

Diploma Thesis

# Identification of HIV-1 transmission clusters in Southeastern Austria by phylogenetic analysis

submitted by

**Karin Weninger**

for receiving the academic degree

**Medical Doctor, MD**

(Dr. med. univ.)

at the

**Medical University of Graz**

**Institute of Hygiene, Microbiology and Environmental Medicine**

First Supervisor:

**Univ.- Prof. Dr. med. univ. Harald H. Kessler**

Second Supervisor:

**Mag. rer. nat. Dr. scient. med. Evelyn Stelzl**

Graz, March 9, 2016

# Declaration of Originality

I, hereby, declare that the following diploma thesis has been written only by the undersigned and without any assistance from third parties. Furthermore, I confirm that no sources have been used in the preparation of this thesis other than those indicated in the thesis itself.

Graz, \_\_\_\_\_

Place

March 9, 2016 \_\_\_\_\_

Date

Karin Weninger e.h. \_\_\_\_\_

Signature

## **Acknowledgements**

Foremost, I would like to thank my first supervisor Univ.- Prof. Dr. med. univ. Harald H. Kessler, who was always available for my questions and gave generously of his time and knowledge. He supported me throughout this thesis with ideas and guidance.

I would like to thank my second supervisor Mag. rer. nat. Dr. scient. med. Evelyn Stelzl for giving me deeper insight in the interesting field of molecular diagnostics and providing detailed explanations.

Furthermore, I would like to thank my parents for their support throughout all my studies at university.

Also I would like to thank my sister Astrid for encouraging me to finish this thesis.

Last, but not least I would like to thank my partner Oliver for his patience and support.

## Abstract

**Background:** Although the awareness of the Human immunodeficiency virus type 1 (HIV-1) is increasing on population level, the number of newly-infected persons per year continues to be high in Southeastern Austria. Therefore, deeper insight in the epidemiology of HIV-1 transmission in Southeastern Austria is necessary to elaborate targeted prevention programs.

**Objectives:** Aim of this study was to reconstruct the HIV-1 transmission network in Southeastern Austria. Demographic data, risk factors for HIV-1 infection, and HIV-1 subtypes were compared between individuals who clustered within the transmission network and those who did not cluster within the transmission network.

**Material and Methods:** Partial HIV-1 *polymerase (pol)* sequences obtained from 259 individuals newly diagnosed with HIV infection from 2008 to 2014 and found to reside in Southeast Austria were analyzed for genetic relatedness. A putative transmission linkage between two sequences was established if a genetic distance of <1.5% was found.

**Results:** Analysis revealed that 45.6% of the sequences were genetically linked to at least one other sequence. Therefore 34 putative transmission clusters were identified. Clustering individuals were more likely to be men having sex with men (MSM) ( $P<0.001$ ), and infected with subtype B ( $P=0.003$ ) or subtype F ( $P<0.033$ ). Among the clustering males reporting heterosexual (HSX) transmission as their only risk for HIV-1 infection, 47% clustered closely with MSM.

**Conclusion:** A high degree of clustering was found. The high amount of males that reported HSX transmission as their only risk for HIV-1 infection but clustered closely with MSM suggests either missing individuals in the transmission network linking these individuals or misrepresentation of sexual risk behaviors by individuals.

**Keywords:** HIV-1, transmission network, Southeastern Austria, MSM.

## Kurzfassung

**Hintergrund:** Obwohl das Bewusstsein um HIV-1 in der Bevölkerung zunimmt, bleibt die Anzahl der neuinfizierten Personen pro Jahr in Südost-Österreich konstant hoch. Daher ist es von großer Bedeutung einen tieferen Einblick in die Epidemiologie der Übertragung in dieser Gegend zu erhalten.

**Zielsetzung:** Das Ziel dieser Studie war es, das HIV-1 Übertragungsnetzwerk in Südost-Österreich zu rekonstruieren. Demographische Daten, HIV-1 Risikofaktoren und HIV-1 Subtypen von Personen, die sich innerhalb der gefundenen Gruppen des Übertragungsnetzwerkes befanden, wurden mit denen von Personen verglichen, die sich nicht in Gruppen befanden.

**Material und Methoden:** Es wurden partielle HIV-1 *pol* Sequenzen von 259 Personen, bei denen zwischen 2008 und 2014 HIV-1 diagnostiziert wurde und die in Südost-Österreich leben, auf genetische Ähnlichkeit analysiert. Es wurde eine vermeintliche nahe Zusammengehörigkeit zwischen zwei Sequenzen angenommen, wenn der genetische Unterschied  $<1.5\%$  betrug.

**Ergebnisse:** Die Analyse ergab, dass 45.6% der partiellen *pol* Sequenzen große genetische Ähnlichkeit zu zumindest einer anderen Sequenz aufwiesen. Dadurch konnte auf 34 Gruppen rückgeschlossen werden, innerhalb derer vermeintliche Übertragungseignisse stattgefunden haben. Personen innerhalb dieser Gruppen waren vor allem homosexuelle Männer ( $P<0.001$ ) und mit Subtyp B ( $P<0.003$ ) oder F ( $P<0.033$ ) infiziert. Von jenen Männern, die sich in Gruppen befanden und heterosexuellen Sex als ihr einziges Risikoverhalten für eine HIV-1 Infektion angaben, befanden sich 47% in Gruppen gemeinsam mit homosexuellen Männern.

**Schlussfolgerung:** Es konnte eine starke Gruppenbildung beobachtet werden. Die große Anzahl an heterosexuellen Männern, die sich in Gruppen mit vorwiegend homosexuellen Männern wiederfand, lässt entweder auf fehlende Personen im Übertragungsnetzwerk oder Verheimlichung der wahren sexuellen Orientierung schließen.

**Schlüsselwörter:** HIV-1, Übertragungsnetzwerk, Südost-Österreich, homosexuelle Männer.

# Contents

<b>1</b>	<b>Introduction</b>	<b>1</b>
1.1	The Human Immunodeficiency Virus . . . . .	2
1.2	Structure and Morphology . . . . .	4
1.3	Transmission . . . . .	5
1.4	Life Cycle . . . . .	7
1.5	Diagnosis of Infection . . . . .	8
1.6	Clinical Manifestation . . . . .	10
1.6.1	Category A . . . . .	10
1.6.2	Category B (AIDS non-defining diseases) . . . . .	11
1.6.3	Category C (AIDS defining diseases) . . . . .	11
1.7	Antiretroviral Therapy . . . . .	12
1.8	Mutations of the Viral Genome . . . . .	14
<b>2</b>	<b>Objectives</b>	<b>16</b>
<b>3</b>	<b>Material and Methods</b>	<b>17</b>
3.1	Study Population . . . . .	17
3.2	Area of Investigation . . . . .	19
3.3	Population Sequencing . . . . .	19
3.4	Genetic Network Analysis . . . . .	20
3.5	Phylogenetic Reconstruction and Bayesian Analyses . . . . .	21
3.6	Statistical Analysis . . . . .	22

<b>4 Results</b>	<b>23</b>
4.1 Study Population . . . . .	23
4.2 Transmission Cluster Composition . . . . .	30
4.3 Variables associated with Membership in Clusters . . . . .	34
4.4 Phylogenetic Reconstruction and Migration Analysis . . . . .	37
4.5 Evolution in the Mode of Transmission . . . . .	37
<b>5 Discussion</b>	<b>40</b>
<b>6 Conclusion</b>	<b>44</b>
<b>7 Annotation</b>	<b>45</b>
<b>List of Figures</b>	<b>51</b>
<b>List of Tables</b>	<b>52</b>

# Abbreviations

**AIDS** acquired immunodeficiency syndrome.

**ART** antiretroviral therapy.

**cDNA** complimentary DNA.

**CI** confidence interval.

**CRF** circulating recombinant form.

**D** genetic distance.

**DNA** deoxyribonucleic acid.

**EDTA** ethylenediamine tetraacetic acid.

**ELISA** enzyme linked immunosorbent assay.

**env** envelope.

**gag** group specific antigen.

**HIP** HIV patient management system.

**HIV** Human immunodeficiency virus.

**HIV-1** Human immunodeficiency virus type 1.

**HIV-2** Human immunodeficiency virus type 2.

**HSX** heterosexual.

**IDU** injection drug user.

**IgG** immunoglobulin G.

**IQR** interquartile range.

**LAS** Lymphadenopathy syndrome.

**MCMC** Markov Chain Monte Carlo.

**MRCA** most recent common ancestor.

**mRNA** messenger-RNA.

**MSM** men having sex with men.

**MTCT** mother to child transmission.

**MUG** Medical University of Graz.

**NHP** nonhuman primates.

**NNRTI** Non-nucleoside reverse transcriptase inhibitors.

**NRTI** Nucleoside analog reverse transcriptase inhibitors.

**PI** Protease inhibitors.

**pol** polymerase.

**RNA** ribonucleic acid.

**rt-PCR** real-time polymerase chain reaction.

**RTI** Reverse transcriptase inhibitors.

**SCUEAL** Subtype Classification using Evolutionary Algorithms program.

**SIV** Simian immunodeficiency virus.

**STD** sexually transmitted disease.

**TN93** Tamura-Nei 93.

**UCSD** Division of Infectious Diseases, University of California San Diego.

**UNAIDS** Joint United Nations Programme on HIV/AIDS.

**URF** unique recombinant form.

# 1 Introduction

The Human immunodeficiency virus (HIV) causes the acquired immunodeficiency syndrome (AIDS) in humans, a condition causing a progressive failure of the immune system. Hitherto there is no cure for this disease. However, effective antiretroviral therapy (ART) can extend the lifespan of patients with AIDS by many years.

Although the awareness of HIV is increasing on population level, the number of newly-infected persons per year is still high. According to the Joint United Nations Programme on HIV/AIDS (UNAIDS), an estimated number of approximately 35.0 million individuals were living with HIV worldwide in 2013 [1]. The vast majority of infected people is living in developing countries. The lack of medical help and resources in these countries contribute to the fast and ongoing spread of the HIV, making it one of the five deadliest infectious diseases worldwide. The number of newly infected individuals in Austria is around 500 in recent years, and an estimated amount of 8000 people infected with HIV are currently living here [2].

Due to the fact that there is no cure available, it is crucial that the general public is informed about transmission ways and risk behavior to reduce the number of newly-infected people. There is a need to define risk groups who fuel the ongoing epidemic and to elaborate adequate prevention programs.

## 1.1 The Human Immunodeficiency Virus

HIV developed from the Simian immunodeficiency virus (SIV), a virus which causes an AIDS-like disease in primates. This virus has crossed the species barrier from nonhuman primates (NHP) to humans on several occasions leading to the development of different types and groups of HIV [3]. These zoonotic transmissions took place in West and Central Africa in the early 1900s. Today, two types of the HIV are known, the Human immunodeficiency virus type 1 (HIV-1) and the Human immunodeficiency virus type 2 (HIV-2). Both types can be subdivided into groups, subtypes and sub-subtypes. In comparison to HIV-1, HIV-2 has only spread to some extent in Western Africa [3]. It is considered that the species barrier from primates to humans was at least crossed four times leading to the four groups of HIV-1 and another several times leading to the subtypes of HIV-2 [3].

The four major genetic groups of HIV-1 are: Group M (main), group N (non-M/non-O), group O (outliers), and group P. Within the group M, nine subtypes (A, B, C, D, F, G, H, J, and K) and six sub-subtypes (A1-A4, F1 and F2) have been identified [3]. HIV-1 subtypes can be defined as subgroups (or clades) of genetically similar HIV-1 viruses. Subtypes are genetically closer related to other subtypes within the same group, than to subtypes within other groups. Figure 1 gives an overview of the major groups and subtypes.

Group M is responsible for the majority of HIV-1 infections worldwide, whereas the groups N and O are mostly found in West Africa and play a minor role in the worldwide HIV pandemic. Group P, which remains especially rare despite extensive screening, was up to date only found in two non-related patients of Cameroonian origin [4].

Subtypes show a geographical association. The predominant subtype of group M in Europe, North and South America is subtype B, whereas subtypes A, C, and D are most common in Africa and subtypes C and E are most common in Asia. Furthermore, an increasing amount of recombinant forms can be found, which are thought to have originated

in individuals multiply infected with viruses of two or more subtypes [5]. Recombination between different HIV strains is a relatively common occurrence, especially in regions where several divergent strains co-circulate. In order to define a new circulating recombinant form (CRF), representative strains must be identified in at least three individuals with no direct epidemiological linkage. Recombinant forms, which do not fulfill these criteria, are classified as unique recombinant form (URF) [6]. So far about 70 CRFs have been described. These high recombination rates, together with high rates of mutation and viral replication lead to the genetic variability of HIV-1.

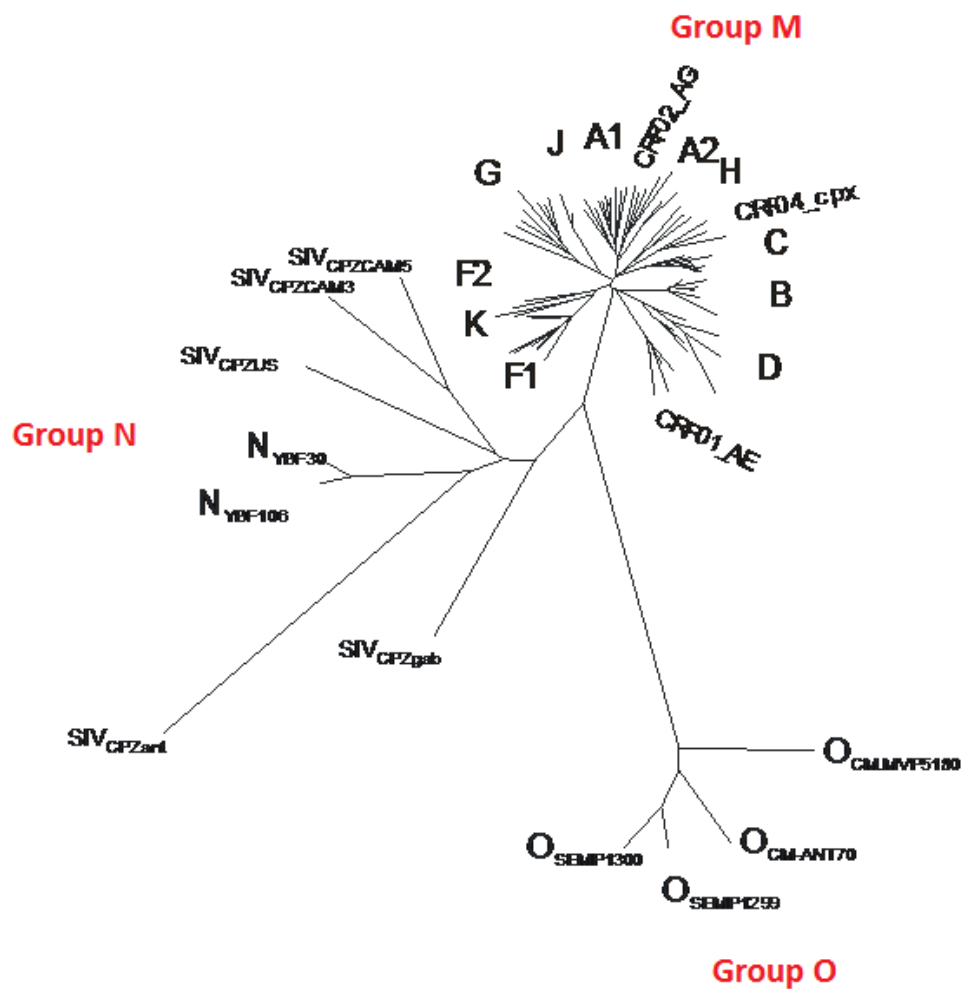


Figure 1: Phylogenetic tree of HIV-1.

## 1.2 Structure and Morphology

HIV belongs to the family *Retroviridae* in the genus *Lentivirus*. The genus *Lentivirus* is characterized by a long incubation period and the unique ability among retroviruses of being able to infect non-dividing cells. HIV consists of a single plus-stranded genome. It replicates in host cells through reverse transcription, which means, that the viral ribonucleic acid (RNA) strand is transcribed into a complimentary DNA (cDNA) strand, allowing the integration of the viral genome into the host desoxyribonucleic acid (DNA) genome.

The viral RNA, together with three enzymes (*protease*, *integrase*, and *reverse transcriptase*) is enclosed by a nucleocapsid, consisting of *p24* proteins. The outermost layer of the virus particle consists of a lipid bilayer of host-cell origin. This envelope is formed when the virus unseals from the infected host-cell. The external glycoprotein *gp120* and transmembrane glycoprotein *gp41* are attached to the envelope (see Figure 2). These envelope proteins are necessary for the binding and the fusion with *CD4*+cells of the host.

The HIV genome contains nine genes. Three of them are genes encoding for structure proteins, the other six encode regulatory proteins. The viral structural proteins are encoded in the *envelope (env)*, *group specific antigen (gag)*, and *pol* genes. The *pol* gene stores the genetic information of the viral enzymes including *reverse transcriptase*, *integrase*, and *protease*, while *env* encodes for the envelope glycoproteins *gp120* and *gp41*. The *gag* gene is transcribed and translated into nucleoproteins for the nucleocapsid and proteins for the viral matrix.

The regulation proteins *tat* and *rev* are essential for the regulation of the HIV gene expression. *Vif*, *vpr*, *vpu*, *vpx*, and *nef* are referred to as auxiliary or accessory genes, since they are not necessary for viral replication in tissue culture. However, their role in vivo is indispensable [7]. *Nef* is necessary for the maintenance of high viral loads. Viruses with defective *nef* were detected in some HIV-1 infected long term survivors [7].

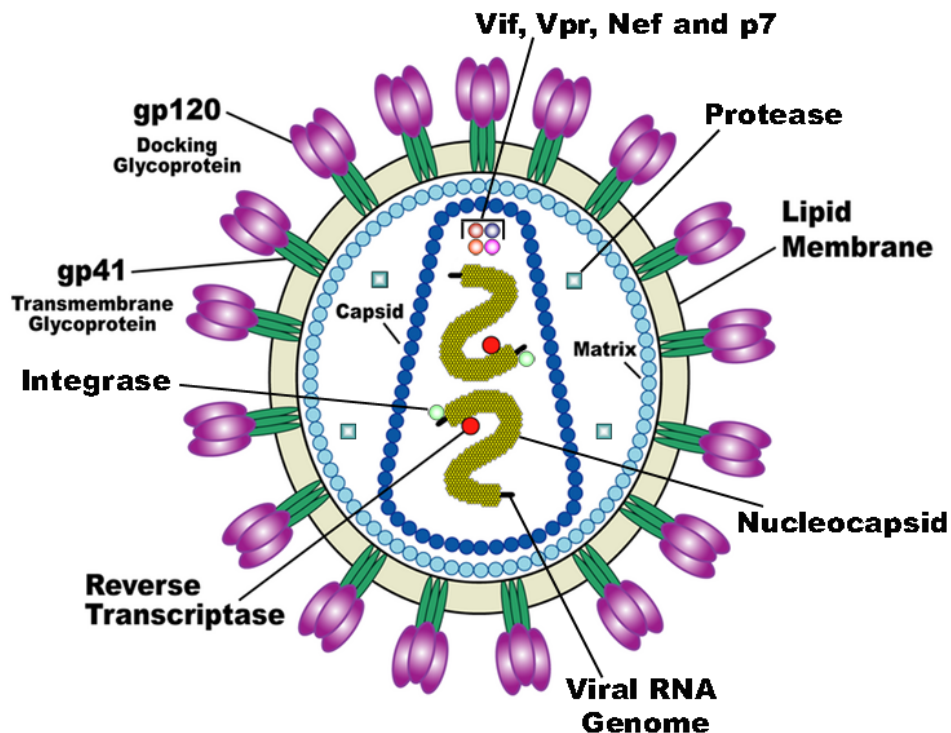


Figure 2: Organization of the HIV-1 virion. Image credit: US National Institute of Health. Available from: <https://commons.wikimedia.org/wiki/File:HI-Virion-en.png>

### 1.3 Transmission

In contrast to other infectious diseases, a relatively high viral load is required to cause an infection with HIV. However, the minimal infective dose is unknown. There are three main ways of transmission: Parenteral (infectious blood and blood products), perinatal/vertical (from mother to child especially during delivery), and sexual (hetero- and homosexual contact).

Parenteral infections are caused by transfusion of infected blood, needle-sharing and accidental needle-stick injuries. In the 1980s, several cases of HIV infections in hemophiliacs were observed, who had received contaminated blood products. Nowadays blood products are screened for HIV-1, leading to an extremely low infection risk. Needle-sharing is a commonly used practice within injection drug user (IDU)s. After a peak in the 1980s,

the number of HIV-positive IDUs in Germany decreased between 1990 and 2010, leading to a constantly low number of infected persons in this risk group [8]. A similar trend can be assumed for Austria [2]. Further prevention programs and the supply with sterile needles and syringes will lead to a further reduction of new infections in this risk group. However, worldwide coverage of HIV prevention, treatment, and care services for IDUs is not sufficient [9]. The infection risk for accidental needle-stick injuries is about 0.3%. To avoid these injuries recapping of used injection needles should be avoided and correct disposal is obligatory.

HIV transmission from an infected mother to her child can take place during pregnancy (rarely, most likely in the last trimester of pregnancy), during delivery, and through breastfeeding. Antiretroviral therapy administered to the mother during pregnancy can reduce the risk for the child effectively. Babies of untreated infected mothers should be delivered by scheduled Caesarean section. There may be an option of vaginal delivery, if the viral load is completely suppressed through ART [10]. Antiretroviral prophylaxis should be administered to the newborn. Breastfeeding must be avoided. Since 2010, testing for HIV-1 is offered to all pregnant women in Austria as a part of the Austrian mother-child-care program [11].

Concerning sexual contacts several co-risk factors for HIV-1 infection can be monitored: The presence of another sexually transmitted disease (STD), the concentration of HIV RNA in the blood plasma, and promiscuity. Male circumcision effectively reduces the risk for sexually acquired HIV-1 infections in males [12]. Unprotected receptive anal intercourse is a special risk behavior indicated by the high per-act and per-partner transmission rates [13]. The per-act transmission rate in anal sex is higher than in vaginal sex due to differences in the rectal and vaginal mucosal tissue. Therefore, MSM are twice as likely infected by more than one viral strain compared to heterosexuals [14].

## 1.4 Life Cycle

To gain entrance in human cells, the HIV is dependent on the presence of *CD4*-receptors. These surface cell receptors are expressed by T-lymphocytes, dendritic cells, monocytes, macrophages, and microglia cells. The binding of the viral surface glycoprotein *gp120* to *CD4*-receptors leads to a conformational change in *gp120*, which exposes an additional binding site for a chemokine co-receptor [15]. The most important chemokine receptors are *CCR5* and *CXC4*. Virus strains can be classified as *R5*, *X4* or *R5X4* according to their use of one or both of the chemokine receptors *CCR5* and *CXC4*. *R5* viruses are predominant in early infection, while viruses that use the co-receptor *CXC4* (*X4* and *R5X4* viruses) emerge later in the course of infection [16]. Binding to these receptors leads to a structural rearrangement in *gp41*, initiating the fusion of the virion with the cellular plasma membrane and release of genetic material into the host cell. The *reverse transcriptase* starts to transcribe viral RNA into cDNA, which is integrated into the human genome by the viral enzyme *integrase*. The integrated viral DNA remains dormant until certain cellular transcription factors, such as *NF-κB*, are present. *NF-κB* becomes highly expressed, when the T-lymphocyte is activated in the course of infection defense. The activation of the HIV-infected T-lymphocyte results in transcription of the viral DNA into viral messenger-RNA (mRNA), which is performed by the *RNA polymerase II* of the host cell. Early in the course of infection the viral mRNA is spliced in order to produce *tat* and *rev* proteins. These proteins promote further transcription of viral mRNA and are essential at this stage of infection. *Tat* is needed for the elongation of the transcription and *rev* is responsible for the transportation of incompletely spliced and unspliced viral mRNA outside of the cellular nucleus, which is necessary for the translation into proteins in the cytoplasm [17]. The viral mRNA is translated into long chains of polypeptides. The enzyme *protease* then cleaves the long polypeptide chains into smaller fragments. This leads to the production of *env*, *gag* and, *pol* proteins. Upon completion of all viral components, including two single plus-stranded RNA particles per virion, *gag* initiates their assembly. The assembled particles bind to the plasma membrane of the host-cell

and the formation of a spherical envelope is initiated. The virion unseals from the host cell, now carrying an envelope consisting of a phospholipid bilayer, in a process called budding. Many virions can bud from one cell before the descent of the plasma membrane leads to disintegration of the cell. The unsealed virions undergo a process of maturation and are then prepared to infect further cells of the host or to get passed on to another host, through the transmission ways described above. The several steps of the HIV lifecycle are shown in Figure 3.

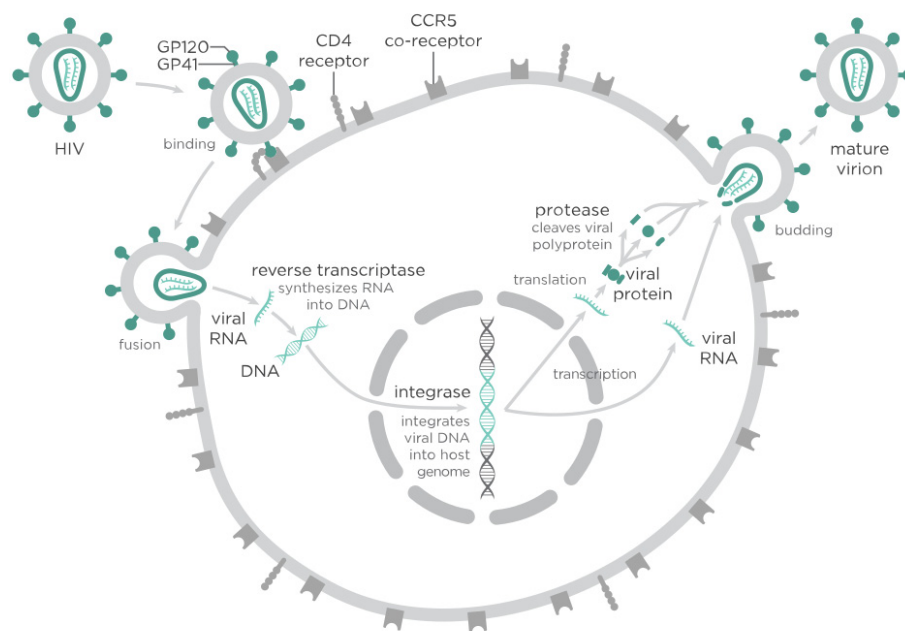


Figure 3: Key aspects of the HIV life cycle. Image credit: Thomas Spletstoesser. Available from: <https://commons.wikimedia.org/wiki/File:HIV-replication-cycle.svg>

## 1.5 Diagnosis of Infection

Testing for HIV-infection can be done through indirect or direct methods. Indirect methods detect HIV antibodies, whereas direct methods test for the presence of viral RNA in the plasma of the patient. Indirect methods include the *enzyme linked immunosorbent assay (ELISA)* for detection of antibodies directed against HIV. Direct methods include

the real-time polymerase chain reaction (rt-PCR) for detection and quantitation of nucleic acids (HIV RNA).

After infection, it takes some time until antibodies against the virus are generated. The period between infection and the first possible detection of antibodies is called serologic window period. Usually, *immunoglobulin G (IgG)* antibodies appear in the patient's blood not earlier than 4-12 weeks after infection. However, in rare cases, it may take up to 6 months until antibodies are detectable. The *p24* antigen is detectable approximately 15-20 days after infection, but only temporarily until antibody production starts [18]. Today, combined antibody/antigen *ELISAs* are used routinely for diagnosis of HIV infection. These tests detect both antibodies directed against HIV (HIV-1 and HIV-2) and the *p24* antigen. In Austria, every positive antibody/antigen *ELISA* result must be confirmed through an immunoblot (most commonly through a recombinant immunoblot) to exclude patients which were tested false-positive. However, a negative antibody/antigen *ELISA* does not always exclude an existing HIV infection because the person tested might be infected with HIV at a very early stage.

For diagnosis of an HIV infection at the earliest stage, a molecular test (nowadays mainly based on a rt-PCR) which is able to shorten the window period may be applied. HIV RNA is detectable about ten days after infection [18]. Besides diagnosis of HIV infection, detection and quantitation of HIV RNA is used for monitoring purposes. The viral load is found to be especially high at the acute stage of infection and at the late stage of untreated AIDS disease [19]. Furthermore, newborns of HIV-positive mothers are tested with a rt-PCR, because antibody tests might lead to false-positive results due to the presence of maternal antibodies in the newborn's blood sample.

## 1.6 Clinical Manifestation

It is of crucial importance to identify persons recently infected with HIV. During the acute stage of infection, the viral load is very high; however, no antibodies are detectable. These persons might be unaware of their infection and engage in risk behavior, which leads to a further transmission of the virus [20]. According to the Center for Disease Control and Prevention (CDC) the HIV disease can be classified into three clinical categories (A, B, C) and three  $CD4+$  cell count categories.  $CD4+$  cell counts less than 200 cells/ $\mu$ l and the presence of AIDS-defining diseases (patients in the categories A3, B3, and C1-C3) mark the onset of AIDS [21]. This classification can be used for adults and adolescents (see Table 1). It does not match with the progression of HIV in children.

Table 1: Classification for HIV-infected adults and adolescents.

CD4+ cell count category	Clinical category		
	A	B	C
>500/ $\mu$ l	A1	B1	C1
200-499/ $\mu$ l	A2	B2	C2
<200/ $\mu$ l	A3	B3	C3

### 1.6.1 Category A

*Acute retroviral syndrome:*

About one to six weeks after infection, unspecific symptoms might appear, which differ in the severity from patient to patient. Mild and uncharacteristic appearance might be misdiagnosed as an influenza infection. About 30% of patients develop symptoms similar to mononucleosis including fever, lymphadenopathy, and occasionally myalgia, splenomegaly, and exanthema [19]. Furthermore, a loss of appetite, night sweat, arthralgia, and headache may occur. Regarding diagnosis of an acute retroviral syndrome, fever and malaise are the most sensitive, and weight loss and oral ulcers are the most specific symptoms [20].

*Latency period:*

During the latency period hardly any symptoms are present although the virus still replicates and the patients remain infectious. Usually, this period starts one to three months after infection and can last up to ten years without antiretroviral therapy, depending on the physiological status of the patient. An effective ART prolongs this period for several decades. However, malnutrition, immunodeficiency, and infection at early age may shorten this period [19].

*Lymphadenopathy syndrome (LAS):*

About 40% of patients develop the LAS, which is characterized as lymphadenopathy persisting longer than three months [19]. General symptoms are usually missing during this period.

**1.6.2 Category B (AIDS non-defining diseases)**

Category B is associated with susceptibility to infection. The increase of immunodeficiency promotes the appearance of opportunistic infections. The  $CD_4+$  cell count decreases while the viral load increases. Diseases indicating category B include oropharyngeal and vulvovaginal candidiasis, idiopathic thrombocytopenic purpura, pelvic inflammatory disease, oral hairy leukoplakia, bacillary angiomatosis, peripheral neuropathy, diarrhea (lasting >1 month), and subfebrile temperature ( $<38,5^\circ\text{C}$ ) [19, 21].

**1.6.3 Category C (AIDS defining diseases)***Wasting syndrome:*

The unintended loss of more than 10% of weight combined with chronic diarrhea ( $>30$

days) and fever or weakness is defined as wasting syndrome [21].

*AIDS defining malignant tumors:*

The destruction of the immune system increases the likelihood of malignant tumors. AIDS-associated malignant tumors are for example the Kaposi sarcoma, cervical carcinoma, and Non-Hodgkin-lymphoma.

*HIV-associated encephalopathy:*

The infection of microglia with HIV leads to atrophy and destruction of the central nervous system, which can cause severe mental and physical impairment.

*AIDS defining opportunistic infections:*

Opportunistic infections are infections which do not harm people with healthy immune system. However, in individuals with impaired immune system, they can cause severe symptoms and might lead to death. ART can reduce the incidence of these infections, which are caused by bacteria (e.g. *Mycobacterium tuberculosis*), viruses (e.g. *herpes simplex virus*), fungi (e.g. *Pneumocystis jiroveci*, *Candida albicans*), and protozoa (e.g. *Toxoplasma gondii*) [19]. These infections are difficult to treat; however, they are rarely the cause of death for AIDS patients nowadays due to ART.

## 1.7 Antiretroviral Therapy

ART is a combinational therapy using at least three different classes of antiretroviral drugs. The combination of drugs leads to an effective suppression of replication at different stages of the replication cycle. The aim of this therapy is to decrease the viral load under the detection level (defined as  $<50$  copies/ml) within the initial six months of therapy. This leads to a slowing down in the progression of the infection, recovery of the  $CD_4^+$

cell count, and improvement of the immune system.

The German and Austrian guidelines for antiretroviral therapy recommend the treatment initiation for ART according the following scheme [22]:

1. Acute HIV-1 infection: ART is indicated for patients with pronounced and long-lasting symptomatology.
2. Symptomatic HIV-1 infection: Indication for ART.
3. Asymptomatic HIV-1 infection: Recommendations for this group depend on the  $CD_4+$  cell count.
  - a)  $CD_4+$  cell count  $<350/\mu\text{l}$ : Indication for ART.
  - b)  $CD_4+$  cell count  $350/\mu\text{l} - 500/\mu\text{l}$  + additional criteria: Recommendation for ART.
  - c)  $CD_4+$  cell count  $>500/\mu\text{l}$  + additional criteria: Decision for ART depends on the individual case.

Additional criteria include high viral load ( $>100.000$  copies/ml), co-infection with hepatitis B or hepatitis C, old age, rapid decrease of the  $CD_4+$  cell count, high cardiovascular risk, HIV-associated neoplasia, and patient's wish [22].

Today, the following groups of drugs are mainly used for antiretroviral therapy:

*Reverse transcriptase inhibitors (RTI):*

These agents interfere in the process of reverse transcription. *Nucleoside analog reverse transcriptase inhibitors (NRTI)* compete with physiologic nucleosides for placement in the newly transcribed cDNA. Their incorporation in the cDNA leads to interruptions in the progress of transcription. The *Non-nucleoside reverse transcriptase inhibitors (NNRTI)* affect the reverse transcription through non-competitive binding to the *reverse transcriptase*.

*Protease inhibitors (PI):*

These agents inhibit the viral enzyme *protease*.

*Integrase inhibitors:*

This class of antiretroviral drugs inhibits the integration of the viral cDNA into the genome of the host cell.

*Entry inhibitors:*

Prior to the use of this substance group testing of the co-receptor status is obligatory. Co-receptor antagonists blocking the chemokine co-receptor *CCR5* are the only available drugs in this group.

Usually, the combinational therapy consists of two *NRTIs* and one *NNRTI* or one *PI* [19]. An increase of the viral load and a decrease of the *CD4+* cell count under therapy can be evidence for a suboptimal treatment response and development of viral resistance. Poor adherence of the patient can lead to multi-drug resistant viral strains and can be monitored through determination of drug levels in the plasma.

## 1.8 Mutations of the Viral Genome

One of the main reasons contributing to the worldwide spread of the HIV is its enormous genetic variability due to the poor fidelity of the *reverse transcriptase*. This leads to a high rate of genomic evolution of the viral strains within and among hosts. However, not only the *reverse transcriptase*, which transcribes the viral RNA into cDNA, has a poor fidelity, but also the *RNA polymerase II*, which transcribes the cDNA into viral mRNA.

Mutations occur with a high frequency in the course of viral replication and can be disad-

vantageous, silent/neutral, or occasionally favorable for the viral population. Disadvantageous mutations interfering with essential functions of the virus (attachment to host cells, penetration, uncoating, replication, assembly, and budding) are rapidly lost from the viral population [23]. The majority of the mutations are silent/neutral, which means that they do lead to a change in the DNA sequence but not to a change in the aminoacid sequence encoded by the gene [23]. Mutations leading to resistance to antiretroviral drugs are especially favorable for viral populations. They often occur under the selective pressure of ART when HIV replication is not completely suppressed but can also be transmitted from one person to another via a new infection or superinfection (infection of an already infected individual) [16].

HIV-1 drug resistance is therefore tested in patients presenting with a primary infection, untreated patients who are chronically infected at the time of diagnosis or before treatment initiation, and patients under antiretroviral drug therapy with detectable plasma HIV-1 RNA [16]. Mutations in the viral genome can be helpful when it comes to population sequencing and the evaluation of transmission networks. The fact, that mutations can be transmitted from one person to another, makes transmission events traceable, as the genome of the viral population in the newly infected person has a high similarity to the genome of the viral population in the person who transmitted the infection. This can be used to establish transmission clusters. It is recommended to focus on the HIV-1 gene *pol* to evaluate transmission events, as it has a relative slow intra-host mutation rate and therefore allows the establishment of linkage, even after long periods between diagnosis of infection and putative transmission events [24].

## 2 Objectives

The aim of this diploma thesis was to reconstruct the HIV-1 transmission network in Southeastern Austria. The HIV-1 epidemic in this region may be characterized by significant ethnical and gender disparities. Furthermore population groups engaging in risk behavior may be targeted. Through homology analysis of HIV-1 partial *pol* sequences (including the regions encoding for *reverse transcriptase* and *protease* of the viral genome) combined with epidemiological data, trends in transmission on population level may be revealed.

## 3 Material and Methods

This study was performed in collaboration with the Division of Infectious Diseases, University of California San Diego (UCSD), United States.

### 3.1 Study Population

In this study 259 individuals living in Southeastern Austria are included, who were diagnosed with an HIV-1 infection between 2008 and 2014. Patient related data was collected retrospectively at the Department of Infectious Diseases, Landeskrankenhaus Graz Süd-West, Standort West, through access to the HIV patient management system (HIP). This management system serves as a database for the Austrian HIV cohort and stores patient related demographic data, treatment strategies, and clinical information. Therefore, this database can be used for HIV surveillance studies in Austria.

All patient related data presented was de-identified and is therefore not traceable to individual patients. Data collected included sex, ethnical origin, year of diagnosis, age at diagnosis, duration of infection, transmission risk, and area of residence. Regarding sex, there were two categories, female and male. Ethnical origin was stratified into Caucasian, Asian, and Black. The duration of infection was stratified into acute and chronic infection. Acute HIV-1 infection is defined as seroconversion with negative (non-reactive) HIV antibody testing not more than three years before the first positive HIV antibody test [2].

The categories concerning transmission risk were adjusted according to the pre-defined categories of the HIP: men having sex with men (MSM), injection drug user (IDU), man having sex with men plus injection drug user (MSM plus IDU), heterosexual (HSX) transmission, hemophilia/coagulation disorder, nosocomial infection, mother to child transmission (MTCT), and others/unknown. HSX transmission was subdivided into: Sex with an IDU, sex with a bisexual man, sex with a hemophiliac, patient was born in/comes from a high-prevalence country, sex with a person from a high-prevalence country, sex with an HIV-positive person who does not belong to any of the risk groups mentioned before, and patient sure of HSX transmission and no further information about risk factors/risk behavior available. High prevalence countries are defined as countries with an HIV infection rate for adults of  $\geq 1\%$  [2].

Information about the viral load (copies/ml) at the date of investigation (date on which the RNA sequence was generated) and the HIV subtype was extracted from the database at the Institute of Hygiene, Microbiology and Environmental Medicine, Medical University of Graz, Austria.

In this study, 245 of 259 individuals (94.6%) attended the Department of Infectious Diseases, Landeskrankenhaus Graz Süd-West, Standort West for clinical care. Nine individuals (3.5%) attended the Barmherzige Brüder Hospital Graz, the Landekrankenhaus Graz Süd-West, Standort Süd (formerly known as Landesnervenklinikum Sigmund-Freud), or the practice of a local specialist in internal medicine. The remaining five individuals (1.9%) attended departments of the Medical University of Graz (MUG) including the Department of Internal Medicine, the Department of Dermatology and Venereology, the Department of Psychiatry, the Department of Nephrology, and the Department of Ophthalmology.

## 3.2 Area of Investigation

The area of investigation included the Austrian province of Styria (13 districts, approximately 1.2 million inhabitants) and the neighboring area of Southern Burgenland (3 districts, approximately 100.000 inhabitants). The southern part of Burgenland was included in the area of investigation because it is a catchment area of the Landeskrankenhaus Graz Süd-West, Standort West concerning HIV treatment and clinical care.

For better de-identification and to avoid the risk of re-identification of the study population, the 18 districts were grouped into 5 larger areas according to the topography including the urban area of Graz and the rural areas of Graz Country, Northern Styria, Southern/Western Styria, and Eastern Styria (which included the three southern districts of Burgenland). Graz as a city with more than 300.000 inhabitants represents the only urban area in this investigation. The area of Northern Styria included the districts of Liezen, Murau, Murtal, Leoben, and Bruck-Mürzzuschlag, while the area of Southern/Western Styria included the districts of Voitsberg, Deutschlandsberg, and Leibnitz. The area of Eastern Styria included the districts of Südoststeiermark, Weiz, Hartberg- Fürstenfeld, Jennersdorf, Güssing, and Oberwart.

## 3.3 Population Sequencing

Population sequencing was performed at the Molecular Diagnostics Laboratory, Institute of Hygiene, Microbiology and Environmental Medicine, Medical University of Graz. A total of 259 clinical samples obtained from patients with HIV-1 infection were investigated. Peripheral venous blood had been collected in blood collection tubes containing the anticoagulant ethylenediamine tetraacetic acid (EDTA) (Greiner Bio-One, Kremsmünster, Austria). Upon six hours of drawing the blood, the samples were centrifuged at  $1200\times g$  for 12min at room temperature. After centrifugation, plasma aliquots were frozen im-

mediately at  $-70^{\circ}\text{C}$  until tested. In the routine diagnostic laboratory, an aliquot of each sample was analyzed using the COBAS AmpliPrep/COBAS TaqMan HIV-1 Quantitative Test (Roche Molecular Systems, Inc., Branchburg, NJ, USA). The eluates obtained from extraction on the COBAS AmpliPrep platform (Roche Diagnostics International, Rotkreuz, Switzerland) were taken for population sequencing. Aliquots of  $17\ \mu\text{l}$  were used for reverse transcription PCR. Population sequencing was performed with the TRUGENE HIV-1 Genotyping Kit (Siemens Healthcare Diagnostics Inc., Tarrytown, NY, USA) according to the manufacturer's guidelines. This assay consists of the TRUGENE<sup>®</sup> HIV-1 Genotyping Kit and the OpenGene<sup>®</sup> DNA Sequencing System. The TRUGENE<sup>®</sup> HIV-1 Genotyping Kit is designed for sequencing the *protease* and *reverse transcriptase* coding regions of the HIV-1 *pol*-gene [25]. The OpenGene<sup>®</sup> DNA Sequencing System includes hardware and interpretive software for detection of mutations and drug resistances in the genome of HIV-1. For resistance analysis, the GuideLines<sup>™</sup> rules, version 17.0, integrated in the system software were used. This assay can also be used to determine the HIV subtype. For this study, the HIV subtype was determined by using the Subtype Classification using Evolutionary Algorithms program (SCUEAL) program [26].

Sequences generated within the period from October 2007 to December 2014 were included in the sequence analysis. If multiple sequences were available, the earliest sequence available was used for analysis. All patients were ART-naïve at the time of blood sample collection.

### 3.4 Genetic Network Analysis

The genetic network analysis was performed at UCSD, United States. The partial *pol* sequences were used including the regions encoding for *reverse transcriptase* and *protease*.

Transmission networks directly link genetically similar HIV-1 RNA sequences. A linkage

between two sequences was established, when a Tamura-Nei 93 (TN93) genetic distance (D) of  $<1.5\%$  was found [27]. In this study, the threshold of  $<1.5\%$  was used based on results of previous publications, which show that *pol* sequences typically do not diverge more than 1% from baseline in a mono-infected person over a decade [24]. Through these established linkages putative transmission clusters were inferred. Clusters were defined as connected components of the transmission network comprising two or more individuals. The inferred transmission clusters included all connected individuals. Therefore, all individuals within a cluster had a D of  $<1.5\%$  from at least one other individual in the cluster but not necessarily a D of  $<1.5\%$  from all other individuals within the cluster.

### 3.5 Phylogenetic Reconstruction and Bayesian Analyses

Phylogenetic analyses were performed by first inferring a time-scaled maximum clade credibility tree including all 259 of the HIV RNA sequences by using a Markov Chain Monte Carlo (MCMC) framework as implemented in BEAST v1.8.1 [28] with BEAGLE [29]. A discretized gamma distribution was used (GTR + 4 $\Gamma$ ) to account for among-site rate variation. Time scales of the trees were calibrated with the sampling dates available.

An uncorrelated lognormal (U 169 CLN) molecular clock with a gamma distribution prior model that allows rates to vary among the branches of the inferred phylogenies to infer the timescale of HIV evolution for each subject, with a gamma distribution prior on the mean clock rate (shape = 0.001, scale = 1000), was specified. A skygrid tree was used as a coalescent demographic model with time-aware smoothing [30, 31]. Monte Carlo simulations were run for 100 million steps, sub-sampling parameters every 50,000 steps. Convergence of the chains was inspected using Tracer v1.6. Maximum clade credibility trees were obtained with TreeAnnotator v1.8.1 [32]. The HIV-1 viral migration patterns between Graz City and the surrounding rural areas were investigated. The historical inference of these viral migration events from Graz City to rural areas and from rural areas

to Graz City was obtained using estimated Markov jump counts between regions along the branches of the posterior tree distribution as implemented in the BEAST package v1.8.1 [28].

## **3.6 Statistical Analysis**

Statistical analysis was performed using the statistical analysis software SPSS (IBM SPSS Statistics version 22). Demographic data was displayed as absolute numbers plus percentages, as rational numbers plus percentages, as medians plus interquartile range (IQR) or means plus 95% confidence interval (CI). Demographic data, HIV risk factors, and HIV subtypes between individuals that clustered within the transmission network were compared to those that did not cluster within the transmission network by using two tailed Chi-squared or Fishers exact test for categorical variables, and Mann-Whitney-U test for continuous variables.

## 4 Results

### 4.1 Study Population

The demographic characteristics of the study population are shown in Table 2. The study population consisted of 259 patients, 193 (75%) males and 66 (25%) females. A total of 248 (95.7%) specimens had been collected during the chronic stage of infection while 11 (4.3%) specimens had been collected during the acute stage of infection (nucleic acid amplification testing result positive, antibody testing result negative). The mean age at the time of HIV diagnosis was 34 years (range, 4-73 years). For women, the mean age was 31 years (range, 4-63) and for men 35 years (range, 8-73 years). Concerning the ethnicity the study population consisted of 221 (85.3%) individuals of Caucasian origin, 34 (13.1%) individuals of African origin, and four (1.5%) individuals of Asian origin. The 221 Caucasians were subdivided into 179 males and 42 females. Of the 34 individuals of African origin 13 were males while 21 were females and of the four individuals of Asian origin one was male and three were female.

Table 2: Baseline demographic characteristics.

Characteristics	Study population (%)	Individuals clustering within the network (%)	Non-clustering individuals(%)	<i>P</i> value
All individuals	259 (100)	118 (45.6)	141 (54.4)	
Male	193 (74.5)	99 (83.9)	94 (66.7)	<i>P</i> =0.002
Female	66 (25.5)	19 (16.1)	47 (33.3)	<i>P</i> =0.002
Mean age (yrs)	34	33	34	
Age range (yrs)	4-73	17-65	4-73	
Age <24 yrs	53 (20.5)	24 (20.3)	29 (20.6)	n.s.
Age >50 yrs	27 (10.4)	10 (8.5)	17 (12.1)	n.s.
Caucasian	221 (85.3)	110 (93.2)	111 (78.8)	<i>P</i> =0.001
Black	34 (13.1)	8 (6.8)	26 (18.4)	<i>P</i> =0.006
Asian	4 (1.5)	0	4 (2.8)	n.s.
Urban residence	117 (45.2)	50 (42.4)	67 (47.5)	n.s.
Rural residence	142 (54.8)	68 (57.6)	74 (52.5)	n.s.

A high heterogeneity of subtypes could be observed. In Figure 4, the distribution of HIV-1 subtypes and CRFs in the study population is shown. The sequences represented several subtypes of HIV-1 group M including subtypes A, B, C, D, F, and G, and two CRFs (CRF01\_AE and CRF02\_AG) as shown in Table 3.

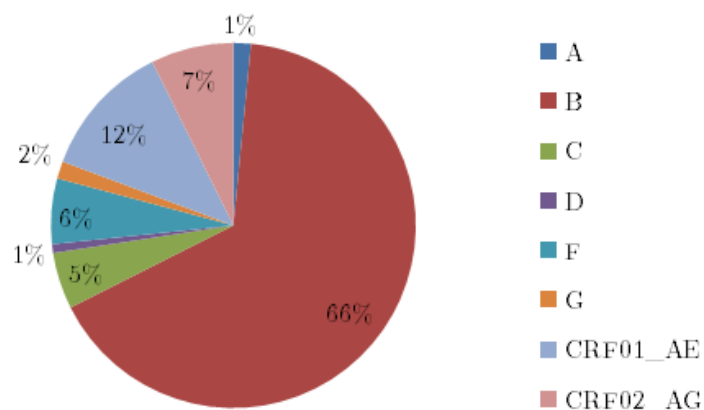


Figure 4: Distribution of HIV-1 subtypes / CRFs in the study population.

Table 3: Baseline viral characteristics.

	Study population (%)	Individuals clustering within the network (%)	Non-clustering individuals(%)	<i>P</i> value
HIV-1 subtype A	4 (1.5)	0	4 (2.8)	n.s.
HIV-1 subtype B	171 (66.0)	90 (76.3)	81 (57.4)	<i>P</i> =0.003
HIV-1 subtype C	13 (5.0)	0	13 (9.2)	<i>P</i> <0.001
HIV-1 subtype F	15 (5.8)	11 (9.3)	4 (2.8)	<i>P</i> =0.033
HIV-1 CRF01_AE	32 (12.4)	11 (9.3)	21 (14.9)	n.s.
HIV-1 CRF02_AG	18 (6.9)	4 (3.4)	14 (9.9)	<i>P</i> =0.049
Other HIV-1 subtype <sup>a</sup>	6 (2.3)	2 (1.7)	4 (2.8)	n.s.

<sup>a</sup>Including subtypes D and G.

Table 4: Self- reported risk factors for HIV-1 .

	Study population (%)	Individuals clustering within the network (%)	Non-clustering individuals(%)	<i>P</i> value
MSM	122 (47.1)	70 (59.3)	52 (36.9)	<i>P</i> <0.001
HSX	115 (44.4)	37 (31.4)	78 (55.3)	<i>P</i> <0.001
IDU	15 (5.8)	10 (8.5)	5 (3.5)	n.s.
MSM plus IDU	2 (0.8)	1 (0.8)	4 (0.7)	n.s.
Other risk factor <sup>a</sup>	5 (1.9)	0	5 (3.5)	n.s.

<sup>a</sup>Including MTCT and nosocomial infection.

Information on the transmission risk for HIV-1 infection was available for all individuals within our study population. Of 193 males, 122 (63%) self-identified as MSM, while 54 (28%) self-reported HSX transmission as their risk factor. Another 13 (7%) reported to be IDUs. Two individuals (1%) reported to be MSM plus IDUs. Furthermore, two (1%) individuals reported other risk factors, MTCT and nosocomial transmission, respectively. Among 54 men reporting HSX transmission as their risk factor, 13 (24%) reported sex with a female from a high prevalence country, while 11 (20%) reported origin from a high prevalence country. Two individuals (4%) reported sex with a female known to be HIV positive. The remaining 28 (52%) males reporting HSX transmission did not provide further information. In Table 4 the self-reported risk factors for individuals clustering in the transmission network are compared to individuals not clustering. In Figure 5, the self-reported transmission risk for HIV-1 infection for males is shown, in Figure 6, the self-reported subcategory of HSX transmission for males.

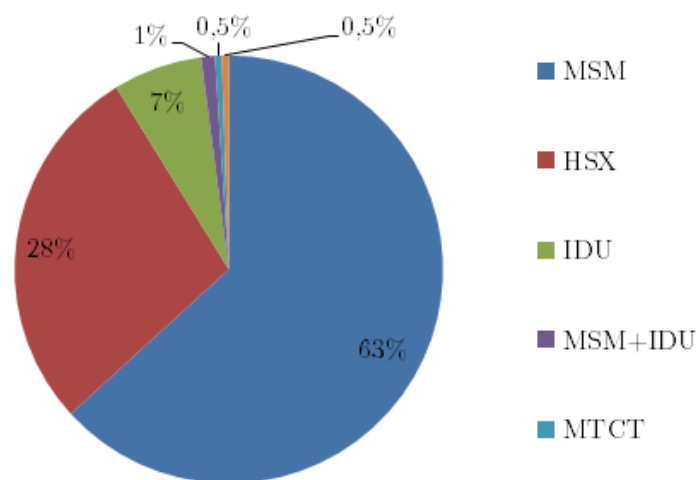


Figure 5: Self-reported transmission risk for males.

Of 66 females, 61 (92%) reported HSX transmission as their only risk factor for HIV-1 infection, while two (3%) reported to be IDUs. Nosocomial transmission was reported by three (5%). Among 61 females reporting HSX transmission, 24 (39%) reported sex with a male from a high prevalence country, four (7%) reported sex with a male known to be

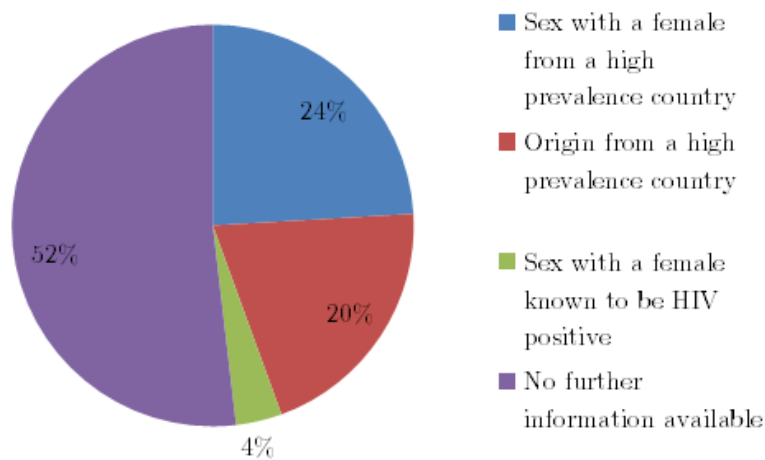


Figure 6: Self-reported subcategory of HSX transmission for males.

HIV positive, another four (7%) sex with an IDU male, and two (3%) sex with a bisexual male. The remaining 27 (44%) females reporting HSX transmission did not provide further information. In Figure 7, the self-reported transmission risk for HIV-1 infection for females is shown, in Figure 8, the self-reported subcategory of HSX transmission for females.

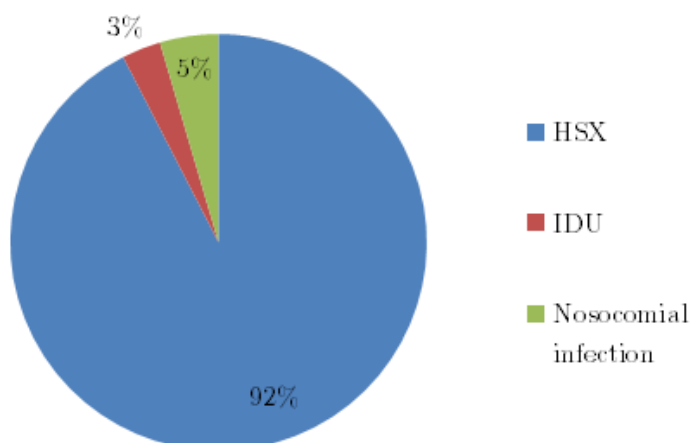


Figure 7: Self-reported transmission risk for females.

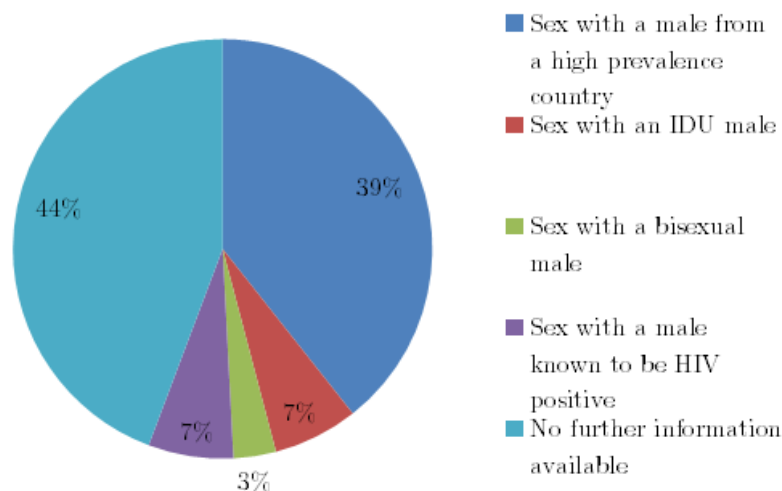


Figure 8: Self-reported subcategory of HSX transmission for females.

A total amount of 117 individuals was found to reside in the urban region of Graz, while 142 individuals were found to reside in rural areas (61 in Northern Styria, 31 in Southern/Western Styria, 31 in Eastern Styria, and 19 in Graz County). In Figure 9, the distribution of HIV-1 subtypes according to the place of residence is shown. Figure 10 shows the distribution of risk factors for HIV-1 infection by place of residence.

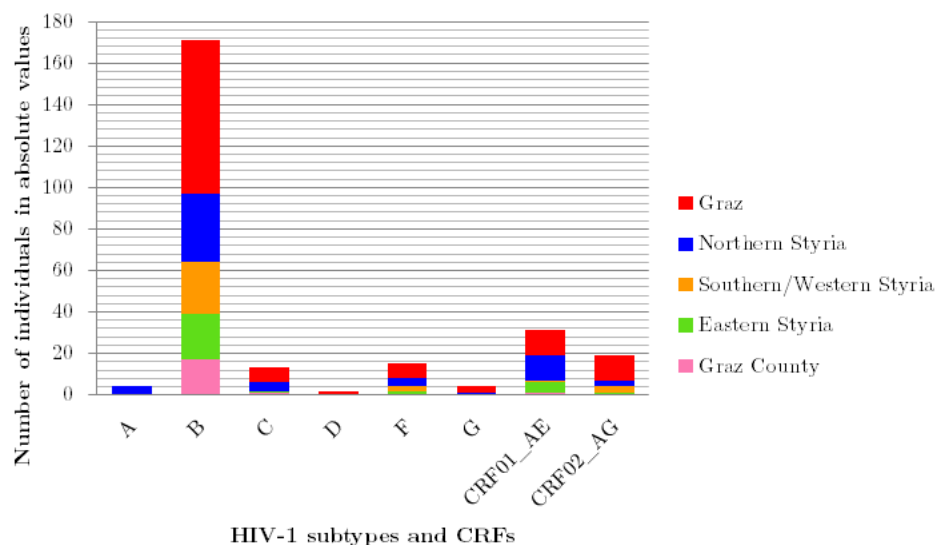


Figure 9: Distribution of the HIV-1 subtypes and CRFs by place of residence.

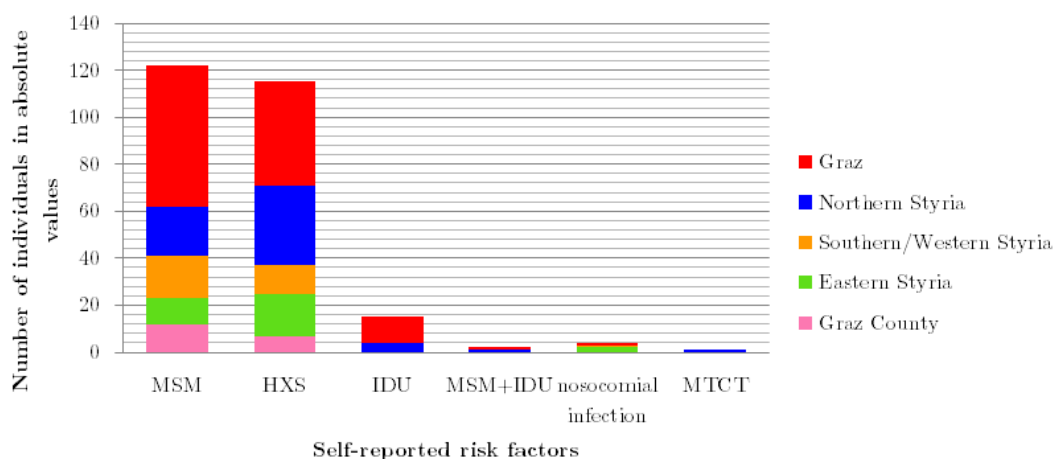


Figure 10: Distribution of risk factors for HIV-1 infection by place of residence.

## 4.2 Transmission Cluster Composition

Inclusion in a transmission cluster required a D of  $<1.5\%$  to another sequence within the set of sequences. Using a threshold of  $1.5\%$ , a total number of 118 (45.6%) individuals was connected to at least one other person of our study population. Linkages could be established for 99 males and 19 females, resulting in 34 putative transmission clusters. The mean number of members per cluster was 3.5 (range, 2-12). Of these established clusters, 19 (55.9%) were pairs, seven (20.6%) included three individuals, and two (5.9%) included four individuals. One cluster with five, one with six, one with seven, and one with nine members could be established. Two large clusters with 12 members each were found. Table 5 shows demographic and viral characteristics according to the cluster size. Figures 11 and 12 give an overview on the established transmission clusters.

Table 5: Baseline characteristics by cluster size.

Cluster size	Pairs (n=19)	Triplets (n=7)	Clusters including $\geq 4$ individuals (n=8)
Median age (years)	36	31	33
Age range (years)	22- 64	22-40	25-53
MSM exclusive (%)	6 (31.6)	3 (42.9)	0
HSX exclusive (%)	8 (42.1)	2 (28.6)	0
IDU exclusive (%)	1 (5.3)	1 (14.3)	2 (25.0)
HSX+MSM mixed (%)	4 (21.1)	1 (14.3)	6 (75.0)
HIV-1 Subtype B (%)	11 (57.9)	6 (85.7)	6 (75.0)
HIV-1 Subtype F (%)	2 (10.5)	1 (14.3)	1 (12.5)
HIV-1 Subtype D (%)	1 (5.2)	0	0
CRF01_AE (%)	3 (15.8)	0	1 (12.5)
CRF02_AG (%)	1 (5.3)	0	0

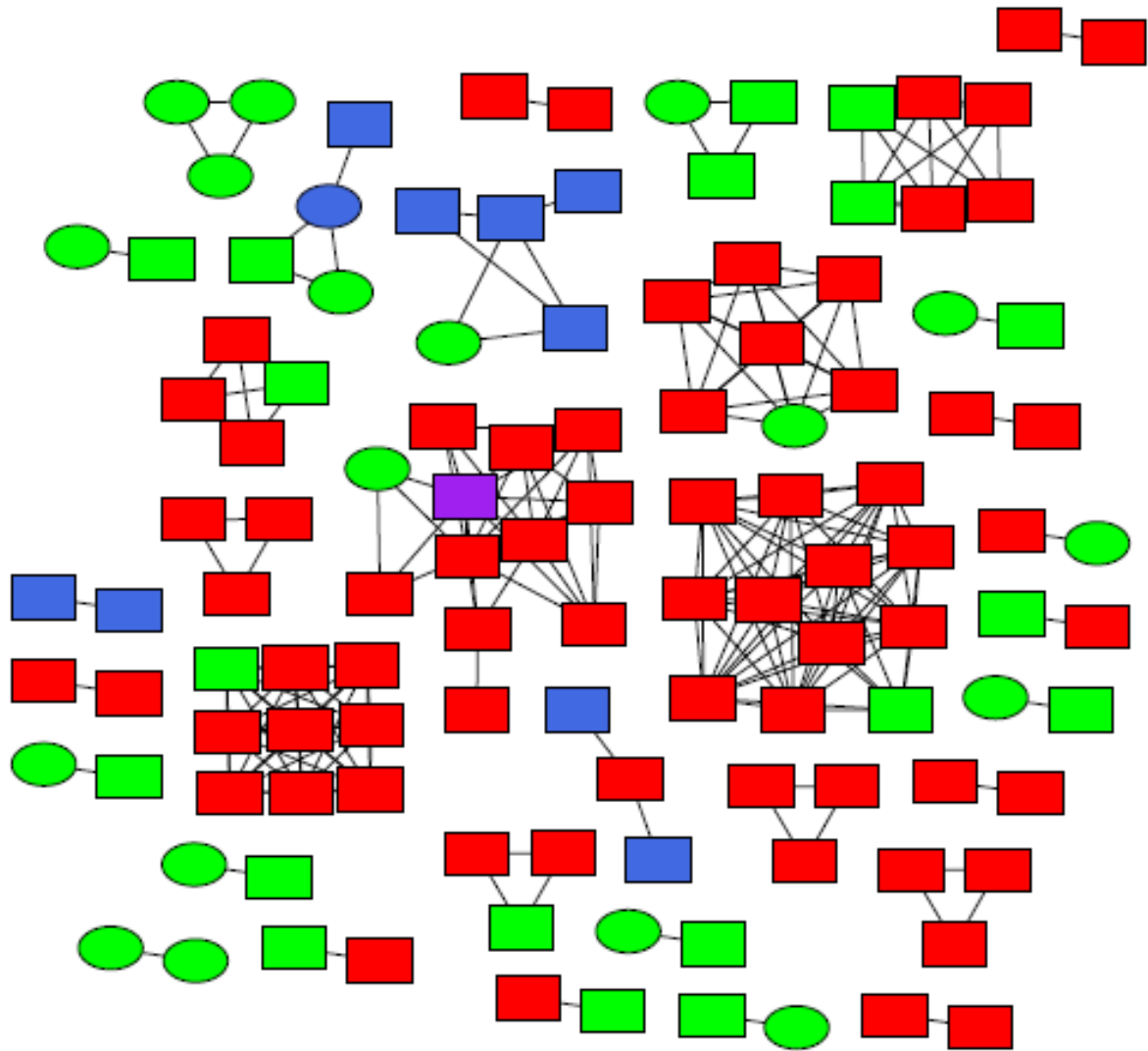


Figure 11: The inferred transmission network according to self-reported risk factors for HIV-1 infection. Only clustered individuals (nodes) are shown. The transmission network is color coded according to the self-reported risk factors: red for MSM, green for HSX transmission, blue for IDU, purple for MSM plus IDU. Rectangles represent males, circles represent women. Lines represent putative transmission linkages between individuals.

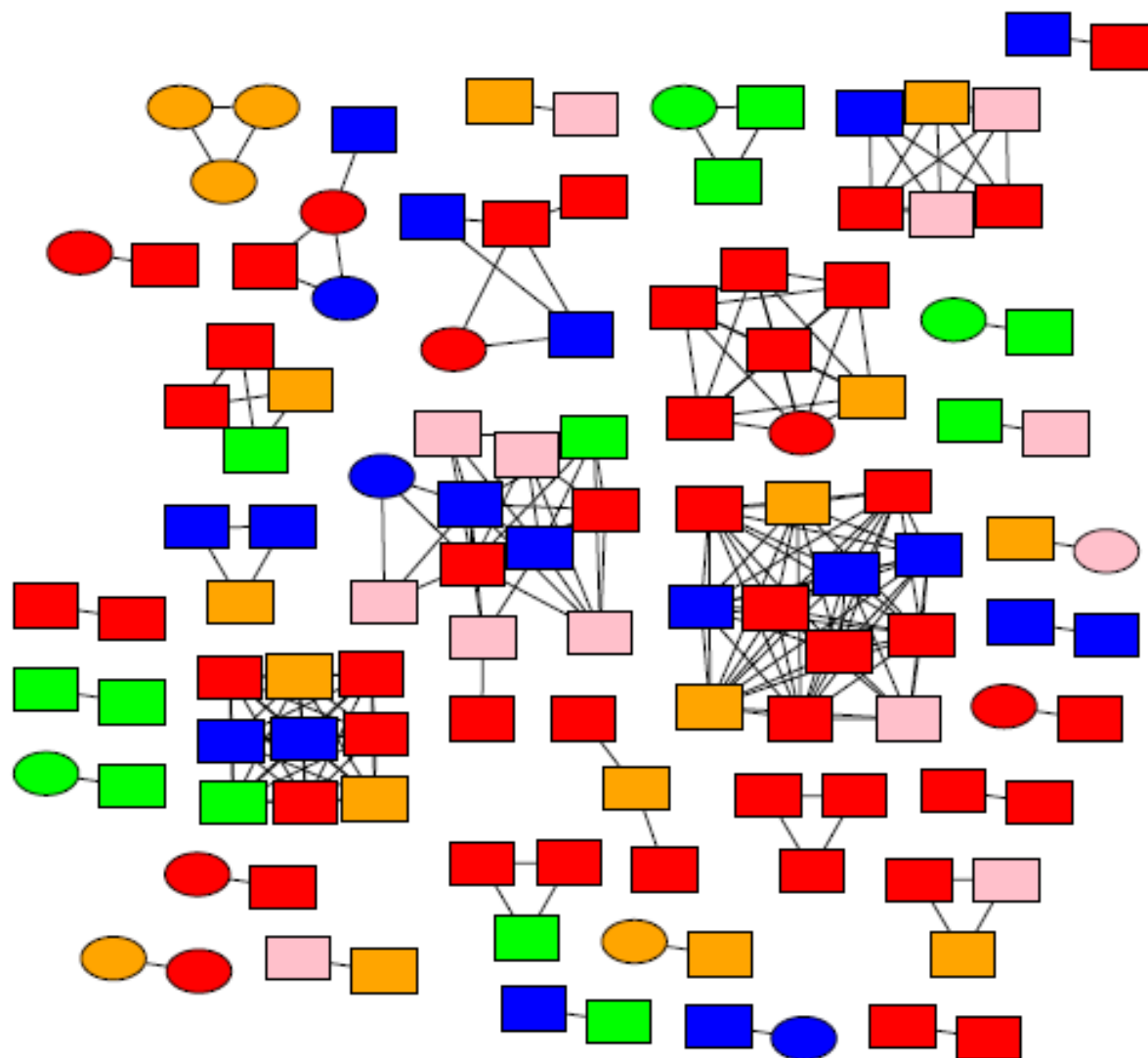


Figure 12: The inferred transmission network according to the place of residence. Only clustered individuals (nodes) are shown. The transmission network is color coded according to the area of residence: red for Graz, blue for Northern Styria, orange for Southern/Western Styria, green for Eastern Styria, and pink for Graz Country. Rectangles represent males, circles represent women. Lines represent putative transmission linkages between individuals.

### 4.3 Variables associated with Membership in Clusters

The likelihood of clustering for an individual differed by sex. Males were more likely to cluster than females ( $P=0.002$ ). The mean age for clustering individuals was 33 years (range, 17-65). No significant difference in clustering could be established for the place of residence (urban versus rural). Of the clustering individuals, 50 were found to reside in Graz, 13 in Graz County, 15 in Northern Styria, 21 in Southern/Western Styria, and 19 in Eastern Styria. It was significantly more likely for individuals of Caucasian origin to cluster within the transmission network ( $P=0.001$ ) than for individuals of African origin ( $P=0.006$ ). In this study, individuals of Asian origin did not cluster at all (Figure 13). Of the eight individuals of African origin, four (50%) clustered in pairs with individuals from the identical ethnical origin (one MSM and one HSX cluster). The remaining four clustered with Caucasians (three HSX and one MSM cluster).

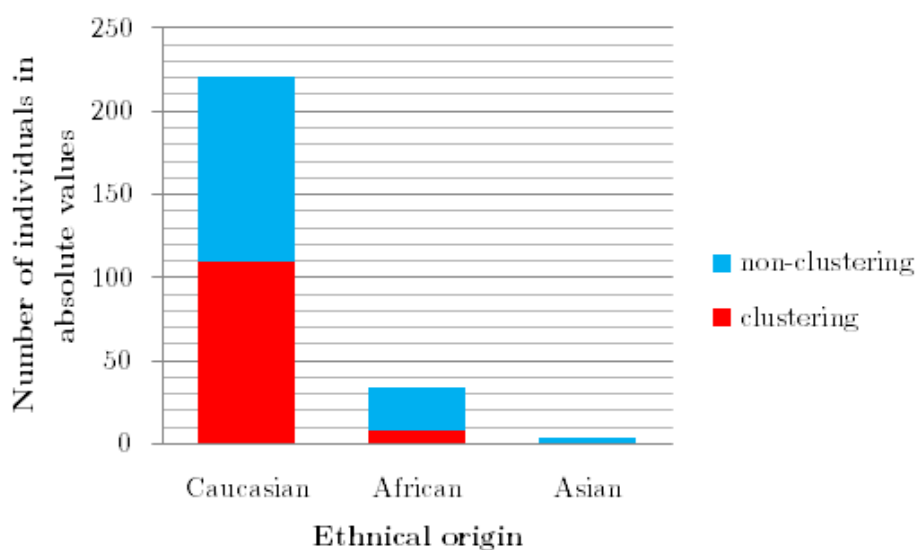
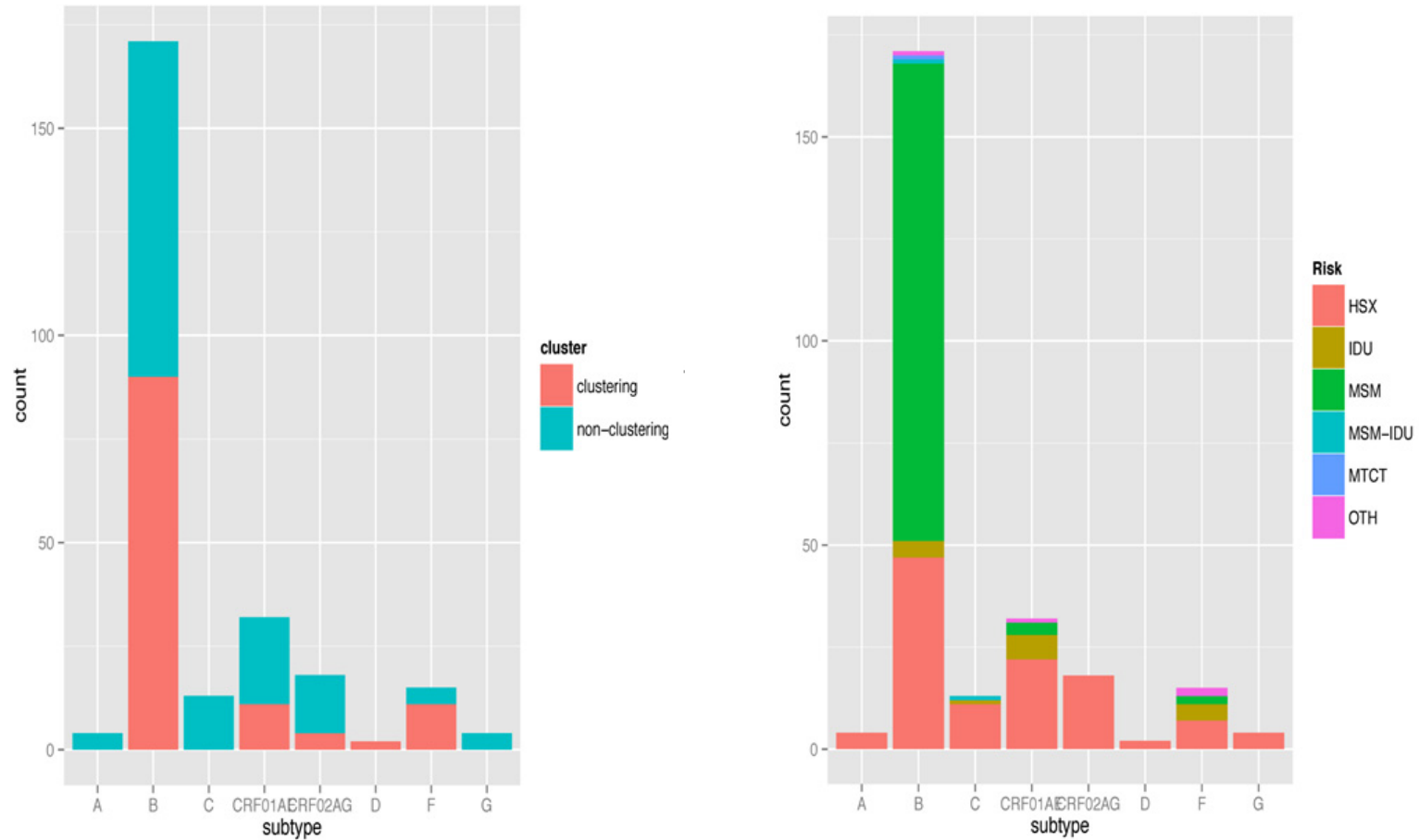


Figure 13: Comparison of clustering and non-clustering individuals according to their ethnical origin.

Clustering individuals reported significantly more likely to be MSM ( $P<0.001$ ) when compared to non-clustering individuals. The large clusters were predominantly composed of MSM (see Figure 11). Of 19 individuals self-reporting HSX transmission as their only

risk factor, nine (47%) clustered closely to MSM. Of those nine males, seven had reported that they could not specify their risk other than having HSX contacts, while two had reported HSX transmission by having sex with a female from a high prevalence country. As shown in Figure 11, three of these nine males clustered in pairs with males reporting to be MSM. Another one was found together with two MSM in a cluster consisting of three individuals. The remaining five clustered within the four large predominantly MSM clusters. One of these clusters included four males reporting to be MSM and two reporting HSX transmission. The remaining three males reporting HSX transmission represented the only non-MSM within their respective transmission clusters. A higher prevalence of subtype B ( $P<0.001$ ) and subtype F ( $P=0.033$ ) was found in all 118 clustering individuals (see Figure 14a). It was significantly more likely for non-clustering individuals to be infected with subtype C ( $P<0.001$ ). MSM were significantly ( $P<0.001$ ) more likely infected with subtype B, whereas HSXs ( $P<0.001$ ) and IDUs ( $P=0.002$ ) were more likely infected with non-B subtypes (see Figure 14b).



(a) Comparison between clustering and non-clustering individuals by HIV-1 subtype/CRF.

(b) HIV-1 subtypes/CRFs according to the self-reported risk factors for HIV-1 infection.

Figure 14: Distribution of the HIV-1-subtypes/ CRFs in the study population.

## 4.4 Phylogenetic Reconstruction and Migration Analysis

With phylogenetic methods, the time of most recent common ancestor (MRCA) for each individual cluster was inferred, showing that large clusters were significantly older. A time-scaled phylogenetic reconstruction (see Figure 15) of clustering individuals according to their HIV-1 subtype was inferred. For each HIV-1 subtype a phylogenetic tree was constructed.

## 4.5 Evolution in the Mode of Transmission

The evolution of the mode of transmission among HIV-1 infected individuals in Southeastern Austria is shown in Figure 16. The relative frequency of the reported risk is indicated. MSM and HSX transmission account for 40-60% of new HIV-1 infections in Southeastern Austria within the observation period each year. After a peak in 2011, a slight tendency of decrease in HSX transmission can be observed. In contrast, a tendency of increase in MSM transmission occurred from 2008 to 2010 and from 2011 (following the trough in that year) onwards. MSM plus IDU, MTCT, and nosocomial infection account for constantly low numbers of new HIV-1 cases.

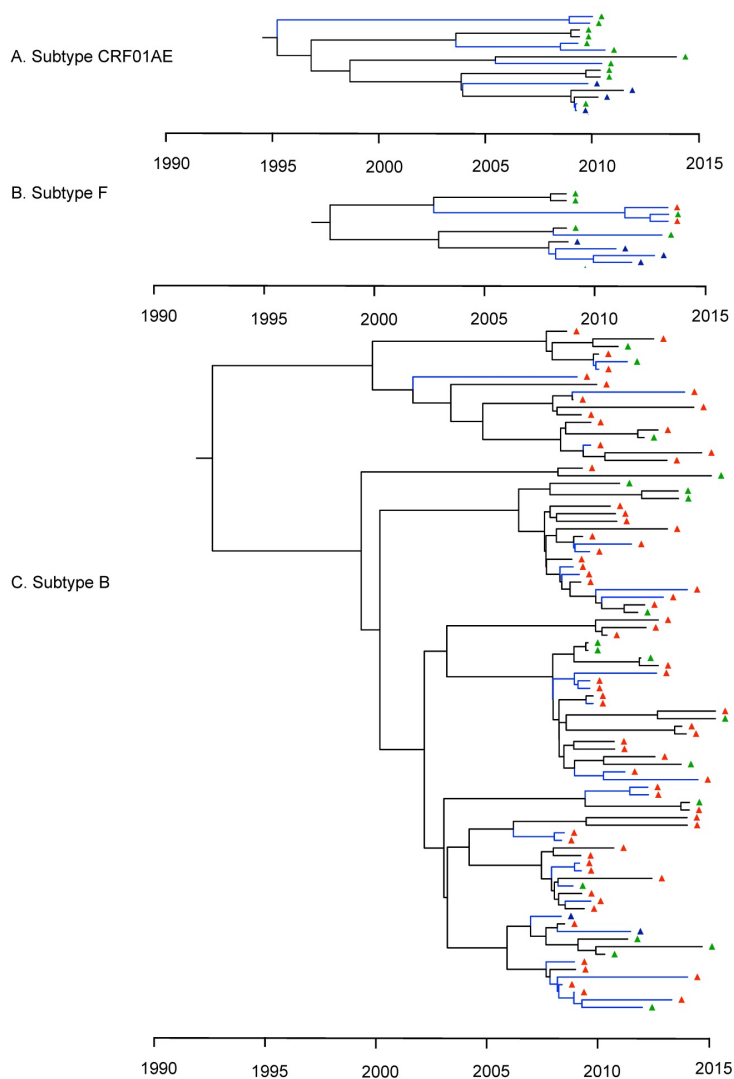


Figure 15: Time-scaled phylogenetic reconstruction of the clustering individuals. Tips are colored by their reported risk factor in red (MSM), green (HSX transmission), blue (IDU), and purple (MSM plus IDU). Branch colors represent the most probable place of residence of the parental node of each branch. Sequences originating from Graz are depicted in blue, sequences originating from rural areas in black.

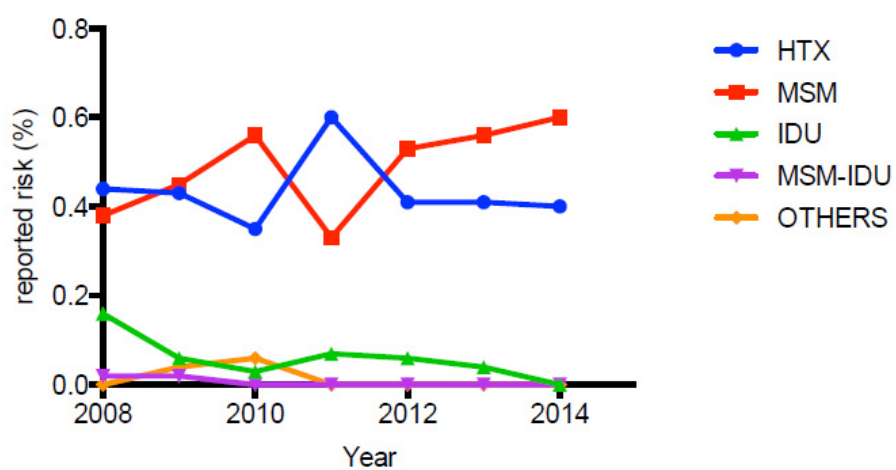


Figure 16: Change in the prevalent mode of transmission between 2008 and 2014 in South-eastern Austria. The category “Others” includes MTCT and nosocomial infection.

## 5 Discussion

In this study, partial *pol* sequences obtained from 259 individuals living in Southeastern Austria, who had been diagnosed with HIV-1 between 2008 and 2014, were analyzed. Female and male individuals of diverse ethnical origin, infected with a variety of HIV-1 group M subtypes and heterogeneous transmission risks were included. Through the combination of methods from molecular epidemiology and demographic data it was possible to infer and characterize the HIV-1 transmission network for Southeastern Austria. Information obtained from the phylogenetic analysis of HIV-1 partial *pol* sequences was supplemented with data about sex, age at diagnosis, subtype, transmission risk, and place of residence to gain deeper insight in the local transmission network. This study represents the first phylogenetic analysis of HIV-1 partial *pol* sequences in Southeastern Austria. The sequences included in the analysis represent over 95% of the new HIV-1 diagnoses for Southeastern Austria and about 7.4% of the new diagnoses across Austria for the period between 2008 and 2014 [2].

Local transmission networks for individuals infected with HIV-1 can never be assumed to be complete. Through incomplete sampling potential cluster members may remain unidentified and putative linking chains may be incorrect. A putatively inferred linkage between two individuals is not equivalent to an actual event of transmission, as additional unsampled individuals may have been members of the transmission chain linking the two observed putatively linked individuals. Individuals infected but not yet diagnosed may be a reason for undersampling. This is a special problem with infectious diseases having long

incubation periods and a late onset of symptoms like the HIV-1 infection. Individuals diagnosed but not linked to care may be another reason. A bias towards the preferential inclusion of individuals, who present for regular screening in clinical care, might have been created because a minimum of demographic data per individual was required to be enrolled in this study. Individuals were excluded from this study, if the minimum dataset required was not complete.

In comparison with recent reports in this field [33, 34], clustering was found at a higher degree in Southeastern Austria. Putative transmission linkages could be inferred for 45.6% of the analyzed sequences. The number of males in our study population was three times higher as the number of females (75% vs. 25%). This is similar to recent reports and also representative for the Austrian HIV-cohort [33, 2]. It was more likely for males ( $P=0.002$ ) to cluster in the transmission network. This may suggest that males are more likely to engage in risk behavior like unprotected sex.

Individuals of African origin were less likely to cluster in the local transmission network ( $P=0.006$ ) than individuals of Caucasian origin ( $P=0.001$ ). Infection with HIV-1 in high-prevalence countries and immigration to Austria without knowledge of the existing infection may be assumed as a reason. Further studies focusing on migration in conjunction with phylogenetic analysis may help to better understand the way of transmission in this group. Individuals of Asian origin did not cluster at all. This might be due to the low number of Asians in the study population.

The mean age at the time of HIV-1 diagnosis was 34 years (range, 4 – 73 years). This is a representative number for Austria, as the median age at diagnosis is between 30 and 36 years in Austria [11].

Subtype B accounts for the majority of HIV-1 infections in Southeastern Austria. Within the study period, it accounted for 66% of the HIV-1 infections. The epidemic in Austria is characterized by the predominance of subtype B since its beginning [2]. The present

study shows that more and larger clusters were composed of individuals infected with subtype B than with any other subtype.

A substantial mixing between MSM and individuals infected through HSX transmission was observed, as 47% of the clustering males reporting HSX contacts as their only risk factor for HIV-1 infection clustered closely with MSM. Those men clustered either in pairs with MSM or in larger transmission clusters consisting of MSM predominantly. This observation may be caused by missing individuals in the study population or, more likely, an incorrect self-reported risk factor. HSX transmission might be reported by homosexual or bisexual males as their risk factor, because they fear discrimination. A survey conducted by the European Union in 2012 revealed that only 26% of Austrian lesbian, gay, bisexual, and transgender individuals reported their sexual orientation or gender identity to the majority of people in their private or professional environment, while 29% reported that they never come out [35].

As no significant relationship between clustering individuals and their place of residence could be established and several closely clustering individuals were found to reside in different areas (see Figure 12), a high degree of migration was suspected within the area of investigation. Regarding to analysis on directionality and history of transmissions, the time-scaled phylogenetic reconstruction suggests a predominant migration from rural areas to Graz that took place particularly in the late 1990's and early years of the new millennium.

Although the number of newly diagnosed individuals remained constantly high over the observation period, a change in the most prevalent transmission mode within the study population can be observed between 2008 and 2014. However, this finding may not be representative for whole Austria where MSM account for the majority of new infections diagnosed since 2008 with a slight decrease of infections [2]. MSM show a relatively high awareness of HIV compared to other risk groups and have good knowledge about HIV-1 testing [36]. This may have led to a bias, as MSM may more likely present early after

infection. Individuals belonging to other risk groups might have been infected in the same period of time, but are underrepresented in the study population as they might not yet have been diagnosed. Therefore, the prevalent mode of transmission in diagnosed individuals may not represent the most prevalent mode of transmission events for a distinctive period of time.

Within the transmission network, 54.4% of the sequences analyzed did not cluster. This observation suggests that many transmission events did not occur locally. Sex tourism to foreign countries as well as migration of individuals infected with HIV-1 from foreign countries to Southeastern Austria might be assumed as possible reasons for not clustering, as a variety of HIV-1 subtypes (A, C, F, G, and CRF02\_AG) were detected among the study population, which are not endemic in Austria. Furthermore, individuals infected but not yet diagnosed might link those individuals to the transmission network.

## 6 Conclusion

In this study, the majority of individuals within transmission clusters were found to be males. The predominant factors associated with clustering were MSM transmission as risk factor for HIV-1 infection, Caucasian ethnical origin, and infection with HIV-1 subtype B or subtype F. Collection of demographic data in combination with results obtained from the phylogenetic analysis of HIV-1 partial *pol* sequences has an added value to the understanding of local HIV-1 epidemics and gives insights into the transmission network. This unique view into the complex structure of the local transmission network may help to elaborate targeted prevention programs to reduce HIV-1 transmissions in Southeastern Austria. However, further studies may be beneficial to expand the transmission network and connect it with other networks from remaining parts of Austria.

## 7 Annotation

This study was accepted for publication in PLOS ONE as “Characterization of HIV Transmission in South-East Austria” by M. Hönigl, A. Chaillon, HH. Kessler, B. Haas, E. Stelzl, K. Weninger, SJ. Little and SR. Mehta (date of acceptance: 01-Mar-2016).

# Bibliography

- [1] UNAIDS. *Epidemiology publications*. Available: <http://www.unaids.org/en/dataanalysis/knowyourepidemic/epidemiologypublications>; Updated: July 2014; Accessed: November 10, 2014.
- [2] M Giesinger, M Kitchen, et al. *26 th Report of the Austrian HIV Cohort Study*. Available: <http://dermatologie.tirol-kliniken.at/data.cmf?vpath=dokumente/kohortenbericht>; Updated: September 2014; Accessed: November 10, 2014.
- [3] M Peeters, M D’Arc, and E Delaporte. “Origin and diversity of human retroviruses.” In: *AIDS reviews* vol.16.no.1 (Jan. 2014), pp. 23–34.
- [4] A Vallari, V Holzmayer, et al. “Confirmation of putative HIV-1 group P in Cameroon.” In: *Journal of virology* vol.85.no.3 (Feb. 2011), pp. 1403–7.
- [5] FE McCutchan. “Global epidemiology of HIV.” In: *Journal of Medical Virology* vol.78.SUPPL. 1 (Apr. 2006), pp. 7–12.
- [6] DL Robertson, JP Anderson, et al. “HIV-1 nomenclature proposal.” In: *Science* vol.288.no.5463 (Apr. 2000), pp. 55–6.
- [7] HIV sequence database Development. *HIV-1 gene map*. Available: <http://http://www.hiv.lanl.gov/content/sequence/HIV/MAP/landmark.html>; Updated: May 2014; Accessed: November 22, 2014.

- 
- [8] Robert Koch Institut. “Schätzung der Prävalenz und Inzidenz von HIV-Infektionen in Deutschland (Stand: Ende 2013).” In: *Epidemiologisches Bulletin* vol.44 (Nov. 2014), pp. 429–37.
- [9] BM Mathers, L Degenhardt, et al. “HIV prevention, treatment, and care services for people who inject drugs: a systematic review of global, regional, and national coverage.” In: *Lancet* vol.375.no.9719 (Mar. 2010), pp. 1014–28.
- [10] S Sollai, A Noguera-Julian, et al. “Strategies for the prevention of mother to child transmission in Western countries: an update.” In: *The Pediatric infectious disease journal* vol.34.no.5 (May 2015), pp. 14–30.
- [11] UNAIDS Austria. *Country Progress Report 2012*. Available: [http://www.unaids.org/sites/default/files/country/documents/ce\\_AT\\_Narrative\\_Report.pdf](http://www.unaids.org/sites/default/files/country/documents/ce_AT_Narrative_Report.pdf); Updated: July 2014; Accessed: November 10, 2014.
- [12] RH Gray, G Kigozi, et al. “Male circumcision for HIV prevention in men in Rakai, Uganda: a randomised trial.” In: *Lancet* vol.369.no.9562 (Feb. 2007), pp. 657–66.
- [13] RF Baggaley, RG White, and MC Boily. “HIV transmission risk through anal intercourse: systematic review, meta-analysis and implications for HIV prevention.” In: *International journal of epidemiology* vol.39.no.4 (Aug. 2010), pp. 1048–63.
- [14] H Li, KJ Bar, et al. “High Multiplicity Infection by HIV-1 in Men Who Have Sex with Men.” In: *PLoS pathogens* vol.6.no.5 (May 2010), e1000890.
- [15] H Salimi, M Roche, et al. “Macrophage-tropic HIV-1 variants from brain demonstrate alterations in the way gp120 engages both CD4 and CCR5.” In: *Journal of leukocyte biology* vol.93.no.1 (Jan. 2013), pp. 113–26.
- [16] J Izopet. “Human immunodeficiency virus”. In: *Molecular Diagnostics of Infectious Diseases*. Ed. by HH Kessler. 3rd ed. Walter DeGruyter, pp. 97–112.
- [17] J Karn and CM Stoltzfus. “Transcriptional and posttranscriptional regulation of HIV-1 gene expression.” In: *Cold Spring Harbor perspectives in medicine* vol.2.no.2 (Feb. 2012), a006916.

- 
- [18] CDC Centers for disease control and prevention. *Laboratory Testing for the Diagnosis of HIV Infection: Updated Recommendations*. Available: <http://www.cdc.gov/hiv/pdf/HIVtestingAlgorithmRecommendation-Final.pdf>; Updated: June 2014; Accessed: November 22, 2014.
- [19] G Herold. “HIV Infektion [Z21] und AIDS [B24]”. In: *Innere Medizin*. Ed. by G Herold. Edition of 2014, pp. 895–903.
- [20] FM Hecht, MP Busch, et al. “Use of laboratory tests and clinical symptoms for identification of primary HIV infection.” In: *AIDS* vol.16.no.8 (May 2002), pp. 1119–29.
- [21] “HIV Classification: CDC and WHO Classification systems”. In: *Guide for HIV/AIDS Clinical Care*. Ed. by S Coffey. Available: <http://aidsetc.org/guide/hiv-classification-cdc-and-who-staging-systems>; Published: April 2014; accessed: November 22, 2014.
- [22] DAIG Deutsche Aids-Gesellschaft ev. *Deutsch-Österreichische Leitlinien zur antiretroviralen Therapie der HIV-Infektion*. Available: [http://www.daignet.de/site-content/hiv-therapie/leitlinien-1/Deutsch\\_Osterreichische%20Leitlinien%20zur%20antiretroviralen%20Therapie%/der%/HIV\\_Infektion.pdf](http://www.daignet.de/site-content/hiv-therapie/leitlinien-1/Deutsch_Osterreichische%20Leitlinien%20zur%20antiretroviralen%20Therapie%/der%/HIV_Infektion.pdf); Updated: May 2014; Accessed: November 28, 2014.
- [23] WR Fleischmann Jr. “Viral Genetics”. In: *Medical Microbiology*. Ed. by S Baron. 4th ed. University of Texas Medical Branch at Galveston, 1996.
- [24] GK Hightower, SJ May, et al. “HIV-1 clade B pol evolution following primary infection.” In: *PloS one* vol.8.no.6 (Jan. 2013), e68188.
- [25] DR Kuritzkes, RM Grant, et al. “Performance characteristics of the TRUGENE HIV-1 Genotyping Kit and the Opengene DNA Sequencing System.” In: *Journal of clinical microbiology* vol.41.no.4 (Apr. 2003), pp. 1594–9.

- [26] SL Kosakovsky Pond, D Posada, et al. “An evolutionary model-based algorithm for accurate phylogenetic breakpoint mapping and subtype prediction in HIV-1.” In: *PLoS computational biology* vol.5.no.11 (Nov. 2009), e1000581.
- [27] K Tamura and M Nei. “Estimation of the number of nucleotide substitutions in the control region of mitochondrial DNA in humans and chimpanzees.” In: *Molecular biology and evolution* vol.10.no.3 (May 1993), pp. 512–26.
- [28] AJ Drummond and A Rambaut. “BEAST: Bayesian evolutionary analysis by sampling trees.” In: *BMC evolutionary biology* vol.7 (Jan. 2007), p. 214.
- [29] DL Ayres, A Darling, et al. “BEAGLE: an application programming interface and high-performance computing library for statistical phylogenetics.” In: *Systematic biology* vol.61.no.1 (Jan. 2012), pp. 170–3.
- [30] AJ Drummond, GK Nicholls, et al. “Estimating mutation parameters, population history and genealogy simultaneously from temporally spaced sequence data.” In: *Genetics* vol.161.no.3 (July 2002), pp. 1307–20.
- [31] VN Minin, EW Bloomquist, and MA Suchard. “Smooth skyride through a rough skyline: Bayesian coalescent-based inference of population dynamics.” In: *Molecular biology and evolution* vol.25.no.7 (July 2008), pp. 1459–71.
- [32] AJ Drummond, MA Suchard, et al. “Bayesian phylogenetics with BEAUti and the BEAST 1.7.” In: *Molecular biology and evolution* vol.29.no.8 (Aug. 2012), pp. 1969–73.
- [33] AM Dennis, S Hué, et al. “Phylogenetic insights into regional HIV transmission.” In: *AIDS* vol.26.no.14 (Sept. 2012), pp. 1813–22.
- [34] JL Aldous, S Kosakovsky Pond, et al. “Characterizing HIV transmission networks across the United States.” In: *Clinical infectious diseases* vol.55.no.8 (Oct. 2012), pp. 1135–43.

- [35] European Union Agency for Fundamental Rights. *European union lesbian, gay, bisexual and transgender survey- Results at a glance*. Available: [http://fra.europa.eu/sites/default/files/eu-lgbt-survey-results-at-a-glance\\_en.pdf](http://fra.europa.eu/sites/default/files/eu-lgbt-survey-results-at-a-glance_en.pdf); Published: 2013; Accessed: February 01, 2016.
- [36] The EMIS Network. *EMIS 2010: The European Men-Who-Have-Sex-With-Men Internet Survey. Findings from 38 countries*. Available: [http://www.emis-project.eu/sites/default/files/public/publications/emis-2010\\_european\\_msm\\_internet\\_survey\\_38\\_countries\\_v5.pdf](http://www.emis-project.eu/sites/default/files/public/publications/emis-2010_european_msm_internet_survey_38_countries_v5.pdf); Updated: June 12, 2013; Accessed: February 01, 2016.

# List of Figures

1	Phylogenetic tree of HIV-1. . . . .	3
2	Organization of the HIV-1 virion. . . . .	5
3	Key aspects of the HIV life cycle. . . . .	8
4	Distribution of HIV-1 subtypes / CRFs in the study population. . . . .	25
5	Self-reported transmission risk for males. . . . .	27
6	Self-reported subcategory of HSX transmission for males. . . . .	28
7	Self-reported transmission risk for females. . . . .	28
8	Self-reported subcategory of HSX transmission for females. . . . .	29
9	Distribution of the HIV-1 subtypes and CRFs by place of residence. . . . .	29
10	Distribution of risk factors for HIV-1 infection by place of residence. . . . .	30
11	The inferred transmission network according to self-reported risk factors for HIV-1 infection. . . . .	32
12	The inferred transmission network according to the place of residence. . . . .	33
13	Comparison of clustering and non-clustering individuals according to their ethnic origin. . . . .	34
14	Distribution of the HIV-1-subtypes/ CRFs in the study population. . . . .	36
15	Time-scaled phylogenetic reconstruction of the clustering individuals. . . . .	38
16	Change in the prevalent mode of transmission between 2008 and 2014 in Southeastern Austria. . . . .	39

# List of Tables

1	Classification for HIV-infected adults and adolescents. . . . .	10
2	Baseline demographic characteristics. . . . .	24
3	Baseline viral characteristics. . . . .	26
4	Self- reported risk factors for HIV-1 . . . . .	26
5	Baseline characteristics by cluster size. . . . .	31