

# Looking at HIV-1 Transmission in Quebec Through a Phylogenetic Lens

by

**Ernesto Armando Cuadra Foy**

Department of Microbiology and Immunology

McGill University, Montreal, Canada

January 2020

A thesis submitted to McGill University in partial fulfillment  
of the requirements of the degree of  
Master of Science

© Ernesto Armando Cuadra Foy, 2020

## Contents

Abstract.....	4
Résumé .....	6
Acknowledgments.....	8
Chapter 1 – Introduction.....	13
1.1 Origins of the AIDS epidemic and HIV discovery .....	13
1.2 Diversity in HIV .....	13
1.3 HIV-1 Genome.....	14
1.4 HIV-1 Structure and Life Cycle .....	14
1.4.2 Binding and entry .....	16
1.4.3 Reverse Transcription and Integration .....	16
1.4.3 Viral Replication .....	18
1.4.4 Host-mediated translation of HIV-1 proteins .....	18
1.4.5 Viral Assembly, Budding and Maturation .....	19
1.5 HIV-1 Infection and Treatment .....	20
1.5.1 Integrase Inhibitors .....	22
1.5.2 Integrase Drug Resistance Mutations and the development of Second Generation INSTIs .....	24
1.6 – Viral Phylodynamics of HIV .....	28
1.6.2 HIV Transmission Networks .....	29
1.6.3 Reconstruction of Networks and Epidemiological Assumptions .....	30
1.7 Transmission/Founder Virus.....	31
1.8 The HIV-1 Epidemic in Quebec.....	32
Chapter 2 – Rationale, Hypotheses, and Objectives.....	34
2.1 Objectives.....	35
Chapter 3 – Materials and Methods .....	35
3.1 Data Analysis .....	35
3.3 Tropism Assay .....	35
3.3 Washout of Second-Generation INSTIs.....	36
Chapter 4 – Results .....	38
4.1 Reconstruction of the Dynamics of Spread of HIV-1 .....	38
4.1.2 Age Differences Between Classes .....	41
4.1.3 Characterization of Large Clusters .....	45
4.1.4 Transmission of Drug-Resistant Mutations.....	47

4.2 Tropism of T/F Viruses and their Viral Replication .....	49
4.3 Susceptibility of Second-Generation INSTIs on C185 Isolates .....	52
4.3.1 Determining Inhibitory Concentration Values .....	55
4.3.2 Determining effects of INSTI washout on NL4.3 and C185 isolates.....	57
Chapter 5 – Discussion .....	61
5.1 Small Networks Model Can Explain HIV-1 Expansion and Clustering.....	61
5.3 Large Cluster Species Maintain Similar Tropism and Replicative Capacity.....	64
5.4 Cluster 185 Viruses Show Decreased Susceptibility to Second-Generation INSTIs .....	66
Chapter 6 – Conclusions.....	70
References .....	72

## Abstract

**Introduction:** The year 2020 will mark 40 years since AIDS was first considered a global epidemic. An ambitious program by UNAIDS, called 90-90-90 initiative, was put in place in 2013. This initiative specifies that by 2020, there will be a significant decrease in HIV incidence. The objectives of the 90-90-90 program will not be realized in this time frame, and it is unlikely that progress can be made unless there can be an understanding of the transmission networks by which the virus continues to propagate.

**Objectives:** A) Analyze the dynamics of spread of the HIV-1 MSM epidemic in Quebec using phylogenetic and clinical data. B) To determine the tropism of Transmission/Founder viruses from HIV infected patients. C) Determine the susceptibility of Cluster 185 (C185) viruses on integrase strand transfer inhibitors: Dolutegravir (DTG), Cabotegravir (CAB), and Bictegravir (BIC).

**Methods:** Statistical analysis of datasets from the genotyping program was performed using SAS and Python. Tropism was determined by infecting U87 astrogloma cells expressing CD4 and CCR5 or CXCR4 co-receptors with HIV-1 primary isolates. Viral replication was monitored by measuring reverse transcriptase (RT) levels. MT-2 cells were treated with DTG, CAB or BIC were infected first with HIV lab strains NL4.3<sub>WT</sub>, NL4.3<sub>IN(R263K)</sub>, NL4.3<sub>IN(G118R)</sub>, or NL4.3<sub>IN(G140S/Q148H)</sub>. Similarly, this procedure was conducted on C185 viruses C185.09<sub>WT</sub>, C185.15<sub>WT</sub>, and their respective R263K mutations. Drugs were removed from the media on the peak day of infection. Viral replication was monitored by measuring RT and integration was measured by qPCR.

**Results:** Out of 6,600 patients, 41% were found to be part of a large cluster class (LC, 6 or more paired sequences), compared to 37% belonging to singletons (no pairing). Singletons were observed prominently during the 2002-2009 period; however, large cluster cases increased by 60% between 2010-2017. LC viruses were found to be significantly more recent infections ( $p < 0.0001$ ) than singletons, showing a percent median diversity of 0.4% (IQR 0.2-1.3). Age differences also were noted between the classes. The median age of LC members appeared to vary equally between 28 and 48 years old, whereas the median age for singleton patients was 48. Analysis of the median age per year as a function of year of visit showed a negative correlation ( $r^2 = -0.488$ ) for LC patients compared to a positive correlation for singletons patients ( $r^2 = 0.315$ ). This was also observed on their clinical status, a strong negative correlation ( $r^2 = -0.840$ ) was noted for patients who had been recently infected compared to a positive correlation ( $r^2 = 0.646$ ) for patients who had been infected longer than six months. It was also found that 50% of patients belonging to an LC class showed some form of drug-resistance related polymorphism. The tropism assay did not reveal any striking differences between LC viruses and singletons, showing mostly R5 tropism. It was noted that 68% of LC viruses had a better replication capacity than singletons. C185 isolates preserved the same dual-tropic properties and were selected for further experiments. The drug washout experiments reveal that DTG and BIC outperformed CAB in the highly resistant variant NL4.3<sub>IN(G140S/Q148H)</sub>. The

C185.15<sub>IN(R263K)</sub> virus could rebound under BIC treatment and showed 15-fold resistance. The presence of natural polymorphism M50I could explain these observations.

**Conclusions:** The fast spread of LC viruses can be attributed to quick transmission events happening in a determined local structure that enhances the ability of infection. It is possible that the mixing of susceptible populations might be age-dependent, as noted in some of the yearly trends. LC viruses clearly show more selective advantage than singletons, including lower susceptibility to some INSTI, higher replicative capacity, and increased drug-resistance transmission.

## Résumé

**Introduction:** L'année 2020 marquera le quarantième anniversaire depuis que le SIDA fut considéré une pandémie. Une Initiative de l'ONUSIDA nommée 90-90-90, fut mis en place en 2013. Cette initiative précise que d'ici 2020, l'incidence de nouveaux cas de VIH diminuera considérablement. Les objectifs du programme 90-90-90 visiblement ne seront pas atteints, et il est peu probable que des progrès puissent être accomplis sans une compréhension des réseaux de transmission par lesquels le virus continue de se propager.

**Objectifs:** A) Analyser la dynamique de propagation de l'épidémie de VIH-1 HSH au Québec à l'aide de données phylogénétiques et cliniques. B) Déterminer le tropisme des virus de transmission / fondateur provenant de patients infectés par le VIH. C) Déterminer la sensibilité des virus du groupe 185 (C185) aux inhibiteurs de transfert de l'intégrase: dolutegravir (DTG), cabotegravir (CAB) et Bictegravir (BIC).

**Méthodes:** L'analyse statistique des jeux de données du programme de génotypage a été réalisée à l'aide de SAS et Python. Le tropisme a été déterminé en infectant des cellules d'astroglome U87 exprimant les co-récepteurs CD4 et CCR5 ou CXCR4 avec des isolats primaires de VIH-1. La réplication virale a été contrôlée en mesurant les taux de transcriptase inverse (RT). Les cellules MT-2 traitées avec DTG, CAB ou BIC ont d'abord été infectées avec les souches VIH NL4.3WT, NL4.3IN (R263K), NL4.3IN (G118R) ou NL4.3IN (G140S / Q148H). De manière similaire, cette procédure a été réalisée sur les virus C185 C185.09WT, C185.15WT et leurs mutations respectives R263K. Les médicaments ont été retirés des médias le jour de pointe de l'infection. La réplication virale a été contrôlée en mesurant la RT et l'intégration par qPCR.

**Résultats:** Sur 6 600 patients, 41% appartenaient à une grande classe de cluster (LC, 6 séquences appariées ou plus), contre 37% appartenant à des singletons (aucun appariement). Les singletons ont été observés de manière évidente au cours de la période 2002-2009; Cependant, les cas de grands clusters ont augmenté de 60% entre 2010 et 2017. Les virus à LC étaient significativement plus récents ( $p < 0,0001$ ) que les singletons, montrant une diversité médiane de 0,4% (IQR 0,2-1,3). Des différences d'âge ont également été notées entre les classes. L'âge médian des membres des LC semblait également varier entre 28 et 48 ans, alors que l'âge médian des patients célibataires était de 48 ans. L'analyse de l'âge médian par an en fonction de l'année de la visite a montré une corrélation négative) pour les patients LC, comparée à une corrélation positive pour les patients uniques ( $r^2 = 0,315$ ). Cela a également été observé sur leur état clinique, une forte corrélation négative ( $r^2 = -0,840$ ) a été notée pour les patients récemment infectés par rapport à une corrélation positive ( $r^2 = 0,646$ ) pour les patients infectés plus de six mois. Il a également été constaté que 50% des patients appartenant à une classe de LC présentaient une certaine forme de polymorphisme lié à la résistance aux traitement antiviraux. Le test de tropisme n'a révélé aucune différence frappante entre les virus LC et les singletons, montrant principalement le tropisme R5. Il a été noté que 68% des virus LC avaient une meilleure capacité de réplication que les singletons. Les isolats

C185 ont conservé les mêmes propriétés bi-tropiques et ont été sélectionnés pour des expériences ultérieures. Les expériences d'élimination du médicament ont révélé que le DTG et le BIC étaient plus performants que le CAB dans le variant hautement résistant NL4.3IN (G140S / Q148H). Le virus C185.15IN (R263K) pourrait rebondir sous traitement par BIC et présenter une résistance de 15 fois. La présence du polymorphisme naturel M50I pourrait expliquer ces observations.

**Conclusions:** La propagation rapide des virus du type LC peut être attribuée à des événements de transmission rapides se produisant dans une structure locale déterminée qui améliore la capacité d'infection. Il est possible l'âge au moment de l'infection soit un facteur important à considérer, comme indiqué dans certaines des tendances annuelles. Les virus de LC montrent clairement un avantage plus sélectif que les singletons, notamment une sensibilité plus faible à certains INSTI, une capacité de réplication plus élevée et une transmission accrue de la résistance aux antiviraux.

## List of Abbreviations

Abbreviation	Full Term
AIDS	Acquired immunodeficiency syndrome
ART	Antiretroviral therapy
BIC	Bictegravir
CAB	Cabotegravir
CCA	Capsid domain
CCR5	C-C chemokine receptor type 5
cDNA	Complementary DNA
CRF	Circulating recombinant forms
CT	Chronic treated
CUN	Chronic untreated
CXCR4	C-X-C chemokine receptor type 4
DDE	D64, D116, E152
DKA	$\alpha$ , $\gamma$ -diketoacids
DMEM	Dulbecco's modified eagle medium
DTG	Dolutegravir
ECDF	Empirical cumulative distribution functions
EFV	Efavirenz
Env	Envelope glycoprotein
ETR	Etravirine
EVG	Elvitegravir
Gag	Group-specific antigen
gRNA	Genomic RNA
HAART	Highly active antiretroviral therapy
HET	Heterosexual Transmission
HIV-1	Human immunodeficiency virus
IC <sub>50</sub>	Half maximal inhibitory concentration
IN	Integrase
INSTI	Integrase strand transfer inhibitors
IN-vDNA	Integrase-viral DNA
IRES	Internal ribosomal entry site
KDP	Kernel density plot
LC	Large clusters
MA	Membrane binding domain
MOI	Multiplicity of infection
MSM	Men-Having-Sex-With-Men
NC	Nucleocapsid domain
NNRTI	Non-nucleoside reverse transcriptase inhibitors
NPC	Nuclear pore complexes
NRTI	Nucleoside reverse transcriptase inhibitors
NSI	Non-syncytia induction
NVP	Nevirapine
PBMC	Peripheral blood mononuclear cells
PCR	Polymerase chain reaction
PHA-P	Phytohaemagglutinin P
PHI	Primary HIV infections

PI	Protease inhibitors
Pol	Polymerase
PR	Viral protease
PWID	People Who Inject Drugs
qPCR	Quantitative PCR
RAL	Raltegravir
RPV	Rilpivirine
RT	Reverse transcriptase
SC	Small clusters
SI	Syncytia inducing
SIR	Susceptible-Infectious-Recovered
SIV	Simian immunodeficiency virus
STD	Sexually transmitted diseases
T/F Virus	Transmission/Founder Virus
TAR	Transactivation-responsive region
U87	Human primary glioblastoma cell line
Vpr	Viral protein R
vRNA	Viral RNA
WT	Wild type

## Acknowledgments

I still remember those early days when I before I joined the Lady Davis Institute. It was April 2017, a student from the lab had previously sent me a text message explaining to me the unbelievable news and told me to think thoroughly about the next steps I should take since I got recently admitted to the Microbiology and Immunology program. It hadn't struck me particularly hard until I saw it in the news that day while I was at Station Vendome looking for a new supervisor. I looked up at the monitors in the metro, Mark Wainberg, aged 71, had passed away in an accident while on vacation in Florida. To say that this was an unfair turn of events would be an understatement. Previously Dr. Wainberg and I met and had a good discussion about all the different projects I could undertake and his interest in my work with the Canadian Forces Primary Reserve. He was supportive of my involvement with the Reserves and told me not to worry about starting in September, as I was planning to leave for Nova Scotia in May to complete my training as an Intelligence Operator. Short of time, I asked Dr. Bluma Brenner if she could take me into her group and complete my master's degree under her supervision. I don't think that I am able to convey enough words in this section to thank her for allowing me to work alongside her and her team. Through her vision and dedication, I have been able to learn a great deal not only about HIV but also how to be a scientist and a better person. Thank you very much, sincerely, from the bottom of my heart.

Dr. Brenner wouldn't be able to get the job done if it weren't for the people that work behind the scenes. I would like to thank Illinca Ibanescu and Maureen Oliveira for showing me the ropes in the early days and to have patience as I was struggling to find my way through my project.

I would also like to thank Nathan Osman for being a supportive sidekick and friend during my time at the lab. Most people found our strange sense of humour baffling, but I knew that it would be the ideal catalyst for our partnership. I am thankful for your selfless dedication in helping me build my project and showing me how to be an inquisitive and resourceful student. Looking forward to calling you Dr. Osman (ironically).

I don't want to admit that the presence of a fellow Peruvian is the reason why I joined the lab, however I think it might have had an impact in making my experience more enjoyable. I thank Cesar for his support and all the good talks we've had about how much of a circus our homeland still is after all these years.

To Drs. Chen Liang and Anne Gatignol for acting as committee members for my project. Thank you for your direction and advice during my project.

Completing my degree would not have been possible without the financial support of the Microbiology and Immunology Department during the first year of my degree. Aspects of research in the thesis were conducted with funds that Dr. Brenner received from the Fonds de Recherche du Quebec (FRQS) – Reseau SIDA, Gilead Sciences and the Canadian Institutes of Health Research.

A special thanks go to Rebecca and Miko Robertson for their moral support throughout these two years. They are the pillars at home that keep me sane and are my inspiration for working towards my goals. Miko in particular who asks to be patted while I'm trying to type. He knows what priorities are (He is a cat, by the way).

Lastly, none of this would've happened without the unconditional support and love of my parents. I am always thinking of the sacrifice they did 15 years ago when they left behind jobs, friends, and family to allow me and my brothers and sister to have a future they could only dream of. Muchas gracias papa y mama, los quiero mucho.

## Chapter 1 – Introduction

### 1.1 Origins of the AIDS epidemic and HIV discovery

The origins of the acquired immunodeficiency syndrome (AIDS) pandemic can be traced back to the summer of 1981 when the disease was first recognized in young homosexual men who were falling ill and then succumbing to unusual opportunistic infections (1, 2). Doctors inferred that some sexually transmitted infectious agent had a critical role in the pathogenesis of this immunodeficient state (3). Two years later, it was found that the virus belonged to a family of T-lymphotropic retroviruses that are horizontally transmitted in humans (4). Since its discovery as the cause of AIDS, the human immunodeficiency virus (HIV) has spread globally, killing an estimated 25 million people and currently infecting 33 million people living with the virus (5). According to the World Health Organization (WHO), about 1.8 million new cases of infection were reported in 2016.

### 1.2 Diversity in HIV

HIV is separated into two major types. HIV subtype 1 is responsible for the ongoing epidemic, while HIV subtype 2 is considered to have lower pathogenic potential and not recognized outside of Africa. These two viruses belong to the *Lentivirus* genus (6). Zoonotic transmissions of the simian immunodeficiency virus (SIV) from non-human primates to humans in West and Central Africa are believed to be the origin of the virus, and it is also these transmission events that are responsible for the creation of HIV-1 groups M, N, O, and P (7). One of the most important factors in spreading HIV is the ability to diversify thanks in part due to the lack of a proof-reading mechanism in RNA polymerase reverse transcriptase (RT) enzyme (8). Viruses from group M account for more than 90% of HIV infections worldwide. Furthermore, group M viruses are divided into subtypes (A-D, F-H, J, and K) along with 55 circulating recombinant forms (CRFs) (9).

Subtypes can also be linked geographically or epidemiologically. For example, viruses belonging to the B subtype are predominant in patients within North American and Western European populations, while CRF14\_BG subtype is found among injection-drug users in Spain and Portugal. These differences in viral subtypes may influence transmission, disease progression, and treatment effectiveness (10).

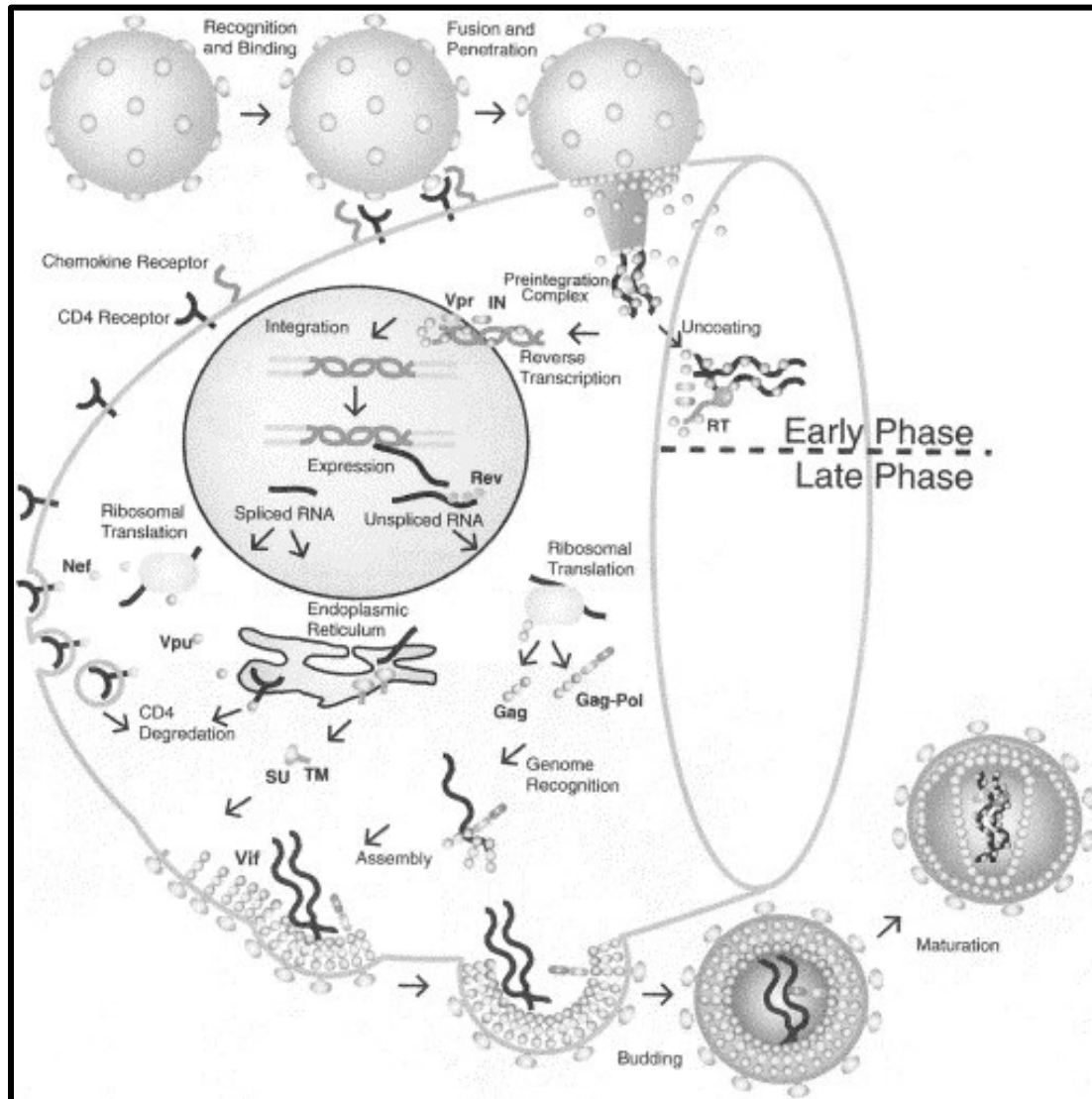
### 1.3 HIV-1 Genome

HIV-1 virus belongs to the *Retroviridae* family showing a linear, single-stranded, and non-segmented genome of positive polarity. Viruses from this family contain at least three genes: *gag* (group-specific antigen), *pol* (polymerase), and *env* (envelope glycoprotein), forming the body of structural proteins(11). The HIV-1 genome is a primarily coding RNA of 9 kb in length, which contains nine open reading frames. These reading frames generate several proteins that assist in transcription, genomic dimerization, packaging, RNA nuclear export, signal polyadenylation, and interact with viral and host proteins (12). Additionally, HIV contains a set of regulatory proteins: *tat* and *rev*, necessary for transcriptional activity. A group of genes called “accessory” proteins are also included in the genome, which at the time of their discovery were considered dispensable, and only recent research has started to show interesting properties for crucial functions *in vivo* (13).

### 1.4 HIV-1 Structure and Life Cycle

The HIV replication process is divided into “early” and “late” phases. During the early events, binding to target cells occurs as well as fusion, reverse transcription activity of RNA into DNA, and nuclear import of the provirus into the nucleus. On late events, integration in the nucleus

takes place and is followed by the production of viral proteins and their assembly, leading to maturation of the virion and budding (14).



**Figure 1: Retrovirus Life Cycle.** The viral core enters by virus-cell membrane fusion. From genomic RNA (gRNA), double-stranded DNA copies are synthesized by virion-associated reverse transcriptase. This core is transported to the nucleus, and the provirus is formed by integration into the host genome, completing the early phase of infection. Viral RNA (vRNA) can be expressed from the provirus leading to the synthesis of viral proteins on host ribosomes. Progeny virion particles are formed by budding through the membrane of infected cells (15).

#### 1.4.2 Binding and entry

HIV-1 *env* gene encodes for the trimeric glycoprotein gp120/gp41, which targets CD4<sup>+</sup> lymphocytes and whose binding activity establishes infection (16). Comparisons between different gp120 sequences have allowed the identification of five variable regions (V1-V5) forming surface exposing loops that contain disulphide bonds at their bases. Conformational changes in gp120 ensue, involving the exposure of a binding site for specific chemokine receptors, which serve as obligate coreceptors for virus entry (17). Viral strains use either the CCR5 and or CXCR4 chemokines as coreceptors, with the V3 loop acting as an important determinant of HIV-1 tropism (18). Gp120 then dissociates from gp41, who promotes fusion between viral and cellular membranes resulting in the release of viral contents into the cell (19). Following the entry into the cytoplasm, the virus undergoes a process known as uncoating. In this process, the capsid composed of viral proteins capsid (CA) and nucleoprotein (NC) must deliver the reverse transcriptase (RT), integrase (IN), and the viral genome to the nuclear pore of the host. Premature uncoating or unstable capsid mutants have shown to cause abortive infection (20).

#### 1.4.3 Reverse Transcription and Integration

HIV contains a single-stranded RNA genome that requires reverse transcription to a double-stranded DNA to produce more copies of itself. This step is catalyzed by RT, which plays a central part in HIV replication (21). HIV-1 RT has been characterized as a heterodimer consisting of 66 and 51 kDa chains performing polymerization and RNase H activities, respectively (22). RT activity occurs within the capsid structure as uncoating occurs. A primer binding site (PBS) on the viral genome neighbouring a tRNA acts as the primer for DNA synthesis. RT starts transcription from the U5 long terminal repeat (LTR) forming a minus-strand DNA, while RNase

H degrades the viral RNA template until reaching the tRNA region, exposing the minus DNA single strand. The repeat (R) region on the 5' end of the minus DNA strand then hybridizes to the 3' R region of the RNA template. Once this first strand transfer is completed, RT proceeds to full-length genome transcription. RNase H degradation degrades RNA from the hybrid except for the poly-purine tract (PPT) region that is resistant to degradation. DNA synthesis of the plus strand continues using the ppt-DNA hybrid as a template. Once the U3 terminal is reached, RNase H proceeds to degrade PPT and the tRNA. The PBS primer then proceeds to hybridize with the complementary sequence of the new DNA strand, causing a second strand transfer. These minus- and plus-DNA strands are extended and create the full-length viral cDNA genome (23). Once the transcription is completed, IN proceeds to hydrolyze extremities of viral DNA to generate reactive  $\text{CA}_{\text{OH}}\text{-}3'$  ends. The combination of the cDNA, viral protein R (Vpr), IN, matrix protein (MA), and a triple-stranded cDNA flap create the HIV-1 pre-integration complex (PIC) (24). The PIC, which carries the viral cDNA, needs to be actively imported into the nucleus, as it is significantly larger than the nuclear pore complexes (NPC) (28 nm vs. 9 nm) (25). Through a combination of viral and host factors that remain incompletely understood, the PIC is capable of associating with NPCs and cause uncoating of CA, followed by the release of IN-associated cDNA into the nucleus (26). The first step to integration occurred in the PIC with the 3'-end processing. For the second step, IN catalyzes DNA strand transfer reaction when 3'-ends of viral DNA covalently bind a pair of phosphodiester bonds in host DNA through a nucleophilic attack. Two unpaired 5'-ends on the viral DNA are removed and then integration is completed by cellular repair enzymes. Incorporation of viral DNA takes place preferably within intronic regions of actively transcribed genes. This integrated genome is called the provirus (27).

#### 1.4.3 Viral Replication

Once the provirus has been integrated, a latency stage sets in following acute infection where it remains transcriptionally silent. This is due in part to the nature of CD4<sup>+</sup> T cells, which remain dormant until an antigen is encountered. Methylation of the provirus can also further improve latency (28). HIV-1 can produce short mRNA that encodes the viral regulatory proteins Tat and Rev. Tat activates transcription by stimulating elongation from the LTR. LTR acts as the viral promoter, the transactivation-responsive region (TAR) is found downstream from the initiation site and acts as the binding site for Tat (29). Tat can then recruit the positive transcription elongation factor (P-TEFb) CDK9/CycT1, followed by phosphorylation of the RNA polymerase II (RNAP II) C-terminal domain. This hyperphosphorylated RNAP II allows enhanced transcription of the full HIV-1 genome (29, 30). Pre-RNA transcripts that were generated are then spliced. The catalytic reaction is performed by spliceosomes that contain six small nuclear ribonucleoproteins (snRNPs) U1 to U6. Over 40 mRNA species are produced as a result of this reaction. However, about half of these transcripts are not spliced and become genomic RNA used for replication and integration. (31, 32). Only spliced mRNA can be exported through normal nuclear pores. Interaction between the RNA-binding Rev protein and its site the Rev response element (RRE) allows for the export of viral mRNA and pre-mRNA to the cytoplasm. This RRE site is present in all unspliced and partially spliced HIV-1 mRNA species (33), making the Rev-RRE complex capable of being exported to the cytoplasm.

#### 1.4.4 Host-mediated translation of HIV-1 proteins

HIV-1 mRNA is divided into three groups: 1) full-length transcripts that are unspliced and encode the Gag and Gag-Pol polyproteins 2) Singly spliced transcripts responsible for generating proteins Env, Vif and Vpu, and 3) fully spliced transcripts that express Rev, Tat, Vpr

and Nef. HIV-1 relies entirely on host cell mechanisms to translate virion proteins (34).

Translation of eukaryotic mRNAs generally involves the aggregation of ribosomal units alongside some initiation factors (eIFs) that help in the recruitment, unwinding, and translation of 5' capped mRNA (35). Since HIV-1 can produce unspliced and uncapped transcripts, cap-independent mechanisms become necessary, by which HIV-1 is capable of hijacking host translation systems using a structural element called internal ribosomal entry site (IRES), which promotes the recruitment of the initiation complex and redirect host translational machinery toward viral mRNA (36). To maximize the production of viral proteins, HIV-1 can manipulate the host lifecycle by arresting cells in the G2/M phase using Vpr, and this causes a global reduction of host protein synthesis while maintaining the translation of HIV-1 structural proteins (37).

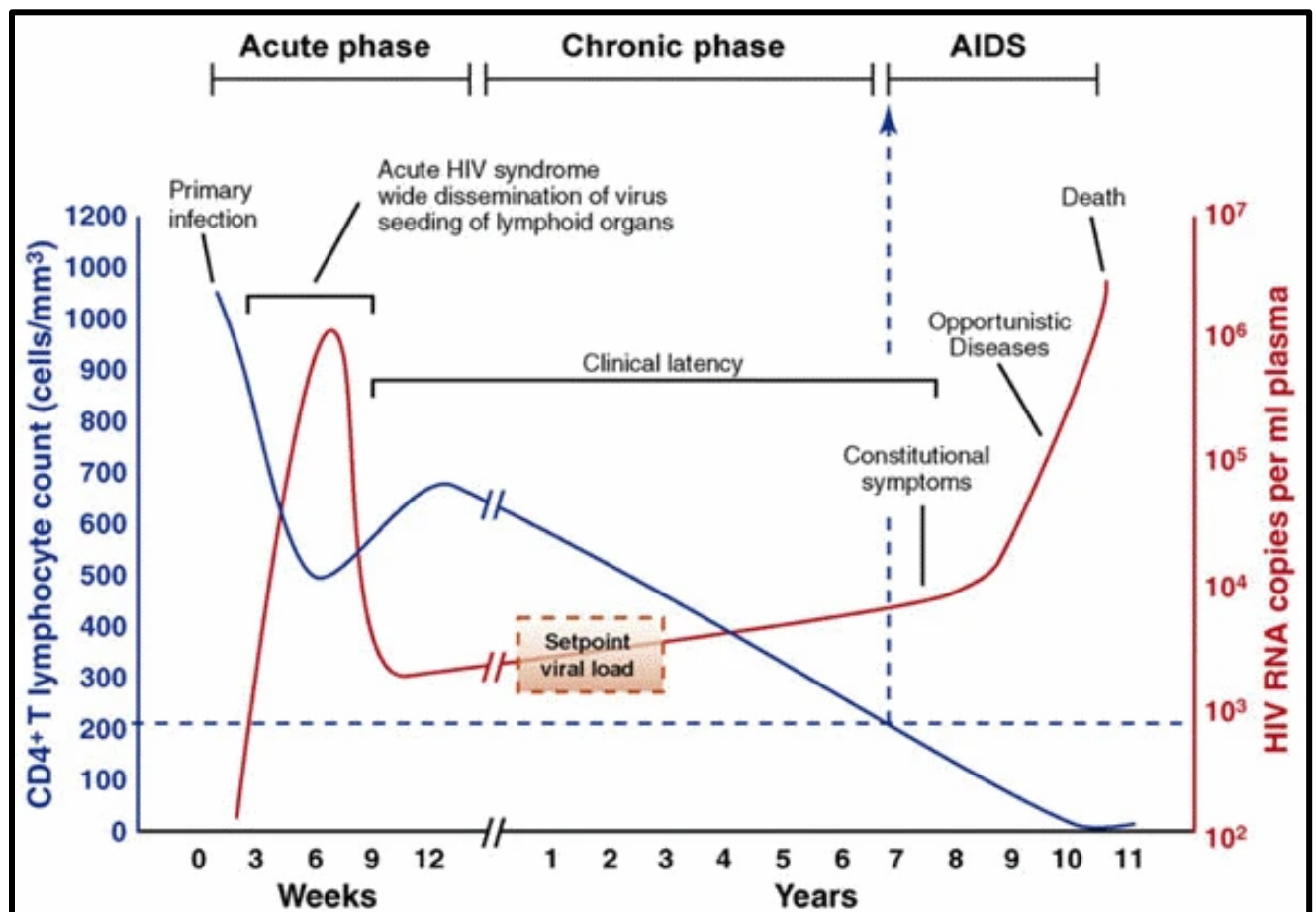
#### 1.4.5 Viral Assembly, Budding and Maturation

The formation of virions capable of carrying infection corresponds to the late stages of the virus lifecycle. These three stages are all coordinated by the Gag polyprotein and its proteolytic maturation products (38). The Gag polyprotein is composed of three folded polypeptides (MA, the membrane-binding domain, CA, the capsid domain and NC, the nucleocapsid domain) and three smaller peptides (SP1, SP2, and p6). Genomic RNA (gRNA) recruitment by Gag is performed before assembly, possibly acting as a seed for an assembly site (39). During budding, viral protease (PR) cleaves the Gag and Gag-Pro-Pol polyproteins at ten sites. These morphological changes are required for the creation of infectious virions (40). Following maturation, vRNA is condensed with CA forming a conical core. MA protein remains associated with the viral envelope(41).

### 1.5 HIV-1 Infection and Treatment

CD4<sup>+</sup>T lymphocytes are the major target for HIV-1 infection (42). Other targets include macrophages and dendritic cells, which can contribute to the establishment of the viral reservoir (43). HIV-1 infection response consists of three stages: initial acute/primary infection (0-6 months), a long asymptomatic period (1-9 years), and a final symptomatic period causing a collapse in healthy CD4<sup>+</sup>T cell counts and the onset of AIDS (44). Early (acute) stages of HIV are associated with high transmission rates and viral loads (45), usually occurring within 2 to 4 weeks after infection. During this time, HIV remains hidden from detection for a duration of 7 to 21 days in what is called an “eclipse phase,” whose detection is only possible through the usage of fourth-generation tests (46). Following primary HIV infection (6 weeks to 6 months), the patients enter the asymptomatic stage of the infection (viral set-points below 10,000 copies/mL), also known as chronic HIV infection. At this point, infected patients may not show any HIV-related symptoms for many years while the virus continues to replicate slowly in the body. People who take antiretroviral therapy (ART) can stay at this stage for decades, to the point of becoming virally suppressed and less likely to transmit HIV (47). At the AIDS stage, HIV has severely damaged the immune system, and the body can become the target of opportunistic infections and infection-related cancers. (48). As CD4 count drops the body's ability to renew T-cells decreases while antigen-experienced CD8 T-cells increase. These early-differentiated CD8 T-cells maintain the ability to proliferate and offer protective immunity, however as they become highly differentiated they start to exhibit characteristics associated with replicative senescence and are unable to continue proliferation (49). The cytotoxic potential of CD8 T-cells is capable of controlling the virus during acute infection, however as the disease progresses their potential dramatically decreases and are no longer capable of exerting

an appropriate antiviral response (50). Advances in ART development has allowed HIV-infected patients to manage the disease and maintain undetectable viral levels in plasma, thus preventing further spread of the virus. Timely diagnosis, initiation of appropriate therapy, and access to such medication could improve the lifestyle of affected patients and help mitigate the effects of the ongoing pandemic (51). Current drugs are divided into six categories: nucleoside and nucleotide reverse transcriptase inhibitors (NRTI) non-nucleoside reverse transcriptase inhibitors (NNRTI), protease inhibitors (PI), CCR5 inhibitors and fusion inhibitors (52). The rest of this work will focus primarily on integrase inhibitors.



**Figure 2: Viral load and lymphocyte during AIDS progression.** Once an individual becomes infected, HIV-1 seeks out and overtakes CD4-bearing T-lymphocytes, which begin to struggle to control viral

proliferation in the beginning weeks. More than 1 billion new HIV copies are produced each day by an acutely infected person. The acute HIV infection, also known as primary HIV infection stage, occurs as early as two of four weeks after infection and last up to six months. During the chronic phase, it can take years until HIV-1 completely depletes CD4 cells leading to AIDS and the appearance of opportunistic infections (53).

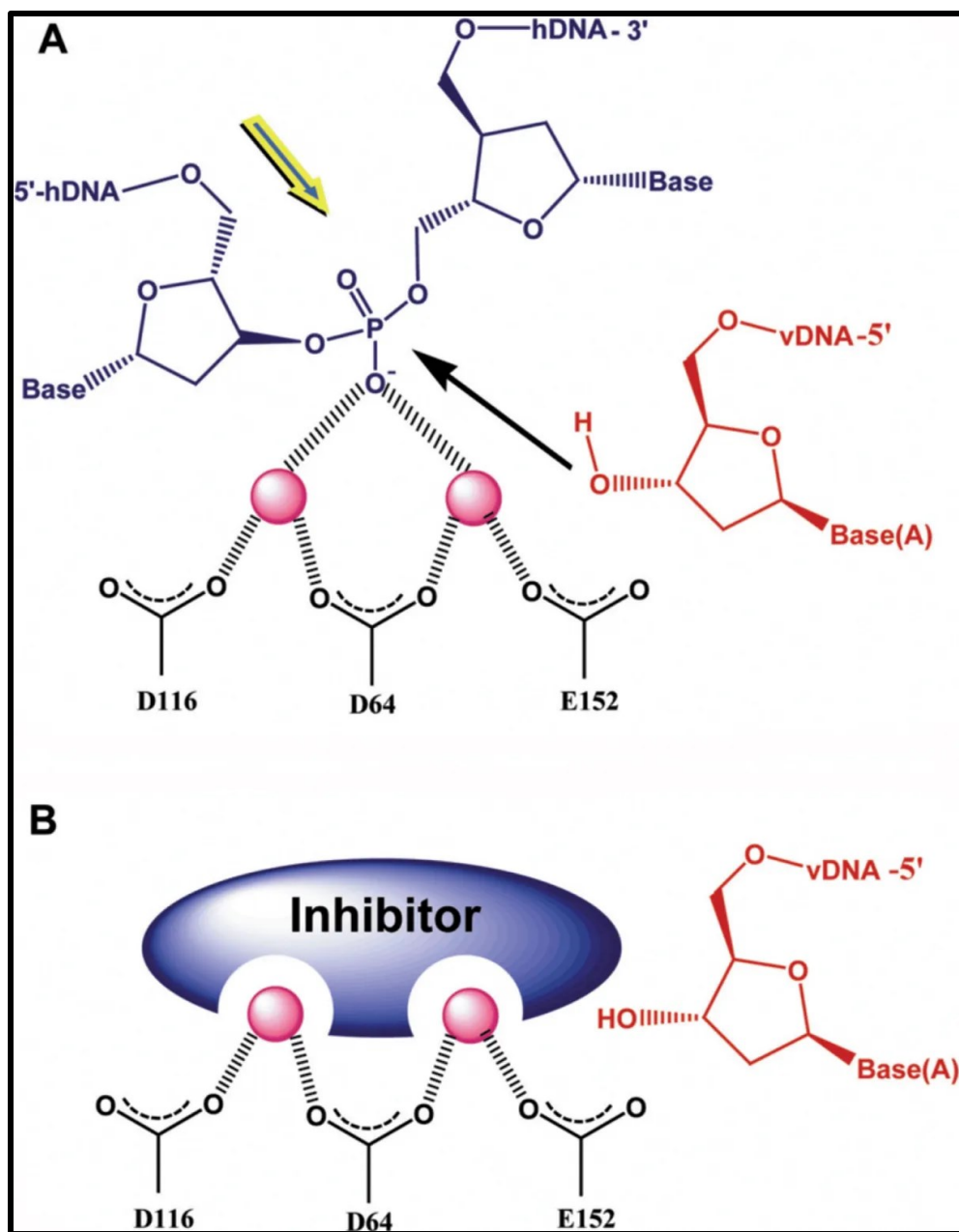
#### 1.5.1 Integrase Inhibitors

Early highly antiretroviral therapy consisted of NNRTI, NRTI, and PI. The integrase inhibitor class of drugs, introduced in 2007, is the treatment option of choice in first-line therapy.

Raltegravir was the first of its kind (MK-0518, Merck) and received approval by the FDA in 2007.

These drugs intervene during the strand transfer step of the integrase when the provirus is introduced to the host genome. As this enzyme is required for replication and is not endogenous to humans, the structure of these integrase strand transfer inhibitors (INSTI) is unlikely to cause off-target effects when compared to nucleoside derived therapies from NRTI or NNRTI (54). Integrase catalytic activity requires the presence of a metallic cationic cofactor, which is coordinated by the catalytic triad (DDE motif D64, D116, E152) within the HIV-1 integrase. Raltegravir and other  $\alpha$ ,  $\gamma$ -diketoacids (DKA) off which INSTIs are based, compete with target DNA by binding to the IN-vDNA complex. Recognition of these compounds is specific to the catalytic triad, which can efficiently chelate the  $Mg^{2+}$  cation required for integrase activity (55). These inhibitors selectively inhibit during the strand transfer step, as observed in Grobler et al., where  $IC_{50}$  values for tested DKA differed between strand transfer (100 nM) and 3' processing (5  $\mu$ M) (56). Due to the ability of HIV to mutate in the presence of ART, drug resistance is a recurrent issue that can compromise the effectiveness of therapeutic options. Thus, antiviral drug cocktails-high active anti-retroviral therapy (HAART) was conceived in hopes that with enough mutations, the virus would become unfit and reach undetectable levels throughout the patient's lifetime (57). As of 2017, resistance to NNRTI

(efavirenz/nevirapine) was significantly higher among first-line treatment initiations compared with ART drug-naïve individuals. The annual incremental increase of NNRTI resistance is expected to increase mostly in African countries as of 2017 (58). The World Health Organization (WHO) expects that combining INSTI with two NRTIs in countries with high levels of NNRTI resistance can be used as an alternative for first-line treatment (59).



**Figure 3 Strand Transfer Reaction and Mechanism of IN Inhibition by INSTIs.** Strand transfer enzymatic activity occurs through the attack of the phosphodiester backbone in target DNA by the 3' hydroxyl groups of the processed viral DNA. In this figure, host DNA and vDNA are shown in blue and red, respectively. INSTIs may chelate the two metal ions in the catalytic site of integrase, blocking the binding of host DNA (53).

1.5.2 Integrase Drug Resistance Mutations and the development of Second Generation INSTIs  
Despite RAL remarkable antiviral efficacy and minimal side effects, its usefulness tends to be

limited by the rapid occurrence of drug-resistant mutations. Primary mutations involving drug

resistance to RAL are via the N155H, Q148H/RK, and less frequent Y143/R/C/H pathways.

Q148H is generally associated with a secondary mutation at position G140S that is capable of rescuing the negative effects of the primary Q148H mutation (60). Pharmacokinetics of RAL might be the main reason for the quick appearance of resistance mutations due to its limited intestinal absorption leading to low drug levels in plasma and allowing the virus to escape drug pressure. This resistance cannot be overcome by increasing dosage (61).

The emergence of a new class of drugs pushed for the development of more efficient alternatives. Gilead developed Elvitegravir (EVG JTK-303/GS-9137) and then approved by the FDA in 2012. EVG makes use of CYP3A inhibitors acting as pharmacokinetic boosters to achieve adequate serum levels and prolongs elimination half-life from 3.6 to 9.5 hours, offering an advantage over RAL, which demands twice-daily dosing regimens compared to once-daily for EVG (62). Drug resistance profiles for the drug were identified not long before the drug was put in the market. Selective drug pressure of EVG generated T66I and E92Q drug resistance mutations after a few passages and later Q148H (63). As a result, EVG susceptibility is also diminished in patients with RAL failure. Cross-resistance between RAL and EVG is high, with drug susceptibility being more frequently reduced for EVG than for RAL (50).

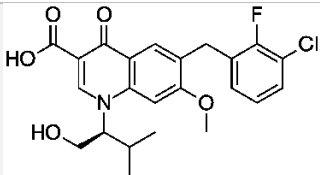
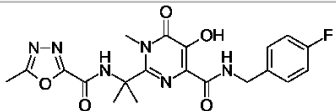
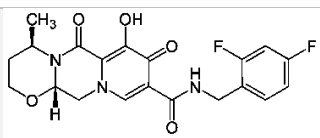
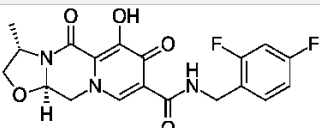
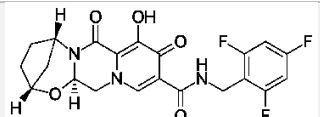
The ability of HIV to escape drug pressure remains a significant challenge even today, 30 years after zidovudine's (AZT) introduction to the market as the first therapy of its kind in the fight against the AIDS epidemic. Development of second-generation INSTIs should be focused on preventing the resurgence of cross-resistance mutations like in the case of RAL and EVG, as well as provide longer elimination half-lives without the need for adjuvants. Dolutegravir (DTG, S/GSK1349572) was the first INSTI to be recognized as a "next-generation" INSTI thanks to its

unique properties, including unboosted daily dosing, high resistance barrier and low-cross resistance to RAL and EVG mutations (64). One of the problems with the first-generation drugs was that its inhibition depended on the displacement of the viral 3'-adenosine from the active site which also correlates with slow association and dissociation, explaining the competition between the viral DNA and the drug for the active site (intasome) (65). Dolutegravir maintains a similar structure to other DKA looking to chelate  $Mg^{2+}$  cations. Improvements in these structures include a 2,4 difluorobenzyl group that occupies a pocket vacated by the adenine base of 3' terminal viral DNA that allows better displacement of this terminal nucleotide (66). Also, hydrophobic interactions of the tricyclic core from DTG are essential for the correct positioning of the drug molecule (67). The SAILING study provided proof of DTG's superiority as a safe and efficient therapy for patients that had experienced viral failure. The results from the randomized, double-blind study compared DTG with RAL on treatment-experienced, integrase-inhibitor naïve patients. After four months, 71% of those assigned DTG vs. 64% of those assigned RAL had significantly lower viral loads. The conclusions of this study showed that once-daily DTG, in combination with two other ART drugs, was better tolerated with improved virologic effect compared to the twice-daily regime of RAL (68).

Further studies were conducted on patients that had previously developed INSTI drug resistance mutations. The VIKING-3 study showed that 50 mg of DTG were efficacious in treatment-experienced populations (69). Our group previously performed *in vitro* selections using DTG on viruses of subtypes B, C, and A/G recombinants. After 20 weeks, the R263K mutation was observed in the C-terminal portion of the protein, causing a 30% decrease in the biochemical activity of integrase as well as diminishing viral fitness (70).

Today INSTIs represent the most efficient way of treating new HIV infections. Their high genetic barrier means that viruses have a low chance of developing mutations that would result in viral failure and possible transmission of resistance. Bictegravir (BIC, GS-9883) is another second-generation therapy approved in 2018 for experienced or naïve patients (71). Another investigational INSTI drug on the pipeline, developed by ViiV, couples Cabotegravir (CAB, S/GSK1265744) with rilpivirine as a monthly-injectable regime. Currently, this drug combination is being evaluated as part of a long-acting formulation for intramuscular injection that would confer an exceptionally long biological half-life of 21 to 50 days following a single dose (72).

**Table 1 Summary of known major INSTI resistance mutations.**

INSTI	Structure	Major Resistance Mutations	Reduced Susceptibility
<b>Elvitegravir</b>		T66I/K E92Q G140S/A/C S147G Q148H/K/R/N N155H E92Q	10 to 40-fold ~30-fold 3 to 5-fold 5-fold >100-fold with G140 30-fold ~30-fold
<b>Raltegravir</b>		T66K Y143C/R/H/K/S/G/A Q148H/K/R/N N155H	>10-fold ~5 to 20-fold >100-fold with G140 10-fold
<b>Dolutegravir</b>		G118R E138K+Q148 R263K S153Y/F G140+Q148	~5-fold 10-fold 2-fold 3-fold 10-fold
<b>Cabotegravir</b>		G149A	
<b>Bictegravir</b>		G118R G180+Q148 S153Y/F	~2 to 3-fold 10-fold 3-fold

## 1.6 – Viral Phylodynamics of HIV

To fully understand the constant expansion of new cases of HIV infection from the last 40 years, it is not enough to only comprehend how an infectious agent works in isolation. The classic host-agent-environment triangle of epidemiology can be used as a tool to better explain the underlying mechanisms of the HIV outbreak (73). The first part of the triangle, the agent, was explored in section 1.4. Environment and host are two items that tend to overlap with each other. Susceptible hosts are likely to be infected by the agent in an environment that fails them to provide the necessary access to treatment or information that would prevent future

infections. It is important to take these factors into consideration when studying the evolution of HIV and the transmission dynamics that drive the epidemic within a certain population, as it would allow a certain frame by which host-environment relations can be studied. Using viral phylodynamics, the epidemiological, immunological, and evolutionary processes can be studied to have an understanding of how viral phylogenies shape and impact viral genetic variation (74). Droplet or sexually transmitted viruses, like HIV, spread most rapidly through complex human networks. Conventional epidemiological methods like interview or classical phylogenetic analysis of viral gene sequences have inherent limitations that fail to detect “clustering” and transmission connections. Learning more about the behaviour of pathogens concerning their host population is the main goal of viral phylogenetics (75).

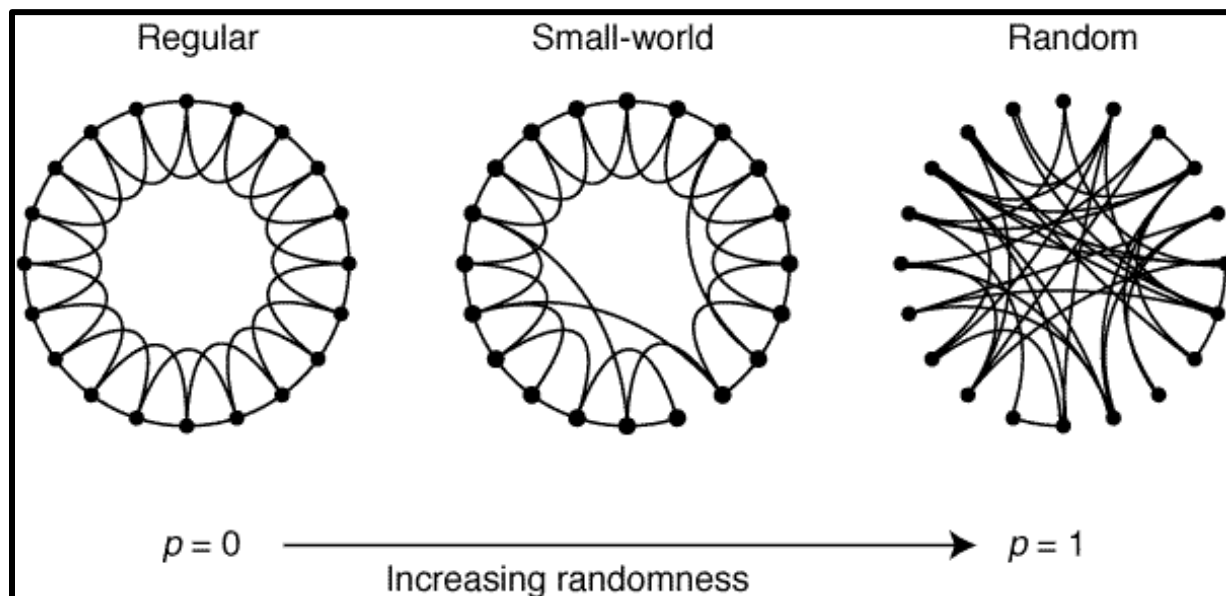
#### 1.6.2 HIV Transmission Networks

More recently, HIV outbreaks could be described as small endemic events wherein a determined setting, the mix of epidemiology, viral characteristics, host immunology, social networks, demographic phenomena, and social determinants can influence transmission, prevalence, and mortality. Transmission networks provide a way to organize these multifaceted issues (76). Population-based analysis of viral sequences has yielded different outcomes according to the risk group, leading to the discovery of clustering based on the genetic relatedness of their HIV viruses (77). Understanding the events that lead to an infection event can help understand the transmission history of the virus. As a result, it is possible to reconstruct the transmission network to predict the risk factors associated with transmission and to plan future intervention strategies (78).

### 1.6.3 Reconstruction of Networks and Epidemiological Assumptions

As mentioned in section 1.5, because of the high mutation rate of the virus caused by selective drug pressure, the need to monitor drug-resistant strains has led to the creation of HIV genotyping programs (79). Analysis of *pol* sequences can be used to define these transmission networks within a population (80). Routine sequencing of the conserved protease and RT genes is performed in the clinical context of genotyping drug resistance, and it allows for less variability than the V3 loop in *env*. *Pol* has shown to have the lowest rate of substitution and has been used consistently in the creation of phylogenetic trees that permit reconstruction of transmission histories (81). Reconstruction of the network requires more than relying exclusively on sequences and needs to consider some assumptions predicted through epidemiological models. Classical epidemic models cannot be applied to HIV in the evaluation of the extent of the epidemic, however. These models rely heavily upon random mixing assumption, wherein each individual within a population, has a small and equal chance of coming into contact with any other individual. Within the context of HIV, this assumption fails since subjects have a fixed number of contacts. These contacts are not sampled uniformly among individuals; as a result, employing networks constraining transmissions to the neighbourhood of infected subjects becomes necessary (82). Network structure allows public health agencies for better management of the epidemic (83). Another advantage of reconstructing networks is that it yields novel insights into the drivers of transmission near real-time. HIV is capable of evolving faster than transmission occurs so that viral sequences obtained

from an individual tend to be characteristic of that individual weeks after infection (84).

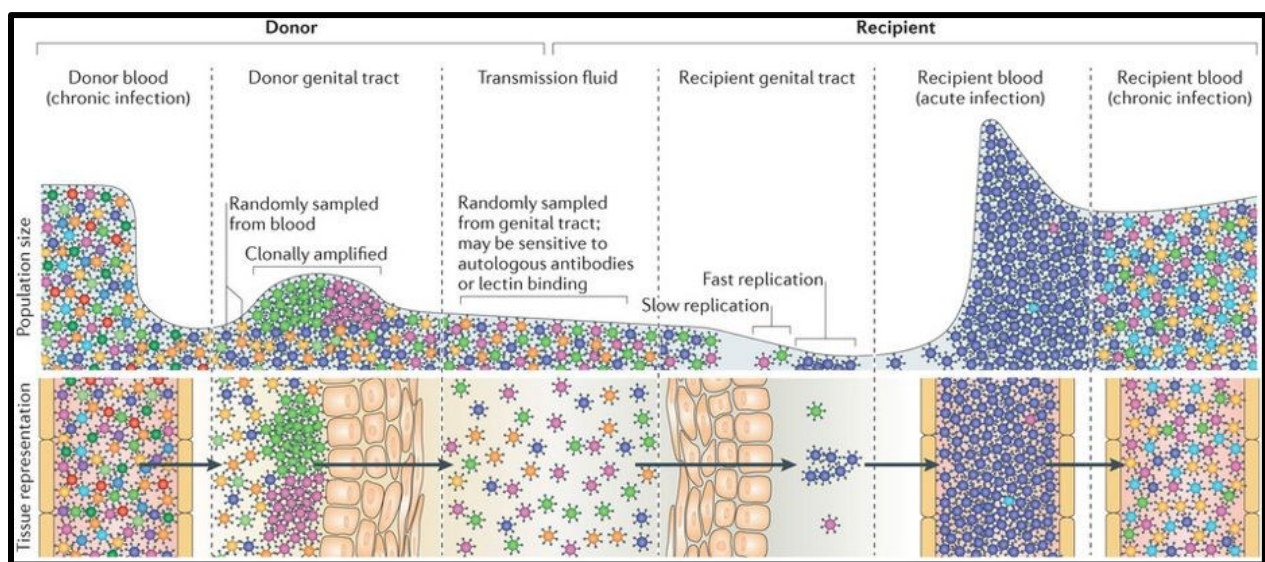


**Figure 4 Graphical Representation of ‘small-world’ Dynamics.** A ring of vertices, each connected to its nearest neighbours by undirected edges. With probability  $p$ , one of the vertices connects to an edge chosen uniformly at random over the entire ring. Duplicate edges are forbidden. The process is repeated clockwise around the ring. As  $p$  increases, the graph becomes increasingly disordered until  $p = 1$ , all edges are rewired randomly. Intermediate values of  $p$  show small-world networks: clustered like a regular graph, with small characteristic path length, like a random graph (81).

### 1.7 Transmission/Founder Virus

The existence of a possible common ancestor found in transmission networks and the tendency to form clusters within a certain population raises the question about a possible virus variant that could be selected as being better suited to overcome transmission barriers and harbour properties that might provide certain advantages. Establishment of virus infection results from this variant termed the transmitted/founder virus (85). A limited number of founder viruses usually establishes infection. In Men-Having-Sex-With-Men (MSM) populations, about 83% of patients exhibit a single founder virus, which is on levels similar to heterosexual transmission (HET). The transmitted HIV-1 ‘quasispecies’ can, therefore, undergo selective pressure by way of a population bottleneck upon transmission that results from a combination of host factors

such as mucosal barriers, target cell availability, and levels of immune activation (86). Studying these bottlenecks could provide critical information about selective pressures acting on the virus during transmission and reception and prove advantageous in shaping efficient prevention strategies. Elucidating what events shape them remains an arduous task, and it is unsure whether these T/F viruses are the result of stochastic forces which act on all the viruses or whether there are selective pressures that favour certain phenotypes (87).



**Figure 5 Transmission/Founder Viruses are Shaped by Genetic Bottlenecks.** Viruses from the blood of chronically infected patients show extremely diverse populations, and some of them manage to seed the genital tract of the donor, where the resulting viral population is less diverse than in the blood. Crossing the genital tract requires the virus to go a selection process that could select for specific phenotypes that enhance its survival and replication. Typically when a systemic infection is established, the initial viral population of the recipient's blood will be genetically homogeneous because it was established from a single viral genotype that was able to replicate in the genital tract (86).

### 1.8 The HIV-1 Epidemic in Quebec

By the end of 2016, about 63,110 people in Canada were living with HIV. The current prevalence rate in the country is 173 per 100,000 people. Today, the epidemic continues to expand, albeit slowly, with recent data showing a 3% increase in 2017 from 2016. In 2017, Ontario continued to account for the highest number of reported HIV cases (38.9%), followed

by Quebec (27.9%), Alberta (11.7%), and British Columbia (7.8%) (88). Since 2001, the province of Quebec has been collecting HIV-1 Pol sequences as part of their drug resistance testing program. Genotyping is typically carried out as part of a clinical follow-up of HIV infected patients to determine possible ART resistance. Two main laboratories oversee HIV genotyping in the province, the Centre Hospitalier de l'Université de Montréal (CHUM), and the McGill AIDS Centre at the Jewish General Hospital (JGH) (89). Pioneering studies by our laboratory have combined sequence datasets from both sites for population-level surveillance of transmission dynamics of the HIV-1 epidemic within the province. The building of phylogenetic trees using sequences originating from the Quebec genotyping program has uncovered transmission dynamics of first genotypes from all treatment-naïve and newly infected persons, including Men having Sex with Men (MSM), People Who Inject Drugs (PWID), and new migrants harbouring viruses from their endemic country of origin (79).

Phylogenetic tree analysis has uncovered transmission dynamics of HIV-1 spread, using first genotypes of newly infected, treatment-naïve MSM. Three different patterns of viral spread occur among MSMS in newly infected persons: Unique 'dead-end' PHI infections (Singletons), small clusters (2-4 PHI), and large clusters (5+) networks (90). In recent years, cases of patients infected by large cluster species have been on the rise. Not only are these viruses becoming more prevalent in new PHI cases, but they also appear to show accelerated development of drug resistance and viral escape from INSTIs. Large Cluster isolates can select for resistance to DTG, EVG, BIC, and lamivudine faster than small cluster strains. In drug selections, the development of R263K mutations by DTG and BIC appeared 8-12 weeks as compared to 24 weeks in small cluster lineages or 20 weeks in laboratory strains. Furthermore, it was noted

that some large cluster species appear to harbour a commonly shared integrase, as observed in clusters C118, C045 and C185 (91).

## Chapter 2 – Rationale, Hypotheses, and Objectives

The year 2020 will mark 40 years since AIDS was considered a global epidemic. Understanding how the virus operates has been crucial in the design of strategies currently employed that have allowed patients to survive HIV as a chronic infection. An ambitious program by UNAIDS was put in place in 2013, called the 90-90-90 initiative, which states that by 2020, 90% of people living with HIV will know their status, 90% will receive sustained ART, and 90% will have achieved viral suppression (92). As mentioned in section 1.1, about 1.8 million new cases of HIV are recorded every year, with that rate showing minimal decrease since 2010. It is doubtful that those 90-90-90 targets will be met by next year unless there can be an understanding of the transmission networks by which the virus continues to propagate. Phylogenetic studies made by our group have allowed us to reconstruct clusters of infected patients harbouring similar viruses. And better characterize large cluster outbreaks. In 2019, the plan to end the HIV epidemic in the United States has added phylogenetic analysis to tailor interventions to avert large cluster outbreaks.

As a result, we postulate that large cluster strains have unique properties that offer phenotypic, genotypic, or demographic differences that provide them with a selective advantage and can distinguish them from small cluster outbreaks. Of particular interest for this project is the existence of a subset of clusters (clusters 185, 45, and 118) that share a common integrase.

## 2.1 Objectives

**Aim 1:** To characterize and describe the spread dynamics of HIV-1 in the province of Quebec, analyzing phylogenetic data on sequences from the provincial drug resistance genotyping program.

**Aim 2:** Determine cell tropism of transmission/founder viruses isolated from HIV-infected patients.

**Aim 3:** Determine the relative susceptibility of isolated C185 viruses on second-generation INSTIs: DTG, CAB, and BIC in drug washout experiments.

## Chapter 3 – Materials and Methods

### 3.1 Data Analysis

SAS programming language (SAS Studio) was used to analyze a dataset containing about 20,000 patients from the provincial genotyping program. For this study, we focused primarily on using data from patients whose samples were collected between the years 2002 and 2017 as they represent the most up-to-date information currently available. Clean-up, data preparation, and visualizations were conducted using several Python packages (Pandas, Seaborn, Numpy, Scikit-learn) to ensure uniformity throughout the process. The code used for this project can be accessed from the following Github repository: <https://github.com/ecuadrafoy/ThesisMsc>

### 3.3 Tropism Assay

Viruses belonging to large clusters, small clusters, or singletons were isolated from patient blood of 45 patients. Peripheral blood mononuclear cells (PBMC) were isolated using Ficoll-Hypaque density gradient centrifugation. Isolated PMBCs were placed in T-25 flasks containing

10% FBS RPMI media with 10 µg/ml of PHA-P. CD4<sup>+</sup> cells were enriched by the depletion of CD8<sup>+</sup> cells, which involved the usage of Dynabead M-450 magnetic beads. Virally infected isolates were titered and stored at -80 °C.

Isolated clinical samples infected CD4<sup>+</sup>/U87 cells expressing the CXCR4 or the CCR5 coreceptors in a 96-well format. These adherent cells cultivated in separate flasks containing 15% FBS DMEM, Pen-Strep & Glutamine, 300 µg/ml G418, and 1 µg/ml puromycin (R15 media). At the time of preparation, the cells had to be at about 70% confluency before infection.

Approximately  $1 \times 10^4$  U87 cells were distributed in 100 µL of R15 media per well in a 96 well plate. Cells were incubated overnight at 37 °C. On the next day, viruses were added to respective wells at an MOI of 0.1. Each virus was done in triplicates on either U87.CD4.R5 or U87.CD4.X4 cells. After three days of infection, samples from each well was collected and tested for RT activity. Visualization and analysis were performed using the clustermap option on the Python package Seaborn. Values were normalized to pNL, and BaL controls for columns X4 and R5, respectively.

### 3.3 Washout of Second-Generation INSTIs

NL4.3 viruses were obtained by site-directed mutagenesis on pNL4.3 vector (using QuickChange II kit, Agilent) and then transfected in HEK 293T cells. These viruses were then tittered at equal amounts. Viruses from primary isolates of patients belonging to C185 were extracted as described previously in 3.3.

Equal amounts of viruses were used to infect 300,000 MT-2 cells. Samples were collected for 14 days, and replication was measured using an RT assay.

In order to assess IC50 values, infections were carried out in MT-2 cells in serial dilutions of drug concentrations for DTG, CAB, and BIC (500 – 0 nM) on NL4.3 mutants and C185 viruses. Once the IC50 values for each drug were calculated using GraphPad Prism, MT-2 infections were treated using 20 times IC90 values of each specific drug based on WT virus results. WT values were used because drug doses are fixed at a specific value and not distributed according to a resistance mutation a patient might carry.

To washout the drug from the media, cells were pelleted by using a centrifuge, the supernatant was discarded, and then the cells were re-suspended in RPMI 10% FBS media. This was repeated three times to ensure complete removal of the drug. These infected cells were plated again in wells; one part of the cells was treated again with the same amount of drugs, and the other was kept without any drugs. Samples were collected to measure their replicative capacity using an RT assay on days 3, 7, and 11 for NL4.3 viruses; for C185 viruses, samples were collected on days 5, 8, and 14. Our results were analyzed using GraphPad Prism. All viruses were normalized to WT on the first day of sample collection when replication is at its peak, as seen before by Osman et al., 2018. Integrated DNA was measured using a two-step Alu-mediated quantitative PCR. Total cellular DNA was extracted from MT-2 cells from pellets at the end of days 3, 7, and 11 using a DNeasy blood and tissue kit (Qiagen). Following DNA purification, 65 ng of genomic DNA was amplified by PCR in a first step using primers: sense, 5'-GCCTCCCAAAGTGCTGGGATTACAG-3', and antisense, 5'-GTTCTGCTATGTCACTTCC-3'. The PCR products were then used for quantitative PCR using the primers: sense, 5'-TTAAGCCTCAATAAAGCTTGCC-3', and antisense, 5'-GTTCTGGGCGCCACTGCTAGA-3'. Cycling was performed on a Corbett RotorGene 6000 thermocycler, as follows: (i) 50°C for 2 min; (ii) 95°C

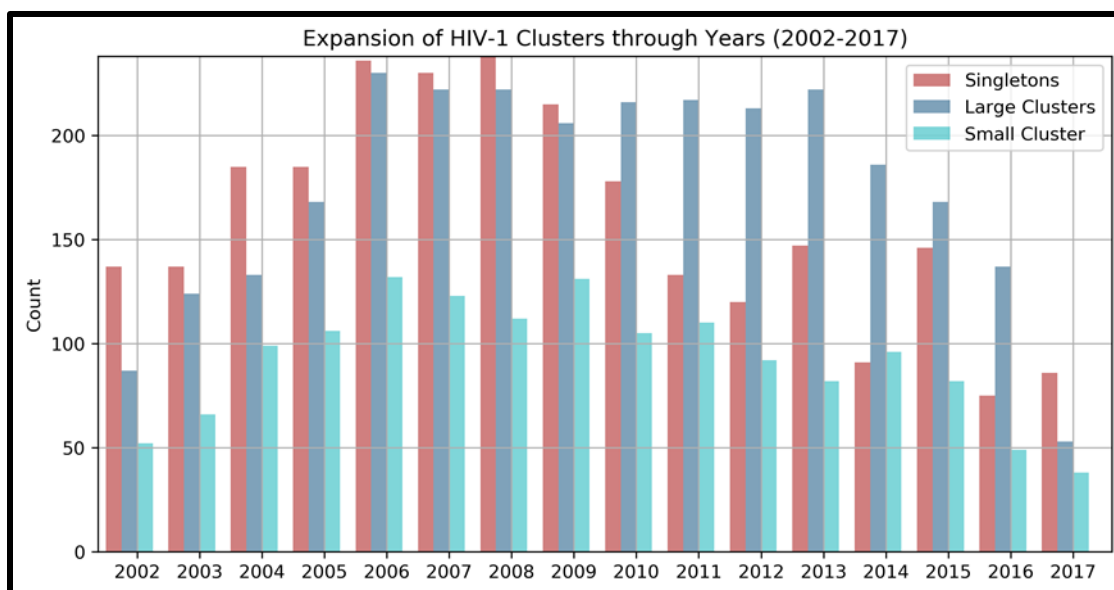
for 2 min; (iii) 50 cycles, with 1 cycle consisting of 95°C for 10 s, 60°C for 10 s, and 72°C for 45 s acquiring on the green channel; and (iv) 50°C for 2 min. The probe sequence for integrated DNA was 5'-FAM-CCAGAGTCACACAACAGAGGGGCACA-TAMRA-3' (6-carboxytetramethylrhodamine).

## Chapter 4 – Results

### 4.1 Reconstruction of the Dynamics of Spread of HIV-1

In this analysis, only the samples of patients who visited a clinic between 2002 and 2017 were considered. Furthermore, samples with missing years of visit were dropped, and only patients older than 18 were chosen for this analysis. This subsetting resulted in 6,600 patients of the total 10,000 being used in the analysis

The building of phylogenetic trees was previously reported in Brenner et al. (2017). HIV-1 *pol* sequences were aligned with consensus HXB2 sequences. Phylogenies were constructed using neighbour-joining models in MEGA 7. A cluster was considered to be where two strains show high bootstrap support values (median 98.5%) and short genetic distances (<1.5%) (93). Viral transmissions were stratified as unpaired Singleton strains, Small Clusters (SC) with 2-5 members, and Large Clusters (LC) having 6 or more linked strains. A summary of the results was compiled (Table 1). Between the first eight years of the epidemic, most cases corresponded to people falling within the Singleton groups; during the next two quadrennial periods, LC viruses started to become more prevalent. As of 2017, two main driving forces were responsible for continuing the epidemic; 41 % of all cases fall under LC, and 37% in Singletons. It is likely that SC cases will shift towards large clusters as more sequences are catalogued and added to the tree, which could explain why their numbers remain relatively constant over the whole period.



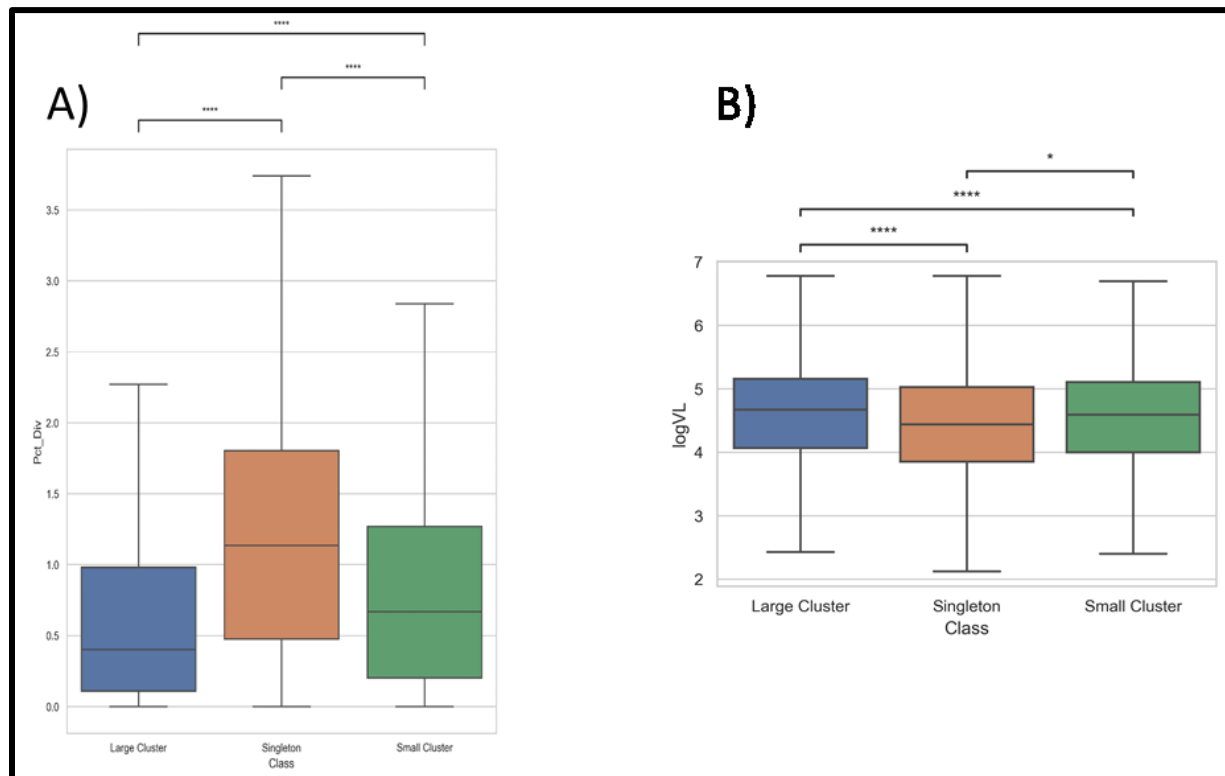
**Figure 6 Timeline of the MSM HIV-1 Epidemic in Quebec.** *Pol* sequences have been collected for newly admitted patients in the province for the purpose of monitoring drug resistance mutations. From the phylogenetic studies performed in Brenner et al. (2017), three classes were found, Singletons, Large Clusters (LC) and Small Clusters (SC). From 2002 until 2008, Singletons were responsible for the majority of newly recorded HIV cases; however, LC cases had been about equal numbers since 2006 and finally surpassing singletons in 2010. SC cases remain relatively in stable numbers as they represent an intermediate state between Singletons and LC.

**Table 2 Summary of Variables Observed**

	Total N(%)	Large Cluster N(%)	Small Cluster N(%)	Singleton N(%)
<b>Total per row</b>	6600 (100)	2729 (41.3)	1425 (21.6%)	2446 (37%)
<b>Quadrennial Period</b>				
<b>2002-2005</b>	1431 (21.7)	495 (7.50)	315 (4.7)	621 (9.4)
<b>2006-2009</b>	2228 (33.8)	862 (13.1)	480 (7.2)	886 (13.4)
<b>2010-2013</b>	1777 (26.9)	844 (12.8)	376 (5.7)	567 (8.4)
<b>2014-2017</b>	1164 (17.6)	528 (8.0)	254 (3.85)	382 (5.8)
<b>Age Group</b>				
<b>18-31</b>	1704 (25.8)	856 (13.0)	352 (5.34)	496 (7.5)
<b>32-47</b>	3346 (50.74)	1344 (20.4)	740 (11.2)	1262 (19.1)
<b>&gt;47</b>	1545 (23.4)	529 (8.0)	331 (5.02)	685 (10.4)
<b>Log10 HIV RNA median (IQR)</b>	4.6 (4.0-5.1)	4.7 (4.1-5.2)	4.6 (4.0-5.1)	4.4 (3.9-5.0)
<b>Recency %Diversity median (IQR)</b>	0.7 (0.2-1.3)	0.4 (0.11-0.98)	0.67 (0.2-1.27)	1.14 (0.53-1.8)
<b>Subtype</b>				
<b>B</b>	5215 (79.0)	2588 (39.21)	1096 (16.6)	1531 (23.2)
<b>Non B</b>	1385 (21.0)	141 (2.14)	329 (4.9)	915 (13.8)

A closer look at the timeline (Figure 6) gives a better perspective of the distribution of the virus through the years. Prevalence of Singleton cases peaked around 2008, with LC-infected patients following closely. LC and Singleton cases appear to mirror each other in terms of time duration, which suggests that these two classes could operate in waves. As of 2017, singletons appear to be on the rise, however at much lower total numbers than when the epidemic started being tracked. It is likely that samples from this period are still being collected, thus skewing the total number of cases. Furthermore, properties between classes are markedly different from one another; the percent diversity of viral species used to estimate recency of infection was found to be significantly different ( $p < 0.0001$ ) among classes, as shown in the boxplots (Figure 7). Percent diversity median values were calculated to be 0.4% for LC, 0.7% for SC, and 1.14% for Singletons. Likewise, the viral load for each of these groups was markedly different; LC and singletons median values varied significantly ( $p < 0.0001$ ) as well as LC from SC ( $p < 0.0001$ ). Less

apparent differences in viral load were noted between SC and Singletons ( $p < 0.01$ ).

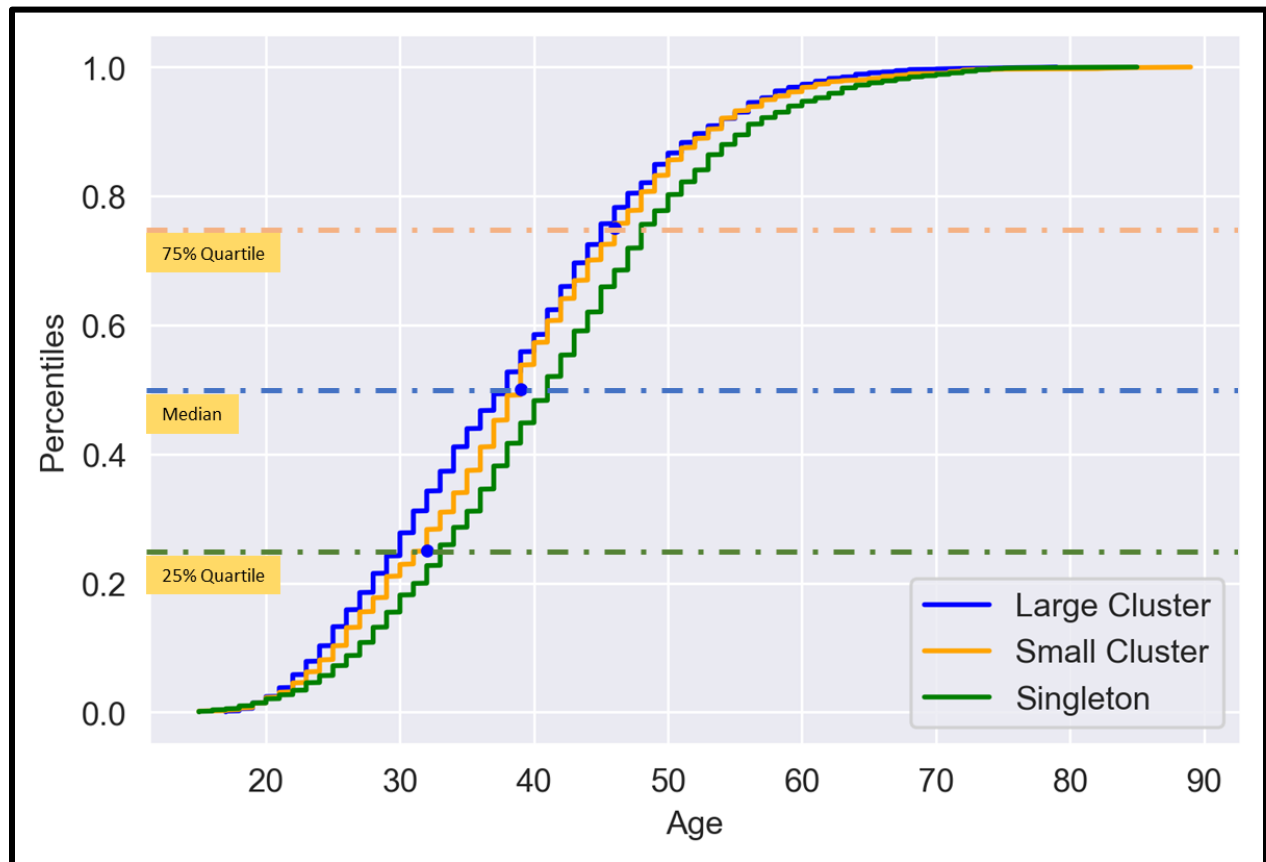


**Figure 7 Comparing Recency and Viral Load between Classes.** Percent diversity and viral load represent the only two variables that are continuous in nature and were compared to observe if there were significant differences between the classes. A) Percent diversity boxplots between classes. B) Viral load boxplot separated by class.

#### 4.1.2 Age Differences Between Classes

Previously in Table 2, age was segregated in groups according to quartiles for the total distribution of ages for each patient. Quartiles were determined by computing empirical cumulative distribution functions (ECDF) in Python (Figure 8). Here, it was noted that the median age between LC and Singletons was different by 3 years; 38 and 41 years old for LC and Singletons, respectively. Through this figure, it was determined that the distribution of age

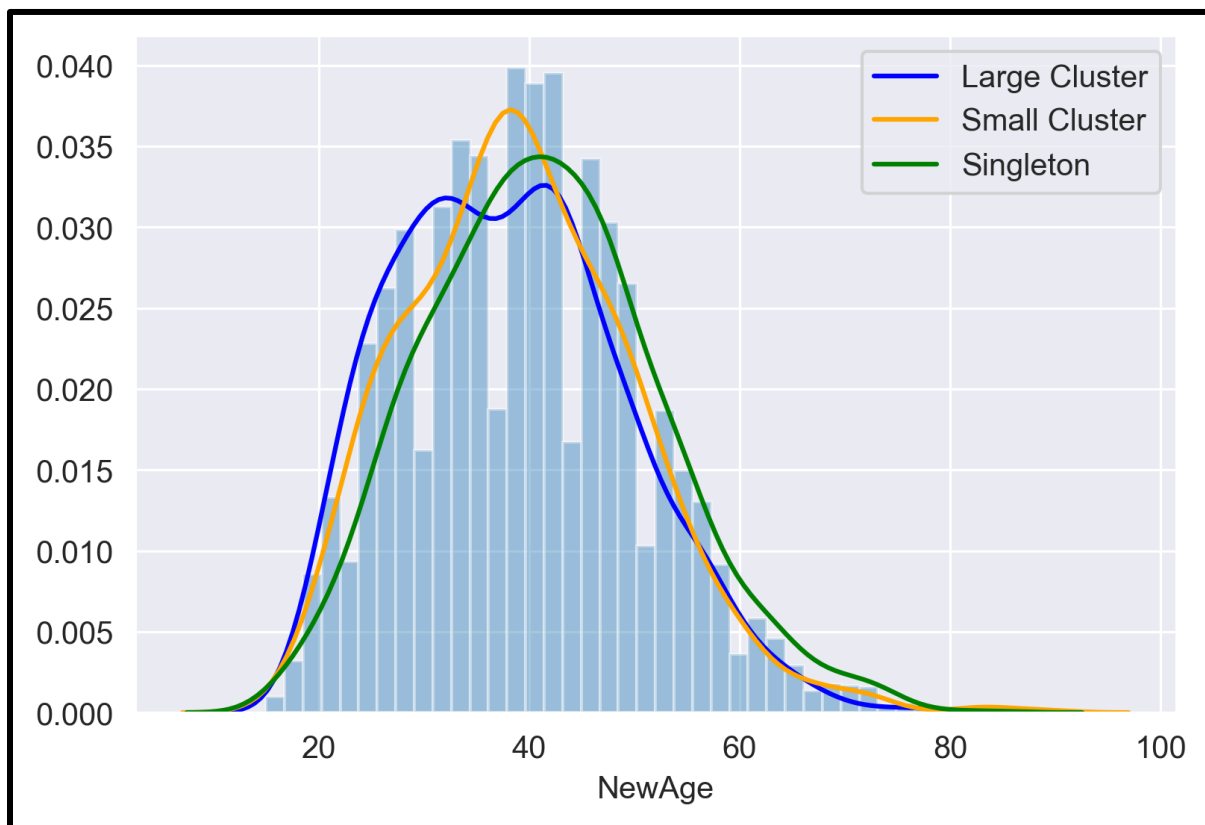
among cluster classes is close to a normal distribution, and to verify this, the data was transposed to a histogram.



**Figure 8 ECDF of Ages by Class.** Building cumulative distribution functions allow easy visualization of differences between classes with a high number of values. Median values are different for each class, 38 years old in LC, 41 years old for Singletons.

At first view, the histogram provides an indication that age groups from the ECDF are mirrored in the shape of the histogram where the dips are observed. However, histograms can be unreliable when large amounts of data are present and face the problem of binning bias. Computing a kernel density plot (KDP) of each group provides us to compensate from the inherently discrete limitations of histograms that are sensitive to bin size. Analysis of the KDP

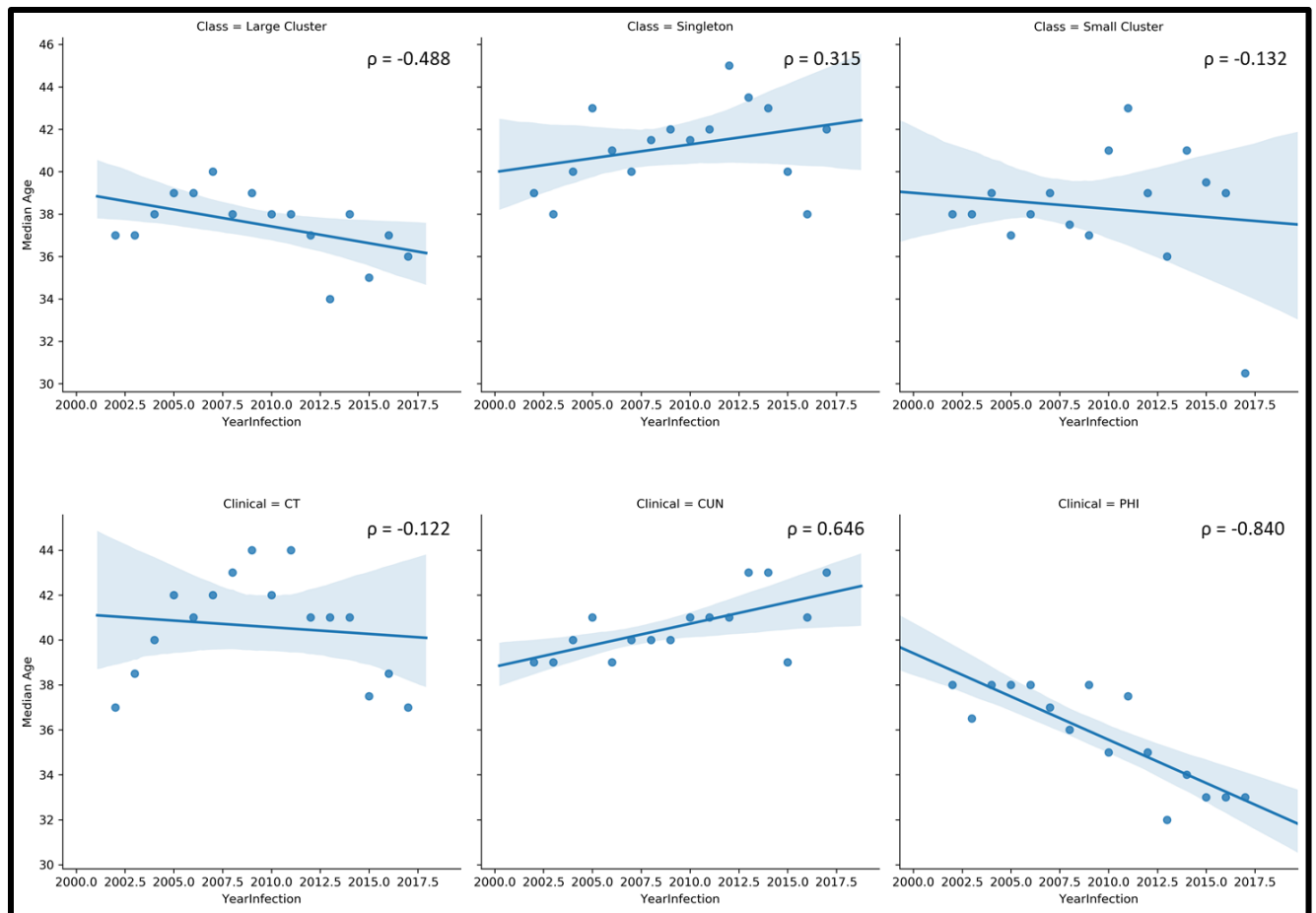
results (Figure 9) show that SC and Singleton classes produced smooth single peaked curves with medians at 39 and 43 years old respectively. Conversely, the KDP for LC class members showed two peaks, one with a median of 28 and the other at 43, about the same value as the Singleton class. This suggests that the age of LC members falls under a slight bimodal distribution.



**Figure 9 Histogram of age Distribution and Kernel Density Plots of each Class.** The peaks on KDP display where values are concentrated over the interval and are independent of bin selection.

A more in-depth analysis was carried out to determine how age among the members of the different groups could have influenced the dynamics of spread through the observed period; to accomplish this, LC, SC, and Singleton cases were grouped by year of visit, and then for each

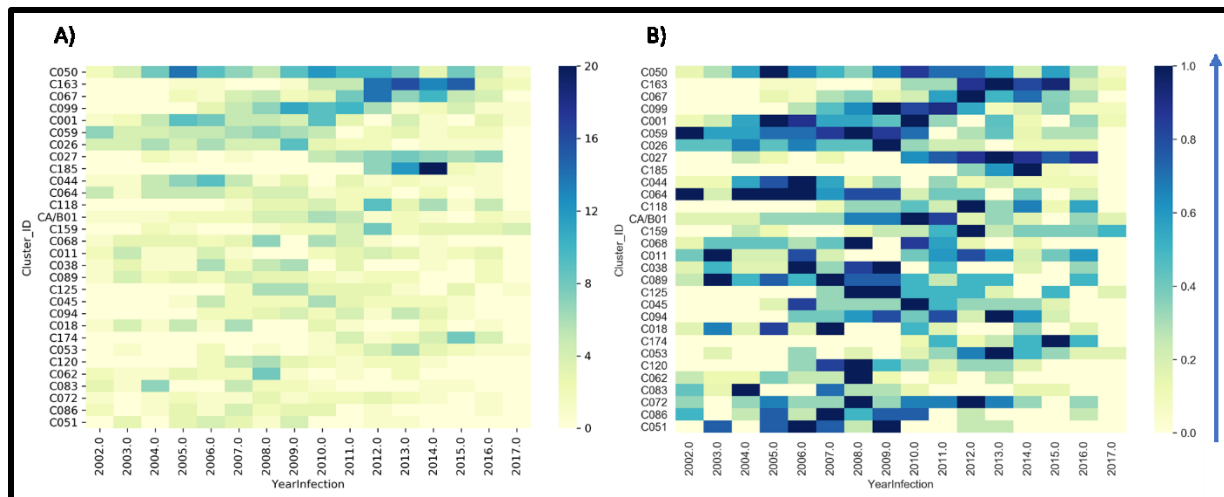
group, their median age was calculated on that specific year. Groups were also segregated according to their clinical status at the time of diagnosis. Clinical status included: Primary HIV-1 infection (PHI, less than 6 months post-infection, treatment naïve), chronic untreated (CUN more than 6 months post-infection, treatment naïve), and chronic treated (CT, more than 6 months infected, experienced). At first, plotting the median age either by the total Class or Clinical group didn't provide any useful information, and no clear correlation could be observed between median age and the year of visit. However, separating them by individual Class or Clinical groups provided a clear trend. It appears that members falling into the LC class have been consistently getting younger as time moves forward, showing a strong negative correlation ( $p = -0.488$ ). Meanwhile, the Singleton class is continuously recruiting more aged members. A similar trend was observed for CUN and PHI patients. For these two groups, there was a robust correlation between age and year of visit showing Person correlation coefficients of 0.68 and -0.88 for CUN and PHI, respectively.



**Figure 10 Median Age Varies in Time According to Class or Clinical Status.** The age of visit for each class or clinical status was grouped, the median measured for each one, and spread over time.

#### 4.1.3 Characterization of Large Clusters

For this study, LC members with more than 20 members in each cluster were selected during the 2002-2017 period. These were considered to be the most significant clusters.



**Figure 11 Timeline of individual Large Clusters.** Clusters with more than 20 members were selected and plotted in decreasing order. A) Clusters were normalized in total by the highest number of cases in a year (C185, n=20, 2015). B) Each individual row was normalized by their unique maximum and minimum values over the observed period.

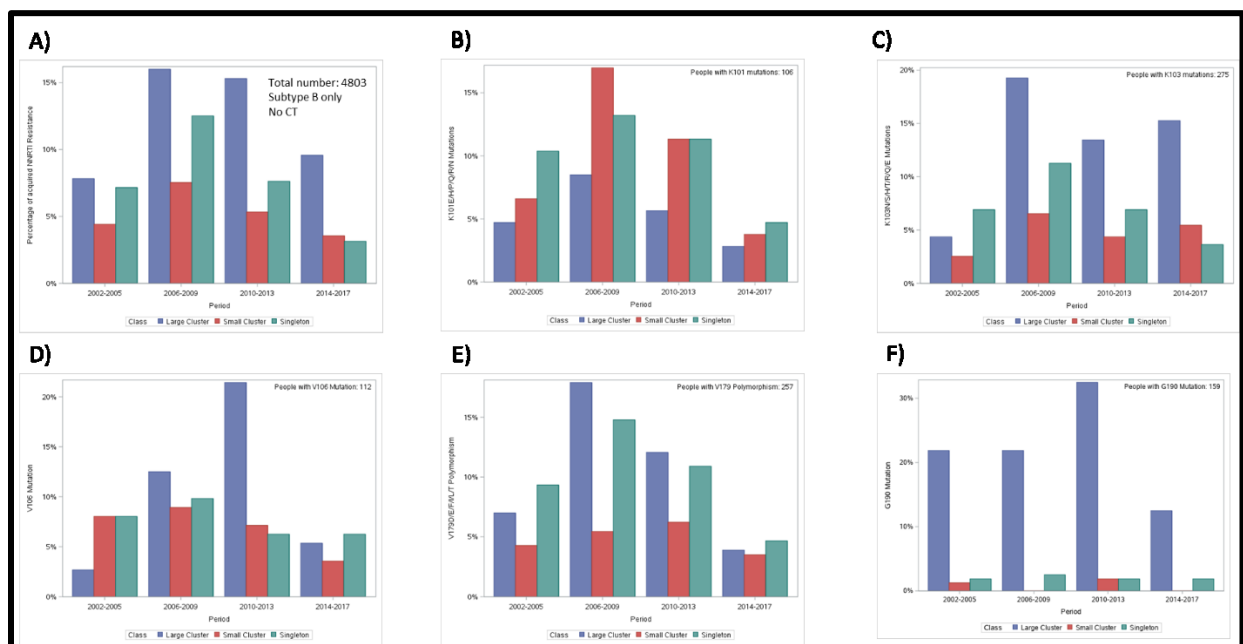
A heatmap was built to analyze the spread of large clusters through time (Figure 11). The order of Cluster\_ID is sorted in descending order by the number of patients found in each cluster.

One of the images builds a gradient from the smallest to the largest number of affected patients per year in the whole set. From the gradient, it can be observed that in 2014, 20 people were estimated to be infected with C185 in that specific year. Figure 11B calculates the minimum and maximum for each row and normalizes each cluster group individually instead of by the total. One of the patterns noticed is that most clusters share a typical lifecycle. In Figure 11A, it can be seen how clusters appear to reach a peak, as evidenced by showing one strong band per row. Following this peak, the virus slowly disappears and stops spreading in the population. C050 is unique in that it emerges in two peak years, in 2004 and 2010. C001 appears almost to follow the same pattern as C050 since both peaked in the same years; however, C001 stopped spreading instantly after its second peak but managed to show some signs of rebound around 2015. C163 is a more recent cluster that coincided with the slow down

of C099. C163 infected a significant number of patients between 2012 and 2015, but it remains to be seen if more could be affected as more samples become available.

#### 4.1.4 Transmission of Drug-Resistant Mutations

The original goal of the provincial genotyping program was to monitor the acquisition and possible transmission of drug resistance mutations in patients who recently were admitted into a clinic. The sequences obtained through Sanger sequencing were put through the Stanford HIV Drug Resistance Database. These sequences were translated and provided an array of natural polymorphisms and possible ART resistance mutations of each individual sample.



**Figure 12 Distribution of Natural Polymorphisms (NPs) and NNRTIs Drug-Resistant mutations by Class.** PHI, CUN, and Subtype B patients were selected for this analysis. Overall, the LC class appears to harbour more NPs and resistance mutations than Singletons or SC. A) overall distribution, normalized by each group. B) Distribution of K101 mutations, C) K103 mutations, D) V106 mutations, E) V179 polymorphism, F) G190 mutation

In this analysis (Figure 12), only Subtype B samples and treatment-naïve patients were chosen, ignoring CT samples. In total, 4,803 samples were selected using this criterion. Figure 12A considered all the natural polymorphisms and resistance mutations that could be associated with decreased susceptibility to NNRTIs. It was noted that LC viruses consistently contained the highest frequency of NNRTIs-associated polymorphisms in comparison with SC and Singletons. However, these observations were related directly to the number total of patients that increased per year, and as a result, the fluctuations in these classes were directly proportional to this trend. Therefore, it was decided to systematically look for drug resistance mutations known to cause decreased susceptibility to NNRTIs. From the total sample of patients selected previously, 13% of these were found to carry a specific NNRTI-resistance mutation. Some of these mutations included K101/E/H/P/Q/R/N, K103N/S/H/T/R/Q/E, V106A/M, and G190A/S. V179I (Figure 12E) was found to be a natural polymorphism; however, there were 257 patients associated with it, which the Stanford database considered to reduce susceptibility to Nevirapine (NVP) and Efavirenz (EFV) by 2 to 5-fold. K101 codons, harbouring natural resistance polymorphisms (Figure 12B) were found present mostly in Singletons and SC patients. These mutations may reduce susceptibility to NVP by 3 to 10-fold, EFV 1 to 5-fold, and to Etravirine (ETR) and Rilpivirine (RPV) about 2-fold. The primary/major resistance mutations K103N (Figure 12C) and V106A (Figure 12D) and G190A (Figure 12F) showed patterns of growth that differed significantly from the global trend. K103 resistance mutations spiked in the 2006-2009 period and were followed closely by Singletons, following a decrease in the next period, K103 had a resurgence in the 2014-2017 period while the other two classes were on the decline. K103 offers a significant >100-fold reduced susceptibility to the first generation NNRTIs, NVP and

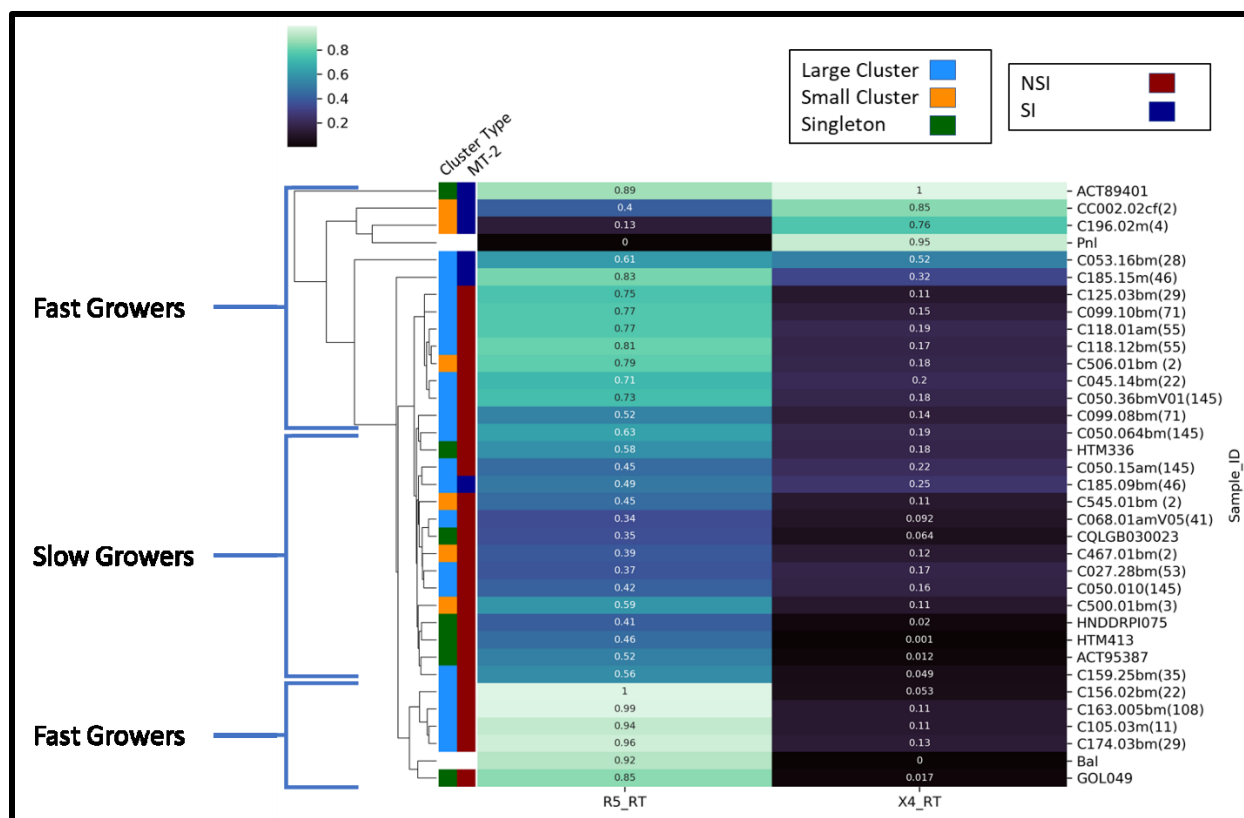
EFV. V106 mutations showed a marked rise from 3% of all V106 cases in the first period to 13% in the second period and then peaking to 25% in 2010-2013. During that time, Singletons and SC cases followed the observed global trend and were unable to match LC rapid increase in new cases. V106A mutations are known to cause a 50-fold reduction in NVP susceptibility and about 5-fold in EFV. Finally, G190 (Figure 12F) mutations were observed exclusively on LC viruses. The number of cases associated with these mutations was kept constant between the first two quadrennial periods at around 23% of the total number of G190A cases. For the third period, a spike of 10% was recorded and then decreased to 12% by the end. G190A reduces NVP susceptibility by more than 50-fold and EFV susceptibility 5 to 10-fold.

This analysis was focused primarily on polymorphisms and mutations associated only with NNRTIs. Eventually, further work could be expanded on other ART therapies like NRTIs and PIs and observe their contribution to either LC, SC or Singletons. It was observed that LC clusters are responsible for carrying a significant number of NNRTIs-related polymorphisms and mutations and where some of these were independent of the total number of patients observed in the global trend. It is likely that the transmission of drug resistance variants is being carried by recent cases in short time spans. In the future, with the use of better computer algorithms, it would be possible to associate natural occurring polymorphisms and predict a possible patient's drug regime for a more personalized approach.

#### 4.2 Identifying Tropism of Transmission/Founder Viruses and Viral Replication

An important aim of our studies was to ascertain the differential tropism of representative Large Cluster variants (n=3.5% of lineages accounting for 40% of transmissions) as compared to

the 77% of lineages resulting in 35% of Singleton “dead-end” transmissions. Tropism of clinical isolates was performed to investigate if there were any marked differences between LC, SC, and Singleton viruses. Results showed that most transmitted viruses appeared to be R5-tropic, and there isn’t a marked difference between classes, as evidenced by the lack of any observable grouping on the Cluster Type column. It was observed that most of the samples tested appeared to grow well on the R5 line (Figure 13), which is expected from most primary isolates. While these experiments were carried out at similar titers for all viruses, it was observed that not all isolates grew equally well. From the analysis of the clustermap, a cutoff for less than 0.25 normalized units can be assumed to be not replicating. The results from this figure can be interpreted as follows; at the top of the gradient, mostly X4 tropic viruses were found able to show syncytia, as evidenced by the MT-2 columns showing positive results. After this comes a mixture of possible dual tropic viruses that show signs of replication on both lines and syncytia induction. There is a portion of slow R5-tropic growers that are on the middle part of the colour scale, showing values between 0.2 and 0.5. At the bottom of the gradient are potent R5-tropic viruses that show high replication capacity.



**Figure 13 Determining Viral Tropism of LC, SC and Singleton Viruses.** Besides their tropism, the cluster map also separated the viruses by how efficiently they replicate on each cell line. It was observed that some viruses are good growers, positioning themselves in peripheral positions, and others were slower growers, found at the centre of the cluster map.

C050 viruses showed growth in the middle to low end of the gradient. In fact, only 4 out of 15 C050 viruses were successful in registering anything above mock values. Viruses for C185, C118, and C045 appear to be on the same clade as calculated by the script, which is at the highest end of the gradient. Both C118 viruses that were tested had similar levels of replicative capacity on R5 cells, paired next to each other on the gradient. About 5 Singleton viruses from the 7 tested fell into the middle portion of the gradient range, suggesting low replication compared to LC variants. One of the isolates from C196 was fully X4 tropic, growing better than the NL-4.3 WT control. Evidence of dual-tropism in some of the viral isolates was noted. ACT89401 appears to be a fully dual-tropic species, growing well in both cell lines. CC002.02, a subtype C virus, was

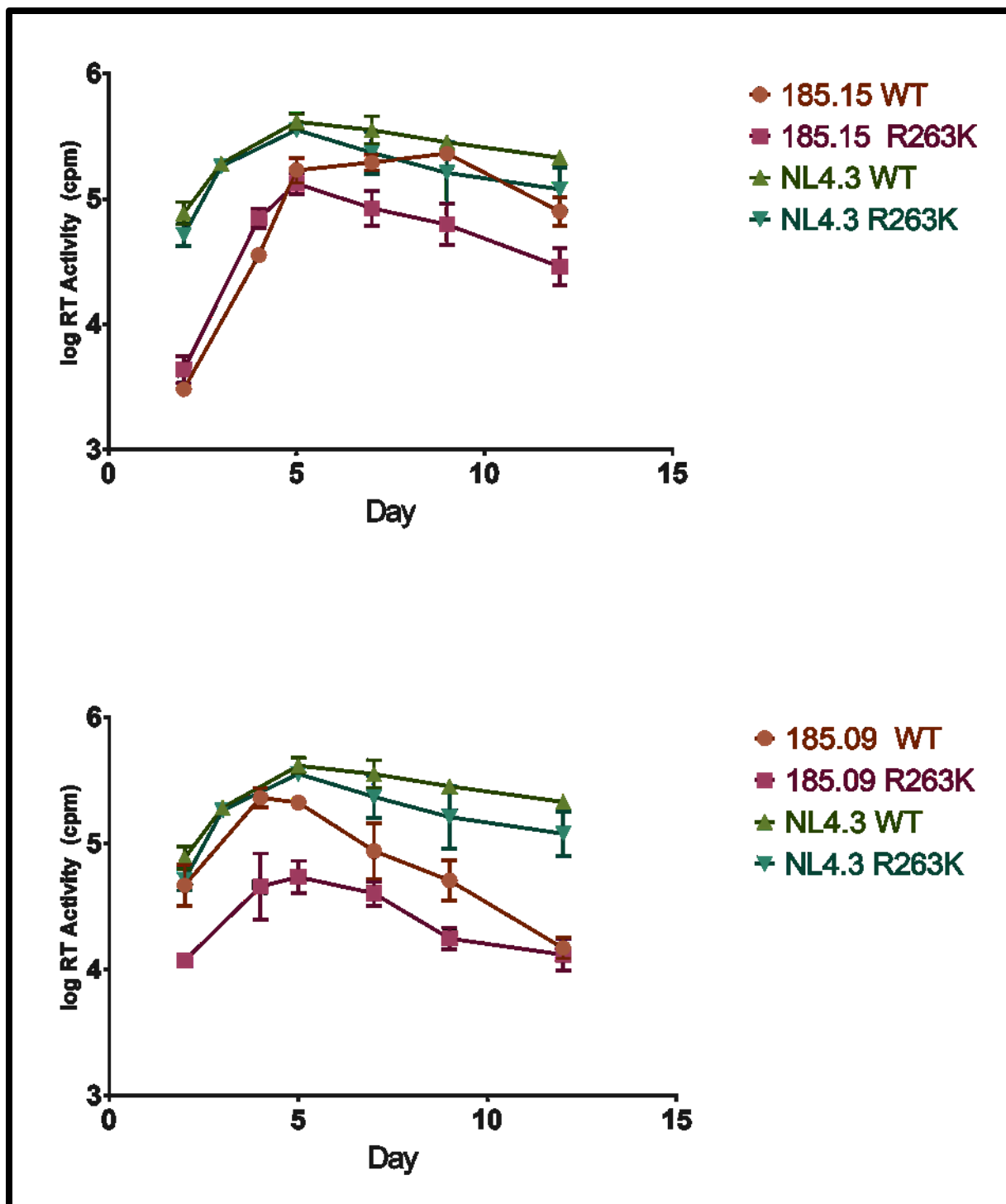
more R5-tropic than X4-tropic but still showed syncytia formation as well as significant RT values on the X4 cell line. While C185 viruses were syncytia inducing (SI), their levels of replication varied greatly. C185.15 was present in the high end of the gradient, but C185.09 was found in the middle alongside other slow growers. However, both showed SI and similar traces of X4 replication levels. If X4 RT values had not been accounted for, it is likely that both C185 viruses would be found next to each other as C118 viruses did.

Overall, no striking differences between LC, SC, and Singletons were found in terms of their tropism. Based on our findings, it would seem that LC viruses show better replication capacity than Singletons since most of them are found on the fast grower ends; however, further analysis of a bigger panel of samples would be required to confirm that hypothesis. It was unexpected to find a fully X4 tropic virus from a primary isolate, and it would be of interest to see if the other isolates from this small cluster are also X4-tropic. C185 viruses appear to be somewhere between being R5-tropic and dual-tropic. The fact that they can show syncytia reveals that they are capable of being X4-tropic while preferring to depend on CCR5 coreceptors, as evidenced by their better growth on this cell line. For this reason, both isolates from this cluster were selected for further experiments.

#### 4.3 Susceptibility of Second-Generation INSTIs on C185 Isolates

The drug washout of INSTIs was adapted from Osman et al., 2018 (94). Previously it had been demonstrated that DTG was better than the first-generation INSTIs RAL and EVG at maintaining high levels of viral suppression following removal of the drug, on drug-resistant variants. We were interested in repeating the analysis using only second-generation INSTIs on drug-resistant

mutants from the pNL4-3 backbone as well as on two C185 isolates. Site-directed mutagenesis was performed on pNL4-3 plasmids, which generated INSTI-resistance mutations, R263K, G118R, and G140S/Q148H. The transfection of these plasmids was done on 293T cells. Viruses were then aliquoted, quantified, and stored at -80 °C. Optimal titers were determined by diluting the virus at several concentrations in 300,000 MT-2 cells.



**Figure 14 Comparing Replicative Capacity of C185 isolates and NL 4.3 viruses.** NL4.3 viruses show better replication capacities than clinical isolates, including the R263K mutant.

#### 4.3.1 Determining Inhibitory Concentration Values

First, the half-maximal inhibitory concentration ( $IC_{50}$ ) was calculated for each virus against DTG, BIC and CAB to determine the optimal amount of drug to be used for washouts. Serial dilutions for BIC and DTG were done from 0-100 nM, and CAB from 0 to 250 nM. The reason for this difference in concentrations is that previously, CAB was unable to reach full suppression of some variants using this range, and it was challenging to estimate  $IC_{50}$  values for this drug. Tables 3 and 4 summarize the results for NL 4.3 and C185 viruses. A first observation was that despite DTG and BIC sharing similar structures, BIC requires about 50% less drug to reach  $IC_{50}$ . Surprisingly, BIC continued to show improved performance over DTG for every drug-resistant variant tested, as evidenced by the fold change differences between the compounds. BIC showed a higher genetic barrier to resistance than DTG and CAB, based on  $IC_{50}$  values. The R263K G118R, G140S/Q148H clones showed 1-, 1.4-, and 3.5-fold resistance to BIC relative to WT, respectively. This compares to 3.5-, 1.7- and 6.6-fold resistance to DTG and 0.8-, 6.4-, and 6.8-fold resistance to CAB against R263K, G118R, and Q140S/Q148H clones, respectively. The same assay was performed on C185 isolates 09 and 15 and their respective R263K isolates. These isolates had developed resistance to DTG following drug pressure selectivity assays. Here, it was observed that C185 WT viruses were almost equally susceptible to all the drugs, displaying  $IC_{50}$  between 0.24 and 0.6 nM. However, R263K mutants differed significantly between species; 185.09 variants registered minimal fold change resistance values among the different drugs, while C185.15 mutants were almost 16-fold resistant to BIC and CAB. With these results, 20 times the  $IC_{90}$  for each drug was calculated, which is required to perform the washout experiments. It was assumed that at 20 times the  $IC_{90}$ , the virus would be unable to

escape drug pressure and that following clearance of the drug from the media, there should be no rebound.

**Table 3 IC50 values for Second-Generation INSTIs on NL 4.3 viruses**

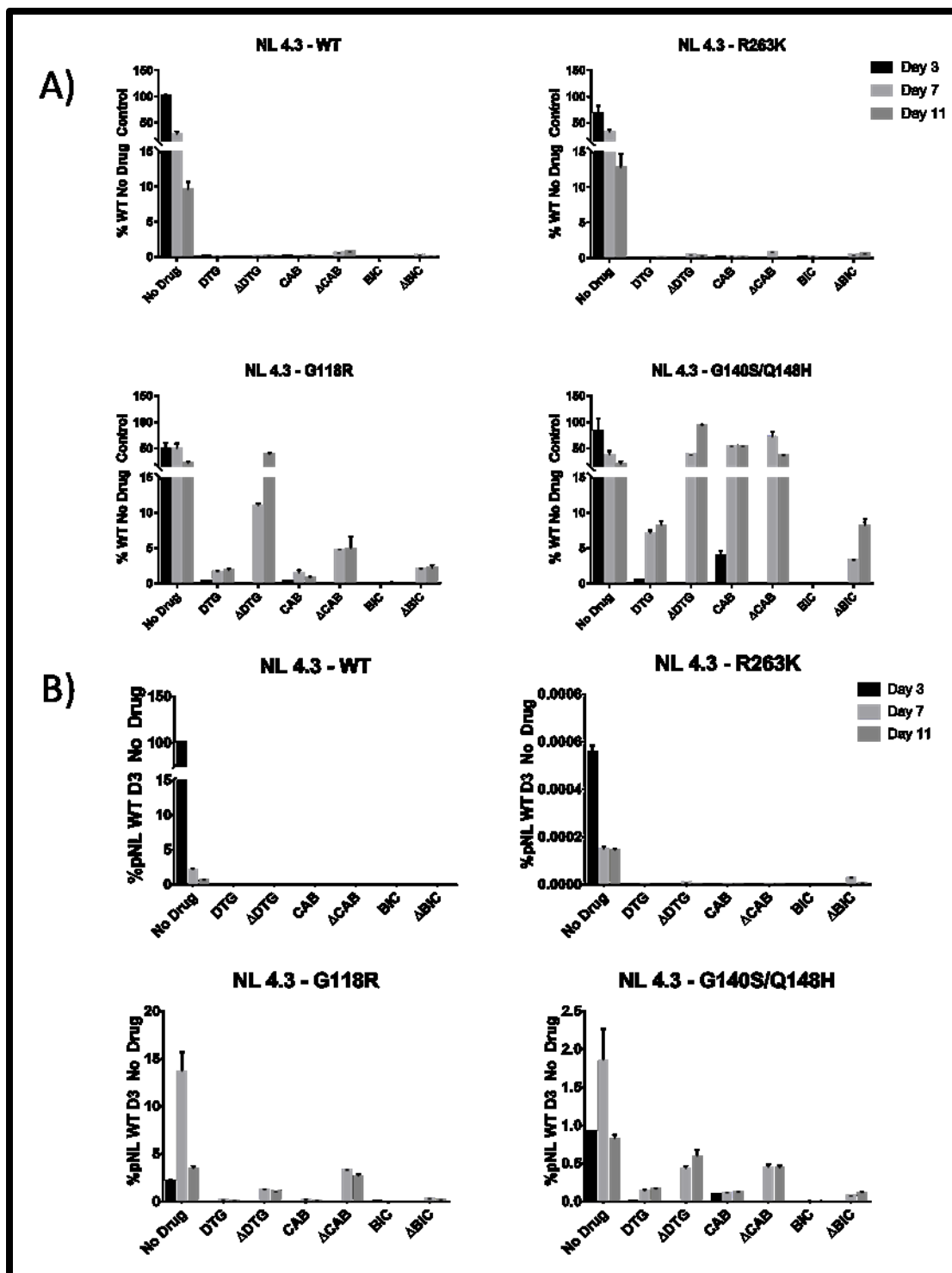
	DTG			CAB			BIC		
	IC50 (nM)	95% CI	FC	IC50 (nM)	95% CI	FC	IC50 (nM)	95% CI	FC
<b>WT</b>	1.091	0.997 to 1.19	1	2.582	2.066 to 3.225	1	0.522	0.4618 to 0.5891	1
<b>R263K</b>	3.855	3.225 to 4.609	3.53	2.026	1.713 to 2.395	0.78	0.540	0.4894 to 0.5942	1.03
<b>G118R</b>	1.838	1.466 to 2.305	1.68	16.64	12.64 to 21.75	6.44	0.708	0.5642 to 0.8882	1.36
<b>G148S/Q148H</b>	6.476	5.039 to 8.322	5.94	16.05	12.33 to 20.85	6.22	1.808	1.585 to 2.063	3.47

**Table 4 IC50 values for Second-Generation INSTIs on C185 isolates**

	DTG			CAB			BIC		
	IC50 (nM)	95% CI	FC	IC50 (nM)	95% CI	FC	IC50 (nM)	95% CI	FC
<b>185.09 WT</b>	0.24	0.22 to 0.26	1	0.24	0.22 to 0.26	1	0.280	0.23 to 0.33	1
<b>185.09 R263K</b>	0.38	0.32 to 0.45	1.58	0.4	0.31 to 0.50	1.67	0.560	0.43 to 0.73	2
<b>185.15 WT</b>	0.11	0.097 to 0.13	1	0.6	0.48 to 0.75	1	0.590	0.47 to 0.74	1
<b>185.15 R263K</b>	2.84	2.01 to 4.00	25.30	9.11	5.19 to 16.7	15.18	9.020	5.12 to 16.6	15.29

#### 4.3.2 Determining effects of INSTI washout on NL4.3 and C185 isolates

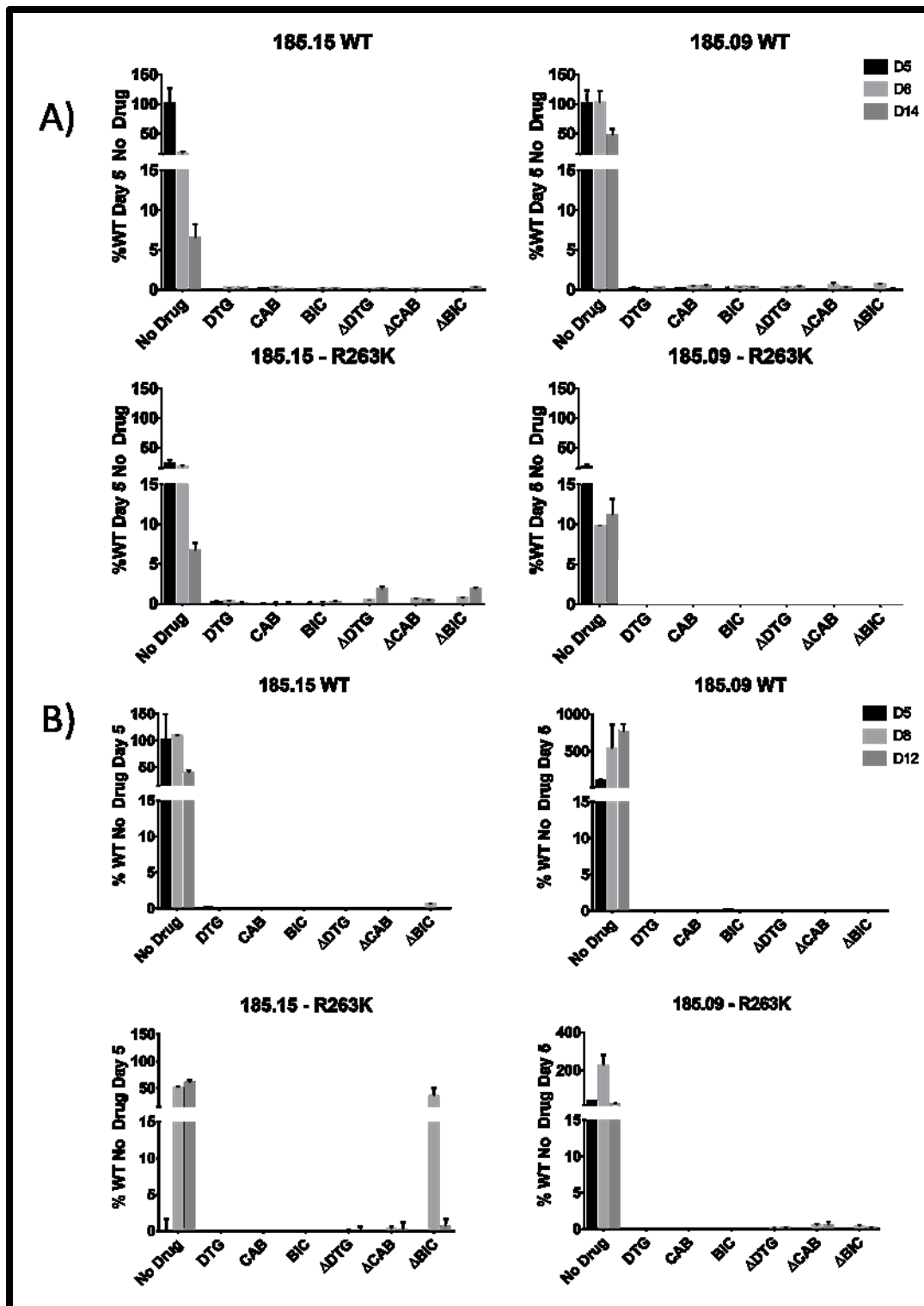
MT-2 cells were infected at equal titers with NL4-3 and C185 isolates in 24-well plates in triplicates. On the same day of the infection, all the viruses were treated with DTG, CAB or BIC at 20 times their IC90. There was also a No Drug control for each virus. Drugs were cleared out from the cells on the peak day of infection. For NL4.3 viruses, it was determined to be three days after infection, while C185 isolates this was determined to be on day 5. DTG successfully suppresses WT virus replication when present and after washout of the drug. Likewise, all the other INSTIs successfully maintained suppression after the drug had been removed on the WT variant. Replication levels of R263K were lower than WT, as expected since this mutation has been proven to have an impact on integrase activity (67). Similarly to WT viruses, R263K mutants were unable to rebound on any of the drugs tested following washout. With G118R infected cells, viral rebound occurred following DTG washout with minimal increase in replication following CAB washout and no rebound following BIC washout. The G140S/Q148H clones were not susceptible to CAB prior to and following drug washout. While DTG could suppress replication of G140S/Q148H infected cells, viral rebound occurred following washout (day 7). In contrast, BIC successfully suppressed replication through the 11 days of infection, showing minimal rebound after drug removal on day 11.



**Figure 15 Washout experiment on NL 4.3 viruses.** NL 4.3 viruses were used to infect MT-2 at equal titers and treated with 20 times the IC<sub>90</sub> values for each drug. All these values were normalized to WT on day 3. A) The replication capacity of the viruses was measured using an RT assay. WT and R263K viruses demonstrated susceptibility to all drugs. G118R viruses were susceptible to DTG but were able to escape pressure following washout. The double mutant G140S/Q148H showed lower susceptibility to CAB and

DTG while staying suppressed under BIC pressure. B) Integrated viral DNA was measured using an Alu based assay to determine the number of copies integrated into the host DNA.

This assay was replicated on WT C185 viruses and their respective mutations. Unlike NL4.3, C185 viruses require a more extended incubation period. From Figure 14, the peak day was estimated to be 5 days after infection. Values were normalized to day 5 infection for each wild type virus. Both C185 isolates were equally susceptible to DTG, CAB and BIC. R263K values were expected to follow the same pattern as their NL 4.3 results showed; however, it was noted that there was minimal rebound on C185.15 mutants following washout of DTG and BIC (Figure 16A). Some of the observations in C185.09 mutants suggest that the viruses are less susceptible to all the drugs, which is unlikely, and it is likely that these results are noise from the RT assay. A qPCR assay used to measure integrated copy numbers was used to verify these results. As expected, WT viruses were susceptible to the drugs showing no signs of integrated provirus. Surprisingly, the C185.15 mutant could resume integration following BIC washout, whereas none of the C185.09 mutants were able to do so.



**Figure 16 Washout experiment on C185 isolates and their R263K mutants.** The washout experiment was performed using the same method as in Figure 15, with day 5 being the peak day of infection. A) C185 viruses were equally susceptible to all drugs in WT. After the washout of BIC and DTG, C185.15 mutant appeared to rebound slightly. B) The Alu assay using qPCR allowed to verify the previous

observation. After removal of BIC, C185.15 could rebound after washout and resume integration on day 7.

## Chapter 5 – Discussion

### 5.1 Small Networks Model Can Explain HIV-1 Expansion and Clustering

Over the last decade, new cases of HIV remain steady at 1.7-1.8 million despite advances in Treatment-as-Prevention. The virus not only affects the immune system of the unsuspecting host but also leads to a domino effect on healthcare infrastructure and society as a whole. The first part of the project consisted of analyzing some of the demographics and attributes that characterize the MSM HIV epidemic in Quebec from the last 15 years. Data from different subjects consisted of sex, age, viral load (viral RNA/mL), percent diversity, clinical status, the year of visit to a clinic, type of class (LC, SC, Singleton), risk group, and possible polymorphisms (natural or drug-resistance related).

Sexually transmitted diseases (STD), like HIV, do not behave the same way as other pathogens. The virus transmits through networks of people by way of a particular contact, including sexual contact and intravenous drug use (95). Therefore, profiling the structure of a social network in a specific setting is essential for understanding the nature of the transmission dynamics of HIV (96). Susceptible-Infectious-Recovered (SIR) models, where the population is assumed to mix homogeneously, have been traditionally used to quantify the spread of HIV. However, these models do not quantify the clustering characteristics of real networks and also do not account for time, which is considered to be the duration of partnerships. The small-world phenomenon can be used to represent dynamic network interactions and the spread of STDs. In simple

terms, small-world models mean that almost every element of the network is somehow close to almost every other element (97). Currently, the methods used to analyze the Quebec HIV sequence dataset relied heavily on human supervision, which could ignore some more discrete events; in the future, as more data can be obtained from these samples, it is expected that unsupervised methods and machine learning techniques will be able to compensate for these shortcomings. In the Quebec MSM epidemic (2004-2018), it can be observed that during the first four years Singleton cases prevailed and were then overtaken by LC cases for the rest of the observed period. It was shown that members from LC class show significantly different properties. PHI patients and those with low recency appear to aggregate mostly on LC classes, while the opposite is true for Singletons. This further proves the likelihood that LC cases are linked to quick transmission chain events, in which groups can only be formed through a series of infection events close in time (98). This was observed by comparing percent diversity values on box plots for each of the three different classes.

While the goal of viral phylogenetics is not to determine partnership, it provides further evidence in the formation of clustering in the spatial structure of a population. As mentioned previously, HIV infection events do not take place within well-mixed populations usually skewed towards specific agents in a determined setting. Using network models provides an individual-based approach to studying the impact of spatial structure. Spatial structure is not defined solely in geographical terms, but it also includes variables that can describe specific communities. Previously, the concept of networks was introduced in section 1.6.3. Networks can vary between regular lattices ( $p=0$ ) and random graphs ( $p=1$ ). The major difference is that interactions are purely local in the first and then global in the second. The Small-World model

takes a moderate stance where the possibility of regular lattices (clusters) rewiring to random global edges (99). The data shown here proves the existence of a small-world type of network. Mathematical models usually do not account for heterogeneity, meaning that the discrete nature of each individual patient can influence the overall structure of the network. As a result, it is expected that clusters will drive the epidemic forward because of the ease at which HIV can spread in spatial structures showing similar traits.

One of the potential risk factors implicated in viral spread was age of the patient at first diagnosis. Age is an interesting variable because, unlike percent diversity or class, age is an inherent discrete value from the patient at the time of visit. The plotted KDP revealed differences among the clusters classes clearly. It was observed that the median age of Singleton cluster groups was larger than SC or LC groups. However, LC showed mixing of two peaks of equal heights at different median ages. On the ECDF, it was observed that LC is consistently composed mostly of younger members and that SC appears to hover somewhere in between. Modelling is often complicated because it sometimes excludes variables like time and the different factors that go with it. As a result, it was decided to estimate the median age per year and look at how the trend varies during the observed period.

The correlation between PHI patients and the year of visit to a clinic was surprising. It appears that new cases of infections are carried out by increasing numbers of younger people as the years go by. In contrast, CUN patients are increasingly older by the year. A similar observation was noted in Dennis et al., 2018. Their investigation of HIV transmission patterns in Tennessee revealed that younger age is associated with cluster membership and that as time progresses, new clusters become populated by younger median age patients than non-clustered members

(100). A possible explanation of this trend could relate to the ease by which sexual partners can be found via the Internet using applications that facilitate proximity-based sexual partnering, such as GRINDR (101). As a result, technology and commercial resources far outpace public health efforts and the ability to target populations engaging in risky behaviour (102). Thus, young MSM populations who represent the most significant users of new technologies, engage in risky behaviour that have led to an expansion of the HIV epidemic and helps explain the trend observed from the past 15 years of collected data.

### 5.3 Large Cluster Species Maintain Similar Tropism and Replicative Capacity

For the viral tropism experiment (Figure 13), representative viral species from large clusters were chosen and compared against a panel of Singletons and other small cluster isolates. One of the first observations noted was that all of specimens, except for the pNL control and C196.02, grew in the R5-tropic cell line. This is representative of what is expected for viruses that are isolated from peripheral blood of individuals shortly after infection and during the asymptomatic phase (103). C050 represents the largest cluster in our phylogenetic tree, with 150 members since 2002 and still expanding. During that time, C050 had a peak in number of cases in 2005 and then followed by a slight resurgence in 2010. In the tropism assay, four samples belonging to C050 were tested, half of them came from visits in 2005 (C050.015 and C050.010) and the other half from visits in 2013 (C050.064 and C050.036). All these samples maintained R5-tropism despite the 8-year difference between them, which shows that following the onset of infection, clones belonging to transmission/founder viruses are overrepresented in the overall viral population. That was noted as well in C185 viruses, where both species maintained dual tropism as evidenced in their ability to show growth in the R5 and

X4 cell lines and syncytia induction in MT-2 cells. Secondly, an unexpected observation was noted when the program clustered the different strains by their ability to replicate in U87 cells. The code written in Python to create Figure 13 was built using a hierarchical clustering method based on minimal supervision. In the figure, clustering can be observed using values from the RT assays and then normalized to the highest grower for either R5 or X4. The program detected two subgroups, slow growers and high growers. Fast growers show high RT activity on R5 and X4 cell lines and are found either at the top or the bottom of the gradient, while slow growers are grouped in the middle of the histogram. It was estimated that 68% of the large cluster viruses fell in the high grower category, and those that fell in the slow grower subgroup were mostly represented by C050 viruses. HIV tropism has been largely associated with disease progression, where viruses that exclusively use CCR5 co-receptors are generally predominant during the early stages of infection (104). Evolution to the CXCR4 co-receptor usage is associated with low CD4<sup>+</sup> T-cell counts and accelerated disease progression. One method to determine HIV co-receptor usage is by evaluating the third variable region (V3) from the gp120 envelope glycoprotein. The presence of basic residues at positions 11 and 25 of this loop is strongly predictive of X4-tropism. It is possible that CXCR4 evolution could be influenced by the immune pressure that promotes host-specific adaptation, which limits the incidence of the R5 to X4 co-receptor switch (105). X4 viruses have been shown to be more pathogenic and virulent than R5 (106) and were also observed on the viruses tested in Figure 13. From our tested panel of viruses that were exclusively X4-tropic, they had been found only in the high grower bracket, thus proving their higher replication capacity. In conclusion, most of the viruses belonging to Large Cluster groups maintain similar tropism and replicative capacities. These properties could

play a part in the way the virus propagates within a community. As an example, C050 was found in the group of slow growers, yet has been able to maintain a continuous growth for almost 15 years uninterrupted. The properties its envelope offers could make it more fit to get past the epithelial barrier and other immune factors to find a susceptible cell in the host (107). The fact that only 3.5% of the total amount of quasispecies collected is responsible for half the total number of cases is remarkable, proving that viruses are being selected for special attributes. While not all these attributes are similar, these large cluster species are able of continuing their propagation by employing different strategies that enhance their survivability and transmissibility. Experiments in the future could focus on comparing the ability of large cluster viruses to be more infectious than singletons, or their susceptibility to interferons and other innate immune responses.

#### 5.4 Cluster 185 Viruses Show Decreased Susceptibility to Second-Generation INSTIs

Our work on the washout experiments using second-generation INSTIs represents an extension of work that had been performed previously in Osman et al., 2018. Originally the experiment was designed to compare the ability of RAL, EVG, and DTG against a panel of NL4.3 mutant strains. Results from DTG treatment and washout (Figure 15) compare to those obtained previously by our group. DTG was designed with a higher barrier to resistance and excellent potency showing near wild-type activity *in vitro* against resistance mutations like Q148H, N155H and Y143 (65). Not only is DTG stronger against INSTI-related resistance mutations than RAL or EVG, but it has also shown to dissociate slowly from an IN-DNA complex, which provides further evidence of their higher genetic barrier (108). This dissociation value could act as a proxy to determine the effects of a patient potentially failing to follow therapy and thus

develop resistance mutations. As explained previously, R263K offers low-level resistance and decreased replicative capacity than WT virus, therefore if a patient decides to stop taking DTG for a certain amount of time and return to it, even if the mutation emerges, it would not have a great impact upon resuming treatment. That is what was observed in Figure 15. First it was noted that R263K is capable of replicating 50% less efficiently than WT without any drug present. After drug washout, no traces of viruses harbouring the R263K were noted in the presence of all drugs or following washout on RT activity or proviral integration in qPCR. While most of the mutations that confer high-level resistance are found near the active site, the R263K mutation emerging in the presence of DTG is an exception as it's located near the C-terminal end of the protein. The mutation G118R challenged the efficiency of DTG following washout, managing to escape drug pressure to almost no drug levels on day 11. This pales in comparison to the efficiency of CAB and BIC in maintaining suppression in G118R. Both drugs successfully prevented replication of the virus even after the removal of the drug owing probably to better binding kinetics than DTG. G118R has previously been observed to emerge on DTG monotherapy, managing to reduce susceptibility >5-fold (109). CAB and BIC have been shown to have higher potency than RAL or EVG independent of the subtype (110). As a result, it was surprising to see that CAB was unable to suppress G140S/Q148H replication. The double mutant has been associated primarily with resistance to RAL, with Q148H acting as the resistance mutation and G140S rescuing viral fitness (111). The results obtained compare with previous observations where it was observed that G140S/Q148H manages to show replication levels on par with WT viruses (111), as seen in CAB treatment failure and after DTG removal. Our results for these mutations show that BIC represents the drug of choice for INSTI-

experienced patients who might have failed therapy on EVG or RAL regimes. Not only can BIC maintain viral suppression days after removal, but is also more efficient for IC50 values, showing a 3.50-fold increase when compared to 5.94-, 6.20-fold values for DTG and CAB respectively when facing the G140S/Q148H mutation.

A different picture is painted when facing the C185 isolates. As was expected, primary isolates tend to replicate at lower levels than laboratory strains, which was observed in Figure 14.

C185.09<sub>R263K</sub> isolates were remarkably slow in replicating, unable to achieve WT levels of replication at any point. One of the reasons why C185.09 viruses did not grow as well as C185.15 could be due to their inherently slow growth, as seen from the heatmap produced from the tropism experiment. The R263K mutation made C185 isolates less fit than they were already since this mutation is known to decrease replicative capacity by almost 50%. It was noted that C185.15<sub>R263K</sub> isolates showed abnormally high fold change values in comparison to WT. In previous experiments, R263K mutations in pNL4-3 backbone viruses have been shown to decrease susceptibility of DTG, RAL and EVG by 1.95, 1.23 and 3.28, respectively (112).

However, it was noticeable that C185.15<sub>R263K</sub> viruses were between 25 and 15-times less susceptible to better performing second-generation INSTIs. C185.15<sub>R263K</sub> was also capable of resuming integration of the proviral after drug removal of BIC and in some cases, capable of resuming minimal replication levels on DTG and BIC washouts. Previous experiments by our group have demonstrated that R263K, in combination with a naturally occurring polymorphism in position M50 increased resistance to DTG in tissue culture while also having an overlapping effect on BIC (113). Amino acid changes in this region are not very common, only observed in

13% of patients with a subtype B integrase (114); however both of our tested isolates harboured an M50I polymorphism, which could explain in part their tolerance to DTG and BIC.

**Table 5 Primary Isolate Integrases from Large Clusters with M50 Polymorphisms**

<b>Total LC Integrase Samples (n)</b>	<b>LC with M50 (n)</b>	<b>Singletons with M50 (n)</b>	<b>C045 with polymorphism</b>	<b>C118 with M50 polymorphism</b>	<b>C185 with M50 polymorphism</b>
<b>674</b>	149	22	5/7	19/22	22/23

In our database of Subtype B integrases, after selecting only untreated patients, it was observed that from a total of 674 patients considered to fall in a Large Cluster group, only 149 have an associated M50 polymorphism, representing 22% of the total sample. Looking at the three clusters shown to have a shared integrase (91), we observe that they represent 30% of the total number of M50 cases in large clusters. Thus, M50 polymorphism is over-represented in these three different clusters and could potentially mark a significant selection event that could enhance their solvability against integrase strand transfer inhibitors.

## Chapter 6 – Conclusions

HIV-1 epidemics are more complex to predict and evaluate than other diseases. By reconstructing the transmission networks from which the virus can spread, it is possible not only to forecast new cases of infection but also employ resources more efficiently by designing more dynamic public health campaigns. Phylodynamics complement classical epidemiological methods by associating each individual medical history with the type of viral species they are carrying. Through the reconstruction of phylogenies, we were able to see that there is some level of clustering among the MSM population and that these individuals show similar characteristics as expected from 'small-world' phenomena. Clustering was observed mostly in patients who are in primary infection status, and the trend has shown that recruitment into these clusters is driven by younger populations over the last 15 years. We showed that Large Clusters possess distinctive phenotypic properties that give them an advantage over singleton or small cluster variants. Large Clusters viruses contribute to larger viral loads and better replicating than singletons and transmit resistance mutations to antiretroviral therapy at a much higher frequency.

We demonstrated that BIC represents the best alternative for INSTI-experienced patients failing therapy and is capable of long-term suppression in highly resistant Q148H/Q140S mutants. The R263K mutation is known for not being a highly resistant mutation, which provides decreased replication levels when compared to WT virus. In our lab strains, R263K showed minimal fold-change values for second-generation INSTIs when compared to WT, however the C185.15<sub>R263K</sub> virus showed high values for IC<sub>50</sub> on all tested INSTIs. This virus was also capable of rebound at a certain level after BIC removal, as shown by the amount of proviral measured after 8 days

post-washout. It was found that the natural occurring polymorphisms in M50 residues are more present in C185 viruses and other species that share similar integrases. These polymorphisms could impact the success of therapy on DTG or BIC if R263K emerges at some point during treatment, and as a result it becomes more important than ever to properly genotype patients as soon as they go into the clinic so that they can be assigned a more personalized ART regime. Overall phylogenetics provides a different way of looking at HIV epidemics, where viral history and the populations involved can be studied and understood for better management of resources that could lead to an end in new cases of HIV.

## References

1. Sharp PM, Hahn BH. Origins of HIV and the AIDS Pandemic. Cold Spring Harbor Perspectives in Medicine. 2011;1(1).
2. C. GW. A history of AIDS: Looking back to see ahead. European Journal of Immunology. 2007;37(1):S94-S102.
3. Gottlieb MS, Schroff R, Schanker HM, Weisman JD, Fan PT, Wolf RA, et al. Pneumocystis carinii Pneumonia and Mucosal Candidiasis in Previously Healthy Homosexual Men. New England Journal of Medicine. 1981;305(24):1425-31.
4. Barre-Sinoussi F, Chermann J, Rey F, Nugeyre M, Chamaret S, Gruest J, et al. Isolation of a T-lymphotropic retrovirus from a patient at risk for acquired immune deficiency syndrome (AIDS). Science. 1983;220(4599):868-71.
5. Smith JH, Whiteside A. The history of AIDS exceptionalism. Journal of the International AIDS Society. 2010;13(1):47.
6. Kanki PJ, Travers KU. Slower heterosexual spread of HIV-2 than HIV-1. Lancet. 1994;343(8903):93.
7. Hemelaar J. Implications of HIV diversity for the HIV-1 pandemic. Journal of Infection. 2013;66(5):391-400.
8. Korber B, Gaschen B, Yusim K, Thakallapally R, Kesmir C, Detours V. Evolutionary and immunological implications of contemporary HIV-1 variation. British Medical Bulletin. 2001;58(1):19-42.
9. Lau KA, Wong JLL. Current Trends of HIV Recombination Worldwide. Infectious Disease Reports. 2013;5(Suppl 1):e4.
10. Taylor BS, Sobieszczyk ME, McCutchan FE, Hammer SM. The Challenge of HIV-1 Subtype Diversity. New England Journal of Medicine. 2008;358(15):1590-602.
11. Coffin JM. Structure, Replication, and Recombination of Retrovirus Genomes: Some Unifying Hypotheses. Journal of General Virology. 1979;42(1):1-26.
12. Watts JM, Dang KK, Gorelick RJ, Leonard CW, Bess JW, Swanstrom R, et al. Architecture and Secondary Structure of an Entire HIV-1 RNA Genome. Nature. 2009;460(7256):711-6.
13. Trono D. HIV accessory proteins: Leading roles for the supporting cast. Cell. 1995;82(2):189-92.
14. Freed EO. HIV-1 Replication. Somatic Cell and Molecular Genetics. 2001;26(1):13-33.
15. Ferguson MR, Rojo DR, von Lindern JJ, O'Brien WA. HIV-1 replication cycle. Clinics in Laboratory Medicine. 2002;22(3):611-35.
16. Bour S, Geleziunas R, Wainberg MA. The human immunodeficiency virus type 1 (HIV-1) CD4 receptor and its central role in promotion of HIV-1 infection. Microbiological Reviews. 1995;59(1):63-93.
17. Kwong PD, Wyatt R, Robinson J, Sweet RW, Sodroski J, Hendrickson WA. Structure of an HIV gp120 envelope glycoprotein in complex with the CD4 receptor and a neutralizing human antibody. Nature. 1998;393:648.
18. Moore JP, Trkola A, Dragic T. Co-receptors for HIV-1 entry. Current Opinion in Immunology. 1997;9(4):551-62.
19. Chan DC, Fass D, Berger JM, Kim PS. Core Structure of gp41 from the HIV Envelope Glycoprotein. Cell. 1997;89(2):263-73.
20. Arhel N. Revisiting HIV-1 uncoating. Retrovirology. 2010;7(1):96.
21. Telesnitsky A, Goff SP. Reverse Transcriptase and the Generation of Retroviral DNA. Cold Spring Harbor Laboratory Press, Cold Spring Harbor (NY); 1997.
22. Jacobo-Molina A, Arnold E. HIV reverse transcriptase structure-function relationships. Biochemistry. 1991;30(26):6351-61.
23. Hu W, Temin H. Retroviral recombination and reverse transcription. Science. 1990;250(4985):1227-33.

24. Zhan P, Liu X, De Clercq E. Blocking Nuclear Import of Pre-Integration Complex: An Emerging Anti-HIV-1 Drug Discovery Paradigm. *Current Medicinal Chemistry*. 2010;17(6):495-503.
25. Miller MD, Farnet CM, Bushman FD. Human immunodeficiency virus type 1 preintegration complexes: studies of organization and composition. *Journal of Virology*. 1997;71(7):5382-90.
26. Matreyek KA, Engelman A. Viral and Cellular Requirements for the Nuclear Entry of Retroviral Preintegration Nucleoprotein Complexes. *Viruses*. 2013;5(10):2483-511.
27. Jegede O, Babu J, Di Santo R, McColl DJ, Weber J, Quinones-Mateu M. HIV type 1 integrase inhibitors: from basic research to clinical implications. *AIDS reviews*. 2008;10(3):172-89.
28. Ruelas Debbie S, Greene Warner C. An Integrated Overview of HIV-1 Latency. *Cell*. 2013;155(3):519-29.
29. Karn J, Stoltzfus CM. Transcriptional and Posttranscriptional Regulation of HIV-1 Gene Expression. *Cold Spring Harbor Perspectives in Medicine*. 2012;2(2):a006916.
30. Zhou M, Deng L, Kashanchi F, Brady JN, Shatkin AJ, Kumar A. The Tat/TAR-dependent phosphorylation of RNA polymerase II C-terminal domain stimulates cotranscriptional capping of HIV-1 mRNA. *Proceedings of the National Academy of Sciences*. 2003;100(22):12666-71.
31. Hryckiewicz K, Bura M, Kowala-Piaskowska A, Bolewska B, Mozer-Lisewska I. HIV RNA splicing. *HIV & AIDS Review*. 2011;10(3):61-4.
32. Stoltzfus CM, Madsen JM. Role of Viral Splicing Elements and Cellular RNA Binding Proteins in Regulation of HIV-1 Alternative RNA Splicing. *Current HIV Research*. 2006;4(1):43-55.
33. Fischer U, Huber J, Boelens WC, Mattajt LW, Lührmann R. The HIV-1 Rev Activation Domain is a nuclear export signal that accesses an export pathway used by specific cellular RNAs. *Cell*. 1995;82(3):475-83.
34. Ohlmann T, Mengardi C, López-Lastra M. Translation initiation of the HIV-1 mRNA. *Translation (Austin, Tex)*. 2014;2(2):e960242-e.
35. Jackson RJ, Hellen CUT, Pestova TV. The mechanism of eukaryotic translation initiation and principles of its regulation. *Nature Reviews Molecular Cell Biology*. 2010;11:113.
36. Pelletier J, Sonenberg N. Internal initiation of translation of eukaryotic mRNA directed by a sequence derived from poliovirus RNA. *Nature*. 1988;334:320.
37. Sharma A, Yilmaz A, Marsh K, Cochrane A, Boris-Lawrie K. Thriving under Stress: Selective Translation of HIV-1 Structural Protein mRNA during Vpr-Mediated Impairment of eIF4E Translation Activity. *PLOS Pathogens*. 2012;8(3):e1002612.
38. Sundquist WI, Kräusslich H-G. HIV-1 assembly, budding, and maturation. *Cold Spring Harbor perspectives in medicine*. 2012;2(7):a006924-a.
39. Jouvenet N, Simon SM, Bieniasz PD. Imaging the interaction of HIV-1 genomes and Gag during assembly of individual viral particles. *Proceedings of the National Academy of Sciences*. 2009;106(45):19114-9.
40. Anderson J, Schiffer C, Lee S-K, Swanstrom R. Viral Protease Inhibitors. In: Kräusslich H-G, Bartenschlager R, editors. *Antiviral Strategies*. Berlin, Heidelberg: Springer Berlin Heidelberg; 2009. p. 85-110.
41. de Marco A, Müller B, Glass B, Riches JD, Kräusslich H-G, Briggs JAG. Structural Analysis of HIV-1 Maturation Using Cryo-Electron Tomography. *PLOS Pathogens*. 2010;6(11):e1001215.
42. Biancotto A, Iglehart SJ, Vanpouille C, Condack CE, Lisco A, Ruecker E, et al. HIV-1-induced activation of CD4+ T cells creates new targets for HIV-1 infection in human lymphoid tissue ex vivo. *Blood*. 2008;111(2):699-704.
43. Fauci AS. HIV and AIDS: 20 years of science. *Nature Medicine*. 2003;9:839.
44. Hernandez-Vargas EA, Middleton RH. Modeling the three stages in HIV infection. *Journal of Theoretical Biology*. 2013;320:33-40.

45. Hollingsworth TD, Anderson RM, Fraser C. HIV-1 Transmission, by Stage of Infection. *The Journal of Infectious Diseases*. 2008;198(5):687-93.
46. Cohen MS, Shaw GM, McMichael AJ, Haynes BF. Acute HIV-1 Infection. *New England Journal of Medicine*. 2011;364(20):1943-54.
47. Control CfD. Stages of HIV Infection 2016 [cited 2019. Available from: [https://wwwn.cdc.gov/hivrisk/what\\_is/stages\\_hiv\\_infection.html](https://wwwn.cdc.gov/hivrisk/what_is/stages_hiv_infection.html).
48. Services UDoHaH. The Stages of HIV Infection. 2019.
49. T-Cell Differentiation and Progression of HIV Infection. *PLOS Biology*. 2004;2(2):e41.
50. Garrido C, Villacian J, Zahonero N, Pattery T, Garcia F, Gutierrez F, et al. Broad Phenotypic Cross-Resistance to Elvitegravir in HIV-Infected Patients Failing on Raltegravir-Containing Regimens. *Antimicrobial Agents and Chemotherapy*. 2012;56(6):2873-8.
51. Cohen MS, Chen YQ, McCauley M, Gamble T, Hosseinipour MC, Kumarasamy N, et al. Prevention of HIV-1 Infection with Early Antiretroviral Therapy. *New England Journal of Medicine*. 2011;365(6):493-505.
52. Volberding PA, Deeks SG. Antiretroviral therapy and management of HIV infection. *The Lancet*. 2010;376(9734):49-62.
53. O'Brien SJ, Hendrickson SL. Host genomic influences on HIV/AIDS. *Genome Biology*. 2013;14(1):201.
54. Zhao H, Neamati N, Sunder S, Hong H, Wang S, Milne GW, et al. Hydrazide-containing inhibitors of HIV-1 integrase. *Journal of medicinal chemistry*. 1997;40(6):937-41.
55. Mouscadet J-F, Tchertanov L. Raltegravir: molecular basis of its mechanism of action. *European journal of medical research*. 2009;14 Suppl 3(Suppl 3):5-16.
56. Grobler JA, Stillmock K, Hu B, Witmer M, Felock P, Espeseth AS, et al. Diketo acid inhibitor mechanism and HIV-1 integrase: Implications for metal binding in the active site of phosphotransferase enzymes. *Proceedings of the National Academy of Sciences*. 2002;99(10):6661-6.
57. Da-Yong L, Hong-Ying W, Nagendra Sastry Y, Bin X, Jian D, Ting-Ren L. HAART in HIV/AIDS Treatments: Future Trends. *Infectious Disorders - Drug Targets*. 2018;18(1):15-22.
58. World Health Organization USCfDCaP, The Global Fund to Fight AIDS, Tuberculosis and Malaria. HIV Drug Resistance Report 2017. World Health Organization Publications 2017.
59. Organization WH. Global Action Plan on HIV Drug Resistance 2017-2021: 2018 progress report. World Health Organization Publications 2018.
60. Métifiot M, Marchand C, Maddali K, Pommier Y. Resistance to integrase inhibitors. *Viruses*. 2010;2(7):1347-66.
61. da Silva D, Van Wesebeeck L, Breilh D, Reigadas S, Anies G, Van Baelen K, et al. HIV-1 resistance patterns to integrase inhibitors in antiretroviral-experienced patients with virological failure on raltegravir-containing regimens. *Journal of Antimicrobial Chemotherapy*. 2010;65(6):1262-9.
62. Wills T, Vega V. Elvitegravir: a once-daily inhibitor of HIV-1 integrase. *Expert Opinion on Investigational Drugs*. 2012;21(3):395-401.
63. Abram ME, Hluhanich RM, Goodman DD, Andreatta KN, Margot NA, Ye L, et al. Impact of Primary Elvitegravir Resistance-Associated Mutations in HIV-1 Integrase on Drug Susceptibility and Viral Replication Fitness. *Antimicrobial Agents and Chemotherapy*. 2013;57(6):2654-63.
64. Dow DE, Bartlett JA. Dolutegravir, the Second-Generation of Integrase Strand Transfer Inhibitors (INSTIs) for the Treatment of HIV. *Infectious Diseases and Therapy*. 2014;3(2):83-102.
65. Hare S, Smith SJ, Métifiot M, Jaxa-Chamiec A, Pommier Y, Hughes SH, et al. Structural and Functional Analyses of the Second-Generation Integrase Strand Transfer Inhibitor Dolutegravir (S/GSK1349572). *Molecular Pharmacology*. 2011;80(4):565-72.
66. Bailly F, Cotellet P. The preclinical discovery and development of dolutegravir for the treatment of HIV. *Expert Opinion on Drug Discovery*. 2015;10(11):1243-53.

67. Schreiner E, Richter F, Nerdinger S. Development of Synthetic Routes to Dolutegravir. In: Časar Z, editor. *Synthesis of Heterocycles in Contemporary Medicinal Chemistry*. Cham: Springer International Publishing; 2016. p. 187-208.
68. Cahn P, Pozniak AL, Mingrone H, Shuldyakov A, Brites C, Andrade-Villanueva JF, et al. Dolutegravir versus raltegravir in antiretroviral-experienced, integrase-inhibitor-naïve adults with HIV: week 48 results from the randomised, double-blind, non-inferiority SAILING study. *The Lancet*. 2013;382(9893):700-8.
69. Castagna A, Maggiolo F, Penco G, Wright D, Mills A, Grossberg R, et al. Dolutegravir in Antiretroviral-Experienced Patients With Raltegravir- and/or Elvitegravir-Resistant HIV-1: 24-Week Results of the Phase III VIKING-3 Study. *The Journal of Infectious Diseases*. 2014;210(3):354-62.
70. Quashie PK, Mesplède T, Han Y-S, Oliveira M, Singhroy DN, Fujiwara T, et al. Characterization of the R263K Mutation in HIV-1 Integrase That Confers Low-Level Resistance to the Second-Generation Integrase Strand Transfer Inhibitor Dolutegravir. *Journal of Virology*. 2012;86(5):2696-705.
71. Inc GS. Gilead Presents New Data on Biktarvy 2019 [Available from: <https://www.gilead.com/news-and-press/press-room/press-releases/2019/3/gilead-presents-new-data-on-biktarvy-bictegravir-emtricitabine-and-tenofovir-alafenamide-and-tafbased-regimens-for-the-treatment-of-hiv1-in-chil>].
72. Healthcare V. ViiV HEALTHCARE SUBMITS NEW DRUG APPLICATION TO US FDA FOR THE FIRST MONTHLY, INJECTABLE, TWO-DRUG REGIMEN OF CABOTEGRAVIR AND RILPIVIRINE FOR TREATMENT OF HIV 2019 [Available from: <https://www.viivhealthcare.com/en-gb/media/press-releases/2019/april/viiv-healthcare-submits-new-drug-application-to-us-fda-for-the-first-monthly-injectable-two-drug-regimen-of-cabotegravir-and-rilpivirine-for-treatment-of-hiv/>].
73. Gulis G, Fujino Y. Epidemiology, population health, and health impact assessment. *J Epidemiol*. 2015;25(3):179-80.
74. Volz EM, Koelle K, Bedford T. Viral Phylodynamics. *PLOS Computational Biology*. 2013;9(3):e1002947.
75. Shiino T. Phylodynamic analysis of a viral infection network. *Frontiers in Microbiology*. 2012;3(278).
76. Rothenberg R. HIV transmission networks. *Current opinion in HIV and AIDS*. 2009;4(4):260-5.
77. Lewis F, Hughes GJ, Rambaut A, Pozniak A, Leigh Brown AJ. Episodic Sexual Transmission of HIV Revealed by Molecular Phylodynamics. *PLOS Medicine*. 2008;5(3):e50.
78. Ragonnet-Cronin M, Hu YW, Morris SR, Sheng Z, Poortinga K, Wertheim JO. HIV transmission networks among transgender women in Los Angeles County, CA, USA: a phylogenetic analysis of surveillance data. *The Lancet HIV*. 2019;6(3):e164-e72.
79. Brenner BG, Roger M, Stephens D, Moisi D, Hardy I, Weinberg J, et al. Transmission Clustering Drives the Onward Spread of the HIV Epidemic Among Men Who Have Sex With Men in Quebec. *The Journal of Infectious Diseases*. 2011;204(7):1115-9.
80. Aldous JL, Pond SK, Poon A, Jain S, Qin H, Kahn JS, et al. Characterizing HIV Transmission Networks Across the United States. *Clinical Infectious Diseases*. 2012;55(8):1135-43.
81. Hué S, P Clewley J, Cane P, Pillay D. HIV-1 pol gene variation is sufficient for reconstruction of transmissions in the era of antiretroviral therapy. *AIDS (London, England)*. 2004;18:719-28.
82. Watts DJ, Strogatz SH. Collective dynamics of 'small-world' networks. *Nature*. 1998;393(6684):440-2.
83. Keeling Matt J, Eames Ken TD. Networks and epidemic models. *Journal of The Royal Society Interface*. 2005;2(4):295-307.
84. Ratmann O, Grabowski MK, Hall M, Golubchik T, Wymant C, Abeler-Dörner L, et al. Inferring HIV-1 transmission networks and sources of epidemic spread in Africa with deep-sequence phylogenetic analysis. *Nature Communications*. 2019;10(1):1411.

85. Parker ZF, Iyer SS, Wilen CB, Parrish NF, Chikere KC, Lee F-H, et al. Transmitted/Founder and Chronic HIV-1 Envelope Proteins Are Distinguished by Differential Utilization of CCR5. *Journal of Virology*. 2013;87(5):2401-11.
86. Tully DC, Ogilvie CB, Batorsky RE, Bean DJ, Power KA, Ghebremichael M, et al. Differences in the Selection Bottleneck between Modes of Sexual Transmission Influence the Genetic Composition of the HIV-1 Founder Virus. *PLOS Pathogens*. 2016;12(5):e1005619.
87. Joseph SB, Swanstrom R, Kashuba ADM, Cohen MS. Bottlenecks in HIV-1 transmission: insights from the study of founder viruses. *Nature Reviews Microbiology*. 2015;13:414.
88. Haddad N LK, Totten S, McGuire M. HIV In Canada - Surveillance Report, 2017. In: Canada PHAo, editor. *Can Commun Dis Rep* 2017. p. 8.
89. Canada PHAo. Update on HIV-1 Strain and Transmitted Drug Resistance in Canada: 2012-2013. In: *Control CfCDal*, editor. 2017. p. 48.
90. Brenner Bluma GB. Future of phylogeny in HIV prevention. *JAIDS Journal of Acquired Immune Deficiency Syndromes*. 2013;63:248-54.
91. Brenner BG, Ibanescu R-I, Oliveira M, Roger M, Hardy I, Routy J-P, et al. HIV-1 strains belonging to large phylogenetic clusters show accelerated escape from integrase inhibitors in cell culture compared with viral isolates from singleton/small clusters. *Journal of Antimicrobial Chemotherapy*. 2017;72(8):2171-83.
92. UNAIDS. 90-90-90 An Ambitious Treatment Target to help end the AIDS Epidemic. Joint United Nations Programme on HIV/AIDS; 2017.
93. Brenner BG, Ibanescu R-I, Hardy I, Stephens D, Otis J, Moodie E, et al. Large cluster outbreaks sustain the HIV epidemic among MSM in Quebec. *AIDS*. 2017;31(5):707-17.
94. Osman N, Mesplède T, Oliveira M, Hassounah S, Wainberg MA, Brenner BG. Durable suppression of HIV-1 with resistance mutations to integrase inhibitors by dolutegravir following drug washout. *AIDS*. 2018;32(13):1773-80.
95. Klov Dahl AS, Potterat JJ, Woodhouse DE, Muth JB, Muth SQ, Darrow WW. Social networks and infectious disease: The Colorado Springs study. *Social Science & Medicine*. 1994;38(1):79-88.
96. Rothenberg RB, Potterat JJ, Woodhouse DE, Muth SQ, Darrow WW, Klov Dahl AS. Social network dynamics and HIV transmission. *AIDS*. 1998;12(12):1529-36.
97. Muñoz-Martínez EJ. *Small Worlds: The Dynamics of Networks Between Order and Randomness*, by Duncan J. Watts, (Princeton Studies in Complexity), Princeton University Press, 1999. \$39.50 (hardcover), 262 pp. ISBN: 0-691-00541-9. (Book Reviews). *Bulletin of Mathematical Biology*. 2000;62(4):794-6.
98. Brenner BG, Roger M, Routy J-P, Moisi D, Ntemgwa M, Matte C, et al. High Rates of Forward Transmission Events after Acute/Early HIV-1 Infection. *The Journal of Infectious Diseases*. 2007;195(7):951-9.
99. Gumel AB, Castillo-Chavez C, Mickens RE, Clemence DP. *Mathematical studies on human disease dynamics: emerging paradigms and challenges*: American Mathematical Society; 2006.
100. Dennis AM, Volz E, Frost ASMSDW, Hossain M, Poon AFY, Rebeiro PF, et al. HIV-1 Transmission Clustering and Phylodynamics Highlight the Important Role of Young Men Who Have Sex with Men. *AIDS research and human retroviruses*. 2018;34(10):879-88.
101. Hunter P, Oyervides O, Grande KM, Prater D, Vann V, Reitzl I, et al. Facebook-augmented partner notification in a cluster of syphilis cases in Milwaukee. *Public Health Rep*. 2014;129 Suppl 1(Suppl 1):43-9.
102. Hightow-Weidman LB, Muessig KE, Bauermeister J, Zhang C, LeGrand S. Youth, Technology, and HIV: Recent Advances and Future Directions. *Curr HIV/AIDS Rep*. 2015;12(4):500-15.
103. Berger EA, Murphy PM, Farber JM. CHEMOKINE RECEPTORS AS HIV-1 CORECEPTORS: Roles in Viral Entry, Tropism, and Disease. *Annual Review of Immunology*. 1999;17(1):657-700.

104. Seclén E, Soriano V, González MM, Martín-Carbonero L, Gellermann H, Distel M, et al. Impact of Baseline HIV-1 Tropism on Viral Response and CD4 Cell Count Gains in HIV-Infected Patients Receiving First-line Antiretroviral Therapy. *The Journal of Infectious Diseases*. 2011;204(1):139-44.
105. Sede MM, Moretti FA, Laufer NL, Jones LR, Quarleri JF. HIV-1 Tropism Dynamics and Phylogenetic Analysis from Longitudinal Ultra-Deep Sequencing Data of CCR5- and CXCR4-Using Variants. *PLOS ONE*. 2014;9(7):e102857.
106. Hazenberg MD, Otto SA, Hamann D, Roos MT, Schuitemaker H, de Boer RJ, et al. Depletion of naive CD4 T cells by CXCR4-using HIV-1 variants occurs mainly through increased T-cell death and activation. *AIDS*. 2003;17(10):1419-24.
107. Burgener A, McGowan I, Klatt NR. HIV and mucosal barrier interactions: consequences for transmission and pathogenesis. *Current Opinion in Immunology*. 2015;36:22-30.
108. Hightower KE, Wang R, Deanda F, Johns BA, Weaver K, Shen Y, et al. Dolutegravir (S/GSK1349572) exhibits significantly slower dissociation than raltegravir and elvitegravir from wild-type and integrase inhibitor-resistant HIV-1 integrase-DNA complexes. *Antimicrobial agents and chemotherapy*. 2011;55(10):4552-9.
109. Rhee S-Y, Grant PM, Tzou PL, Barrow G, Harrigan PR, Ioannidis JPA, et al. A systematic review of the genetic mechanisms of dolutegravir resistance. *J Antimicrob Chemother*. 2019;74(11):3135-49.
110. Neogi U, Singh K, Aralaguppe SG, Rogers LC, Njenda DT, Sarafianos SG, et al. Ex-vivo antiretroviral potency of newer integrase strand transfer inhibitors cabotegravir and bictegravir in HIV type 1 non-B subtypes. *AIDS (London, England)*. 2018;32(4):469-76.
111. Delelis O, Malet I, Na L, Tchertanov L, Calvez V, Marcelin A-G, et al. The G140S mutation in HIV integrases from raltegravir-resistant patients rescues catalytic defect due to the resistance Q148H mutation. *Nucleic Acids Res*. 2009;37(4):1193-201.
112. Mesplède T, Quashie PK, Osman N, Han Y, Singhroy DN, Lie Y, et al. Viral fitness cost prevents HIV-1 from evading dolutegravir drug pressure. *Retrovirology*. 2013;10(1):22.
113. Pham HT, Mesplède T. Bictegravir in a fixed-dose tablet with emtricitabine and tenofovir alafenamide for the treatment of HIV infection: pharmacology and clinical implications. *Expert Opinion on Pharmacotherapy*. 2019;20(4):385-97.
114. Wares M, Mesplède T, Quashie PK, Osman N, Han Y, Wainberg MA. The M50I polymorphic substitution in association with the R263K mutation in HIV-1 subtype B integrase increases drug resistance but does not restore viral replicative fitness. *Retrovirology*. 2014;11:7-.