intermediates, from a person with acquired drug resistance. Polymorphisms were excluded, although some of it confers diminished susceptibility to antiretrovirals, because they represent mutations emerging frequently in virus not exposed to selective drug pressure. Only mutations conferring resistance to protease inhibitors (PI), nucleoside reverse transcriptase inhibitors (NRTI) and non nucleoside reverse transcriptase inhibitors (NNRTI) were considered, as integrase strand transfer inhibitors were infrequently used during the study period.

RNA extraction, amplification and sequencing

Viral RNA was extracted using Magna Pure Total Nucleic Acid Isolation Kits (Roche Applied Science). RT-PCR and DNA sequencing were performed with Trugene HIV-1 Genotyping System (Siemens Healthcare Diagnostics). The sequenced regions include part of the coding sequences of *gag* (492 to 501), p6 (44 to 53), *pol* (60 to 402), p2p7p1p6 (129 to 138), Protease (4 to 99) and RT (1 to 127, reported positions are amino acid positions relative to protein start in the HXB2 reference genome, GenBank: K03455.1). The subtyping of the 289 sequences was made using REGA 3.0[348] and non-automatic phylogenetic analysis. Non-automatic bootscan analysis was also done with the program SimPlot to confirm selected subtypes using the F84 nucleotide substitution model and a sliding window of 200-bp, a 40-

bp step[232]. Detection of recombination was confirmed using the program RDP[349]. Sequences were uploaded to GenBank and assigned the following accession numbers: KM205831-KM206119.

Phylogenetic analysis and identification of clusters

The 289 HIV-1 sequences obtained in this study and 88 sequences from the databases including the M group consensus and a previously defined set of subtype reference sequences[348] including at least two reference sequences from each M group subtype (A1, A2, B, C, D, F1, F2, G, H, J and K) and from 26 CRF (CRF01_AE, CRF02_AG, CRF03_AB, CRF04_CPX, CRF05_DF, CRF06_CPX, CRF10_CD, CRF11_CPX, CRF12_BF, CRF13_CPX, CRF14_BG, CRF18_CPX, CRF19_CPX, CRF20_BG, CRF24_BG, CRF25_CPX, CRF27_CPX, CRF29_BF, CRF31_BC, CRF35_AD, CRF37_CPX, CRF39_BF, CRF40_BF, CRF42_BF, CRF47_BF) were aligned using MUSCLE[350]. The phylogenetic analysis of the 377 sequences was conducted using RAxML 7.0.3 to produce a maximum likelihood tree using 1000 bootstrapping replicates[351]. Analysis was repeated with PhyML[352] computing the aLRT support of all tree branches and by Bayesian analysis using BEAST[353]. The best fitting nucleotide-substitution model for the Bayesian analysis was estimated using jModeltest v2.1.2[354] to be the general time reversible (GTR) model with a proportion of invariant site (I) and gamma distribution of rates (G), selected among 88 different models according to the Akaike Information Criterion (AIC), the Bayesian Information Criterion (BIC), and the Decision Theoretic Framework (DT). An eventual bias introduced by convergent evolution due to the presence of drug resistant mutations was discarded by repeating the analysis after removal of codons associated with drug resistance in the standardized list of mutations for surveillance of transmitted drug resistance established by the World Health Organization[355]. The general topology of the trees and identification of clustering remained unchanged. Clusters of at least three individuals and pairs of transmission were identified based on a ML bootstrap support > 95%, a Bayesian posterior probability >0.95.

Evaluation of cluster dynamics

For each large cluster (≥5 elements) cluster growth factor was defined for the interval 2008-2012 as the number of new sequences per sequence present in 2008 by the formula: ————, where *n2012* is the number of sequences present in 2012, *n2008* is the number of sequences present in 2008. For example, if a cluster contains 2 sequences in 2008 and 11

in 2012, the growth factor is—— which represent the addition of 4.5 sequences for each original seed member.

Statistical analysis

Comparisons between groups were analyzed using the chi-squared test for categorical variables or Fisher's exact test if assumptions to apply the chi-squared were violated, and the Mann-Whitney nonparametric test for continuous variables. The chi-squared test for trend was used to assess linear trend of drug resistance between 2005 and 2012. Logistic regression was performed to assess the predictive factors on the likelihood to belong to a transmission cluster. Statistical analysis and graphical presentation was done using GraphPad Prism version 6.04 for Windows, GraphPad Software, La Jolla California USA. All *p* values were two-sided and considered to be statistically significant if <0.05.

Ethical considerations

The project was approved by the Ethics Committee of the Hospital de Braga. Written consent was obtained for all the patients enrolled in the study. All clinical and laboratorial data were anonymized prior to analysis to ensure confidentiality of the patients.

Results

Study population

Overall, 76.8 % of cases were male with a median age of 34.0 years (interquartile range (IQR), 27.0-42.3). 94.8% of the patients were of white ethnicity and 90.0% were of Portuguese nationality. No demographic significant differences were found when the study population was compared to the whole HIV-1 infected population attending Hospital de Braga in the same period. However, when compared to Portugal HIV infected population, this cohort have a higher prevalence of heterosexual transmission (55.0%, 95% CI 49.3-60.7 vs. 43.3%, 95% CI 41.8-44.7; p<0.0001) and a lower prevalence of MSM (9.3%, 95% CI 6.5-13.3 vs. 13.7%, 95% CI 12.7-14.8; p=0.027). The median CD4+ cell count in initial assessment was 303 cells/mm^3 (IQR, 145.7-524.0) and the median initial viral load was 35890 copies/ml. Based on *pol* gene sequencing, subtypes distribution showed subtype G as the most frequently found (29.4%), followed by subtype B (27.0%) and subtype C (14.5%). The other subtypes present were F1 (7.6%), CRF14_BG (5.2%), A1 (4.2%) and CRF02_AG (1.4%). Others, including D, J, and various CRFs and URFs represented 10.7%.

Levels and trends of overall HIV-1 drug resistance

People who were ART experienced had significantly higher rates of DRM (50/109, 45.9 %) than naïve patients (17/180, 9.4%; p<0.0001). The rate of drug resistance was also significantly higher in individuals with unknown duration of infection compared with those infected less than one year (24.5% vs. 10.7%; p=0.035). That difference is clearer in patients surely infected more than a year before sequencing, who have higher probability of carrying virus with DRM (31.9% vs. 12.4%; p<0.0001). However, differences between resistance rates according to time of infection were not significant if classes of antiretrovirals were considered separately. Patients with resistance mutations had a lower CD4+ cell count (median of 214 vs. 378; p=0.018) and no significant differences in sex, age, route of transmission and viral load. Frequency of resistance to any drug was higher in subjects infected with subtype B when compared with those infected with non-B subtypes, but not in a significant way (35.8% and 24.3% respectively; p=0.063). No significant difference was observed in resistance frequency between subtypes according to drug class (table 10).

	With any DRM (n=67)	With no DRM (n=222)	P value
Gender			
Male	49 (73.1%)	173 (77.9%)	0.415
Age in years			
Median (IQR)	38.0 (29.3-43.6)	36.0 (27.4-48.0)	0.657
HIV transmission mode			
Heterosexual	33 (49.3%)	126 (56.8%)	0.279
MSM	7 (10.4%)	20 (9.0%)	0.723
IDU	25 (37.3%)	75 (33.8%)	0.595
CD4+ cell initial count in cells/mm^3			
Median (IQR)	214 (26.8-467.0)	378 (154.8-560.5)	0.018*
HIV-1 subtype			
B	24 (35.8%)	54 (24.3%)	0.063
G	18 (26.9%)	67 (30.2%)	0.602
Non B non G	25 (37.3%)	101 (45.5%)	0.237
HIV diagnosis			
Recent infection	3 (4.5%)	25 (11.3%)	0.099
Newly diagnosed[¥]	13 (19.4%)	88 (39.6%)	0.002**
Previously diagnosed[¥]	51 (76.1%)	109 (49.1%)	<0.0001***
Antiretroviral Therapy			
Naïve	17 (25.4%)	163 (73.4%)	<0.0001***

Table 10: Differences among characteristics in the population studied regarding being infected with a virus harboring or not harboring drug resistance mutations (DRM). IQR: interquartile range; MSM: men who have sex with men; IDU: intravenous drug users; [¥]: infection of unknown duration

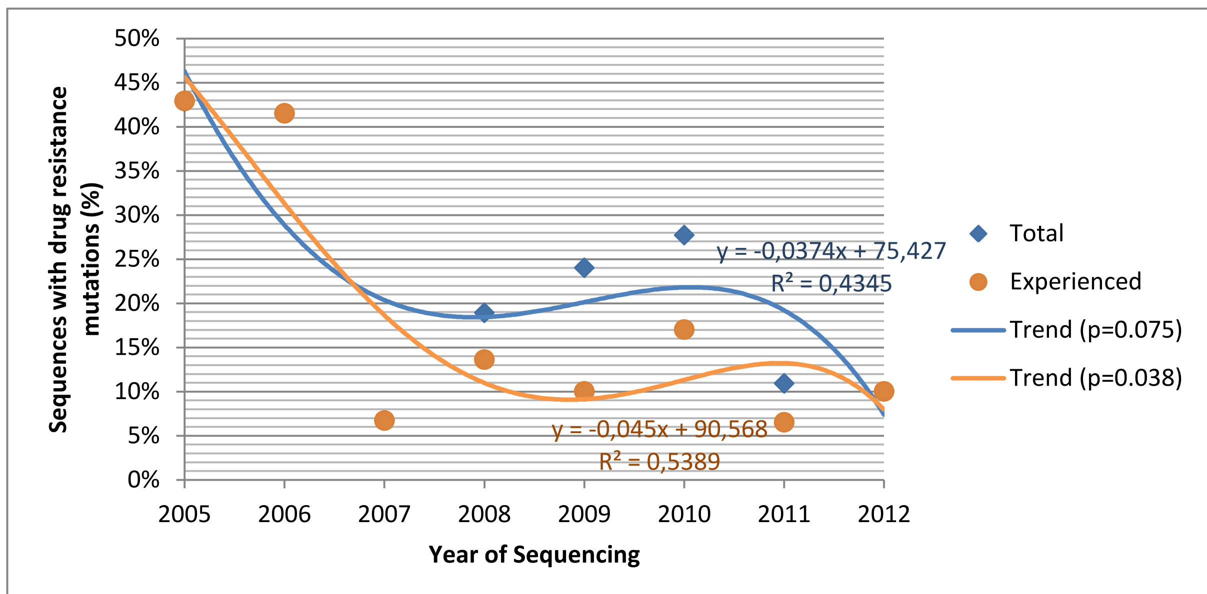

Figure 25: Temporal trends of drug resistance rates between 2006 and 2012. Experienced means patients with previous exposure to antiretroviral therapy.

Overall resistance is decreasing, as it is explicit in figure 25. For ART exposed patients, there is a significant declining temporal trend. By class of drugs this decreasing tendency is significant for NRTI (p=0.012) and for PI (p=0.030), but not for NNRTI.

Transmitted drug resistance

Over the 8 years surveyed, the average rate of drug resistance mutations in drug-naïve patients to any antiretroviral drug was 9.4% (17/180; 95% CI, 5.6% - 14.7%). For protease inhibitors was 3.3% (95% CI, 1.2% - 7.1%), for nucleoside reverse transcriptase inhibitors was 3.3% (95% CI, 1.2% - 7.1%) and for non nucleoside reverse transcriptase inhibitors was 3.9% (95% CI, 1.6% - 7.8%).

	Patient	NRTI		NNRTI		PI		Resistance level		
		n	Mutations	n	Mutations	n	Mutations	Low	Interme-diate	High
Recent Infection	20	1	M41L	0	None	0	None	AZT	None	None
	23	0	None	1	K101E	0	None	EFV, ETR	None	None
	80	2	T69D, M184V	1	K101E	1	N88S	ABC, EFV, ETR, SQV	DDI	3TC, FTC, ATV
Infection of unknown duration	14	0	None	1	K103S	0	None		EFV	NVP
	26	0	None	1	G190A	0	None	ETR, RPV	EFV	NVP
	71	2	T69D, M184V	0	None	0	None	ABC	DDI	3TC, FTC
	109	3	D67N, T215S, K219Q	0	None	0	EFV, ETR	TDF, DDI, ABC	AZT	None
	110	0	None	0	None	1	L90M	ATV, LPV	SQV	None
	118	0	None	0	None	1	M46I	None	None	None
	120	1	T215S	0	None	1	None	AZT	None	None
	141	0	None	1	None	0	M46L	None	None	None
	152	0	None	1	P225H	0	None	None	EFV, NVP	None
	160	0	None	0	None	1	I54v	ATV, LPV, SQV	None	None
	184	0	None	0	None	1	L90M	ATV, LPV	SQV	None
	209	0	None	1	K101E	0	None	EFV, ETR	None	None
	211	1	None	0	K103N	0	None	None	None	EFV, NVP
	270	1	Q151L	0	None	0	None	None	ABC, AZT, DDI	None

Table 11: Transmitted drug resistance among study cohort, resistance level estimated by HIVdb: Genotypic Resistance Interpretation Algorithm, Stanford University. Only first line antiretrovirals were considered. PI: protease inhibitor; NRTI: nucleoside reverse transcriptase inhibitor; NNRTI: non nucleoside reverse transcriptase inhibitor

Of the 17 naïve patients infected with virus carrying at least one mutation associated with drug resistance, 82.4% harbored one mutation, 11.8% two or three mutations and 5.9% over

three mutations. Mutations of resistance to just one pharmacological class were found in 94.1% of subjects and one patient was infected with virus harboring resistance mutations to the three classes tested. More frequently, mutations appeared as singletons. Mutations represented twice were T69D, M18V and T215S against NRTI; K103N/S against NNRTI; and L90M and M46I/L against PI. K101E was the only mutation represented thrice (Table 11). In recently infected patients, the prevalence of overall TDR was 11.5% (3/26; 95% CI 2.4-30.2), no different from patients with unknown duration of infection (14/154, 9.1%; 95% CI 5.1-14.8). However, comparing newly diagnosed with previously diagnosed patients, a significant difference was found regarding existence of TDR: patients diagnosed within one year of sequencing had 12.2% of TDR and patients diagnosed over a year before sequencing had 3.6% ($\chi 2=5.44$, $p=0.02$). By drug class, among recently infected patients TDR was 7.7% against NRTI, 7.7% against NNRTI and 3.8% against PI. In patients with unknown duration of infection, TDR varied from 2.6% against NRTI to 3.2% against NNRTI and PI. 13 out of 17 (76.4%) patients with TDR were clustered, an almost significant difference with patients with no TDR ($p=0.055$). No intravenous drug user had TDR, other modes of transmission showed no differences. There were also no differences between patients with TDR or not regarding sex, age, initial CD4+ cell count and HIV-1 subtype.

Trends of TDR rates

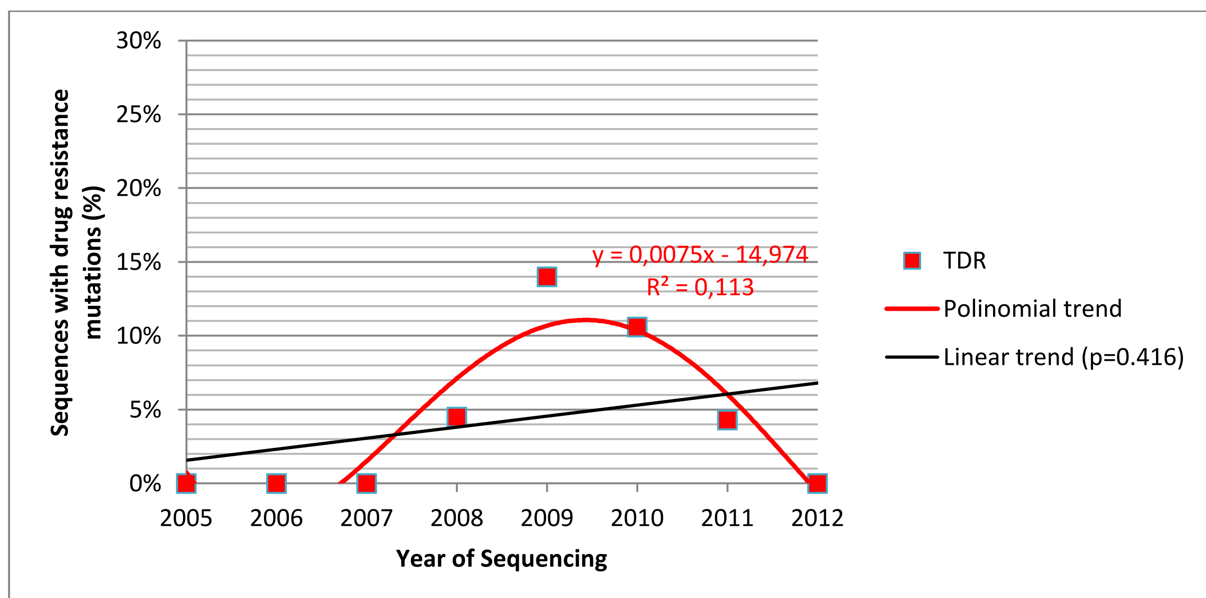

Figure 26: Temporal trends of transmitted drug resistance (TDR) rate in naïve population between 2006 and 2012.

In figure 26 we can see a linear not significant increase in rate of TDR between 2006 and 2012. Clearly, polynomial trend shows a parabolic temporal curve, concentrating all TDR

between 2008 and 2011. Disaggregating by drug class, there were sharp increases of NRTI in 2009 and NNRTI in 2010, shaping this trend.

Transmission clusters and clustering

Among the 289 subjects, 114 (39.4%) were part of 17 transmission clusters, ranging from 3 to 31 members (mean=6.7), named '01' to '17'. 14 pairs of transmission were also identified, named 'a' to 'n'. 28 sequences grouped in 9 small clusters (with 3 or 4 members), distributed by subtype B (n=6), subtype C (n=1), subtype G (n=1), subtype F1 (n=1) and 86 were part of 8 large clusters (with at least 5 individuals), two with 5 (both of subtype B), two with 7 (subtypes G and F1), one with 9 (subtype A1), one with 10 (subtype F1), one with 12 (CRF14_BG) and another with 31 sequences. This largest cluster consisted of individuals infected with a subtype C virus, 87.1% reporting intravenous drug use as mode of transmission. Additional large clusters showed homogeneity in main route of transmission and a tendency to be composed of patients with non recent infections (Table 12). Clusters 5, 14 and 16, formed respectively by subtypes B, F1 and A1 infected individuals, had only naïve patients. On other hand, clusters 11 and 17, formed respectively by subtype C and CRF14_BG infected individuals, had large proportions (48 and 75%) of ART experienced patients.

Cluster	Subtype	n	Main Transmission route	Recent infections	Patients with TDR	Patients with DRM	ART experienced	Maximum Window Period
11	C	31	IDU (87.1%)	1	0	5	15	2005-2011
17	CRF14_BG	12	IDU (83.3%)	0	0	3	9	2006-2011
13	F1	10	HET (70.0%)	2	1	1	3	2007-2011
16	A1	9	HET (66.7%)	2	1	0	0	2008-2011
14	F1	7	HET (85.7%)	0	0	0	0	2008-2011
9	G	7	HET (71.4%)	1	0	1	1	2006-2009
1	B	5	IDU (80%)	0	1	2	2	2007-2011
5	B	5	MSM (60%)	1	0	0	0	2006-2010

Table 12: Characterization of large clusters identified in study patients. Maximum window period is the time interval between the first and last infections within each cluster. TDR: transmitted drug resistance; DRM: drug resistance mutation; IDU: intravenous drug users; HET: heterosexual contact; ART: antiretroviral therapy.

Clustering occurred significantly more in non-B non-G subtypes, specifically in F1 ($p<0.0001$), C ($p=0.0008$) and A1 ($p=0.001$). On other hand, a patient infected with either subtype B or G virus had significantly less probability of being part of a cluster. Regarding mode of transmission, clustering was significantly more probable among intravenous drug users ($p=0.008$). No significant differences were found regarding MSM and heterosexual contact as mode of transmission and clustering (table 13).

According to the duration of infection, transmission clusters were independent from estimated time of infection as 75.0% of recently infected individuals and 64.4% of patients with unknown duration of infection are in a cluster (p=0.698).

Several factors that could enhance the likelihood of belonging to a transmission cluster were investigated by a multivariate regression model. It was found that being infected with a A1 (p=0.01), C (p=0.0004), or F1 (p<0.0001) HIV-1 subtype or had contracted the infection via

	Not in a Cluster (n=175)	In a Cluster (n=114)	P value
Gender			
Male	133 (76.0%)	89 (78.1%)	0.684
Age in years			
Median (IQR)	34.0 (27.0-43.0)	33.0 (26.9-41.1)	0.638
HIV diagnosis			
Recent infection	16 (9.1%)	12 (10.5%)	0.698
Newly diagnosed	53 (30.3%)	48 (42.1%)	0.039*
Previously diagnosed	106 (60.6%)	54 (47.4%)	0.027*
HIV transmission mode			
Heterosexual	103 (58.9%)	56 (49.1%)	0.104
MSM	19 (10.9%)	8 (7.0%)	0.273
IDU	50 (28.6%)	50 (43.9%)	0.008**
CD4+ cell initial count in cells/mm^3			
Median (IQR)	295 (145.5-498.7)	308 (145.6-566.5)	0.670
Initial viral load (log10)			
Median (IQR)	4.533 (3.858-5.172)	4.612 (3.993-5.243)	0.794
HIV-1 subtype			
B	57 (32.6%)	21 (18.4%)	0.008**
G	75 (42.9%)	10 (8.8%)	<0.0001***
Non-B non-G	43 (24.6%)	83 (72.8%)	<0.0001***
Drug Resistance Mutations (DRM)			
Any DRM	45 (25.7%)	22 (19.3%)	0.207
PI	12 (6.9%)	1 (0.9%)	0.017*
NRTI	32 (18.3%)	13 (11.4%)	0.115
NNRTI	26 (14.9%)	17 (14.9%)	0.990
Antiretroviral Therapy			
Naïve	104 (59.4%)	76 (66.7%)	0.215

Table 13: Differences among characteristics in the population studied regarding belonging or not to a cluster. IQR: interquartile range; MSM: men who have sex with men; IDU: intravenous drug users; PI: protease inhibitor; NRTI: nucleoside reverse transcriptase inhibitor; NNRTI: non nucleoside reverse transcriptase inhibitor

use of intravenous drugs (p=0.0005) were significantly related with the risk of belonging to a transmission cluster. On other hand, being infected with subtype G was negatively correlated with clustering (p<0.0001)

Transmission clusters and HIV-1 resistance

Among 17 patients with TDR, 9 (52.9%) were involved in clusters and 4 were part of transmission pairs (Table 14). One of the transmission pairs was formed by 2 TDR homosexual patients, both sharing PI mutation L90M. Cluster 6 is the only one where TDR patients are majority (2/3). Mutation K101E exists in all patients involved in that cluster. In cluster 13, a mutation in codon 103 is present in both patients who have DRM. Altogether it was possible to phylogenetically ascertain a relation between infecting virus in 76.5% of naïve patients. Considering ART experienced patients involved in clusters and transmission pairs, where naïve patients with TDR were identified, as "seeders" (n=11), a comparison between this group of seeders and all other ART experienced individuals showed a strong positive correlation with infection of unknown duration (spearman r=0.938, 95% CI 0.910-0.958, p<0.0001).

Patient	Cluster (size)	Pair	Subtype	Mode of Transmission	Year of Infection (Stage)	TDR Mutations	DRM present in other cluster/pair members (n)
270	01 (5)	No	B	HET	2009 (N)	Q151L	K101E (1)
26	04 (3)	No	B	HET	2010 (N)	G190A	None
80	06 (3)	No	B	HET	2008 (R)	N88S; T69D; M184V; K101E	M184V+K101E (1)
209	06 (3)	No	B	HET	2008 (N)	K101E	
71	07 (3)	No	B	HET	2009 (N)	T69D; M184V	None
211	10 (3)	No	G	HET	2011 (N)	K103N	None
14	13 (10)	No	F1	HET	2010(N)	K103S	K103N (1)
160	15 (4)	No	F1	HET	2009 (N)	I54V	None
109	16 (9)	No	A1	HET	2009 (N)	D67N; T215S; K219Q	None
110	No	a	C	MSM	2005 (P)	L90M	NA
184	No	a	C	MSM	2009 (N)	L90M	NA
118	No	h	B	MSM	2006 (P)	M46I	None
141	No	j	B	HET	2008 (N)	M46L	None

Table 14: Naïve patients with transmitted drug resistance (TDR), integrated in clusters or transmission pairs. DRM: drug resistance mutations; HET: heterosexual contact; MSM: men who have sex with men; R: infections recent (≤ 1 year); N: newly diagnosed patients (sequenced in the first 12 months after diagnosis) with infection of unknown duration; P: previously diagnosed patients with infection of unknown duration; NA: not applicable

A multivariate regression model trying to recognize factors correlated to the risk of TDR failed significance with the following parameters: age, sex, HIV-1 subtype, infection stage and mode of transmission. In univariate analysis, only the fact of being IDU has a negative significant correlation with TDR, as no case of transmitted drug resistance occurred with that route of transmission.

Cluster dynamics

We found 8 transmission clusters with at least 5 elements. The cluster growth factor was estimated for the interval 2008-2012 (table 15). All increased on size, by an average factor of 2.02. An average growth of 1 was used as cutoff to discriminate clusters, based in the assumption that a value over 1 means sustainable growth in the considered period of time. Noticeably, larger average growth rates were related to sexual transmission especially among MSM. On the contrary, IDU exposure was strongly related to slow growth clusters (table 16). Newly diagnosed patients with infection of unknown duration were significantly associated with cluster average growth>1. No differences were detected regarding sex, age, CD4+ cells initial count and recent infections. Initial viral load had a higher not significant median in fastest growing clusters, with p=0.059.

Cluster (subtype)	Elements in 2008	Elements in 2012	Growth factor
11 (C)	18	31	0.72
17 (BG)	8	12	0.50
13 (F1)	5	10	1.00
16 (A1)	1	9	8.00
14 (F1)	2	7	2.50
9 (G)	5	7	0.40
1 (B)	2	5	1.50
5 (B)	2	5	1.50

Table 15: Cluster growth between 2008-2012 for the clusters with n≥5

	Average Growth>1 (n=26)	Average Growth≤1 (n=60)	P value
Gender			
Male	19 (73.1%)	49 (81.7%)	0.369
Age in years			
Median (IQR)	37.0 (22.9-44.2)	32.0 (27.0-40.0)	0.840
HIV transmission mode			
Heterosexual	15 (57.7%)	18 (30.0%)	0.015*
MSM	6 (23.1%)	0 (0%)	0.0001***
IDU	5 (19.2%)	42 (70.0%)	<0.0001***
CD4+ cell initial count in cells/mm^3			
Median (IQR)	301.0 (135.7-591.2)	363.5 (203.8-579.6)	0.602
HIV-1 initial viral load (log10)			
Median (IQR)	4.768 (4.083-5.610)	4.425 (3.694-4.806)	0.059
HIV diagnosis			
Recent infection	3 (11.5%)	4 (6.7%)	0.448
Newly diagnosed[¥]	14 (53.8%)	18 (30.0%)	0.036*
Previously diagnosed[¥]	9 (34.6%)	38 (63.3%)	0.014*

Table 16: Differences among characteristics in patients integrating clusters regarding average growth. IQR: interquartile range; IDU: intravenous drug users; MSM: men who have sex with men; [¥]: infection of unknown duration

Reverse transcriptase A98S polymorphism and protease E35_T insertion

These two non-resistance related mutations were studied because of their strong association with a subtype in a geographic region (subtype G in Iberian Peninsula for A98S) and for its rarity among naïve patients (for protease codon 35 insertion). In our cohort, A98S was found in 104 (36%) sequences. The subtype distribution was disproportional, as in subtype G, 81/85 (95.3%) sequences carried this mutation with extra 7 present in several recombinant forms of subtype G. The remaining 16 A98S polymorphisms were present in subtype B virus (10), subtype C (4), F1 and CRF_47 (1 each). 33/104 of A98S positive sequences clustered (31.7%, comparing with 58.9% clustered A98S negative sequences; p<0.0001), 18 of them in subtype G clusters or pairs of transmission. All of cluster 01 (subtype B) elements (n=5) carried the polymorphism as well as 4 of the 31 members of cluster 11 (subtype C). This polymorphism appeared in 60 sequences obtained from naïve patients (with no significant difference to A98S negative sequences found in naïve patients) and in association with 26 sequences carrying some DRM. Among those, 20 were resistance mutations to NNRTI, although in only two patients these DRM were classified as TDR (K101E and K103N). 25% of recent infections carried A98S mutation, comparing with 8% of infections with unknown duration (p=0.01).

Insertion of a threonine (nucleotides: ACA) at codon 35 of protease (E35E_T) was found in 22 subtype C sequences, all involved in the same cluster, indicating its monophyletic origin. Its prevalence in this cohort was 7.6%, but among naïve patients was 12.2% and in subtype C infected individuals it was 52.4%. This insertion was not found in any other HIV-1 subtype in Braga. CD4+ cell count median was 203 cells/mm3, range 22 to 478. Through coalescent-based analysis it was concluded that the time of the most recent common ancestor was 1994 (1990-1998). Intravenous drug use was the route of transmission for 86.4% and all but three patients, one from Lisbon and two from Oporto, where from Braga region. None was considered a recent infection and 13 were never exposed to ART. No PI resistance mutations were found in any of these patients. In one patient K103N was detected and in another there were 6 RT resistance mutations (K65R, V75M, Y115F, M184V for NRTI and K103N, V106M for NNRTI). Both of these patients were ART experienced.

Six of the subtype C cluster patients were re-sequenced (mean interval between the 2 sequences dates: 25 months, range 8 to 42 months). Three of them already carried E35E_T insertion and maintained it 2 years later. In the other three, that insertion emerged. In the first sequencing, none of these 3 virus exhibited DRM. In the second sequencing, one of it carried

K103N and Y181C against NNRTI, and K65R, L74I, Y115F, M184V against NRTI. This patient initiated ART with efavirenz, tenofovir and emtricitabine 15 months after first sequence and 27 months before the second. No DRM against protease inhibitors were detected in none of these patients.

Discussion

HIV transmission depends on many factors: access to screening, access to antiretroviral therapy, human behavior, viral load, sexually transmitted diseases and other coinfections. That is probably why some studies concluded by a disproportionately high responsibility of early stages for transmission in TDR[395] and others suggest that most transmitters are individuals chronically infected and on antiretroviral therapy[396]. Probably, all stages of infection are contributing to the transmission of HIV-1. Primary infection is surely more contagious but for a shorter period. On other hand, asymptomatic stages, even being less infectious, will typically contribute more to the net transmission of HIV-1 because of its longer extent in community. Transmitted drug resistance has a potential impact on first line regimens of ART, compromising its efficacy and durability. It is of immense importance to take full advantage of this initial treatment offered to a HIV-1 positive person in order to control his viral load (and in that way prevent transmission and setback progression to AIDS), prevent more drug resistance mutations, while maintaining toxicity and costs at a reasonable level. Missing this target means second-line therapy, usually more complex to adhere, iatrogenic and expensive. In that sense, TDR surveillance is a fundamental public health approach to minimize its occurrence. Comparing to data from previous regional or national surveys, TDR prevalence in Braga (9.4%) is in line with overall levels found in Portugal (7.8%)[245], in France (9.0%)[397], in Leuven (9.6%)[398], in Madrid (9.7%)[399] and globally in Europe (8.9%)[382]. Nevertheless, it is somewhat distant from the low value of 5.6% in Sweden[380] and from an elevated 14.6% in United States of America[381]. Regarding routes of transmission, this study cohort is different from Portugal reality – more heterosexual contact and less MSM. It is then possible that TDR prevalence from a region is not representative from a country, as it is demonstrated in several studies[400,401]. These differences are supportive of the importance of studying local and regional epidemics and integrate data in larger multicenter initiatives.

In this study there is no influence of migrations, as more than 90% of the population is native and infected in the region. Other locations have sharp differences between native population and immigrants, with the later contributing in large scale for TDR[380,402].

Several significant associations were found: clustering was significantly related to non B non G subtypes and to transmission via intravenous drug use. In the first case it is believed that

the longer persistency and diffusion of subtypes B and G in this region led to sparse and scattered clustering. Concurrent to that is the fact that subtype B is implicated in 8 clusters, all with 3 to 5 elements, contrasting with larger clusters of subtypes A1, C, and F1. IDU networks of transmission are typically formed in closed subpopulations, while sexual transmission engages in more complex and open network structures, explaining why IDU transmission correlates positively with clustering.

It seems logic that a chronic infection gives birth to more mutations as the positive association found in this study indicate. Late presenters are patients who have a CD4+ cell count under 200/mm^3 in the moment of diagnosis. They have a remarkable influence in the presence of DRM: CD4+ count<200 cells/mm^3 in initial counting significantly correlates with presence of any mutation (p=0.021)

Although 76.5% of patients with TDR were involved in clusters or transmission pairs, none of the studied factors were significantly associated with transmission clusters containing TDR. Therefore, it was impossible to identify a non-sequence-based predictor of being in a transmission cluster with TDR, probably due to small size of most transmission chains in Braga. The same conclusion was reached recently by Yebra and colleagues in Spain[399]. There was a negative predictor, as no patient presumably infected via intravenous drug use exhibited TDR. This finding is coherent with observations made in Madrid, where TDR among IDU is lower than among MSM, even inexistent in some years[403], and with SPREAD Programme, which revealed an IDU TDR prevalence lower than in MSM or heterosexually infected people[382].

Before 2008 and after 2011 no TDR was found in this cohort. The trend of TDR in 2008-2011 is justified with a sharp increase on NRTI resistance in 2009 and on resistance to NNRTI a year later. These peaks were not related to clustering or recent infections. Probably, they translate prescription practices, with use of more potent regimens with large scale use of NNRTI beginning some years earlier, as the peak of resistance to both classes among ART experienced patients occurred in 2006. Trend of DRM in ART exposed individuals is significantly decreasing, explained by better virologic control and better management of drug resistances. This tendency will reflect on TDR with some delay. Consequently, ART experienced patients will tend to be less important in transmitting drug resistance and we are still experiencing a rising trend of naïve patients in perpetuating TDR.

As reported previously[404,405], the majority of patients with TDR (82%) had virus with singleton resistance mutations. Resistance profiles would compromise first line therapeutics if based in NNRTI for patients 14, 26, 152 and 211. In all other, recommended first line regimens were predicted to be effective but the mutations present will lower the threshold for subsequence emergence of resistance and thus affect treatment efficacy. The presence of the mutation pair D67N/K219Q in a patient justifies apprehension because this specific combination has a low fitness cost to the virus, is readily transmitted and is durable in the recipient[406].

Cluster growth factor was used as a measure of cluster dynamics. In the way it was defined in this study, this rate cannot be used as a measure of onward transmissions as our population has a definite date of diagnosis but not an accurate date of infection. However, a cluster still reflects related infections and when sampled over an exact period of time, cluster with larger growth factors can indicate subepidemics with greater relative transmission rates. Larger average growths in a cluster were found to be associated with sexual transmission, particularly among MSM, and newly diagnosed infections. A strong negative association with IDU exposure might denote a deceleration in growth of clusters integrating intravenous drug users, leading these clusters (particularly 11C and 17BG) to unsustainable growth rates and eventually to inactivity. No association with recent infections was established, however infections with evolution superior to one year were more frequently found in slower growing clusters, a logical assumption when we remember that existing clusters informs on historical transmissions, not necessarily active relations between partners.

In an international multicenter study, Kantor and collaborators compared more than 3600 non B subtypes sequences with more than 4700 subtype B sequences to conclude for the existence of subtype-specific polymorphisms[304]. Among those it was A98S, a mutation in HIV-1 reverse transcriptase, described as a subtype G polymorphism. However, when reviewing the data, we found that more than 70% of subtype G sequences used were originated from Portugal and Spain. Querying Stanford HIV Drug Resistance Database and selecting subtype G sequences from drug naïve persons (n=165; 57 from Portugal or Spain and 108 from other countries) it was found that 51/57 (89.5%) of the Iberian sequences harbored the A98S mutation against only 10/108 (9.3%) of sequences from the rest of the world. A paper published in 2001 highlighted the unique occurrence of this polymorphism in samples from Spain and Portugal[407]. Considering that HIV-1 distribution obeys to several geographical bias, it is plausible to consider that patterns of subtype-specific polymorphisms may be

geographic-specific instead. In Braga's cohort it was evident a massive predilection of A98S polymorphism for subtype G and a significant greater occurrence in recent infections. On the contrary, A98S positive sequences clustered less than their counterparts, probably because a patient infected with subtype G virus had significantly less probability of being part of a cluster.

To our knowledge, this is the largest prevalence (7.6%) ever reported of a protease codon 35 insertion. These insertions have been reported since 2001, although with prevalences much smaller than ours, varying between 0.2% and 4.5%[408-411]. There are several variants of this insertion, configuring a heterogeneous group in terms of phenotypic consequences: insertion E35E_G and E35E_E apparently replicates better than the wild type virus[408,409], E35E_TN decreased the replication rate of PI-resistant strains, E35E_TD and E35E_T were associated with decreases in replicative capacity in single case reports[412,413], but to date none of these insertions showed potential to alter drug susceptibility. In any case, the impact of the insertion on viral replication is difficult to predict, depending on the nature of the inserted amino acids and the pattern of drug-resistance-associated mutations. In our cohort it was E35E_T that was isolated, the same insertion described in another Portuguese city, Coimbra, in 2009[414]. In both centers only subtype C was involved and the insertion clustered in one phylogenetic lineage, suggesting the possibility of transmission. In the same way, no mutations conferring resistance to protease inhibitors were found. In Coimbra, the cluster related mainly to sex workers, in Braga the main transmission route was IDU. Equally, CD4+ cell count and viral load varied to a great extent, signaling no influence of codon 35 insertion on disease progression. Overall, protease insertions were observed both in PI-treated patients and in PI-naïve patients and were transmissible. In PI-naïve patients, protease insertion virus persisted for a long time. In addition, codon 35 insertions were able to emerge, even in the absence of PI pressure. As stated earlier, if this insertion decreases viral replicative capacity, although with no impact on resistance level, what permit its persistence? A clue is given by Paolucci and collaborators, who found that E35E_G insertion could recover the viral replication under antiretroviral treatment[409], but further investigation is needed to elucidate the effects of other insertions on viral fitness. In any case, this protease insertion could be useful as epidemiological signature

The main limitation of this study resides in its cross-sectional design, which included convenience sampling, collected as patients were diagnosed or came into clinics for care. This methodology may not truly represent the region's population and may explain why there

were no TDR among IDU, for example. The representativeness of the population studied can be further disturbed by the impossibility of sequencing virus from the persons who remain undiagnosed and unaware of their HIV status. In addition, chronically infected and treated individuals with undetectable viral loads cannot be genotyped, although some of them might have been source of infections before achieving virologic suppression.

In summary, in Braga there is a prevalence of TDR similar to other locations in Europe. In the only published study focusing TDR in Portugal[245],TDR rate was no different (14/180 vs. 17/180), in a sample including more MSM and less IDU, collected in a single year (2003) and trying to reflect the distribution of HIV-1 infections in Portugal. However, local specificities are of utmost relevance to the clinician, as local TDR data accurately reflects the ecology of the virus, driving more adequate use of ART and phylogenetics applied at clinical level may reveal related infections. Subsequently, this data needs to be collated at regional, national and international level for purposes of TDR surveillance and to elucidate transmission patterns and trends. A relation of faster growing clusters with sexual modes of transmission was detected. Opposing, IDU exposure in clusters was related with unsustainable growth rates, indicating network transmissions with little activity.

This study cohort was mainly composed of patients with unknown duration of infection and 76.4% of the patients with TDR were involved in transmission networks. Among patients with unknown duration of infection, being diagnosed less than a year before sequencing was significantly correlated with presence of TDR. In addition, most clusters mix naïve and ART experienced patients, as well as recent and chronic infections. This means, as this study was unable to significantly identify a population that could be targeted for future TDR prevention strategies, that it seems rational to stress two points: first, the importance of universal performing of baseline HIV-1 genotyping; and second, early access to antiretroviral treatment. Achieving these two goals will allow sources of transmission (seeders) identification and offer them better virologic control, lowering the risk of HIV-1 transmission.

CHAPTER 5: GENERAL DISCUSSION AND CONCLUSIONS

CHAPTER 5: GENERAL DISCUSSION AND CONCLUSIONS

HIV-1 elusive origins can be looked as a fable with several moralities attached. By the time physicians realized that HIV-1/AIDS existed it was already well established in the human host, serving as a practical reminder of the conditions that foster the emergence of new infectious diseases: zoonosis, originated in wildlife and correlated with socio-economic, environmental and ecological factors. For decades, HIV-1 infected human populations but had such a small impact that it passed unnoticed. In comparison to pathogens like malaria (which is carried by mosquitoes) and the common cold (which can travel through the air), HIV-1 is quite untransmissible, relying on the direct transfer of body fluids. There was a coincidence of historical events that allowed the window of opportunity for the virus to spread in humans. These include: the practice of hunting chimpanzees; the rise of densely populated cities in Africa; and a correlated increase in high-risk behaviors involving the exchange of body fluids such as injection drug use or prostitution. The fact that changes in human societies were so critical in the rise of the virus raised the awareness to emerging infectious diseases with zoonotic origins as human populations grow and affect the climate and wildlife.

An ancestral virus in chimpanzees existed, it is possible that multiple subsequent cross-over events between non-human primate species occurred or that the virus was carried in humans prior to the expansion of the M group for a large period of time. Chimpanzees might be considered a "passage" for lentiviruses to adapt to humans, given the genetic similarity between humans and chimpanzees. For example, *vpx* gene was lost by deletion upon adaptation to chimpanzees[415]. This genomic deletion resulted in the reconstruction of the overlapping *vif* gene by ‘‘overprinting,’’ creating a unique *vif* that overlaps in its 3’ end with the *vpr* gene and can antagonize hominid APOBEC3 more efficiently than its ancestors[415]. Consequently, viral gene loss and adaptation in chimpanzees predated the origin of HIV-1, which was originated through a series of gene loss and adaptation events in its chimpanzee precursors, inducing viral adaptations to new hosts proteins. This allowed for efficient infection in the human host. This transition between species was facilitated by the proximity between the primate species and by the versatility of the virus.

The generation of genetic variability is a million times faster in HIV-1 than in *Homo sapiens*[82]posing a large challenge for viral recognition by the host immune system and contributing to the fact that HIV-1 is still responsible for thousands of deaths each year.

The high level of HIV-1 genetic diversity has important implications in the diagnostic, treatment and monitoring of the disease[289,290,416]. Questions have been raised on whether diversity may also affect viral transmissibility and pathogenicity[274,417,418]. Since the human immune response to HIV-1 is strain-specific[419]it is consensual that the viral genetic diversity has been a major limitation in the design and development of an effective vaccine. This work mainly focused on the comprehension of how viral diversity and evolution may provide important tools to undergo surveillance and monitoring of a local or regional HIV-1 epidemic.

HIV-1 strains are not randomly distributed across the World but display distinctive geographical distributions, with regional subtype variation[217]. The existing regional differences are thought to be related with the founder effect that occurred when a certain subtype was introduced in a new susceptible population where it initiated a new transmission network with posterior diversification. Eventually, some strains were of poorest fitness and could not spread far from their origin and others could not establish in extensive transmission networks. However, possibly due to increasing migrations and global human circulation, these geographic patterns are becoming imprecise as diversity increases at regional and national levels.

HIV-1 genetic diversity among different clinical isolates, especially when collected from different geographical locations, can be very high. However, even in isolates from the same patient variants are present as relatively similar quasispecies[420]. This population of genetically diverse virus that develops in each infected individual confers to HIV-1 the capacity to rapidly and effectively adapt to changes in host immune responses or other constrains.

Emergence of HIV/AIDS was facilitated by this ability of viral adaptation and by demographic and social changes in the host population that favored transmission such as higher population density, urbanization and massive migrations[421]. Nevertheless, HIV-1 has poor transmissibility when compared with other agents, like hepatitis B virus, influenza virus

or measles virus. Its high latency time and in particular its capacity to undergo sexual transmission explains HIV-1 success as a human pathogen.

It is well established that the expanding HIV-1 diversity in Western Europe is mainly caused by population movements, such as migration and travelling, and sexual contacts among individuals from different countries where different variants are highly prevalent[422-427].In Portugal that must have happened earlier in the epidemic, as demonstrated by the singular genetic diversity of HIV-1 strains circulating in the country. Knowing the close commercial and social relations between Portugal and its former colonies in Africa, studies were performed to evaluate the genetic diversity in those countries. In Cape Verde HIV-1 subtype G was the prevailing subtype in Oliveira et al study, in 2012[265]. They also found that G variants present in Cape Verde were imported mostly from Angola and Portugal where highly divergent subtype G strains prevail. In Angola, the phylogenetic analyses showed extremely high genetic diversity among circulating HIV-1 strains[269]. Only 53% of the circulating virus were pure subtypes with subtype A and its sub-subtypes A1 and A2 predominating over the other subtypes. There is also a significant proportion of subtype C (11.3%). Expansion of these virus to Portugal may have started as early as 1961, due to large migratory movements caused by Angolan independence war. Thus, Portugal is likely one of the entry-points of HIV-1 in Western Europe, especially in the case of non-B subtypes. Subtype B, the dominant HIV-1 subtype in the Occident, is thought to come from the United States via Haiti, where risky behavior of military personnel returning from Central Africa may have facilitated the viral introduction in Americas[7]. In Braga, similarly to Portugal, we found a high diversity in subtypes and CRF circulating, as it is evident in the local epidemic characterization made in this work. The co-circulation of recombinant virus is known to lead to the appearance of URF, in a so called recombination hotspot[428]. In fact, in figure 19B we can see the rise in the incidence with infections with locally rarer subtypes (A1 and F1), as well as several CRF and URF. Actually, URF represented 9% of total infections in Braga. The geographic distribution of subtypes is subject to constant change. With the globalization new HIV-1 strains are emerging in areas where they were originally non-existent[429-431]. Thus importation and exportation of new types, subtypes and even CRF of HIV-1 is possible. Tracking the presence of new HIV-1 strains is important for surveillance purposes, diagnosis and disease monitoring and possibly also effective development of vaccines. Molecular epidemiology as recently emerged as an excellent tool to investigate HIV-1 diversity and to perform disease surveillance and monitoring[432].

There are clinical and biological differences between HIV-1 group M subtypes, justifying on their own an interest in the study of HIV-1 diversity. One emblematic example are the subtype-related differences in disease progression and viral transmission for subtypes C and D that were shown to be more aggressive than subtype A[135,417,433]. It is reassuring that no difference is apparent in response to antiretroviral therapy across a broad spectrum of HIV-1 subtypes[434]. In addition, some patients, called elite controllers, are able to mount effective host immune responses that seem to be dependent on host genetic factors and not on viral subtype[435]. Thus far the usefulness of phylogenetic study of HIV-1 diversity resided mainly on addressing several relevant questions, including the HIV-1 origin, its evolutionary driving forces and intra- and inter-host diversity. The continuous development of robust statistical and informatics tools can further deliver powerful insights into host-related and environmental evolutionary processes. These will likely allow the identification of population-level phylogenetic patterns reflecting both transmission dynamics and genetic change and contributing to the elucidation of viral polymorphisms associated with transmission and to a better characterization of viral evolution at the individual and population levels.

In the current work, the option was to study diversity at regional level, addressing three dimensions: phylodynamic analysis of a local HIV-1 epidemic; classification of the duration of infection; and assessment of transmitted drug resistance and polymorphisms.

In chapter 2, the existence of high HIV-1 diversity in Minho was confirmed. Investigation of this diversity revealed that heterogeneity of HIV-1 subtypes is attributable to Portuguese-born patients, presumably infected in Portugal. This fact opposes to what is described for the rest of Europe, where immigration is related to most of non-B infections. Just like any other “endemic” infection, this proves ancient more ancestral arrival of non-B HIV-1 subtypes to Portugal when compared with the rest of Western Europe. Combining these data with the fact that Portugal had intense and close relations in African countries with high rates of viral diversity, and with the historical data on intense human migrations between these countries, we can speculate that the non-B HIV-1 epidemic in Portugal dates back to several decades ago and probably was one of first in Western Europe. HIV-1 infection in Minho has a high clustering rate, more pronounced among a specific population (IDU) and among non-B subtypes. In the first case, it is evidence for a closed risk group, in the second it is probably related with the founder effect and more recent viral introduction. Noteworthy, incidence of

these non-B subtypes is increasing, particularly among heterosexual mode of transmission. This finding may contribute to a refocusing of prevention strategies. A specific transmission cluster (17BG), found to be composed almost exclusively by intravenous drug users, illustrates an interesting condition. There are no phylogenetic relations with other CRF14_BG clusters reported in nearby regions and this can be considered as additional evidence of a founder effect and a “closed-circuit” transmission network. The fact that no additional members have been found in the recent years is in favor of the regional success in the implementation of a controlling strategies proposed and applied to control HIV-1 infections among IDU namely based on syringe exchange programs and opioid substitution treatment.

In Chapter 3, we aimed to design a simple yet reliable methodology to differentiate recent from non-recent infections, defined as those occurring more than one year ago. This apparently humble task has a formidable potential in order to determine the HIV-1 incidence, a step of crucial importance to study the evolution of the epidemic and thus validate prevention strategies. Ambiguous nucleotide calls are a byproduct, usually treated as noise when it comes to HIV-1 genotyping. However, as they reflect the variability of the virus and are related to the length of infection, have proven its potential for the construction of an algorithm to estimate if the infection took place more or less than one year from diagnosis. Although with far from ideal sensibility (0.73) and negative predictive value (0.89), it has a specificity of 0.96 and a likelihood ratio of 16. One of the most relevant features of the developed methodology is that it is easily implementable implicating no additional costs and has a robust performance to factors such as HIV-1 subtype and transmission mode. It is not an algorithm to assess individual length of infection, due to lack of accuracy, but it can prove useful at population level. For example, detection of chronic stage infections may contribute significantly to design targeted policies preventing patient’s unawareness of their infection status and poor linkage to care and treatment, by characterizing related demographic or socio-economic factors.

In chapter 4, we found a rate of transmitted drug resistance (TDR) in Braga (9.4%) similar to the ones described in other Portuguese and Europeans studies. This rate can be considered low but remains worrisome, especially because in 4 naïve patients, representing 23.5% of the patients with TDR and 2.2% of the population studied, a first line therapy with NNRTI was predestined to fail. Although reversion of resistance mutations is sometimes observed, it is much slower than the initial installation of a resistant strain[436]. If a resistance mutation does

not compromise fitness, that mutation will tend to persist. Consequently, TDR is a critical feature to control, when ART is indicated. Probably, it will be easier to prevent an epidemic of drug resistant HIV-1 from occurring than to control it once it emerges. Two strategies are being implemented in resourceful locations: “test-and-treat”, which scale-up diagnosis and recommend start of ART as soon as an individual is diagnosed; and pre-exposure prophylaxis. Both strategies count on antiretroviral efficacy, but TDR is more prevalent precisely where ART use is more frequent[391]. For all this, appropriate surveillance of TDR is imposed as a preventive measure. To detect how, when and in whom TDR is more frequent, we can intervene in order to maximize the effectiveness of ART. In Braga, it was impossible to identify a predictor of TDR, except sequencing. Nevertheless, more than 76% of the patients were involved in transmission clusters, mainly associated with non-B and non-G subtypes. A disconcerting finding was that all TDR was concentrated in the period between 2008 and 2011. Drug resistance mutations are decreasing in ART exposed individuals, probably on account of better retention in health care. Eventually, this trend will reflect in naïve patients. Even with small convenience samples, the evaluation of TDR at a local level will unveil specificities that can prove relevant to the treating clinician. These data will reflect the local environment of the virus before integrating larger databases, at national or international level, where it will contribute to TDR global surveillance and to clarify transmission patterns.

Since Kantor’s large scale study[304], it is well recognized the existence of subtype-specific polymorphisms. In our population, a natural polymorphism of the reverse transcriptase gene, A98S, confirmed to be subtype G specific and so far only found in the Iberian Peninsula[407]. Also a protease insertion on the codon 35 proved transmissible and was restricted to a subtype C cluster, mainly composed of intravenous drug users. Both occur naturally, do not entail changes in HIV-1 biology and so may be used as a genetic signature, similar to HIV-1 subtype A, former Soviet Union variant, whose unique genetic features permitted easy tracing in geographically distinct epidemics[238].

The three investigational studies that compose this work provided additional tools to investigate the epidemic at the regional level, characterizing their transmission networks, facilitating the study of the incidence and watching the evolution of drug resistance. Incidentally, they provided clues to genetic signatures that may prove important when

integrating these data in multicenter studies. HIV-1 pandemic cannot be globally curbed without implementation of local actions.

Conclusions

A peculiar conjunction of biological, socio-economic and demographic features gave rise to one formidable pathogen, responsible for a devastating pandemic, still uncontrolled. For this virus it was a matter of the right event (species leap) in the right time (demographic expansion and urbanization in Africa). There is a legitimate concern about the possibility of such coincidence be repeatable, with a new, yet unknown, microbiological agent taking advantage of social and economical disruptions still abundant in those regions. Nowadays, globalization of a potential pathogen surely would be faster. HIV took a few decades to become global and to originate a myriad of local epidemics. The diversity HIV-1 gained along its path made it an undefeatable adversary so far, but this same diversity can work for Man in this two arms race.

In Braga, as well as in Portugal, HIV-1 encompasses a high genetic diversity, probably supported in multiple and ancient introductions of different subtypes, in a remarkable contrast with subtype B dominance in Western Europe, although in decline. Over time, the local incidence of these subtypes varied, according to founder effects and genetic bottlenecks during onward transmission events. Outstandingly, in present days these variations in incidence do not depend on immigration but are verified among native population. The assessment of this diversity enlightened the comprehension of a local HIV-1 epidemic and allowed to:

- Demonstrate an increasing incidence and onward transmission of non predominant subtype virus via sexual exposure with potential for geographic expansion.
- Confirm that diversity positively related with duration of infection and its assessment can be used to estimate length of infection in order to facilitate incidence studies.
- Monitor cluster dynamics and transmitted drug resistance in order to improve efficacy of antiretroviral therapy.
- Detect natural polymorphisms with particularities enabling them to serve as genetic signatures.
- Perform continuous monitoring of epidemic evolution and so reinforce preventive strategies that should not be restricted to specific populations, rather be comprehensive.

These purposes were fulfilled with practical application of 3 simple and undemanding strategies, illustrated in chapters 2, 3 and 4: phylogenetic analysis of clusters and inference of most recent common ancestors, appliance of ambiguities found in standard sequences to distinguish between recent and chronic infections, and surveillance of transmitted drug resistance and natural polymorphisms.

REFERENCES

1. Sharp PM, Hahn BH. The evolution of HIV-1 and the origin of AIDS. *Philos Trans R Soc Lond B Biol Sci.* 2010;365(1552):2487-2494.
2. (CDC) CfDC. Pneumocystis pneumonia--Los Angeles. *MMWR Morb Mortal Wkly Rep.* 1981;30(21):250-252.
3. UNAIDS. *Global Report: UNAIDS report on the global AIDS pandemic.* GENEVA, Switzerland: United Nations;2013.
4. (CDC) CfDC. Opportunistic infections and Kaposi's sarcoma among Haitians in the United States. *MMWR Morb Mortal Wkly Rep.* 1982;31(26):353-354, 360-351.
5. (CDC) CfDC. Pneumocystis carinii pneumonia among persons with hemophilia A. *MMWR Morb Mortal Wkly Rep.* 1982;31(27):365-367.
6. (CDC) CfDC. Update on acquired immune deficiency syndrome (AIDS)--United States. *MMWR Morb Mortal Wkly Rep.* 1982;31(37):507-508, 513-504.
7. Gilbert MT, Rambaut A, Wlasiuk G, Spira TJ, Pitchenik AE, Worobey M. The emergence of HIV/AIDS in the Americas and beyond. *Proc Natl Acad Sci U S A.* 2007;104(47):18566-18570.
8. Masur H, Michelis MA, Wormser GP, et al. Opportunistic infection in previously healthy women. Initial manifestations of a community-acquired cellular immunodeficiency. *Ann Intern Med.* 1982;97(4):533-539.
9. Acquired immune deficiency syndrome (AIDS): precautions for clinical and laboratory staffs. Centers for Disease Control. *Emerg Med Serv.* 1983;12(5):69.
10. (CDC) CfDC. Possible transfusion-associated acquired immune deficiency syndrome (AIDS) - California. *MMWR Morb Mortal Wkly Rep.* 1982;31(48):652-654.
11. (CDC) CfDC. Unexplained immunodeficiency and opportunistic infections in infants--New York, New Jersey, California. *MMWR Morb Mortal Wkly Rep.* 1982;31(49):665-667.
12. Tsoukas C, Gervais F, Shuster J, Gold P, O'Shaughnessy M, Robert-Guroff M. Association of HTLV-III antibodies and cellular immune status of hemophiliacs. *N Engl J Med.* 1984;311(23):1514-1515.
13. van Ameijden EJ, van den Hoek JA, van Haastrecht HJ, Coutinho RA. The harm reduction approach and risk factors for human immunodeficiency virus (HIV) seroconversion in injecting drug users, Amsterdam. *Am J Epidemiol.* 1992;136(2):236-243.
14. Clumeck N, Mascart-Lemone F, de Maubeuge J, Brenez D, Marcelis L. Acquired immune deficiency syndrome in Black Africans. *Lancet.* 1983;1(8325):642.
15. Gottlieb MS, Schroff R, Schanker HM, et al. Pneumocystis carinii pneumonia and mucosal candidiasis in previously healthy homosexual men: evidence of a new acquired cellular immunodeficiency. *N Engl J Med.* 1981;305(24):1425-1431.
16. Goedert JJ, Neuland CY, Wallen WC, et al. Amyl nitrite may alter T lymphocytes in homosexual men. *Lancet.* 1982;1(8269):412-416.
17. Mavligit GM, Talpaz M, Hsia FT, et al. Chronic immune stimulation by sperm alloantigens. Support for the hypothesis that spermatozoa induce immune dysregulation in homosexual males. *JAMA.* 1984;251(2):237-241.
18. Shilts R. And the band played on: politics, people, and the AIDS epidemic. New York: <st1:place w:st="on">St. Martin's Press; 1987.
19. Barré-Sinoussi F, Chermann JC, Rey F, et al. Isolation of a T-lymphotropic retrovirus from a patient at risk for acquired immune deficiency syndrome (AIDS). *Science.* 1983;220(4599):868-871.
20. Gallo RC, Salahuddin SZ, Popovic M, et al. Frequent detection and isolation of cytopathic retroviruses (HTLV-III) from patients with AIDS and at risk for AIDS. *Science.* 1984;224(4648):500-503.

21. Pear R. AIDS blood test to be available in 2 to 6 weeks. the New York Times1985.
22. (CDC) CfDC. Classification system for human T-lymphotropic virus type III/lymphadenopathy-associated virus infections. *MMWR Morb Mortal Wkly Rep.* 1986;35(20):334-339.
23. 1993 revised classification system for HIV infection and expanded surveillance case definition for AIDS among adolescents and adults. *MMWR Recomm Rep.* 1992;41(RR-17):1-19.
24. Clavel F, Guétard D, Brun-Vézinet F, et al. Isolation of a new human retrovirus from West African patients with AIDS. *Science.* 1986;233(4761):343-346.
25. Coffin J, Haase A, Levy JA, et al. What to call the AIDS virus? *Nature.* 1986;321(6065):10.
26. Kanki PJ. Biologic features of HIV-2. An update. *AIDS Clin Rev.* 1991:17-38.
27. Nelson KE, Celentano DD, Eiumtrakol S, et al. Changes in sexual behavior and a decline in HIV infection among young men in Thailand. *N Engl J Med.* 1996;335(5):297-303.
28. *No time to lose: getting more from HIV prevention.* Washington, D.C.: National Academy Press; 2001.
29. (CDC) CfDCaP. Update: mortality attributable to HIV infection among persons aged 25-44 years--United States, 1994. *MMWR Morb Mortal Wkly Rep.* 1996;45(6):121-125.
30. Hogan DR, Salomon JA, Canning D, Hammitt JK, Zaslavsky AM, Bärnighausen T. National HIV prevalence estimates for sub-Saharan Africa: controlling selection bias with Heckman-type selection models. *Sex Transm Infect.* 2012;88 Suppl 2:i17-23.
31. Gouws E, Cuchi P, Transmission ICoEHIbMo. Focusing the HIV response through estimating the major modes of HIV transmission: a multi-country analysis. *Sex Transm Infect.* 2012;88 Suppl 2:i76-85.
32. Warnke D, Barreto J, Temesgen Z. Antiretroviral drugs. *J Clin Pharmacol.* 2007;47(12):1570-1579.
33. *Global report : UNAIDS report on the global AIDS epidemic.* Geneva : UNAIDS, 2010-.
34. Aghokeng AF, Ayouba A, Mpoudi-Ngole E, et al. Extensive survey on the prevalence and genetic diversity of SIVs in primate bushmeat provides insights into risks for potential new cross-species transmissions. *Infect Genet Evol.* 2010;10(3):386-396.
35. Charleston MA, Robertson DL. Preferential host switching by primate lentiviruses can account for phylogenetic similarity with the primate phylogeny. *Syst Biol.* 2002;51(3):528-535.
36. Apetrei C, Sumpter B, Souquiere S, et al. Immunovirological analyses of chronically simian immunodeficiency virus SIVmnd-1- and SIVmnd-2-infected mandrills (Mandrillus sphinx). *J Virol.* 2011;85(24):13077-13087.
37. Wertheim JO, Worobey M. A challenge to the ancient origin of SIVagm based on African green monkey mitochondrial genomes. *PLoS Pathog.* 2007;3(7):e95.
38. Gao F, Yue L, White AT, et al. Human infection by genetically diverse SIVSM-related HIV-2 in west Africa. *Nature.* 1992;358(6386):495-499.
39. Gao F, Bailes E, Robertson DL, et al. Origin of HIV-1 in the chimpanzee Pan troglodytes troglodytes. *Nature.* 1999;397(6718):436-441.
40. Keele BF, Van Heuverswyn F, Li Y, et al. Chimpanzee reservoirs of pandemic and nonpandemic HIV-1. *Science.* 2006;313(5786):523-526.
41. Vidal N, Peeters M, Mulanga-Kabeya C, et al. Unprecedented degree of human immunodeficiency virus type 1 (HIV-1) group M genetic diversity in the Democratic Republic of Congo suggests that the HIV-1 pandemic originated in Central Africa. *J Virol.* 2000;74(22):10498-10507.
42. Keele BF, Jones JH, Terio KA, et al. Increased mortality and AIDS-like immunopathology in wild chimpanzees infected with SIVcpz. *Nature.* 2009;460(7254):515-519.
43. Rudicell RS, Holland Jones J, Wroblewski EE, et al. Impact of simian immunodeficiency virus infection on chimpanzee population dynamics. *PLoS Pathog.* 2010;6(9):e1001116.

44. Bailes E, Gao F, Bibollet-Ruche F, et al. Hybrid origin of SIV in chimpanzees. *Science.* 2003;300(5626):1713.
45. Goodall J. *The chimpanzees of Gombe : patterns of behavior.* Cambridge, Mass. ; London: Belknap Press of Harvard University Press; 1986.
46. Plantier JC, Leoz M, Dickerson JE, et al. A new human immunodeficiency virus derived from gorillas. *Nat Med.* 2009;15(8):871-872.
47. Zhu T, Korber BT, Nahmias AJ, Hooper E, Sharp PM, Ho DD. An African HIV-1 sequence from 1959 and implications for the origin of the epidemic. *Nature.* 1998;391(6667):594-597.
48. Worobey M, Gemmel M, Teuwen DE, et al. Direct evidence of extensive diversity of HIV-1 in Kinshasa by 1960. *Nature.* 2008;455(7213):661-664.
49. Hance WA. *Population, migration, and urbanization in Africa.* New York ; London: Columbia University Press; 1970.
50. McKelvey JJ. *Man against tsetse : struggle for Africa.* Ithaca ; London: Cornell University Press; 1975.
51. Pépin J, Labbé AC, Mamadou-Yaya F, et al. Iatrogenic transmission of human T cell lymphotropic virus type 1 and hepatitis C virus through parenteral treatment and chemoprophylaxis of sleeping sickness in colonial Equatorial Africa. *Clin Infect Dis.* 2010;51(7):777-784.
52. van der Loeff MF, Larke N, Kaye S, et al. Undetectable plasma viral load predicts normal survival in HIV-2-infected people in a West African village. *Retrovirology.* 2010;7:46.
53. 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(4):329-331.
54. Levy JA. HIV pathogenesis and long-term survival. *AIDS.* 1993;7(11):1401-1410.
55. Frankel AD, Young JA. HIV-1: fifteen proteins and an RNA. *Annu Rev Biochem.* 1998;67:1-25.
56. Loftin LM, Kienzle M, Yi Y, Collman RG. R5X4 HIV-1 coreceptor use in primary target cells: implications for coreceptor entry blocking strategies. *J Transl Med.* 2011;9 Suppl 1:S3.
57. Lee B, Sharron M, Montaner LJ, Weissman D, Doms RW. Quantification of CD4, CCR5, and CXCR4 levels on lymphocyte subsets, dendritic cells, and differentially conditioned monocyte-derived macrophages. *Proc Natl Acad Sci U S A.* 1999;96(9):5215-5220.
58. van Rij RP, Blaak H, Visser JA, et al. Differential coreceptor expression allows for independent evolution of non-syncytium-inducing and syncytium-inducing HIV-1. *J Clin Invest.* 2000;106(8):1039-1052.
59. Goodenow MM, Collman RG. HIV-1 coreceptor preference is distinct from target cell tropism: a dual-parameter nomenclature to define viral phenotypes. *J Leukoc Biol.* 2006;80(5):965-972.
60. Didigu CA, Doms RW. Novel approaches to inhibit HIV entry. *Viruses.* 2012;4(2):309-324.
61. Arhel N. Revisiting HIV-1 uncoating. *Retrovirology.* 2010;7:96.
62. Skalka AM. Retroviral DNA integration: lessons for transposon shuffling. *Gene.* 1993;135(1-2):175-182.
63. Temin HM. Sex and recombination in retroviruses. *Trends Genet.* 1991;7(3):71-74.
64. Ellison V, Brown PO. A stable complex between integrase and viral DNA ends mediates human immunodeficiency virus integration in vitro. *Proc Natl Acad Sci U S A.* 1994;91(15):7316-7320.
65. Schröder AR, Shinn P, Chen H, Berry C, Ecker JR, Bushman F. HIV-1 integration in the human genome favors active genes and local hotspots. *Cell.* 2002;110(4):521-529.
66. Craigie R, Bushman FD. HIV DNA integration. *Cold Spring Harb Perspect Med.* 2012;2(7):a006890.
67. Blankson JN, Persaud D, Siliciano RF. The challenge of viral reservoirs in HIV-1 infection. *Annu Rev Med.* 2002;53:557-593.
68. Siliciano JD, Kajdas J, Finzi D, et al. Long-term follow-up studies confirm the stability of the latent reservoir for HIV-1 in resting CD4+ T cells. *Nat Med.* 2003;9(6):727-728.

69. Karn J, Stoltzfus CM. Transcriptional and posttranscriptional regulation of HIV-1 gene expression. *Cold Spring Harb Perspect Med.* 2012;2(2):a006916.
70. Jouvenet N, Bieniasz PD, Simon SM. Imaging the biogenesis of individual HIV-1 virions in live cells. *Nature.* 2008;454(7201):236-240.
71. Skaug B, Chen ZJ. Emerging role of ISG15 in antiviral immunity. *Cell.* 2010;143(2):187-190.
72. Neil SJ, Eastman SW, Jouvenet N, Bieniasz PD. HIV-1 Vpu promotes release and prevents endocytosis of nascent retrovirus particles from the plasma membrane. *PLoS Pathog.* 2006;2(5):e39.
73. Briggs JA, Grünewald K, Glass B, Förster F, Kräusslich HG, Fuller SD. The mechanism of HIV-1 core assembly: insights from three-dimensional reconstructions of authentic virions. *Structure.* 2006;14(1):15-20.
74. Müller-Trutwin MC, Corbet S, Souquière S, et al. SIVcpz from a naturally infected Cameroonian chimpanzee: biological and genetic comparison with HIV-1 N. *J Med Primatol.* 2000;29(3-4):166-172.
75. Silvestri G, Sodora DL, Koup RA, et al. Nonpathogenic SIV infection of sooty mangabeys is characterized by limited bystander immunopathology despite chronic high-level viremia. *Immunity.* 2003;18(3):441-452.
76. Letvin NL, Daniel MD, Sehgal PK, et al. Induction of AIDS-like disease in macaque monkeys with T-cell tropic retrovirus STLV-III. *Science.* 1985;230(4721):71-73.
77. Letvin NL, King NW. Immunologic and pathologic manifestations of the infection of rhesus monkeys with simian immunodeficiency virus of macaques. *J Acquir Immune Defic Syndr.* 1990;3(11):1023-1040.
78. Khabbaz RF, Rowe T, Murphey-Corb M, et al. Simian immunodeficiency virus needlestick accident in a laboratory worker. *Lancet.* 1992;340(8814):271-273.
79. Khabbaz RF, Heneine W, George JR, et al. Brief report: infection of a laboratory worker with simian immunodeficiency virus. *N Engl J Med.* 1994;330(3):172-177.
80. Bibollet-Ruche F, Heigele A, Keele BF, et al. Efficient SIVcpz replication in human lymphoid tissue requires viral matrix protein adaptation. *J Clin Invest.* 2012;122(5):1644-1652.
81. Peeters M, Courgnaud V, Abela B, et al. Risk to human health from a plethora of simian immunodeficiency viruses in primate bushmeat. *Emerg Infect Dis.* 2002;8(5):451-457.
82. Sharp PM, Hahn BH. Origins of HIV and the AIDS pandemic. *Cold Spring Harb Perspect Med.* 2011;1(1):a006841.
83. Djoko CF, Wolfe ND, Aghokeng AF, et al. Failure to detect simian immunodeficiency virus infection in a large Cameroonian cohort with high non-human primate exposure. *Ecohealth.* 2012;9(1):17-23.
84. Ayouba A, Mauclère P, Martin PM, et al. HIV-1 group O infection in Cameroon, 1986 to 1998. *Emerg Infect Dis.* 2001;7(3):466-467.
85. Damond F, Apetrei C, Robertson DL, et al. Variability of human immunodeficiency virus type 2 (hiv-2) infecting patients living in france. *Virology.* 2001;280(1):19-30.
86. Ishikawa K, Janssens W, Banor JS, et al. Genetic analysis of HIV type 2 from Ghana and Guinea-Bissau, West Africa. *AIDS Res Hum Retroviruses.* 2001;17(17):1661-1663.
87. Pandrea I, Sodora DL, Silvestri G, Apetrei C. Into the wild: simian immunodeficiency virus (SIV) infection in natural hosts. *Trends Immunol.* 2008;29(9):419-428.
88. Pandrea I, Apetrei C. Where the wild things are: pathogenesis of SIV infection in African nonhuman primate hosts. *Curr HIV/AIDS Rep.* 2010;7(1):28-36.
89. Wertheim JO, Worobey M. Dating the age of the SIV lineages that gave rise to HIV-1 and HIV-2. *PLoS Comput Biol.* 2009;5(5):e1000377.
90. Ariën KK, Abraha A, Quiñones-Mateu ME, Kestens L, Vanham G, Arts EJ. The replicative fitness of primary human immunodeficiency virus type 1 (HIV-1) group M, HIV-1 group O, and HIV-2 isolates. *J Virol.* 2005;79(14):8979-8990.

91. Schroder K, Hertzog PJ, Ravasi T, Hume DA. Interferon-gamma: an overview of signals, mechanisms and functions. *J Leukoc Biol.* 2004;75(2):163-189.
92. Sheehy AM, Gaddis NC, Choi JD, Malim MH. Isolation of a human gene that inhibits HIV-1 infection and is suppressed by the viral Vif protein. *Nature.* 2002;418(6898):646-650.
93. Koning FA, Newman EN, Kim EY, Kunstman KJ, Wolinsky SM, Malim MH. Defining APOBEC3 expression patterns in human tissues and hematopoietic cell subsets. *J Virol.* 2009;83(18):9474-9485.
94. Harris RS, Petersen-Mahrt SK, Neuberger MS. RNA editing enzyme APOBEC1 and some of its homologs can act as DNA mutators. *Mol Cell.* 2002;10(5):1247-1253.
95. Newman EN, Holmes RK, Craig HM, et al. Antiviral function of APOBEC3G can be dissociated from cytidine deaminase activity. *Curr Biol.* 2005;15(2):166-170.
96. Mariani R, Chen D, Schröfelbauer B, et al. Species-specific exclusion of APOBEC3G from HIV-1 virions by Vif. *Cell.* 2003;114(1):21-31.
97. Goila-Gaur R, Strebel K. HIV-1 Vif, APOBEC, and intrinsic immunity. *Retrovirology.* 2008;5:51.
98. Bogerd HP, Doehle BP, Wiegand HL, Cullen BR. A single amino acid difference in the host APOBEC3G protein controls the primate species specificity of HIV type 1 virion infectivity factor. *Proc Natl Acad Sci U S A.* 2004;101(11):3770-3774.
99. Fourati S, Malet I, Binka M, et al. Partially active HIV-1 Vif alleles facilitate viral escape from specific antiretrovirals. *AIDS.* 2010;24(15):2313-2321.
100. Perez-Caballero D, Zang T, Ebrahimi A, et al. Tetherin inhibits HIV-1 release by directly tethering virions to cells. *Cell.* 2009;139(3):499-511.
101. Jouvenet N, Neil SJ, Zhadina M, et al. Broad-spectrum inhibition of retroviral and filoviral particle release by tetherin. *J Virol.* 2009;83(4):1837-1844.
102. Neil SJ, Zang T, Bieniasz PD. Tetherin inhibits retrovirus release and is antagonized by HIV-1 Vpu. *Nature.* 2008;451(7177):425-430.
103. Zhang F, Wilson SJ, Landford WC, et al. Nef proteins from simian immunodeficiency viruses are tetherin antagonists. *Cell Host Microbe.* 2009;6(1):54-67.
104. Le Tortorec A, Neil SJ. Antagonism to and intracellular sequestration of human tetherin by the human immunodeficiency virus type 2 envelope glycoprotein. *J Virol.* 2009;83(22):11966-11978.
105. Jia B, Serra-Moreno R, Neidermyer W, et al. Species-specific activity of SIV Nef and HIV-1 Vpu in overcoming restriction by tetherin/BST2. *PLoS Pathog.* 2009;5(5):e1000429.
106. Sauter D, Schindler M, Specht A, et al. Tetherin-driven adaptation of Vpu and Nef function and the evolution of pandemic and nonpandemic HIV-1 strains. *Cell Host Microbe.* 2009;6(5):409-421.
107. Sauter D, Hué S, Petit SJ, et al. HIV-1 Group P is unable to antagonize human tetherin by Vpu, Env or Nef. *Retrovirology.* 2011;8:103.
108. Goldstone DC, Ennis-Adeniran V, Hedden JJ, et al. HIV-1 restriction factor SAMHD1 is a deoxynucleoside triphosphate triphosphohydrolase. *Nature.* 2011;480(7377):379-382.
109. Ayinde D, Maudet C, Transy C, Margottin-Goguet F. Limelight on two HIV/SIV accessory proteins in macrophage infection: is Vpx overshadowing Vpr? *Retrovirology.* 2010;7:35.
110. Hofmann H, Logue EC, Bloch N, et al. The Vpx lentiviral accessory protein targets SAMHD1 for degradation in the nucleus. *J Virol.* 2012;86(23):12552-12560.
111. Laguette N, Rahm N, Sobhian B, et al. Evolutionary and functional analyses of the interaction between the myeloid restriction factor SAMHD1 and the lentiviral Vpx protein. *Cell Host Microbe.* 2012;11(2):205-217.
112. Lim ES, Fregoso OI, McCoy CO, Matsen FA, Malik HS, Emerman M. The ability of primate lentiviruses to degrade the monocyte restriction factor SAMHD1 preceded the birth of the viral accessory protein Vpx. *Cell Host Microbe.* 2012;11(2):194-204.

113. Reinhard C, Bottinelli D, Kim B, Luban J. Vpx rescue of HIV-1 from the antiviral state in mature dendritic cells is independent of the intracellular deoxynucleotide concentration. *Retrovirology.* 2014;11:12.
114. Manel N, Hogstad B, Wang Y, Levy DE, Unutmaz D, Littman DR. A cryptic sensor for HIV-1 activates antiviral innate immunity in dendritic cells. *Nature.* 2010;467(7312):214-217.
115. Almond D, Kimura T, Kong X, Swetnam J, Zolla-Pazner S, Cardozo T. Structural conservation predominates over sequence variability in the crown of HIV type 1's V3 loop. *AIDS Res Hum Retroviruses.* 2010;26(6):717-723.
116. Korber B, Gaschen B, Yusim K, Thakallapally R, Kesmir C, Detours V. Evolutionary and immunological implications of contemporary HIV-1 variation. *Br Med Bull.* 2001;58:19-42.
117. Lauring AS, Andino R. Quasispecies theory and the behavior of RNA viruses. *PLoS Pathog.* 2010;6(7):e1001005.
118. Korber B, Muldoon M, Theiler J, et al. Timing the ancestor of the HIV-1 pandemic strains. *Science.* 2000;288(5472):1789-1796.
119. Robertson DL, Anderson JP, Bradac JA, et al. HIV-1 nomenclature proposal. *Science.* 2000;288(5463):55-56.
120. Gao F, Robertson DL, Morrison SG, et al. The heterosexual human immunodeficiency virus type 1 epidemic in Thailand is caused by an intersubtype (A/E) recombinant of African origin. *J Virol.* 1996;70(10):7013-7029.
121. HIV-1 subtypes: implications for epidemiology, pathogenicity, vaccines and diagnostics. Workshop Report from the European Commission (DG XII, INCO-DC) and the Joint United Nations Programme on HIV/AIDS. *AIDS.* 1997;11(15):17-36.
122. Cornelissen M, van Den Burg R, Zorgdrager F, Goudsmit J. Spread of distinct human immunodeficiency virus type 1 AG recombinant lineages in Africa. *J Gen Virol.* 2000;81(Pt 2):515-523.
123. McCutchan FE. Global epidemiology of HIV. *J Med Virol.* 2006;78 Suppl 1:S7-S12.
124. Konings FA, Haman GR, Xue Y, et al. Genetic analysis of HIV-1 strains in rural eastern Cameroon indicates the evolution of second-generation recombinants to circulating recombinant forms. *J Acquir Immune Defic Syndr.* 2006;42(3):331-341.
125. Grenfell BT, Pybus OG, Gog JR, et al. Unifying the epidemiological and evolutionary dynamics of pathogens. *Science.* 2004;303(5656):327-332.
126. Abram ME, Ferris AL, Shao W, Alvord WG, Hughes SH. Nature, position, and frequency of mutations made in a single cycle of HIV-1 replication. *J Virol.* 2010;84(19):9864-9878.
127. Kirschner D, Webb GF, Cloyd M. Model of HIV-1 disease progression based on virus-induced lymph node homing and homing-induced apoptosis of CD4+ lymphocytes. *J Acquir Immune Defic Syndr.* 2000;24(4):352-362.
128. Perelson AS, Neumann AU, Markowitz M, Leonard JM, Ho DD. HIV-1 dynamics in vivo: virion clearance rate, infected cell life-span, and viral generation time. *Science.* 1996;271(5255):1582-1586.
129. Rambaut A, Posada D, Crandall KA, Holmes EC. The causes and consequences of HIV evolution. *Nat Rev Genet.* 2004;5(1):52-61.
130. Shriner D, Rodrigo AG, Nickle DC, Mullins JI. Pervasive genomic recombination of HIV-1 in vivo. *Genetics.* 2004;167(4):1573-1583.
131. Grivel JC, Shattock RJ, Margolis LB. Selective transmission of R5 HIV-1 variants: where is the gatekeeper? *J Transl Med.* 2011;9 Suppl 1:S6.
132. Derdeyn CA, Decker JM, Bibollet-Ruche F, et al. Envelope-constrained neutralization-sensitive HIV-1 after heterosexual transmission. *Science.* 2004;303(5666):2019-2022.
133. Haaland RE, Hawkins PA, Salazar-Gonzalez J, et al. Inflammatory genital infections mitigate a severe genetic bottleneck in heterosexual transmission of subtype A and C HIV-1. *PLoS Pathog.* 2009;5(1):e1000274.

134. Keele BF, Giorgi EE, Salazar-Gonzalez JF, et al. Identification and characterization of transmitted and early founder virus envelopes in primary HIV-1 infection. *Proc Natl Acad Sci U S A.* 2008;105(21):7552-7557.
135. Baeten JM, Chohan B, Lavreys L, 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.* 2007;195(8):1177-1180.
136. Kiwanuka N, Laeyendecker O, Quinn TC, et al. HIV-1 subtypes and differences in heterosexual HIV transmission among HIV-discordant couples in Rakai, Uganda. *AIDS.* 2009;23(18):2479-2484.
137. Nofemela A, Bandawe G, Thebus R, et al. Defining the human immunodeficiency virus type 1 transmission genetic bottleneck in a region with multiple circulating subtypes and recombinant forms. *Virology.* 2011;415(2):107-113.
138. Chaillon A, Samleerat T, Zoveda F, et al. Estimating the timing of mother-to-child transmission of the human immunodeficiency virus type 1 using a viral molecular evolution model. *PLoS One.* 2014;9(4):e90421.
139. Lythgoe KA, Fraser C. New insights into the evolutionary rate of HIV-1 at the within-host and epidemiological levels. *Proc Biol Sci.* 2012;279(1741):3367-3375.
140. Finzi D, Hermankova M, Pierson T, et al. Identification of a reservoir for HIV-1 in patients on highly active antiretroviral therapy. *Science.* 1997;278(5341):1295-1300.
141. Redd AD, Collinson-Streng AN, Chatziandreou N, et al. Previously transmitted HIV-1 strains are preferentially selected during subsequent sexual transmissions. *J Infect Dis.* 2012;206(9):1433-1442.
142. Sagar M, Laeyendecker O, Lee S, et al. Selection of HIV variants with signature genotypic characteristics during heterosexual transmission. *J Infect Dis.* 2009;199(4):580-589.
143. Maljkovic Berry I, Ribeiro R, Kothari M, 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.* 2007;81(19):10625-10635.
144. Watts JM, Dang KK, Gorelick RJ, et al. Architecture and secondary structure of an entire HIV-1 RNA genome. *Nature.* 2009;460(7256):711-716.
145. Bar KJ, Li H, Chamberland A, et al. Wide variation in the multiplicity of HIV-1 infection among injection drug users. *J Virol.* 2010;84(12):6241-6247.
146. Li H, Bar KJ, Wang S, et al. High Multiplicity Infection by HIV-1 in Men Who Have Sex with Men. *PLoS Pathog.* 2010;6(5):e1000890.
147. Abrahams MR, Anderson JA, Giorgi EE, 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.* 2009;83(8):3556-3567.
148. Lyles RH, Muñoz A, Yamashita TE, et al. Natural history of human immunodeficiency virus type 1 viremia after seroconversion and proximal to AIDS in a large cohort of homosexual men. Multicenter AIDS Cohort Study. *J Infect Dis.* 2000;181(3):872-880.
149. Mellors JW, Rinaldo CR, Gupta P, White RM, Todd JA, Kingsley LA. Prognosis in HIV-1 infection predicted by the quantity of virus in plasma. *Science.* 1996;272(5265):1167-1170.
150. Salemi M. The intra-host evolutionary and population dynamics of human immunodeficiency virus type 1: a phylogenetic perspective. *Infect Dis Rep.* 2013;5(Suppl 1):e3.
151. Kwong PD, Doyle ML, Casper DJ, et al. HIV-1 evades antibody-mediated neutralization through conformational masking of receptor-binding sites. *Nature.* 2002;420(6916):678-682.
152. Han C, Kawana-Tachikawa A, Shimizu A, et al. Switching and emergence of CTL epitopes in HIV-1 infection. *Retrovirology.* 2014;11:38.
153. Levesque MC, Moody MA, Hwang KK, et al. Polyclonal B cell differentiation and loss of gastrointestinal tract germinal centers in the earliest stages of HIV-1 infection. *PLoS Med.* 2009;6(7):e1000107.

154. Goonetilleke N, Liu MK, Salazar-Gonzalez JF, et al. The first T cell response to transmitted/founder virus contributes to the control of acute viremia in HIV-1 infection. *J Exp Med.* 2009;206(6):1253-1272.
155. Boutwell CL, Rolland MM, Herbeck JT, Mullins JI, Allen TM. Viral evolution and escape during acute HIV-1 infection. *J Infect Dis.* 2010;202 Suppl 2:S309-314.
156. Streeck H, Jolin JS, Qi Y, et al. Human immunodeficiency virus type 1-specific CD8+ T-cell responses during primary infection are major determinants of the viral set point and loss of CD4+ T cells. *J Virol.* 2009;83(15):7641-7648.
157. De Luca A, Hamers RL, Schapiro JM. Antiretroviral treatment sequencing strategies to overcome HIV type 1 drug resistance in adolescents and adults in low-middle-income countries. *J Infect Dis.* 2013;207 Suppl 2:S63-69.
158. Wainberg MA. The impact of the M184V substitution on drug resistance and viral fitness. *Expert Rev Anti Infect Ther.* 2004;2(1):147-151.
159. Kaslow RA, Rivers C, Tang J, et al. Polymorphisms in HLA class I genes associated with both favorable prognosis of human immunodeficiency virus (HIV) type 1 infection and positive cytotoxic T-lymphocyte responses to ALVAC-HIV recombinant canarypox vaccines. *J Virol.* 2001;75(18):8681-8689.
160. Altfeld M, Kalife ET, Qi Y, et al. HLA Alleles Associated with Delayed Progression to AIDS Contribute Strongly to the Initial CD8(+) T Cell Response against HIV-1. *PLoS Med.* 2006;3(10):e403.
161. Allen TM, Altfeld M. Crippling HIV one mutation at a time. *J Exp Med.* 2008;205(5):1003-1007.
162. Brockman MA, Schneidewind A, Lahaie M, et al. Escape and compensation from early HLA-B57-mediated cytotoxic T-lymphocyte pressure on human immunodeficiency virus type 1 Gag alter capsid interactions with cyclophilin A. *J Virol.* 2007;81(22):12608-12618.
163. Deeks SG, Walker BD. Human immunodeficiency virus controllers: mechanisms of durable virus control in the absence of antiretroviral therapy. *Immunity.* 2007;27(3):406-416.
164. Miura T, Brockman MA, Schneidewind A, et al. HLA-B57/B*5801 human immunodeficiency virus type 1 elite controllers select for rare gag variants associated with reduced viral replication capacity and strong cytotoxic T-lymphocyte [corrected] recognition. *J Virol.* 2009;83(6):2743-2755.
165. Martinez-Picado J, Prado JG, Fry EE, et al. Fitness cost of escape mutations in p24 Gag in association with control of human immunodeficiency virus type 1. *J Virol.* 2006;80(7):3617-3623.
166. Frater AJ, Brown H, Oxenius A, et al. Effective T-cell responses select human immunodeficiency virus mutants and slow disease progression. *J Virol.* 2007;81(12):6742-6751.
167. Prince JL, Claiborne DT, Carlson JM, et al. Role of transmitted Gag CTL polymorphisms in defining replicative capacity and early HIV-1 pathogenesis. *PLoS Pathog.* 2012;8(11):e1003041.
168. Shriner D, Shankarappa R, Jensen MA, et al. Influence of random genetic drift on human immunodeficiency virus type 1 env evolution during chronic infection. *Genetics.* 2004;166(3):1155-1164.
169. Edwards CT, Holmes EC, Pybus OG, et al. Evolution of the human immunodeficiency virus envelope gene is dominated by purifying selection. *Genetics.* 2006;174(3):1441-1453.
170. Salemi M, Burkhardt BR, Gray RR, Ghaffari G, Sleasman JW, Goodenow MM. Phylodynamics of HIV-1 in lymphoid and non-lymphoid tissues reveals a central role for the thymus in emergence of CXCR4-using quasispecies. *PLoS One.* 2007;2(9):e950.
171. Troyer RM, Collins KR, Abraha A, et al. Changes in human immunodeficiency virus type 1 fitness and genetic diversity during disease progression. *J Virol.* 2005;79(14):9006-9018.

172. Frost SD, Wrin T, Smith DM, et al. Neutralizing antibody responses drive the evolution of human immunodeficiency virus type 1 envelope during recent HIV infection. *Proc Natl Acad Sci U S A.* 2005;102(51):18514-18519.
173. Zhu P, Liu J, Bess J, et al. Distribution and three-dimensional structure of AIDS virus envelope spikes. *Nature.* 2006;441(7095):847-852.
174. Poignard P, Saphire EO, Parren PW, Burton DR. gp120: Biologic aspects of structural features. *Annu Rev Immunol.* 2001;19:253-274.
175. Zwick MB, Saphire EO, Burton DR. gp41: HIV's shy protein. *Nat Med.* 2004;10(2):133-134.
176. 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(2):750-758.
177. Richman DD, Wrin T, Little SJ, Petropoulos CJ. Rapid evolution of the neutralizing antibody response to HIV type 1 infection. *Proc Natl Acad Sci U S A.* 2003;100(7):4144-4149.
178. Wei X, Decker JM, Wang S, et al. Antibody neutralization and escape by HIV-1. *Nature.* 2003;422(6929):307-312.
179. Fischer W, Ganusov VV, Giorgi EE, et al. Transmission of single HIV-1 genomes and dynamics of early immune escape revealed by ultra-deep sequencing. *PLoS One.* 2010;5(8):e12303.
180. Carbonetti S, Oliver BG, Glenn J, Stamatatos L, Sather DN. Soluble HIV-1 envelope immunogens derived from an elite neutralizer elicit cross-reactive V1V2 antibodies and low potency neutralizing antibodies. *PLoS One.* 2014;9(1):e86905.
181. Tomaras GD, Yates NL, Liu P, et al. Initial B-cell responses to transmitted human immunodeficiency virus type 1: virion-binding immunoglobulin M (IgM) and IgG antibodies followed by plasma anti-gp41 antibodies with ineffective control of initial viremia. *J Virol.* 2008;82(24):12449-12463.
182. Gray ES, Moore PL, Choge IA, et al. Neutralizing antibody responses in acute human immunodeficiency virus type 1 subtype C infection. *J Virol.* 2007;81(12):6187-6196.
183. Bunnik EM, Pisas L, van Nuenen AC, Schuitemaker H. Autologous neutralizing humoral immunity and evolution of the viral envelope in the course of subtype B human immunodeficiency virus type 1 infection. *J Virol.* 2008;82(16):7932-7941.
184. Overbaugh J, Morris L. The Antibody Response against HIV-1. *Cold Spring Harb Perspect Med.* 2012;2(1):a007039.
185. Rong R, Li B, Lynch RM, et al. Escape from autologous neutralizing antibodies in acute/early subtype C HIV-1 infection requires multiple pathways. *PLoS Pathog.* 2009;5(9):e1000594.
186. Moore PL, Gray ES, Morris L. Specificity of the autologous neutralizing antibody response. *Curr Opin HIV AIDS.* 2009;4(5):358-363.
187. Gray ES, Madiga MC, Hermanus T, et al. The neutralization breadth of HIV-1 develops incrementally over four years and is associated with CD4+ T cell decline and high viral load during acute infection. *J Virol.* 2011;85(10):4828-4840.
188. Stamatatos L, Morris L, Burton DR, Mascola JR. Neutralizing antibodies generated during natural HIV-1 infection: good news for an HIV-1 vaccine? *Nat Med.* 2009;15(8):866-870.
189. Scheid JF, Mouquet H, Feldhahn N, et al. Broad diversity of neutralizing antibodies isolated from memory B cells in HIV-infected individuals. *Nature.* 2009;458(7238):636-640.
190. Kirchherr JL, Hamilton J, Lu X, et al. Identification of amino acid substitutions associated with neutralization phenotype in the human immunodeficiency virus type-1 subtype C gp120. *Virology.* 2011;409(2):163-174.
191. Wu X, Yang ZY, Li Y, et al. Rational design of envelope identifies broadly neutralizing human monoclonal antibodies to HIV-1. *Science.* 2010;329(5993):856-861.
192. Burton DR, Stanfield RL, Wilson IA. Antibody vs. HIV in a clash of evolutionary titans. *Proc Natl Acad Sci U S A.* 2005;102(42):14943-14948.
193. Goepfert PA, Lumm W, Farmer P, et al. Transmission of HIV-1 Gag immune escape mutations is associated with reduced viral load in linked recipients. *J Exp Med.* 2008;205(5):1009-1017.

194. Goulder PJ, Phillips RE, Colbert RA, et al. Late escape from an immunodominant cytotoxic T-lymphocyte response associated with progression to AIDS. *Nat Med.* 1997;3(2):212-217.
195. Goulder PJ, Brander C, Tang Y, et al. Evolution and transmission of stable CTL escape mutations in HIV infection. *Nature.* 2001;412(6844):334-338.
196. Moore CB, John M, James IR, Christiansen FT, Witt CS, Mallal SA. Evidence of HIV-1 adaptation to HLA-restricted immune responses at a population level. *Science.* 2002;296(5572):1439-1443.
197. Bunnik EM, Euler Z, Welkers MR, et al. Adaptation of HIV-1 envelope gp120 to humoral immunity at a population level. *Nat Med.* 2010;16(9):995-997.
198. Paillart JC, Shehu-Xhilaga M, Marquet R, Mak J. Dimerization of retroviral RNA genomes: an inseparable pair. *Nat Rev Microbiol.* 2004;2(6):461-472.
199. Zhuang J, Jetzt AE, Sun G, et al. Human immunodeficiency virus type 1 recombination: rate, fidelity, and putative hot spots. *J Virol.* 2002;76(22):11273-11282.
200. Blackard JT, Cohen DE, Mayer KH. Human immunodeficiency virus superinfection and recombination: current state of knowledge and potential clinical consequences. *Clin Infect Dis.* 2002;34(8):1108-1114.
201. Galetto R, Negroni M. Mechanistic features of recombination in HIV. *AIDS Rev.* 2005;7(2):92-102.
202. Smith DM, Richman DD, Little SJ. HIV superinfection. *J Infect Dis.* 2005;192(3):438-444.
203. Bachmann MH, Mathiason-Dubard C, Learn GH, et al. Genetic diversity of feline immunodeficiency virus: dual infection, recombination, and distinct evolutionary rates among envelope sequence clades. *J Virol.* 1997;71(6):4241-4253.
204. Jin MJ, Hui H, Robertson DL, et al. Mosaic genome structure of simian immunodeficiency virus from west African green monkeys. *EMBO J.* 1994;13(12):2935-2947.
205. Hemelaar J, Gouws E, Ghys PD, Osmanov S, Characterisation W-UNfHIa. Global trends in molecular epidemiology of HIV-1 during 2000-2007. *AIDS.* 2011;25(5):679-689.
206. Buonaguro L, Tornesello ML, Buonaguro FM. Human immunodeficiency virus type 1 subtype distribution in the worldwide epidemic: pathogenetic and therapeutic implications. *J Virol.* 2007;81(19):10209-10219.
207. Piyasirisilp S, McCutchan FE, Carr JK, et al. A recent outbreak of human immunodeficiency virus type 1 infection in southern China was initiated by two highly homogeneous, geographically separated strains, circulating recombinant form AE and a novel BC recombinant. *J Virol.* 2000;74(23):11286-11295.
208. Lau KA, Wang B, Kamarulzaman A, Ngb KP, Saksena NK. Continuous crossover(s) events of HIV-1 CRF01_AE and B subtype strains in Malaysia: evidence of rapid and extensive HIV-1 evolution in the region. *Curr HIV Res.* 2008;6(2):108-116.
209. Guimarães ML, Eyer-Silva WA, Couto-Fernandez JC, Morgado MG. Identification of two new CRF_BF in Rio de Janeiro State, Brazil. *AIDS.* 2008;22(3):433-435.
210. Sanabani SS, Pastena ER, Neto WK, Martinez VP, Sabino EC. Characterization and frequency of a newly identified HIV-1 BF1 intersubtype circulating recombinant form in São Paulo, Brazil. *Virol J.* 2010;7:74.
211. Simon-Loriere E, Holmes EC. Why do RNA viruses recombine? *Nat Rev Microbiol.* 2011;9(8):617-626.
212. Burke DS. Recombination in HIV: an important viral evolutionary strategy. *Emerg Infect Dis.* 1997;3(3):253-259.
213. Kim EY, Busch M, Abel K, et al. Retroviral recombination in vivo: viral replication patterns and genetic structure of simian immunodeficiency virus (SIV) populations in rhesus macaques after simultaneous or sequential intravaginal inoculation with SIVmac239Deltavpx/Deltavpr and SIVmac239Deltanef. *J Virol.* 2005;79(8):4886-4895.

214. Peeters M, Liegeois F, Torimiro N, et al. Characterization of a highly replicative intergroup M/O human immunodeficiency virus type 1 recombinant isolated from a Cameroonian patient. *J Virol.* 1999;73(9):7368-7375.
215. Rousseau CM, Learn GH, Bhattacharya T, et al. Extensive intrasubtype recombination in South African human immunodeficiency virus type 1 subtype C infections. *J Virol.* 2007;81(9):4492-4500.
216. Kiwelu IE, Novitsky V, Margolin L, et al. Frequent intra-subtype recombination among HIV-1 circulating in Tanzania. *PLoS One.* 2013;8(8):e71131.
217. Hemelaar J. The origin and diversity of the HIV-1 pandemic. *Trends Mol Med.* 2012;18(3):182-192.
218. Ariën KK, Vanham G, Arts EJ. Is HIV-1 evolving to a less virulent form in humans? *Nat Rev Microbiol.* 2007;5(2):141-151.
219. Lau KA, Wang B, Saksena NK. Emerging trends of HIV epidemiology in Asia. *AIDS Rev.* 2007;9(4):218-229.
220. Frange P, Galimand J, Vidal N, et al. New and old complex recombinant HIV-1 strains among patients with primary infection in 1996-2006 in France: the French ANRS CO06 primo cohort study. *Retrovirology.* 2008;5:69.
221. Lai A, Riva C, Marconi A, et al. Changing patterns in HIV-1 non-B clade prevalence and diversity in Italy over three decades. *HIV Med.* 2010;11(9):593-602.
222. Ciccozzi M, Lai A, Ebranati E, et al. Phylogeographic reconstruction of HIV type 1B in Montenegro and the Balkan region. *AIDS Res Hum Retroviruses.* 2012;28(10):1280-1284.
223. Yebra G, de Mulder M, Martín L, et al. Most HIV type 1 non-B infections in the Spanish cohort of antiretroviral treatment-naïve HIV-infected patients (CoRIS) are due to recombinant viruses. *J Clin Microbiol.* 2012;50(2):407-413.
224. Liu Y, Li L, Bao Z, et al. Identification of a novel HIV type 1 circulating recombinant form (CRF52_01B) in Southeast Asia. *AIDS Res Hum Retroviruses.* 2012;28(10):1357-1361.
225. Piot P, Quinn TC, Taelman H, et al. Acquired immunodeficiency syndrome in a heterosexual population in Zaire. *Lancet.* 1984;2(8394):65-69.
226. Robbins KE, Lemey P, Pybus OG, et al. U.S. Human immunodeficiency virus type 1 epidemic: date of origin, population history, and characterization of early strains. *J Virol.* 2003;77(11):6359-6366.
227. Junqueira DM, de Medeiros RM, Matte MC, et al. Reviewing the history of HIV-1: spread of subtype B in the Americas. *PLoS One.* 2011;6(11):e27489.
228. Simoes EA, Babu PG, John TJ, et al. Evidence for HTLV-III infection in prostitutes in Tamil Nadu (India). *Indian J Med Res.* 1987;85:335-338.
229. Sehgal S. South African HIV strains in India: did clade C follow the mandrax route? *Indian J Med Microbiol.* 2014;32(2):110-111.
230. Jadhav A, Bhattacharjee P, Raghavendra T, et al. Risky Behaviors among HIV-Positive Female Sex Workers in Northern Karnataka, India. *AIDS Res Treat.* 2013;2013:878151.
231. Rao JV, Ganguly NK, Mehendale SM, Bollinger RC. India's response to the HIV epidemic. *Lancet.* 2004;364(9442):1296-1297.
232. Lole KS, Bollinger RC, Paranjape RS, 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(1):152-160.
233. Bobkov A, Cheingsong-Popov R, Garaev M, et al. Identification of an env G subtype and heterogeneity of HIV-1 strains in the Russian Federation and Belarus. *AIDS.* 1994;8(12):1649-1655.
234. Bobkov A, Cheingsong-Popov R, Selimova L, et al. Genetic heterogeneity of HIV type 1 in Russia: identification of H variants and relationship with epidemiological data. *AIDS Res Hum Retroviruses.* 1996;12(18):1687-1690.

235. Novitsky VA, Montano MA, Essex M. Molecular epidemiology of an HIV-1 subtype A subcluster among injection drug users in the Southern Ukraine. *AIDS Res Hum Retroviruses.* 1998;14(12):1079-1085.
236. Fleck F. Eastern Europe and Russia face world's fastest growing HIV epidemic. *BMJ.* 2004;328(7438):486.
237. Bobkov A, Cheingsong-Popov R, Selimova L, et al. An HIV type 1 epidemic among injecting drug users in the former Soviet Union caused by a homogeneous subtype A strain. *AIDS Res Hum Retroviruses.* 1997;13(14):1195-1201.
238. Bobkov AF, Kazennova EV, Selimova LM, et al. Temporal trends in the HIV-1 epidemic in Russia: predominance of subtype A. *J Med Virol.* 2004;74(2):191-196.
239. Glauser MP, Francioli P. Clinical and epidemiological survey of acquired immune deficiency syndrome in Europe. *Eur J Clin Microbiol.* 1984;3(1):55-58.
240. Casado C, Urtasun I, Saragosti S, et al. Different distribution of HIV type 1 genetic variants in European patients with distinct risk practices. *AIDS Res Hum Retroviruses.* 2000;16(3):299-304.
241. Hué S, Pillay D, Clewley JP, Pybus OG. Genetic analysis reveals the complex structure of HIV-1 transmission within defined risk groups. *Proc Natl Acad Sci U S A.* 2005;102(12):4425-4429.
242. Snoeck J, Van Laethem K, Hermans P, et al. Rising prevalence of HIV-1 non-B subtypes in Belgium: 1983-2001. *J Acquir Immune Defic Syndr.* 2004;35(3):279-285.
243. Aggarwal I, Smith M, Tatt ID, et al. Evidence for onward transmission of HIV-1 non-B subtype strains in the United Kingdom. *J Acquir Immune Defic Syndr.* 2006;41(2):201-209.
244. von Wyl V, Kouyos RD, Yerly S, et al. The role of migration and domestic transmission in the spread of HIV-1 non-B subtypes in Switzerland. *J Infect Dis.* 2011;204(7):1095-1103.
245. Palma AC, Araújo F, Duque V, et al. Molecular epidemiology and prevalence of drug resistance-associated mutations in newly diagnosed HIV-1 patients in Portugal. *Infect Genet Evol.* 2007;7(3):391-398.
246. Yerly S, Jost S, Monnat M, et al. HIV-1 co/super-infection in intravenous drug users. *AIDS.* 2004;18(10):1413-1421.
247. Liitsola K, Ristola M, Holmström P, et al. An outbreak of the circulating recombinant form AECM240 HIV-1 in the Finnish injection drug user population. *AIDS.* 2000;14(16):2613-2615.
248. Paraskevis D, Magiorkinis E, Magiorkinis G, et al. Increasing prevalence of HIV-1 subtype A in Greece: estimating epidemic history and origin. *J Infect Dis.* 2007;196(8):1167-1176.
249. Wu J, Meng Z, Xu J, et al. New emerging recombinant HIV-1 strains and close transmission linkage of HIV-1 strains in the Chinese MSM population indicate a new epidemic risk. *PLoS One.* 2013;8(1):e54322.
250. He X, Xing H, Ruan Y, et al. A comprehensive mapping of HIV-1 genotypes in various risk groups and regions across China based on a nationwide molecular epidemiologic survey. *PLoS One.* 2012;7(10):e47289.
251. Neogi U, Sood V, Banerjee S, et al. Global HIV-1 molecular epidemiology with special reference to genetic analysis of HIV-1 subtypes circulating in North India: functional and pathogenic implications of genetic variation. *Indian J Exp Biol.* 2009;47(6):424-431.
252. van Griensven F, Thienkrua W, McNicholl J, et al. Evidence of an explosive epidemic of HIV infection in a cohort of men who have sex with men in Thailand. *AIDS.* 2013;27(5):825-832.
253. Müller V, Fraser C, Herbeck JT. A strong case for viral genetic factors in HIV virulence. *Viruses.* 2011;3(3):204-216.
254. Shirreff G, Pellis L, Laeyendecker O, Fraser C. Transmission selects for HIV-1 strains of intermediate virulence: a modelling approach. *PLoS Comput Biol.* 2011;7(10):e1002185.
255. Departamento de Doenças Infecciosas UdReVE, Infecciosas NdVLdD. *Infeção VIH/SIDA: a situação em Portugal a 31 de dezembro de 2012.* Lisboa: Instituto Nacional de Saúde Doutor Ricardo Jorge;2013.

256. Janiec J, Haar K, Spiteri G, Likatavicius G, Van de Laar M, Amato-Gauci AJ. Surveillance of human immunodeficiency virus suggests that younger men who have sex with men are at higher risk of infection, European Union, 2003 to 2012. *Euro Surveill.* 2013;18(48):20644.
257. Bártolo I, Abecasis AB, Borrego P, et al. Origin and epidemiological history of HIV-1 CRF14_BG. *PLoS One.* 2011;6(9):e24130.
258. Holguín A, de Mulder M, Yebra G, López M, Soriano V. Increase of non-B subtypes and recombinants among newly diagnosed HIV-1 native Spaniards and immigrants in Spain. *Curr HIV Res.* 2008;6(4):327-334.
259. Martín JC, Holguín A, Soriano V. Prevalence of different HIV-1 subtypes in an urban clinic in Madrid. *Sex Transm Infect.* 2002;78(1):E1.
260. Machuca R, Bøgh M, Salminen M, et al. HIV-1 subtypes in Denmark. *Scand J Infect Dis.* 2001;33(9):697-701.
261. Liitsola K, Holmström P, Laukkanen T, Brummer-Korvenkontio H, Leinikki P, Salminen MO. Analysis of HIV-1 genetic subtypes in Finland reveals good correlation between molecular and epidemiological data. *Scand J Infect Dis.* 2000;32(5):475-480.
262. Hughes GJ, Fearnhill E, Dunn D, et al. Molecular phylodynamics of the heterosexual HIV epidemic in the United Kingdom. *PLoS Pathog.* 2009;5(9):e1000590.
263. CPLP. *Epidemia de VIH nos países de língua oficial portuguesa.* Brazil2010.
264. Esteves A, Parreira R, Piedade J, Venenno T, Canas-Ferreira WF. Genetic characterization of HIV type 1 and type 2 from Bissau, Guinea-Bissau (West Africa). *Virus Res.* 2000;68(1):51-61.
265. Oliveira V, Bártolo I, Borrego P, et al. Genetic diversity and drug resistance profiles in HIV type 1- and HIV type 2-infected patients from Cape Verde Islands. *AIDS Res Hum Retroviruses.* 2012;28(5):510-522.
266. de Pina-Araujo II, Guimarães ML, Bello G, Vicente AC, Morgado MG. Profile of the HIV Epidemic in Cape Verde: Molecular Epidemiology and Drug Resistance Mutations among HIV-1 and HIV-2 Infected Patients from Distinct Islands of the Archipelago. *PLoS One.* 2014;9(4):e96201.
267. Bártolo I, Casanovas J, Bastos R, et al. HIV-1 genetic diversity and transmitted drug resistance in health care settings in Maputo, Mozambique. *J Acquir Immune Defic Syndr.* 2009;51(3):323-331.
268. Barreto J, Liljestrand J, Palha de Sousa C, et al. HIV-1 and HIV-2 antibodies in pregnant women in the City of Maputo, Mozambique. A comparative study between 1982/1983 and 1990. *Scand J Infect Dis.* 1993;25(6):685-688.
269. Bártolo I, Rocha C, Bartolomeu J, et al. Highly divergent subtypes and new recombinant forms prevail in the HIV/AIDS epidemic in Angola: new insights into the origins of the AIDS pandemic. *Infect Genet Evol.* 2009;9(4):672-682.
270. Bártolo I, Epalanga M, Bartolomeu J, et al. High genetic diversity of human immunodeficiency virus type 1 in Angola. *AIDS Res Hum Retroviruses.* 2005;21(4):306-310.
271. Santos-Ferreira MO, Cohen T, Lourenço MH, Almeida MJ, Chamaret S, Montagnier L. A study of seroprevalence of HIV-1 and HIV-2 in six provinces of People's Republic of Angola: clues to the spread of HIV infection. *J Acquir Immune Defic Syndr.* 1990;3(8):780-786.
272. Yebra G, Rivas P, Herrero MD, et al. Clinical differences and viral diversity between newly HIV type 1-diagnosed African and non-African patients in Spain (2005-2007). *AIDS Res Hum Retroviruses.* 2009;25(1):37-44.
273. Williamson C, Engelbrecht S, Lambrick M, et al. HIV-1 subtypes in different risk groups in South Africa. *Lancet.* 1995;346(8977):782.
274. John-Stewart GC, Nduati RW, Rousseau CM, et al. Subtype C Is associated with increased vaginal shedding of HIV-1. *J Infect Dis.* 2005;192(3):492-496.
275. Yang C, Li M, Newman RD, et al. Genetic diversity of HIV-1 in western Kenya: subtype-specific differences in mother-to-child transmission. *AIDS.* 2003;17(11):1667-1674.

276. Renjifo B, Gilbert P, Chaplin B, et al. Preferential in-utero transmission of HIV-1 subtype C as compared to HIV-1 subtype A or D. *AIDS.* 2004;18(12):1629-1636.
277. Martínez AM, Hora VP, Santos AL, et al. Determinants of HIV-1 mother-to-child transmission in Southern Brazil. *An Acad Bras Cienc.* 2006;78(1):113-121.
278. Ahmad N. The vertical transmission of human immunodeficiency virus type 1: molecular and biological properties of the virus. *Crit Rev Clin Lab Sci.* 2005;42(1):1-34.
279. Chohan B, Lang D, Sagar M, et al. Selection for human immunodeficiency virus type 1 envelope glycosylation variants with shorter V1-V2 loop sequences occurs during transmission of certain genetic subtypes and may impact viral RNA levels. *J Virol.* 2005;79(10):6528-6531.
280. Taylor BS, Hammer SM. The challenge of HIV-1 subtype diversity. *N Engl J Med.* 2008;359(18):1965-1966.
281. Kiwanuka N, Robb M, Laeyendecker O, et al. HIV-1 viral subtype differences in the rate of CD4+ T-cell decline among HIV seroincident antiretroviral naive persons in Rakai district, Uganda. *J Acquir Immune Defic Syndr.* 2010;54(2):180-184.
282. Easterbrook PJ, Smith M, Mullen J, et al. Impact of HIV-1 viral subtype on disease progression and response to antiretroviral therapy. *J Int AIDS Soc.* 2010;13:4.
283. Kaleebu P, Nankya IL, Yirrell DL, 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.* 2007;45(1):28-33.
284. Huang W, Eshleman SH, Toma J, et al. Coreceptor tropism in human immunodeficiency virus type 1 subtype D: high prevalence of CXCR4 tropism and heterogeneous composition of viral populations. *J Virol.* 2007;81(15):7885-7893.
285. Ray N, Doms RW. HIV-1 coreceptors and their inhibitors. *Curr Top Microbiol Immunol.* 2006;303:97-120.
286. Kwon JA, Yoon SY, Lee CK, et al. Performance evaluation of three automated human immunodeficiency virus antigen-antibody combination immunoassays. *J Virol Methods.* 2006;133(1):20-26.
287. Plantier JC, Djemai M, Lemée V, et al. Census and analysis of persistent false-negative results in serological diagnosis of human immunodeficiency virus type 1 group O infections. *J Clin Microbiol.* 2009;47(9):2906-2911.
288. Vallari A, Bodelle P, Ngansop C, et al. Four new HIV-1 group N isolates from Cameroon: Prevalence continues to be low. *AIDS Res Hum Retroviruses.* 2010;26(1):109-115.
289. Aghokeng AF, Mpoudi-Ngole E, Dimodi H, et al. Inaccurate diagnosis of HIV-1 group M and O is a key challenge for ongoing universal access to antiretroviral treatment and HIV prevention in Cameroon. *PLoS One.* 2009;4(11):e7702.
290. Holguín A, Gutiérrez M, Portocarrero N, Rivas P, Baquero M. Performance of OraQuick Advance Rapid HIV-1/2 Antibody Test for detection of antibodies in oral fluid and serum/plasma in HIV-1+ subjects carrying different HIV-1 subtypes and recombinant variants. *J Clin Virol.* 2009;45(2):150-152.
291. Beelaert G, Fransen K. Evaluation of a rapid and simple fourth-generation HIV screening assay for qualitative detection of HIV p24 antigen and/or antibodies to HIV-1 and HIV-2. *J Virol Methods.* 2010;168(1-2):218-222.
292. Church D, Gregson D, Lloyd T, et al. Comparison of the RealTime HIV-1, COBAS TaqMan 48 v1.0, Easy Q v1.2, and Versant v3.0 assays for determination of HIV-1 viral loads in a cohort of Canadian patients with diverse HIV subtype infections. *J Clin Microbiol.* 2011;49(1):118-124.
293. Bourlet T, Signori-Schmuck A, Roche L, et al. HIV-1 load comparison using four commercial real-time assays. *J Clin Microbiol.* 2011;49(1):292-297.

294. Katsoulidou A, Rokka C, Issaris C, et al. Comparative evaluation of the performance of the Abbott RealTime HIV-1 assay for measurement of HIV-1 plasma viral load on genetically diverse samples from Greece. *Virol J.* 2011;8:10.

295. Swanson P, Huang S, Abravaya K, et al. Evaluation of performance across the dynamic range of the Abbott RealTime HIV-1 assay as compared to VERSANT HIV-1 RNA 3.0 and AMPLICOR HIV-1 MONITOR v1.5 using serial dilutions of 39 group M and O viruses. *J Virol Methods.* 2007;141(1):49-57.

296. Descamps D, Collin G, Letourneur F, et al. Susceptibility of human immunodeficiency virus type 1 group O isolates to antiretroviral agents: in vitro phenotypic and genotypic analyses. *J Virol.* 1997;71(11):8893-8898.

297. Smith RA, Anderson DJ, Pyrak CL, Preston BD, Gottlieb GS. Antiretroviral drug resistance in HIV-2: three amino acid changes are sufficient for classwide nucleoside analogue resistance. *J Infect Dis.* 2009;199(9):1323-1326.

298. Eshleman SH, Hoover DR, Chen S, et al. Nevirapine (NVP) resistance in women with HIV-1 subtype C, compared with subtypes A and D, after the administration of single-dose NVP. *J Infect Dis.* 2005;192(1):30-36.

299. Bocket L, Cheret A, Deuffic-Burban S, et al. Impact of human immunodeficiency virus type 1 subtype on first-line antiretroviral therapy effectiveness. *Antivir Ther.* 2005;10(2):247-254.

300. Frater AJ, Beardall A, Ariyoshi K, et al. Impact of baseline polymorphisms in RT and protease on outcome of highly active antiretroviral therapy in HIV-1-infected African patients. *AIDS.* 2001;15(12):1493-1502.

301. Geretti AM, Harrison L, Green H, et al. Effect of HIV-1 subtype on virologic and immunologic response to starting highly active antiretroviral therapy. *Clin Infect Dis.* 2009;48(9):1296-1305.

302. Kantor R, Katzenstein D. Drug resistance in non-subtype B HIV-1. *J Clin Virol.* 2004;29(3):152-159.

303. Kantor R. Impact of HIV-1 pol diversity on drug resistance and its clinical implications. *Curr Opin Infect Dis.* 2006;19(6):594-606.

304. Kantor R, Katzenstein DA, Efron B, et al. Impact of HIV-1 subtype and antiretroviral therapy on protease and reverse transcriptase genotype: results of a global collaboration. *PLoS Med.* 2005;2(4):e112.

305. Brenner B, Turner D, Oliveira M, et al. A V106M mutation in HIV-1 clade C viruses exposed to efavirenz confers cross-resistance to non-nucleoside reverse transcriptase inhibitors. *AIDS.* 2003;17(1):F1-5.

306. Deshpande A, Jauvin V, Magnin N, et al. Resistance mutations in subtype C HIV type 1 isolates from Indian patients of Mumbai receiving NRTIs plus NNRTIs and experiencing a treatment failure: resistance to AR. *AIDS Res Hum Retroviruses.* 2007;23(2):335-340.

307. Brenner BG, Oliveira M, Doualla-Bell F, et al. HIV-1 subtype C viruses rapidly develop K65R resistance to tenofovir in cell culture. *AIDS.* 2006;20(9):F9-13.

308. Hosseinipour MC, van Oosterhout JJ, Weigel R, et al. The public health approach to identify antiretroviral therapy failure: high-level nucleoside reverse transcriptase inhibitor resistance among Malawians failing first-line antiretroviral therapy. *AIDS.* 2009;23(9):1127-1134.

309. Coutsinos D, Invernizzi CF, Xu H, Brenner BG, Wainberg MA. Factors affecting template usage in the development of K65R resistance in subtype C variants of HIV type-1. *Antivir Chem Chemother.* 2010;20(3):117-131.

310. Martínez-Cajas JL, Pant-Pai N, Klein MB, Wainberg MA. Role of genetic diversity amongst HIV-1 non-B subtypes in drug resistance: a systematic review of virologic and biochemical evidence. *AIDS Rev.* 2008;10(4):212-223.

311. Poveda E, de Mendoza C, Martin-Carbonero L, et al. Prevalence of darunavir resistance mutations in HIV-1-infected patients failing other protease inhibitors. *J Antimicrob Chemother.* 2007;60(4):885-888.

312. R C, A G, P G, et al. Different substitutions under drug pressure at protease codon 82 in HIV-1 subtype G compared to subtype B infected individuals including a novel I82M resistance mutation. Vol 10. Antivir Ther2005.
313. Calazans A, Brindeiro R, Brindeiro P, et al. Low accumulation of L90M in protease from subtype F HIV-1 with resistance to protease inhibitors is caused by the L89M polymorphism. *J Infect Dis.* 2005;191(11):1961-1970.
314. Casado C, Pernas M, Alvaro T, et al. Coinfection and superinfection in patients with long-term, nonprogressive HIV-1 disease. *J Infect Dis.* 2007;196(6):895-899.
315. Pantaleo G, Koup RA. Correlates of immune protection in HIV-1 infection: what we know, what we don't know, what we should know. *Nat Med.* 2004;10(8):806-810.
316. Hoxie JA. Toward an antibody-based HIV-1 vaccine. *Annu Rev Med.* 2010;61:135-152.
317. Flynn NM, Forthal DN, Harro CD, et al. Placebo-controlled phase 3 trial of a recombinant glycoprotein 120 vaccine to prevent HIV-1 infection. *J Infect Dis.* 2005;191(5):654-665.
318. Pitisuttithum P, Gilbert P, Gurwith M, et al. Randomized, double-blind, placebo-controlled efficacy trial of a bivalent recombinant glycoprotein 120 HIV-1 vaccine among injection drug users in Bangkok, Thailand. *J Infect Dis.* 2006;194(12):1661-1671.
319. Simek MD, Rida W, Priddy FH, et al. Human immunodeficiency virus type 1 elite neutralizers: individuals with broad and potent neutralizing activity identified by using a high-throughput neutralization assay together with an analytical selection algorithm. *J Virol.* 2009;83(14):7337-7348.
320. Euler Z, van Gils MJ, Bunnik EM, et al. Cross-reactive neutralizing humoral immunity does not protect from HIV type 1 disease progression. *J Infect Dis.* 2010;201(7):1045-1053.
321. Scheid JF, Mouquet H, Ueberheide B, et al. Sequence and structural convergence of broad and potent HIV antibodies that mimic CD4 binding. *Science.* 2011;333(6049):1633-1637.
322. McLellan JS, Pancera M, Carrico C, et al. Structure of HIV-1 gp120 V1/V2 domain with broadly neutralizing antibody PG9. *Nature.* 2011;480(7377):336-343.
323. McMichael AJ, Borrow P, Tomaras GD, Goonetilleke N, Haynes BF. The immune response during acute HIV-1 infection: clues for vaccine development. *Nat Rev Immunol.* 2010;10(1):11-23.
324. Gaschen B, Taylor J, Yusim K, et al. Diversity considerations in HIV-1 vaccine selection. *Science.* 2002;296(5577):2354-2360.
325. Currier JR, Dowling WE, Wasunna KM, et al. Detection of high frequencies of HIV-1 cross-subtype reactive CD8 T lymphocytes in the peripheral blood of HIV-1-infected Kenyans. *AIDS.* 2003;17(15):2149-2157.
326. Ranasinghe SR, Kramer HB, Wright C, et al. The antiviral efficacy of HIV-specific $CD8^+$ T-cells to a conserved epitope is heavily dependent on the infecting HIV-1 isolate. *PLoS Pathog.* 2011;7(5):e1001341.
327. Gottlieb GS, Heath L, Nickle DC, et al. HIV-1 variation before seroconversion in men who have sex with men: analysis of acute/early HIV infection in the multicenter AIDS cohort study. *J Infect Dis.* 2008;197(7):1011-1015.
328. Hemelaar J, Gouws E, Ghys PD, Osmanov S. Global and regional distribution of HIV-1 genetic subtypes and recombinants in 2004. *AIDS.* 2006;20(16):W13-23.
329. Morison L, Buvé A, Zekeng L, et al. HIV-1 subtypes and the HIV epidemics in four cities in sub-Saharan Africa. *AIDS.* 2001;15 Suppl 4:S109-116.
330. Hudgens MG, Longini IM, Vanichseni S, et al. Subtype-specific transmission probabilities for human immunodeficiency virus type 1 among injecting drug users in Bangkok, Thailand. *Am J Epidemiol.* 2002;155(2):159-168.
331. Bobkov AF, Kazennova EV, Sukhanova AL, et al. An HIV type 1 subtype A outbreak among injecting drug users in Kazakhstan. *AIDS Res Hum Retroviruses.* 2004;20(10):1134-1136.

332. Atlas A, Granath F, Lindström A, Lidman K, Lindbäck S, Alaeus A. Impact of HIV type 1 genetic subtype on the outcome of antiretroviral therapy. *AIDS Res Hum Retroviruses.* 2005;21(3):221-227.
333. Keller M, Lu Y, Lalonde RG, Klein MB. Impact of HIV-1 viral subtype on CD4+ T-cell decline and clinical outcomes in antiretroviral naive patients receiving universal healthcare. *AIDS.* 2009;23(6):731-737.
334. Santos AF, Soares MA. HIV Genetic Diversity and Drug Resistance. *Viruses.* 2010;2(2):503-531.
335. Theys K, Vercauteren J, Snoeck J, et al. HIV-1 subtype is an independent predictor of reverse transcriptase mutation K65R in HIV-1 patients treated with combination antiretroviral therapy including tenofovir. *Antimicrob Agents Chemother.* 2013;57(2):1053-1056.
336. Descamps D, Chaix ML, André P, et al. French national sentinel survey of antiretroviral drug resistance in patients with HIV-1 primary infection and in antiretroviral-naive chronically infected patients in 2001-2002. *J Acquir Immune Defic Syndr.* 2005;38(5):545-552.
337. J M-P, C G, C DM, I E, P D, al e. Surveillance of drug resistance and HIV subtypes in newly diagnosed patients in Spain during 2004. Vol 10. Antivir Ther2005:S137.
338. Monno L, Brindicci G, Lo Caputo S, et al. HIV-1 subtypes and circulating recombinant forms (CRFs) from HIV-infected patients residing in two regions of central and southern Italy. *J Med Virol.* 2005;75(4):483-490.
339. Chaix ML, Seng R, Frange P, et al. Increasing HIV-1 non-B subtype primary infections in patients in France and effect of HIV subtypes on virological and immunological responses to combined antiretroviral therapy. *Clin Infect Dis.* 2013;56(6):880-887.
340. Ciccozzi M, Santoro MM, Giovanetti M, Andrissi L, Bertoli A, Ciotti M. HIV-1 non-B subtypes in Italy: a growing trend. *New Microbiol.* 2012;35(4):377-386.
341. Esteves A, Parreira R, Venenno T, et al. Molecular epidemiology of HIV type 1 infection in Portugal: high prevalence of non-B subtypes. *AIDS Res Hum Retroviruses.* 2002;18(5):313-325.
342. Esteves A, Parreira R, Piedade J, et al. Spreading of HIV-1 subtype G and envB/gagG recombinant strains among injecting drug users in Lisbon, Portugal. *AIDS Res Hum Retroviruses.* 2003;19(6):511-517.
343. Delgado E, Thomson MM, Villahermosa ML, et al. Identification of a newly characterized HIV-1 BG intersubtype circulating recombinant form in Galicia, Spain, which exhibits a pseudotype-like virion structure. *J Acquir Immune Defic Syndr.* 2002;29(5):536-543.
344. Paraskevis D, Nikolopoulos G, Tsiara C, et al. HIV-1 outbreak among injecting drug users in Greece, 2011: a preliminary report. *Euro Surveill.* 2011;16(36).
345. Hué S, Clewley JP, Cane PA, Pillay D. HIV-1 pol gene variation is sufficient for reconstruction of transmissions in the era of antiretroviral therapy. *AIDS.* 2004;18(5):719-728.
346. Brenner BG, Roger M, Routy JP, et al. High rates of forward transmission events after acute/early HIV-1 infection. *J Infect Dis.* 2007;195(7):951-959.
347. Lewis F, Hughes GJ, Rambaut A, Pozniak A, Leigh Brown AJ. Episodic sexual transmission of HIV revealed by molecular phylodynamics. *PLoS Med.* 2008;5(3):e50.
348. Pineda-Peña AC, Faria NR, Imbrechts S, et al. Automated subtyping of HIV-1 genetic sequences for clinical and surveillance purposes: performance evaluation of the new REGA version 3 and seven other tools. *Infect Genet Evol.* 2013;19:337-348.
349. Martin D, Rybicki E. RDP: detection of recombination amongst aligned sequences. *Bioinformatics.* 2000;16(6):562-563.
350. Edgar RC. MUSCLE: multiple sequence alignment with high accuracy and high throughput. *Nucleic Acids Res.* 2004;32(5):1792-1797.
351. Stamatakis A. RAxML-VI-HPC: maximum likelihood-based phylogenetic analyses with thousands of taxa and mixed models. *Bioinformatics.* 2006;22(21):2688-2690.

352. Guindon S, Dufayard JF, Lefort V, Anisimova M, Hordijk W, Gascuel O. New algorithms and methods to estimate maximum-likelihood phylogenies: assessing the performance of PhyML 3.0. *Syst Biol.* 2010;59(3):307-321.
353. Drummond AJ, Rambaut A. BEAST: Bayesian evolutionary analysis by sampling trees. *BMC Evol Biol.* 2007;7:214.
354. Darriba D, Taboada GL, Doallo R, Posada D. jModelTest 2: more models, new heuristics and parallel computing. *Nat Methods.* 2012;9(8):772.
355. Bennett DE, Camacho RJ, Otelea D, et al. Drug resistance mutations for surveillance of transmitted HIV-1 drug-resistance: 2009 update. *PLoS One.* 2009;4(3):e4724.
356. Torti C, Lapadula G, Izzo I, et al. Heterogeneity and penetration of HIV-1 non-subtype B viruses in an Italian province: public health implications. *Epidemiol Infect.* 2010;138(9):1298-1307.
357. Vermund SH, Leigh-Brown AJ. The HIV Epidemic: High-Income Countries. *Cold Spring Harb Perspect Med.* 2012;2(5):a007195.
358. J. BG, Stanley YP. Whites in Angola on the Eve of Independence: The Politics of Numbers. Vol 21. Africa Today1974:23-37.
359. ECDC. *HIV/AIDS surveillance in Europe 2012.* Stockholm: European Centre for Disease Prevention and Control;2013.
360. Sherr L, Lopman B, Kakowa M, et al. Voluntary counselling and testing: uptake, impact on sexual behaviour, and HIV incidence in a rural Zimbabwean cohort. *AIDS.* 2007;21(7):851-860.
361. Reniers G, Eaton J. Refusal bias in HIV prevalence estimates from nationally representative seroprevalence surveys. *AIDS.* 2009;23(5):621-629.
362. Hallett TB, Zaba B, Todd J, et al. Estimating incidence from prevalence in generalised HIV epidemics: methods and validation. *PLoS Med.* 2008;5(4):e80.
363. Mastro TD, Kim AA, Hallett T, et al. Estimating HIV Incidence in Populations Using Tests for Recent Infection: Issues, Challenges and the Way Forward. *J HIV AIDS Surveill Epidemiol.* 2010;2(1):1-14.
364. Sexton CJ, Costenbader EC, Vinh DT, et al. Correlation of prospective and cross-sectional measures of HIV type 1 incidence in a higher-risk cohort in Ho Chi Minh City, Vietnam. *AIDS Res Hum Retroviruses.* 2012;28(8):866-873.
365. Europe ECfDPaCWROf. *HIV/AIDS surveillance in Europe 2012.* Stockholm: European Centre for Disease Prevention and Control;2013.
366. Ho DD, Neumann AU, Perelson AS, Chen W, Leonard JM, Markowitz M. Rapid turnover of plasma virions and CD4 lymphocytes in HIV-1 infection. *Nature.* 1995;373(6510):123-126.
367. Mansky LM, Temin HM. Lower in vivo mutation rate of human immunodeficiency virus type 1 than that predicted from the fidelity of purified reverse transcriptase. *J Virol.* 1995;69(8):5087-5094.
368. Schlub TE, Smyth RP, Grimm AJ, Mak J, Davenport MP. Accurately measuring recombination between closely related HIV-1 genomes. *PLoS Comput Biol.* 2010;6(4):e1000766.
369. Domingo E, Martin V, Perales C, Grande-Pérez A, García-Arriaza J, Arias A. Viruses as quasispecies: biological implications. *Curr Top Microbiol Immunol.* 2006;299:51-82.
370. Shankarappa R, Margolick JB, Gange SJ, et al. Consistent viral evolutionary changes associated with the progression of human immunodeficiency virus type 1 infection. *J Virol.* 1999;73(12):10489-10502.
371. Hirsch MS, Brun-Vézinet F, Clotet B, et al. Antiretroviral drug resistance testing in adults infected with human immunodeficiency virus type 1: 2003 recommendations of an International AIDS Society-USA Panel. *Clin Infect Dis.* 2003;37(1):113-128.
372. Tamura K, Peterson D, Peterson N, Stecher G, Nei M, Kumar S. MEGA5: molecular evolutionary genetics analysis using maximum likelihood, evolutionary distance, and maximum parsimony methods. *Mol Biol Evol.* 2011;28(10):2731-2739.

373. Nomenclature committee of the International Union of Biochemistry (NC-IUB). Nomenclature for incompletely specified bases in nucleic acid sequences. Recommendations 1984. *J Biol Chem.* 1986;261(1):13-17.
374. Lodi S, Phillips A, Touloumi G, et al. Time from human immunodeficiency virus seroconversion to reaching CD4+ cell count thresholds <200, <350, and <500 Cells/mm³: assessment of need following changes in treatment guidelines. *Clin Infect Dis.* 2011;53(8):817-825.
375. Kouyos RD, von Wyl V, Yerly S, et al. Ambiguous nucleotide calls from population-based sequencing of HIV-1 are a marker for viral diversity and the age of infection. *Clin Infect Dis.* 2011;52(4):532-539.
376. Abecasis AB, Wensing AM, Paraskevis D, et al. HIV-1 subtype distribution and its demographic determinants in newly diagnosed patients in Europe suggest highly compartmentalized epidemics. *Retrovirology.* 2013;10:7.
377. Van der Paal L, Shafer LA, Todd J, Mayanja BN, Whitworth JA, Grosskurth H. HIV-1 disease progression and mortality before the introduction of highly active antiretroviral therapy in rural Uganda. *AIDS.* 2007;21 Suppl 6:S21-29.
378. Wandel S, Egger M, Rangsin R, et al. Duration from seroconversion to eligibility for antiretroviral therapy and from ART eligibility to death in adult HIV-infected patients from low and middle-income countries: collaborative analysis of prospective studies. *Sex Transm Infect.* 2008;84 Suppl 1:i31-i36.
379. Antinori A, Coenen T, Costagiola D, et al. Late presentation of HIV infection: a consensus definition. *HIV Med.* 2011;12(1):61-64.
380. Karlsson A, Björkman P, Bratt G, et al. Low prevalence of transmitted drug resistance in patients newly diagnosed with HIV-1 infection in Sweden 2003-2010. *PLoS One.* 2012;7(3):e33484.
381. Wheeler WH, Ziebell RA, Zabina H, et al. Prevalence of transmitted drug resistance associated mutations and HIV-1 subtypes in new HIV-1 diagnoses, U.S.-2006. *AIDS.* 2010;24(8):1203-1212.
382. Frentz D, Van de Vijver DA, Abecasis AB, et al. Increase in transmitted resistance to non-nucleoside reverse transcriptase inhibitors among newly diagnosed HIV-1 infections in Europe. *BMC Infect Dis.* 2014;14:407.
383. Yerly S, von Wyl V, Ledergerber B, et al. Transmission of HIV-1 drug resistance in Switzerland: a 10-year molecular epidemiology survey. *AIDS.* 2007;21(16):2223-2229.
384. Callegaro A, Svicher V, Alteri C, et al. Epidemiological network analysis in HIV-1 B infected patients diagnosed in Italy between 2000 and 2008. *Infect Genet Evol.* 2011;11(3):624-632.
385. Rothenberg RB, Potterat JJ, Woodhouse DE, Muth SQ, Darrow WW, Klovdahl AS. Social network dynamics and HIV transmission. *AIDS.* 1998;12(12):1529-1536.
386. Potterat JJ, Phillips-Plummer L, Muth SQ, et al. Risk network structure in the early epidemic phase of HIV transmission in Colorado Springs. *Sex Transm Infect.* 2002;78 Suppl 1:i159-163.
387. Resik S, Lemey P, Ping LH, et al. Limitations to contact tracing and phylogenetic analysis in establishing HIV type 1 transmission networks in Cuba. *AIDS Res Hum Retroviruses.* 2007;23(3):347-356.
388. Vandamme AM, Sönnerborg A, Ait-Khaled M, et al. Updated European recommendations for the clinical use of HIV drug resistance testing. *Antivir Ther.* 2004;9(6):829-848.
389. Brenner BG, Roger M, Moisi DD, et al. Transmission networks of drug resistance acquired in primary/early stage HIV infection. *AIDS.* 2008;22(18):2509-2515.
390. Kaye M, Chibo D, Birch C. Phylogenetic investigation of transmission pathways of drug-resistant HIV-1 utilizing pol sequences derived from resistance genotyping. *J Acquir Immune Defic Syndr.* 2008;49(1):9-16.

391. Hué S, Gifford RJ, Dunn D, Fernhill E, Pillay D, Resistance UCGoHD. Demonstration of sustained drug-resistant human immunodeficiency virus type 1 lineages circulating among treatment-naïve individuals. *J Virol.* 2009;83(6):2645-2654.
392. Thomson MM, Vinogradova A, Delgado E, et al. Molecular epidemiology of HIV-1 in St Petersburg, Russia: predominance of subtype A, former Soviet Union variant, and identification of intrasubtype subclusters. *J Acquir Immune Defic Syndr.* 2009;51(3):332-339.
393. Castro E, Khonkarly M, Ciuffreda D, et al. HIV-1 drug resistance transmission networks in southwest Switzerland. *AIDS Res Hum Retroviruses.* 2010;26(11):1233-1238.
394. Skoura L, Metallidis S, Buckton AJ, et al. Molecular and epidemiological characterization of HIV-1 infection networks involving transmitted drug resistance mutations in Northern Greece. *J Antimicrob Chemother.* 2011;66(12):2831-2837.
395. Pilcher CD, Joaki G, Hoffman IF, et al. Amplified transmission of HIV-1: comparison of HIV-1 concentrations in semen and blood during acute and chronic infection. *AIDS.* 2007;21(13):1723-1730.
396. Hollingsworth TD, Anderson RM, Fraser C. HIV-1 transmission, by stage of infection. *J Infect Dis.* 2008;198(5):687-693.
397. Descamps D, Assoumou L, Chaix ML, et al. National sentinel surveillance of transmitted drug resistance in antiretroviral-naive chronically HIV-infected patients in France over a decade: 2001-2011. *J Antimicrob Chemother.* 2013;68(11):2626-2631.
398. Pineda-Peña AC, Schrooten Y, Vinken L, et al. Trends and predictors of transmitted drug resistance (TDR) and clusters with TDR in a local Belgian HIV-1 epidemic. *PLoS One.* 2014;9(7):e101738.
399. Yebra G, Holguín A, Pillay D, Hué S. Phylogenetic and demographic characterization of HIV-1 transmission in Madrid, Spain. *Infect Genet Evol.* 2013;14:232-239.
400. Chalmet K, Staelens D, Blot S, et al. Epidemiological study of phylogenetic transmission clusters in a local HIV-1 epidemic reveals distinct differences between subtype B and non-B infections. *BMC Infect Dis.* 2010;10:262.
401. Weinstock HS, Zaidi I, Heneine W, et al. The epidemiology of antiretroviral drug resistance among drug-naive HIV-1-infected persons in 10 US cities. *J Infect Dis.* 2004;189(12):2174-2180.
402. Yebra G, de Mulder M, Pérez-Elías MJ, et al. Increase of transmitted drug resistance among HIV-infected sub-Saharan Africans residing in Spain in contrast to the native population. *PLoS One.* 2011;6(10):e26757.
403. Yebra G, Delgado R, Pulido F, et al. Different trends of transmitted HIV-1 drug resistance in Madrid, Spain, among risk groups in the last decade. *Arch Virol.* 2014;159(5):1079-1087.
404. Bartmeyer B, Kuecherer C, Houareau C, et al. Prevalence of transmitted drug resistance and impact of transmitted resistance on treatment success in the German HIV-1 Seroconverter Cohort. *PLoS One.* 2010;5(10):e12718.
405. programme S. Transmission of drug-resistant HIV-1 in Europe remains limited to single classes. *AIDS.* 2008;22(5):625-635.
406. García-Lerma JG, MacInnes H, Bennett D, Weinstock H, Heneine W. Transmitted human immunodeficiency virus type 1 carrying the D67N or K219Q/E mutation evolves rapidly to zidovudine resistance in vitro and shows a high replicative fitness in the presence of zidovudine. *J Virol.* 2004;78(14):7545-7552.
407. Thomson MM, Delgado E, Manjón N, et al. HIV-1 genetic diversity in Galicia Spain: BG intersubtype recombinant viruses circulating among injecting drug users. *AIDS.* 2001;15(4):509-516.
408. Kozísek M, Sasková KG, Rezácová P, et al. Ninety-nine is not enough: molecular characterization of inhibitor-resistant human immunodeficiency virus type 1 protease mutants with insertions in the flap region. *J Virol.* 2008;82(12):5869-5878.

409. Paolucci S, Baldanti F, Dossena L, Gerna G. Amino acid insertions at position 35 of HIV-1 protease interfere with virus replication without modifying antiviral drug susceptibility. *Antiviral Res.* 2006;69(3):181-185.
410. Chen JH, Wong KH, Chan KC, et al. Molecular epidemiology and divergence of HIV type 1 protease codon 35 inserted strains among treatment-naive patients in Hong Kong. *AIDS Res Hum Retroviruses.* 2008;24(4):537-542.
411. Kim EY, Winters MA, Kagan RM, Merigan TC. Functional correlates of insertion mutations in the protease gene of human immunodeficiency virus type 1 isolates from patients. *J Virol.* 2001;75(22):11227-11233.
412. Amiel C, Charpentier C, Désiré N, et al. Long-term follow-up of 11 protease inhibitor (PI)-naïve and PI-treated HIV-infected patients harbouring virus with insertions in the HIV-1 protease gene. *HIV Med.* 2011;12(3):138-144.
413. Grotto RM, Corvino SM, Munhoz LaS, Ghedini CG, Pardini MI. A first case of protease codon 35 amino acid insertion in a HIV-1 subtype B sequence detected in the Bauru region, state of São Paulo, Brazil: case report. *Rev Soc Bras Med Trop.* 2011;44(3):392-394.
414. Pereira-Vaz J, Duque V, Trindade L, Saraiva-da-Cunha J, Meliço-Silvestre A. Detection of the protease codon 35 amino acid insertion in sequences from treatment-naïve HIV-1 subtype C infected individuals in the Central Region of Portugal. *J Clin Virol.* 2009;46(2):169-172.
415. Etienne L, Hahn BH, Sharp PM, Matsen FA, Emerman M. Gene loss and adaptation to hominids underlie the ancient origin of HIV-1. *Cell Host Microbe.* 2013;14(1):85-92.
416. Snoeck J, Kantor R, Shafer RW, et al. Discordances between interpretation algorithms for genotypic resistance to protease and reverse transcriptase inhibitors of human immunodeficiency virus are subtype dependent. *Antimicrob Agents Chemother.* 2006;50(2):694-701.
417. Kanki PJ, Hamel DJ, Sankalé JL, et al. Human immunodeficiency virus type 1 subtypes differ in disease progression. *J Infect Dis.* 1999;179(1):68-73.
418. Kaleebu P, French N, Mahe C, 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.* 2002;185(9):1244-1250.
419. McKinnon LR, Ball TB, Kimani J, et al. Cross-clade CD8(+) T-cell responses with a preference for the predominant circulating clade. *J Acquir Immune Defic Syndr.* 2005;40(3):245-249.
420. Sguanci L, Bagnoli F, Liò P. Modeling HIV quasispecies evolutionary dynamics. *BMC Evol Biol.* 2007;7 Suppl 2:S5.
421. Gray RR, Tatem AJ, Lamers S, et al. Spatial phylodynamics of HIV-1 epidemic emergence in east Africa. *AIDS.* 2009;23(14):F9-F17.
422. Couturier E, Damond F, Roques P, et al. HIV-1 diversity in France, 1996-1998. The AC 11 laboratory network. *AIDS.* 2000;14(3):289-296.
423. Böni J, Pyra H, Gebhardt M, et al. High frequency of non-B subtypes in newly diagnosed HIV-1 infections in Switzerland. *J Acquir Immune Defic Syndr.* 1999;22(2):174-179.
424. Dietrich U, Ruppach H, Gehring S, et al. Large proportion of non-B HIV-1 subtypes and presence of zidovudine resistance mutations among German seroconvertors. *AIDS.* 1997;11(12):1532-1533.
425. Fransen K, Buvé A, Nkengasong JN, Laga M, van der Groen G. Longstanding presence in Belgians of multiple non-B HIV-1 subtypes. *Lancet.* 1996;347(9012):1403.
426. Puchhammer-Stöckl E, Kunz C, Faatz E, Kasper P, Heinz FX. Introduction of HIV-1 subtypes C, E and A into Austria. *Clin Diagn Virol.* 1998;9(1):25-28.
427. Parry JV, Murphy G, Barlow KL, et al. National surveillance of HIV-1 subtypes for England and Wales: design, methods, and initial findings. *J Acquir Immune Defic Syndr.* 2001;26(4):381-388.
428. Thomson MM, Nájera R. Molecular epidemiology of HIV-1 variants in the global AIDS pandemic: an update. *AIDS Rev.* 2005;7(4):210-224.

429. Pérez L, Thomson MM, Bleda MJ, et al. HIV Type 1 molecular epidemiology in cuba: high genetic diversity, frequent mosaicism, and recent expansion of BG intersubtype recombinant forms. *AIDS Res Hum Retroviruses.* 2006;22(8):724-733.
430. Yang R, Xia X, Kusagawa S, Zhang C, Ben K, Takebe Y. On-going generation of multiple forms of HIV-1 intersubtype recombinants in the Yunnan Province of China. *AIDS.* 2002;16(10):1401-1407.
431. Tee KK, Saw TL, Pon CK, Kamarulzaman A, Ng KP. The evolving molecular epidemiology of HIV type 1 among injecting drug users (IDUs) in Malaysia. *AIDS Res Hum Retroviruses.* 2005;21(12):1046-1050.
432. Bruisten SM, Schouls L. Molecular typing and clustering analysis as a tool for epidemiology of infectious diseases. In: Springer, ed. *Modern Infectious Disease Epidemiology*2010:117-141.
433. Kiwanuka N, Laeyendecker O, Robb M, et al. Effect of human immunodeficiency virus Type 1 (HIV-1) subtype on disease progression in persons from Rakai, Uganda, with incident HIV-1 infection. *J Infect Dis.* 2008;197(5):707-713.
434. Hamers RL, Schuurman R, Sigaloff KC, et al. Effect of pretreatment HIV-1 drug resistance on immunological, virological, and drug-resistance outcomes of first-line antiretroviral treatment in sub-Saharan Africa: a multicentre cohort study. *Lancet Infect Dis.* 2012;12(4):307-317.
435. Pereyra F, Addo MM, Kaufmann DE, et al. Genetic and immunologic heterogeneity among persons who control HIV infection in the absence of therapy. *J Infect Dis.* 2008;197(4):563-571.
436. Little SJ, Frost SD, Wong JK, et al. Persistence of transmitted drug resistance among subjects with primary human immunodeficiency virus infection. *J Virol.* 2008;82(11):5510-5518.