Genetic diversity and baseline drug resistance of South African HIV-1 Integrase sequences prior to the availability of Integrase strand-transfer inhibitors

–

Genetische Variabilität und medikamentöse Resistenz südafrikanischer HIV-1 Integrase Sequenzen vor der Verfügbarkeit von Integrase Strang-Transfer Inhibitoren

Doctoral thesis for a medical doctoral degree
at the Graduate School of Life Sciences,
Julius-Maximilians-Universität Würzburg,
Section Infection and Immunity
submitted by

Dominik Alexander Brado
from
Heidelberg

Würzburg 2020

Submitted on:

04 March 2020

Office stamp

Members of the Thesis Committee:

Chairperson: Prof. Dr. Klaus Brehm, Institute for Hygiene and Microbiology, Medical faculty (Julius-Maximilians-Universität Würzburg)

Primary Supervisor: Prof. Dr. Thomas Dandekar, Department of Bioinformatics (Julius-Maximilians-Universität Würzburg)

Supervisor (Second): Dr. Graeme Brendon Jacobs Division of Medical Virology, Faculty of Medicine and Health Sciences (Stellenbosch University)

Supervisor (Third): Prof. Dr. Carsten Scheller, Institute for Virology and Immunobiology, Medical faculty (Julius-Maximilians-Universität Würzburg)

Supervisor (Fourth): Prof. Dr. Hartwig Klinker, Department of Medicine II, Division of Infectious Diseases (Julius-Maximilians-Universität Würzburg)

Date of Public Defence: 13 November 2020

Date of Receipt of Certificates:

…….

Affidavit

I hereby confirm that my thesis entitled "Genetic diversity and baseline drug resistance of South African HIV-1 Integrase sequences prior to the availability of Integrase strand-transfer inhibitors – Genetische Variabilität und medikamentöse Resistenz südafrikanischer HIV-1 Integrase Sequenzen vor der Verfügbarkeit von Integrase Strang-Transfer Inhibitoren" is the result of my own work. I did not receive any help or support from commercial consultants. All sources and / or materials applied are listed and specified in the thesis.

Furthermore, I confirm that this thesis has not yet been submitted as part of another examination process neither in identical nor in similar form.

Wuerzburg, 04.03.2020

Place, Date Signature

List of scientific contributions

Scientific publications included in this thesis

- Analyses of HIV-1 integrase sequences prior to South African national HIV treatment program and available of integrase inhibitors in Cape Town, South Africa, ***Dominik Brado****, Adetayo Emmanuel Obasa, George Mondinde Ikomey, Ruben Cloete, Kamalendra Singh, Susan Engelbrecht, Ujjwal Neogi & Graeme Brendon Jacobs,* Scientific reports 2018 Mar 16;8(1):4709. doi: 10.1038/s41598018-22914-5 Sci Rep. 2018 Apr 16;8(1):6262. PMID: 29549274.

Scientific publications not included in this thesis

- HIV-1 Integrase Diversity and Resistance-Associated Mutations and Polymorphisms among Integrase strand transfer Inhibitor-Naïve HIV-1 Patients from Cameroon, *Mr. Sello Given Mikasi, Mr. Otwoma Josiah Gichana, Ms. Cheri Van der walt, Mr.* ***Dominik Brado****, Mr. Adetayo Emmanuel Obasa, Mr. Duncan Njenda, Dr. Martha Messembe, Dr. Emilia Lyonga, Prof. Okomo Assoumou, Dr. Ruben Cloete, Dr. George Mondinde Ikomey, and Dr. Graeme Jacobs,* AIDS Research and Human Retroviruses 2019 Dec https://doi.org/10.1089/AID.2019.0264.

Poster presentations

- 12th international Eureka! symposium 2017 entitled "Prevalence of primary HIV-1 Integrase drug resistance mutations and naturally occurring polymorphisms in a cohort from Cape Town, South Africa, before the introduction of Integrase strand transfer inhibitors to the national HIV-treatment program" ***Brado D.****, Obasa AEA, Scheller C., Preiser W., Jacobs GB.*

Research visits

- Research visit to the Division of Medical Virology, Department of Pathology, University of Stellenbosch and NHLS Tygerberg, South Africa. 27th of August 2016 – 24th of March 2017. The aim of this research visit was to conduct the laboratory work on the provided blood samples.

Summary

Background: Integrase strand transfer inhibitors (INSTIs) are the latest addition to the array of antiretroviral compounds used to treat an infection with Human Immunodeficiency Virus (HIV). Due to their high efficacy and increased tolerability, INSTIs have become an integral part of first-line therapy in most high-income countries over the past years. However, little is known about HIV-1's genetic inter- and intra-subtype diversity on the Integrase (IN)-gene and its impact on the emergence of INSTI-resistance. In the absence of a functional cure, long-term efficacy of first-line compounds remains paramount for reducing virological failure and curbing on-going HIV transmissions. South Africa, harbouring more than 20% of the global HIV burden (7.7 / 37.9 million people), requires international attention in order to globally pursue UNAIDS' (Joint United Nations Programme on HIV/AIDS) 90-90-90 goals and the road to ending the HIV/AIDS (Acquired immunodeficiency syndrome) pandemic by 2030.

Methods: In this study, the prevalence of INSTI-resistance associated mutations (RAM) was investigated in a cohort of 169 archived drug-naïve blood samples from multiple collection sites around Cape Town, South Africa. Viral RNA was isolated from plasma samples, the integrase fragment amplified by RT-PCR and subsequently sequenced by Sanger-sequencing. Additionally, all publicly available drug-naïve, South African IN sequences, isolated before the availability of the first INSTIs in 2007, were retrieved from the Los Alamos HIV sequence database (n=284). All sequences were analysed for RAMs using the Stanford HIV Drug resistance database. The identification of polymorphism in the South African subtype C IN consensus sequence allowed for comparative analyses with global subtype B, as well as subtype C sequences, from countries other than South Africa.

Results: The IN gene could be amplified and sequenced in 95/169 samples (56%). Phylogenetic inference revealed close homology between three sequence-pairs, warranting the exclusion of 3/95 sequences from further analyses. Of the 92 samples used for mutational analyses, 86/92 (93.5%) belonged to subtype C, 5/92 (5.4%) to subtype B and 1/92 (1.1%) to subtype A. The prevalence of major and accessory INSTI RAMs was 0/92 (0%) and 1/91 (1.1%), respectively, similar to the observed rates of 8/284 (2.8%) and 8/284 (2.8%) in the database sequences ($p = 0.2076$ and $p = 0.6944$, Fisher's exact test). Compared to subtype B IN sequences, 15 polymorphisms were significantly enriched in South African subtype C sequences (corrected $p<0.0015$. Fisher's exact test, Bonferroni post-hoc procedure).

Compared to subtype C IN sequences isolated outside South Africa, four polymorphisms were significantly enriched in this study cohort (corrected $p<0.0014$, Fisher's exact test, Bonferroni post-hoc procedure). The highest prevalence margin was observed for the polymorphism Met50Ile being present in 60.1% of South African subtype C sequences, compared to 37% in non-South African subtype C sequences.

Conclusions: The low prevalence of major and minor RAMs in all South African Integrase sequences predicts a high susceptibility to INSTIs, however, the presence of natural polymorphisms, in particular Met50Ile, in the majority of sequences warrants further monitoring under therapeutic pressure, as their role in mutational pathways leading to INSTI-resistance is yet to be determined. Additionally, this study revealed the presence of substantial inter- and intra-subtype diversity within the HIV-1 Subtype C IN-gene. These results implicate the need for more research on a regional, potentially patient-specific level, as mutational insights from other diverse backgrounds may not accurately represent the South African context. The implementation of a national pre-treatment INSTI-resistance screening program may provide necessary insights into the development of mutational pathways leading to INSTI-resistance under therapeutic pressure for the South African context and thereby bring South Africa one step closer to achieving UNAIDS 90-90-90 goals and ending the AIDS epidemic by 2030.

Zusammenfassung

Hintergrund: Integrase Strang-Transfer Inhibitoren (INSTIs) sind die neuste medikamentöse Ergänzung in der Therapie einer HIV-Infektion. Auf Grund ihrer starken Wirksamkeit und eines guten Nebenwirkungsprofils sind INSTIs in den letzten Jahren ein integraler Bestandteil von Erstlinien-Therapieregimen in den meisten wirtschaftlich starken Ländern geworden. Allerdings ist wenig bekannt über die genetische Variabilität des IN-gens und über ihren Einfluss auf die Entwicklung von INSTI-Resistenzen. Mit einem Anteil von über 20% der globalen HIV-Last (7,7 / 37,9 Millionen Menschen) benötigt Südafrika einen internationalen Fokus, um die von UNAIDS formulierten 90-90-90 Ziele und das mögliche Ende der HIV/AIDS Pandemie bis 2030 auf globaler Ebene zu verfolgen.

Methoden: In dieser Arbeit wurde die Prävalenz von INSTI RAMs in einer Kohorte von 169 archivierten, therapie-naiven Blutproben von mehreren Sammelstellen um Kapstadt, Südafrika, untersucht. Virale RNA wurde aus Plasmaproben isoliert, das Integrase-Fragment mittels RT-PCR amplifiziert und anschließend Sanger-sequenziert. Zusätzlich wurden alle in der Los Alamos HIV Sequenz Datenbank verfügbaren, therapie-naive, südafrikanische IN Sequenzen, die vor der Verfügbarkeit von INSTIs im Jahr 2007 isoliert wurden, der Analyse dieser Arbeit hinzugefügt (n=284). Die Interpretation der gefundenen Mutationen erfolgte mittels der HIV Therapie-Resistenz Datenbank der Stanford Universität. Durch Generierung eines südafrikanischen IN Subtyp C Consensus-Stranges und nachfolgendem Vergleich mit öffentlich verfügbaren Subtyp B und Subtyp C Sequenzen, die außerhalb Südafrikas isoliert wurden, erfolgte die Analyse von natürlich vorkommenden Polymorphismen.

Ergebnisse: Das IN-Fragment konnte in 95/169 Plasmaproben (56%) erfolgreich amplifiziert und sequenziert werden. Phylogenetische Analysen zeigten eine enge Homologie zwischen drei Sequenz-Paaren, woraufhin 3/95 Sequenzen von weiteren Analysen ausgeschlossen wurden. Von den übrigen 92 Sequenzen gehörten 86/92 (93,5%) zu dem Subtyp C, 5/92 (5,4%) zu dem Subtyp B und 1/91 (1,1%) zu dem Subtyp A. Die Prävalenz von Haupt- und Nebenresistenz-Mutationen lag bei jeweils 0/92 (0%) und 1/92 (1,1%). Ähnliche Raten hierfür von 8/284 (2,8%) und 8/284 (2,8%) konnten in den Datenbank-Sequenzen beobachtet werden ($p = 0,2076$ und $p = 0,6944$, Fisher's exact test). Im Vergleich zu Subtyp B IN Sequenzen waren 15 Polymorphismen signifikant erhöht in südafrikanischen Subtype C IN Sequenzen (korrigiertes $p<0,0015$, Fisher's exact test, Bonferroni post-hoc Korrektur). Im

Vergleich zu nicht-südafrikanischen Subtyp C Sequenzen zeigten sich vier Polymorphismen signifikant erhöht (korrigiertes $p<0,0014$, Fisher's exact test, Bonferroni post-hoc Korrektur). Der größte Prävalenzunterschied konnte für den Polymorphismus Met50Ile beobachtet werden. Dieser war vorhanden in 217/361 (60,1%) der südafrikanischen Subtyp C Sequenzen, verglichen zu 203/548 (37.0%) der nicht-südafrikanischen Subtyp C Sequenzen.

Schlussfolgerung: Die niedrige Prävalenz von Haupt- und Neben-RAMs in südafrikanischen IN-Sequenzen verspricht ein gutes Ansprechen von INSTIs in diesem Kontext. Allerdings bedingt das Vorhandensein von natürlichen Polymorphismen, insbesondere der Polymorphismus Met50Ile das weitere Beobachten dieser Mutationen unter dem Einfluss von therapeutischem Druck, da deren Bedeutung in der Entwicklung von INSTI-Resistenzen noch nicht abschließend geklärt werden konnte. Zudem impliziert die in dieser Arbeit gezeigte inter- und intra-subtyp Diversität auf dem IN-Gen, die Notwendigkeit von weiterer Forschung auf regionaler Ebene, da Beobachtungen, die auf verschiedenen polymorphistischen Kontexten beruhen, nicht notwendigerweise auf den südafrikanischen Kontext übertragen werden können. Mit der Einführung eines nationalen, prä-therapeutischen Screening-Programms für das Vorhandensein von INSTI-Resistenzen könnte Südafrika wichtige Einblicke in die Entwicklung von INSTI-Resistenzen gewinnen und somit den 90-90-90 Zielen und der Möglichkeit die AIDS-Pandemie bis zum Jahr 2030 zu beenden, einen Schritt näher sein.

List of abbreviations

©	Copyright
®	Registered
(+)	Positively sensed
(-)	Negatively sensed
/r	Ritonavir-boosted
3TC	Lamivudine
°C	Degree Celsius
µl	Microliter
A	Adenosine
Å	Ångström
ABC	Abacavir
AIDS	Acquired Immunodeficiency Syndrome
AHI	Acute human immunodeficiency virus infection
Ala	Alanine
APOBEC	Apolipoprotein B mRNA Editing Catalytic Polypeptide-like
Arg	Arginine
ART	Antiretroviral therapy
Asn	Asparagine
Asp	Aspartic acid
BIC	Bictegravir
BLAST	Basic Local alignment Search Tool
bp	Base pairs
C	Cytidine
cART	Combined antiretroviral therapy
CCD	Catalytic core domain
CCR5	C-C chemokine receptor type 5
cDNA	Complementary DNA
CD	Cluster of differentiation

CPZ	Chimpanzees
CRF	Circulating Recombinant Form
CTD	C-terminal end
CTL	Cytotoxic T-Lymphocytes
CXCR4	CXC chemokine receptor type 4
CYP3A4	Cytochrom P450 3A4
Cys	Cystidine
d4T	Stavudine
DC	Dentritic cell
ddI	Didanosine
ddNTP	Di-deoxynucleoside triphosphate
dNTP	Deoxynucleoside triphosphate
DNA	Deoxyribonucleic acid
DOR	Doravirine
DRC	Democratic Republic of the Congo
DRM	Drug-resistance mutation
DRV	Darunavir
DTG	Dolutegravir
EC_{50}	Half-maximal effective concentrations
EDTA	Ethylene diamine tetra-acetic acid
EFV	Efavirenz
ELISA	Enzyme-linked immunosorbent assay
e.g.	Exempli gratia
env	Envelope
ETR	Etravirine
EVG	Elvitegravir
FDA	U.S. Food and Drug Administration
FTC	Emtricitabine
G	Guanosine
g	Gram
gag	Group-specific antigen
Gln	Glutamine

Glu	Glutamic acid
Gly	Glycine
gp	Glycoproteins
His	Histidine
HIV-1	Human Immunodeficiency Virus 1
HIV-1 A	Human Immunodeficiency Virus 1 subtype A
HIV-1B	Human Immunodeficiency Virus 1 subtype B
HIV-1C	Human Immunodeficiency Virus 1 subtype C
HIV-2	Human Immunodeficiency Virus 2
HIV-1 C_{ZA}	Human Immunodeficiency Virus -1 subtype C from South Africa
HPSC	Hematopoetic-Stemcell
HTLV-III	Human T-cell Lymphotropic Type III
ICTV	International Committee on the Taxonomy of Viruses
IgG	Class G immunoglobulins
IgM	Class M immunoglobulins
Ile	Isoleucine
IN	Integrase
INSTI	Intergrase strand-transfer inhibitor
jpHMM	Jumping profile Hidden Markov Model
kb	Kilo bases
kDa	Kilodalton
kg	Kilogram
LANL	Los Alamos National Laboratory
LAV	Lymphadenopathy Associated Virus
LEDGF/p75	Lens-epithelium-derived growth factor
Leu	Leucine

LPC	Lymphoid progenitor cells
LPS	Lipopolysaccharide
LTR	Long Terminal Repeats
Lys	Lysine
Lys3 tRNA	Lysine's tRNA
MEGA	Molecular Evolutionary Genetics Analysis
Met	Methionine
Mg	Magnesium
MHC-I	Major histocompatibility complex type I
min	Minutes
ml	Mililiter
mM	Millimolar
mRNA	Messenger RNA
MTCT	Mother-to-child-transmission
mV	Milivolt
Nef	Negative Factor
ng	Nanogram
nm	Nanometer
nM	Nanomolar
NNRTI	Non-nucleoside reverse-transcriptase inhibitor
NNIBP	Non-nucleoside reverse-transcriptase inhibitor binding pocket
NRTI	Nucleoside reverse-transcriptase inhibitor
NTD	N-Terminal end
NtRTI	Nucleotide reverse-transcriptase inhibitor
NVP	Nevirapine
ORF	Open Reading Frames
PAMP	Pathogen-associated-molecular-

	pattern
PBS	Primer binding site
PCP	Pneumocystis carinii pneumonia
PCR	Polymerase Chain Reaction
Phe	Phenylalanine
PIC	Pre-Integration complex
PI	Protease inhibitor
PLWH	People living with HIV
pol	Polymerase gene
PR	Protease
Pro	Proline
RAL	Raltegravir
RAM	Resistance-associated mutations
rev	Regulator of Virion Expression
RNA	Ribonucleic acid
RNase H	Ribonuclease H
RPV	Rilpivirine
RT	Reverse Transcriptase
RT-PCR	Reverse transcriptase polymerase-chain reaction
RPM	Revolution per minute
s	see
sec	Seconds
Ser	Serine
SIV	Simian Immunodeficiency Virus
Smm	Sooty mangabeys
SSP	Stable synaptic complex
T	Thymidine
Tat	Transcriptional transactivator gene
TAF	Tenofovir alafenamide
TDF	Tenofovir disoproxil fumarate
TDR	Transmitted drug-resistance
TFV	Tenofovir

Thr	Threonine
Tyr	Tyrosine
tRNA	Transport ribonucleic acid
™	Trademark
TP	Triphosphate
UNAIDS	Joint United Nations Programme on HIV/AIDS
URF	Unique Recombinant Forms
USA	United States of America
UV	Ultraviolet
Val	Valine
Vif	Viral Infectivity Factor gene
Vpr	Viral Protein R gene
Vpu	Viral Protein U gene
WHO	World Health Organisation
ZDV/AZT	Zidovudine
Zn	Zink

List of figures

Fig. 1.1 HIV-1 gene map, displaying HXB2's genomic organisation according to the Los Alamos HIV sequence database (Los Alamos HIV sequence database, 1998) 4
Fig. 1.2 Schematic view of the IN strand-transfer reaction, according to Mouscadet et al. (Mouscadet and Tchertanov, 2009) 7
Fig. 1.3 Chemical structure of the INSTIs RAL, EVG and DTG, according to Thierry et al. (Thierry, Deprez and Delelis, 2017) 21
Fig. 4.4: A section of the TV-cohort alignment in MEGA v7 47
Fig. 4.5: A section of the TV-cohort alignment in MEGA v7 47
Fig. 4.6: HIV-1 Integrase subtype distribution within our study cohort 48
Fig. 4.7: HIV subtype recombination bootscan analysis of TV406 conducted in REGA HIV-1&2 online subtyping tool v.3.41 49
Fig. 4.8: HIV subtype recombination bootscan analysis of TV439 conducted in REGA HIV-1&2 online subtyping tool v.3.41 49
Fig. 4.9: HIV subtype recombination analysis of TV 406 using the jumping profile Hiden Markov Model online tool 49
Fig. 4.10: HIV subtype recombination analysis of TV 439 using the jumping profile Hiden Markov Model online tool 49
Fig. 4.11: Evolutionary relationships of the cohort's sequences 51
Fig. 4.12: Phylogenetic analysis of the HIV-1 Integrase segment inferred by the Maximum-Likelihood method and General Time Reversible model 53
Fig. 4.13: Detailed enlargement of Fig 4.12, depicting evolutionary relationships of samples TV 356, 122, 404, 420, 431, 412, 406, 439, 145, 128, 433 and 405. 54
Fig. 4.14: Comparison between the mutational profiles of HIV-1 C (cohort) sequences and HIV-1 C (LANL-ZA) sequences for the Integrase region 59
Fig. 4.15: Comparison between the mutation profiles of HIV-1 C_{ZA} and HIV-1 B sequences for the Integrase region 60
Fig. 4.16: Comparison between the mutation profiles of HIV-1 C_{ZA} and HIV-1 C (non-ZA) sequences for the Integrase region 61

List of tables

Table 3.1: Search criteria for sequence queries in the Los Alamos HIV Sequence Database (https://www.hiv.lanl.gov/components/sequence/HIV/search/search.html) 44
Table 4.1: Description of the studycohort's demographics 46
Table 4.2: Summary of observed IN mutations in the sequencedTV cohort..... 55
Table 4.3: Summary of observed IN mutations in the database-derivedIN sequences 57
Table 4.4: Prevalence rates for individual polymorphisms in the study cohort, compared to their respective prevalence rates in South African HIV-1 C Integrase sequences 59
Table 4.5: Prevalence rates for individual polymorphisms in the South African HIV-1 C consensus sequences compared to their respective prevalence rates in a drug-naïve HIV-1 subtype B Integrase reference set..... 60
Table 4.6: Prevalence rates for individual polymorphisms in the South African HIV-1 C consensus sequences, compared to their respective prevalence rates in HIV-1 C Integrase sequences sampled in countries other than South Africa 62

Table of contents

Affidavit ... **iii**
List of scientific contributions ... **iv**
Summary ... **v**
Zusammenfassung ... **vii**
List of abbreviations ... **ix**
List of figures ... **xv**
List of tables ... **xvi**
Introduction ... **1**
1 Literature review ... **2**
1.1 HIV **2**
1.1.1 Classifying HIV ... 2
1.1.2 Genomic structure ... 3
1.1.3 Structural proteins ... 4
1.1.4 Enzymatic proteins ... 5
1.1.5 HIV pathogenesis ... 9
1.1.6 Clinical course of infection ... 10
1.1.7 Viruses, groups and subtypes ... 12
1.2 Antiretroviral Therapy ... **15**
1.2.1 Overview 15
1.2.2 Nucleoside reverse-transcriptase inhibitors (NRTIs) ... 16
1.2.3 Non-nucleoside reverse-transcriptase inhibitors (NNRTIs) ... 17
1.2.4 Protease inhibitors (PIs) ... 18
1.2.5 Integrase inhibitors (INSTIs) ... 20
1.3 HIV in South Africa ... **23**
1.3.1 Epidemiology ... 23
1.3.2 National treatment guidelines ... 24
1.4 Aim of this study ... **25**
2 Material ... **26**
2.1 Equipment ... **26**
2.2 Additional materials ... **26**
2.3 Reaction Kits ... **27**
2.4 Reagents ... **27**

2.5 Enzymes **28**
2.6 Primer **28**
2.7 Computer software and online tools **29**
3 Methodology **31**
3.1 Sample collection and storage **31**
3.2 Laboratory analyses **31**
3.2.1 Extraction of viral RNA 31
3.2.2 Photometric measurement of RNA concentration 32
3.2.3 Reverse-Transkriptase Polymerase Chain Reaction (RT-PCR) 34
3.2.4 Nested Polymerase Chain Reaction 36
3.2.5 Agarose Gel electrophoresis 38
3.2.6 PCR product purification 39
3.2.7 Sanger-sequencing reaction 39
3.2.8 Sequence-Assembly 42
3.3 Software analyses **42**
3.3.1 Alignment of sequences 42
3.3.2 Phylogenetic analyses 42
3.3.3 Consensus generation 43
3.4 Database derived sequences **43**
3.4.1 HIV-1_{ZA} sequences 43
3.4.2 HIV-$1C_{za}$ sequences 44
3.4.3 HIV-$1C_{non\text{-}ZA}$ sequences 44
3.4.4 HIV-1B sequences 45
3.5 Statistical and graphic analyses **45**
4 Results **45**
4.1 Study design and patient demographics **45**
4.2 Laboratory results **46**
4.3 Subtyping of HIV-1 Integrase **47**
4.3.1 Cohort sequences 47
4.3.2 Database-derived sequences 49
4.4 Phylogenetic analyses **49**
4.4.1 Neighbour-Joining Tree 49
4.4.2 Maximium-Likelihood Tree 51
4.5 Resistance mutation analyses **54**
4.5.1 Cohort sequences 54
2.4 Reagents **27**

4.5.2 Database-derived sequences....55
4.6 Generation of a South African HIV-1 subtype C IN consensus (HIV-1 C_{ZA})....58
4.6.1 Identification of naturally occurring polymorphisms....58
4.6.2 Naturally occurring polymorphisms HIV-1 C_{cohort} versus HIV-1$C_{ZA(LANL)}$....58
4.6.3 Naturally occurring polymorphisms HIV-1C_{ZA} versus HIV-1B_{global}....59
4.6.4 Naturally occurring polymorphisms HIV-1C_{ZA} versus HIV-1C_{non-ZA}....61
5 Discussion....62
5.1 HIV-1 Integrase subtyping distribution in South Africa....62
5.2 Resistance-associated mutations against INSTIs in South Africa....64
5.2.1 Major and minor RAMs in the study cohort's IN sequences....64
5.2.2 Major and minor RAMs in database-derived IN sequences....69
5.3 South African HIV-1 C IN inter and intra-subtype diversity....72
6 Conclusion....76
7 Bibliography....78
8 Acknowledgments....104

2.4 Reagents....x..v..i.i....27

Introduction

In retrospect, the first blood-sample documented case of human infection with HIV (human immunodeficiency virus) dates back to 1959 (Zhu *et al.*, 1998). Its clinical and societal implications, however, only began to surface in 1981, when underlying infections with HIV caused an upsurge of seldom observed opportunistic infections in otherwise healthy individuals across the USA (Centers for Disease Control (CDC), 1981; Gottlieb *et al.*, 1981). Previously, these conditions have only been described in highly immunosuppressed individuals and comprised either the presentation of uncommon and severe clinical courses or the presence of pathogens that do not cause invasive infections in healthy individuals.

Owing to a common observation in all affected patients, originally described by the CDC as "a defect in cell-mediated immunity, occurring in a person with no known cause for diminished resistance to that disease", the condition was termed AIDS (acquired immunodeficiency syndrome) in 1982 (Centers for Disease Control and Prevention, 1982). The following year, the causative agent, explaining observed horizontal human-to-human transmissions of AIDS via blood-products, sexual contacts, as well as intravenous drug consumption was found in the isolation of a new retrovirus, initially labelled as LAV (Lymphadenopathy-associated virus) (Barre-Sinoussi *et al.*, 1983; CDC, 1983). The name was later changed to HIV by the international committee on taxonomy of viruses (ICTV) in 1986 and what followed over the course of years and decades was an expansion of HIV-infections, turning the manifestation of AIDS from sporadic cases in high-risk populations to a pandemic, having by the end of 2018 infected a total of 74.9 million people across all age groups and all parts of the world, according to UNAIDS (Joint United Nations programme on HIV/AIDS) (UNAIDS and Aidsinfo, 2019).

In 2018, an estimate of 37.9 million people were living with HIV (PLWH), and despite this number having increased constantly since the start of UNAIDS' estimations in 2000, the latest estimates for new infections (1.7 million/year) and AIDS-related deaths per year (0.77 million/year) report the lowest values within the last 18 years. This success can be attributed to a combination of general advances in research for preventive and therapeutic measures as well as global efforts from national and international institutions as well as Non-governmental organisations (NGOs) to raise HIV-awareness and guide monetary and informational assets.

Despite having progressed significantly in the field of HIV to a point, where information on the virus' genomic structure as well as the majority of its molecular targets is readily accessible to the public, a functional or complete cure for the infection remains unknown. Moreover, also

major differences in regional HIV-prevalence and progression-to-AIDS rates remain, which warrants detailed attention, should AIDS be combatted on a global scale.

With 20.6 million PLWH, accounting for 54% of the global HIV-positive population, Eastern and Southern Africa face multiple challenges in containing HIV and achieving WHO (World Health Organization) goals. Currently, South Africa is home to the largest single country HIV-positive population with 7.7 million people carrying the virus in 2018, and prevalence rates of 20.4% in the adult population (Age 15-49). Treating 4.8 million people yearly, South Africa's national HIV treatment program has become one of the most refined national treatment programs in the world and advances in the therapy of HIV have been subsequently implemented in its latest iteration in October 2019. The introduction of new compounds to antiretroviral therapy (ART) programs, however, generally requires information on regional resistance rates in order to ensure optimal long-term treatment efficacy in the population and realistically curve the still expanding HIV epidemic in the country. In this thesis, I aimed to provide a mutational profile for South African HIV-1 Integrase (IN) sequences, in order to assess their genotypic baseline susceptibility to Integrase strand-transfer inhibitors (INSTIs) as well as to be able to track HIV-1's future evolution under therapeutic pressure in the South African context. In the future, data from this work will hopefully help to ultimately identify distinct mutational pathways that confer INSTI-resistance and thereby risk the long-term efficacy of South Africa's latest first-line ART regimen.

1 Literature review

1.1 HIV

1.1.1 Classifying HIV

Based on different approaches to characterising viruses, various viral classification systems have been proposed. Two of the most widely used ones are the ICTV, as well as the Baltimore virus classification. While the ICTV classification is based on multiple criteria including the virus' molecular structure, replication cycle and phylogenetic analyses, the Baltimore classification focuses on the mechanism for creating positively sensed (+) messenger ribonucleic acid (mRNA) within the virus' replicative cycle.

In the latest version of the ICTV taxonomy, 2202 species have been classified in 1 phylum, 2 subphyla, 6 classes, 10 orders, 7 suborders, 89 families, 36 subfamilies, 387 genera and 59

subgenera. HIV, along with Simian Immunodeficiency Virus (SIV) and their seven closest relatives, form the genus of *lentivirus*, a subgroup of the family of *retroviridae*. *Retroviridae*, mostly referred to as retroviruses, are small, 80-100nm sized, Enveloped RNA-viruses that share a mutual mechanism of replication: A unique set of viral proteins transfer the viral single-stranded RNA-genome into double stranded complementary DNA (cDNA), integrate it into the host's chromosomal DNA and exploit the infected cell's physiology to assemble new virions. The name *lentivirus* derives from the Latin '*lentus'* = slow and refers to an unspecific, yet typical clinical hallmark of a *lentivirus*-infection: a slow progression of the infection with a variable latent phase.

1.1.2 Genomic structure

Belonging to the family of *rertoviridae,* HIV shares a similar structural and genomic architecture to other retroviruses: Two copies of the genomic information, encoded in single (+)-RNA-strands are surrounded by a *lentivirus*-specific, cone-shaped nucleocapsid. The genome itself contains approximately 9.7 kilobases (kb) and is organised in open reading frames (ORFs) (s. fig. 1.1). Nine genes code for a total of 15 proteins of which three, namely the *protease* (PR), the *reverse transcriptase* (RT) with the embedded ribonuclease H subunit (RNase H) and the *integrase* (IN) possess enzymatic functions. Five more fulfil structural purposes, whereas seven proteins are defined as accessory (s. Fig. 1.1). Accessory proteins do not play an essential role in replication *in-vitro*, but may facilitate the survival and evolution of the virus *in-vivo*, for instance by providing effective evasion and resistance mechanisms to the host's cellular defence, regulating splicing sites or fine-tuning protein-expression rates (Malim and Emerman, 2008, Strebel, 2013, Kuciak *et al.*, 2008).

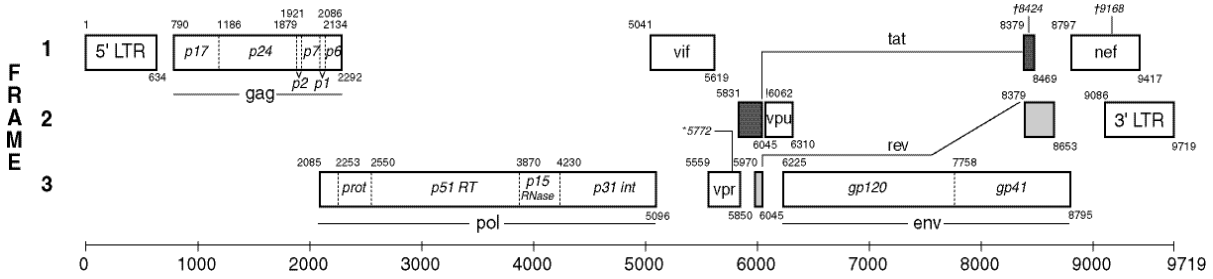

Fig. 1.1 HIV-1 gene map, displaying HXB2's genomic organisation according to the Los Alamos HIV sequence database (Los Alamos HIV sequence databse, 1998). Reading-frames are indicated by the vertically listed numbers 1, 2 and 3. Genes are indicated by rectangles, respectively named below and positioned according to their reading frame. Splicing sites for proteins within genes are indicated by dotted lines with the protein's names displayed in the gene's rectangles, or in the case of p2, p1, tat and rev adjacent to it. Numbering of nucleotide positions is carried out according to the proposal by Korber et al. (Korber *et al.*, 1998)

1.1.3 Structural proteins

Structural proteins p17, p24, p7 are located on the *gag* gene, whereas the remaining two structural glycoproteins gp120 and gp41 are encoded on the *env* gene. After liberation and processing (s. 1.1.4), p17, p24 and p7 remain within the viral particle and provide its structural integrity. P17 coats and stabilises the Envelope's phospholipid bilayer from the inside, whereas oligomers of p24 form HIV's nucleocapsid. P7 acts as a chaperone in binding the genomic RNA and modulating its three-dimensional structure throughout the course of replication. At the same time, the protein's positions are used in an internal guidance system, e.g. directing genomic RNA to future budding points, or linking key reagents within the replicative cycle, such as the host's transport ribonucleic acid (tRNA) for Lysine (Lys3) and the RT/RNA complex (s.1.1.4) (BAMPI *et al.*, 2004; Fiorentini *et al.*, 2006; Grohmann *et al.*, 2008; Zhao *et al.*, 2013).

The remaining structural proteins, gp120 and gp41, are located on the outside of the viral Envelope. After liberation from their mutual precursor protein gp160, they both serve as HIV's cellular receptors. As such, they specifically recognise human cluster of differentiation 4 (CD-4)-receptors as well as co-receptors CXC chemokine receptor 4 (CXCR4) and C-chemokine receptor 5 (CCR5). On a molecular level, viral receptors are composed in a trimeric structure, consisting of three subunits of gp41 and gp120, each. Whereas gp120 is responsible for the attachment to the cellular receptors, gp41 anchors the complex in the viral envelop. Upon binding of gp120 to CD-4 receptors, a complex cascade of consecutive binding processes changes the spike's three-dimensional conformation, bringing the viral and cellular bi-lipid

membranes in close enough contact to allow fusion pores to occur. (Kwong *et al.*, 1998; Blumenthal, Durell and Viard, 2012; Pancera *et al.*, 2014). However, being the virus' sole representatives on its immunologically inert surface, gp120 and gp41 are naturally prone to being recognized by the host's immune system, potentially resulting in the virion's degradation. In order to avoid antibody recognition, both proteins are covered by many N-linked oligosaccharides, a mechanism to hide the immunological epitopes by blocking antibody access (McCaffrey *et al.*, 2004).

1.1.4 Enzymatic proteins

All of HIV-1's enzymes are encoded on the *pol* gene. Immature virions contain precursor polyproteins, whereas only correctly cleft and processed enzymes can fulfil their function to an optimal degree. Moreover, even small amounts of unprocessed proteins are shown to actively impair selective key reactions (Anderson-Daniels *et al.*, 2019). The critical process of virion maturation is orchestrated by the PR, a small (10.8 kDa), homodimeric protein that consists of two identical 99 amino-acid long chains A and B. In its symmetrical centre, two highly conserved aspartic residues (Asp) catalyse a hydrolytic reaction, which is not unique to HIV, but can be observed in many endopeptidases across eu- and procaroytic cells, as well as (retro-)viral organisms. (Meek *et al.*, 1989; Brik and Wong, 2003; Adachi *et al.*, 2009). The exact cleaving site is determined by the binding of PR to an eight amino acid long fragment in the precursor protein, described as P4-P3-P2-P1-P1'-P2'-P3'-P4' in which the cleavage is carried out between P1 and P1'. The specificity of this binding, however, is poorly understood, as known cleavage sites neither share a sequence homology, nor a mutual recognition site, meaning additional factors must be essential for cleavage site recognition (Kontijevskis, Wikberg and Komorowski, 2007; Singh and Su, 2016).

Downstream of the PR-coding sequence lies the RT-gene, coding for HIV's largest enzyme. Apart from the two cleavage sites on both ends of its 1680 nucleotide long RNA sequence, the RT also contains an internal cleavage site at position 3870. Thus, posttranslational processing makes up for three possible amino acid-chains: The uncleft 66 kDa large protein, a 51 kDa reverse-transcriptase and a 15 kDa RNase H subunit. *In-vivo*, P66 and P51 form a heterodimer, known as HIV's final RT (Hottiger and Hübscher, 1996; Frankel and Young, 1998). In order to convert the viral (+)-RNA genome into cDNA, the RT consecutively catalyses three essential reactions: (i) RNA-dependent-DNA polymerisation of a complimentary negatively sensed (-)–

DNA strand, (ii) degradation of the RNA template strand and (iii) DNA-dependent DNA-polymerisation of the (+)-DNA strand. Shortly after the virion's uncoating, (i) is initiated by binding of the host's Lys3 tRNA primer to the viral 5' primer binding site (PBS), being located within the genome flanking long terminal repeat (LTR) sequences. Thereafter, repetitive iterations of RT-catalysed DNA polymerization and RNase H mediated RNA degradation transform HIV's genome into cDNA. Of note is that this reaction includes multiple template-strand switching steps, which allow for the occurrence of recombination events (s.1.1.7) (Sluis-Cremer and Tachedjian, 2008).

As with all DNA–polymerisation-processes, the installation of the correct nucleotides, matching the corresponding base, is paramount for transmitting the correct genomic information. The combination of internal proof-reading and consecutive mismatch repair processes in eukaryotic, DNA-based DNA polymerization results in error rates of approximately $1 * 10^{-9}$ per copied nucleotide (McCulloch and Kunkel, 2008). Due to the lack of proofreading and DNA repair mechanisms, error rates for the reverse-transcription catalysed by HIV's RT range between $5 * 10^{-4}$ and $2.5 * 10^{-4}$ in cell culture experiments, depending on the inserted nucleotide (Preston, Poiesz and Loeb, 1988; Ji and Loeb, 1992; Kati *et al.*, 1992). Coupled with a high turnover rate during an infection, this error-prone replication makes up for a large part of HIV's high genomic variability and the emergence of many new variants, so called *quasispecies* in every patient.

After reverse-transcription, a multi-step procedure then covalently integrates the cDNA intermediate into the host's genome. In this state, the HIV-coding DNA sequence is termed 'pro-viral' (Chun *et al.*, 1995; Finzi *et al.*, 1999). In a first reaction, the IN processes the cDNA's 3'-termini, before mediating their strand transfer into the genomic DNA. Subsequently, cellular enzymes catalyse the ligation of the interrupted DNA strands. Similar to PR and RT, IN is also encoded on the *pol*-region in a 866 nucleotide long gene, which translates into a 288 amino acid long protein. In this, three highly conserved domains, namely the N-terminal domain (NTD) spanning from position 1-46, the catalytic-core domain (CCD) from 56-202 and the C-terminal domain (CTD) from 203-219 are separated by linking sequences in between: (Jaskolski *et al.*, 2009). While the NTD mainly attributes to non-covalently binding DNA via a Zn^{2+} coordinating His-His-Cys-Cys (histidine-histidine-cysteine-cysteine) motif, and the CTD is deemed crucial for orchestrating the steric formation as well as binding of DNA, all enzymatic activities are facilitated by the CCD (Lesbats, Engelman and Cherepanov, 2016). This domain features a Asp-

Asp(35)-Glu (glutamic residue) motif, which can be observed in variant forms in bacteriophages' transposases and other polynucleotide transferases. In case of HIV, it consists of two aspartic residues at position Asp64 and Asp116 as well as one glutamic sidechain Glu152. (Rice and Kiyoshi, 1995; Nowotny, 2009). As the reaction mechanism for both reactions, 3' processing as well as strand transfer, revolves around a nucleophilic attack by a deprotonated oxygen atom on the DNA's phosphodiester bond – either by an activated water molecule, or a deprotonated hydroxyl-group - the coordination of two divalent metal ions is a key component to this catalytic unit. *In-vivo* these ions are most likely to be Mg^{2+} ions (Goldgur *et al.*, 1998; Pommier, Johnson and Marchand, 2005). (s. Fig. 1.2).

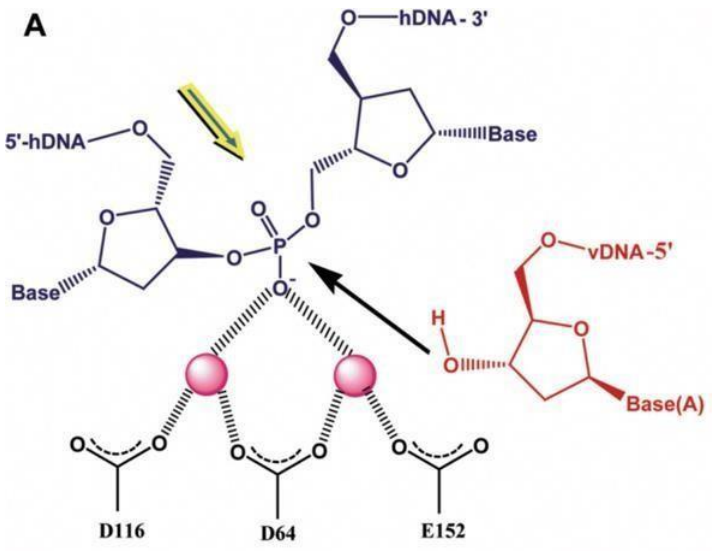

Fig. 1.2 Schematic view of the IN strand-transfer reaction, according to Mouscadet et al. (Mouscadet and Tchertanov, 2009). The figure is used under the CC BY 4.0 license (https://creativecommons.org/licenses/by/4.0/) and cropped to only display part A of the original figure. The black arrow indicates a nucleophilic attack by the viral open (C)A3'-OH terminus (red) on the chromosomal DNA's phosphodiester backbone (blue). Metal ions (pink balls), coordinated by IN's D116, D64 and E152 residues, can be seen to stabilise this transesterification reaction. The green/yellow arrow indicates the scissile posphodiester bond.

IN's first reaction, the 3'-processing of the newly synthesized cDNA strands commences shortly after the reverse transcription within the cell's cytoplasm. Here, a IN dimer binds to a four-nucleotide CAGT (Cytidine, Adenosine, Guanosine, Thymidine) recognition sequence within both the virus' 3' LTRs. Endonucleatic cleavage between the A and the G results in open CA3'-OH and overlapping 5'-CA termini as well as two GT dinucleotides (Engelman and Cherepanov, 2014). Together with multiple copies of IN, the processed cDNA induces the formation of a pre-integration complex (PIC) that consists of viral and cellular co-factors such as MA, NC, VPR,

cellular lens-epithelium-derived growth factor (LEDGF/p75) and others (Engelman, Mizuuchi and Craigie, 1991; Ciuffi *et al.*, 2005; Choi *et al.*, 2018). After translocation into the cell's nucleus, the PIC recognises and binds chromosomal DNA, upon which this complex is termed stable synaptic complex (SSP) (Craigie and Bushman, 2012).

The exact amount of IN copies within this nucleoprotein-complex differs among the family of retroviruses, however, X-ray crystallography, nuclear magnetic resonance spectroscopy and electron microscopy suggest, that the lentiviral IN in its active form, apart from dimeric and octameric configurations, mostly assembles into a tetrameric conformation. Further structural implications can be drawn from the catalytic concept behind the strand-transfer reaction: Open hydroxyl-groups from the processed CA-3' ends perform a nucleophilic attack on the 5' DNA phosphate on both targeted strands. Of note is that both reactions are carried out simultaneously (Pommier, Johnson and Marchand, 2005). In case of HIV, the insertion sites into the two chromosomal strands lie on the DNA's major groove, separating the insertion sites on the respective strands by a 5 base-pair (bp) segment. This is equivalent to a spatial distance of 15 Å (1,5 nm). In HIV IN's soluble dimeric structure, the active sites are separated by at least 30 Å (3 nm), rendering a solely dimeric structure for HIV's strand transfer reaction unlikely (Craigie, 2001). Insights from a recently obtained crystal structure for HIV's close relative 'prototype foamy virus's' SSP, as well as obtained HIV IN two-domain spanning structures, suggested a tetrameric conformation for HIV's SSP, termed dimer-of-dimers. Herein, IN domains from various protomers interlock asymmetrically to create a sterically stable nucleoproteocomplex that conforms to the spatial arrangements of the respective active sites, and allows the enzyme to catalyse reactions on both corresponding DNA strands at the same time (Wang *et al.*, 2001; Faure *et al.*, 2005; Li *et al.*, 2006; Ren *et al.*, 2007; Hare *et al.*, 2009). Following the insertion and alignment of both viral cDNA strands, the strand transfer reaction results in a y-shaped intermediate, in which viral 5' and the host's 3' ends not only remain unlinked, but are also separated by the initial 5 nucleotide gap. Ligation, along with the duplication of the five nucleotide long segment is carried out by cellular enzymes. (Yoder and Bushman, 2000). Selectivity of integration sites within the host's genome is part of on-going research, as successful strand transfer does not appear to be relying on distinctive DNA recognition patterns, however seems to occur with high affinity to active transcription sites. This hints at the importance of cellular DNA-binding proteins in the SSP, such as LEDGF/p75, tethering the

complex to the host's DNA and thereby providing the steric pre-requisites for viral integration (Cherepanov *et al.*, 2003; Ciuffi *et al.*, 2005).

1.1.5 HIV pathogenesis

HIV can be transmitted between humans via various mediums, among others blood and blood-products, breast milk, semen and vaginal as well as rectal secretions. Virions furthermore require a susceptible port-of-entry to a new organism, in case of HIV either mucous membranes, damaged tissue or direct injection into the bloodstream. Hence, horizontal transmission routes for HIV include unprotected sexual contacts and parenteral exposure to HIV-containing body fluids, while the virus is transmitted vertically in pre-, peri- or post-partum mother-to-child-transmissions (MTCT).

Upon sexual transmission of HIV, accounting for the majority of HIV transmissions worldwide, the first immunologic contact points are the new host's dendritic cells (DCs), before reaching secondary lymphatic tissue with a high cumulative density of CD-4^+ T-cells. Owing to the abundance of susceptible target cells, viral replication and budding rates increase drastically at this point, leading to a systemic shedding of infectious viral particles. Initially, HIV spreads to other lymphatic tissue with high CD-4^+ T-cell counts, such as the spleen and tonsils, but soon after any other susceptible organ and cell may become infected as well.

CD-4^+ T-lymphocytes, macrophages and other peripheral blood mononuclear cells genuinely obtain a pivotal role in orchestrating the adaptive immune system. Phagocytosis, antigen presentation as well as induction and amplification of highly specific immune responses are some of the main tasks of these cells (Zhu and Paul, 2008). Their infection not only leads to an impairment of the intended function, but also harbours fatal consequences for the individual cell and, on a larger scale, the entire cell-lineage. On a molecular level, a cell's infection with HIV results in the corruption of its biochemical processes and subsequent viral replication. Activated immune cells' physiological functions on the other hand include the secretion of pro-inflammatory cytokines, presentation and cross-presentation of HIV fragments, as well as direct stimulation of additional immune mediators. All of these functions by default initiate an up-regulation of the cell's gene-expression rates in order to successfully orchestrate and execute an efficient immunologic response (Alberts *et al.*, 2002; Joffre *et al.*, 2012). However, harbouring pro-viral DNA in their genome, increased transcription inevitably leads to viral replication and thus to subsequent infection and recruitment of more immunologic cells. Therefore, the

immunologic response to HIV creates a self-enhancing immunogenic cycle that ultimately promotes higher virological replication rates. Moreover, the natural course of a HIV infection gives rise to super-infections by fellow HIV strains, but also by any other pathogen, potentially leading to an increase in systemic immune activation. Thereby viral replication rates are further increased, causing an overall acceleration in CD-4^+ T-cell depletion and the disease's subsequent progression to AIDS (s. below) (Deeks *et al.*, 2004).

Perishing of infected CD-4^+ lymphocytes mostly occurs via two pathways: Either (i) by a cell's physiological exhaustion due to an over-replication of viral proteins, or (ii) via cellular defence mechanisms. In (ii), activated CD-8^+ cytotoxic T-lymphocytes (CTLs) act as the main facilitator in controlling early and in rare cases long-term infection by specifically identifying and subsequently eliminating infected cells via major histocompatibility complex I (MHC-I) bound epitopes (Cao *et al.*, 1995). Four to twelve weeks after initial infection, seroconversion takes place, induced by the humoral immune system's production of anti-HIV antibodies. However, with some exceptions, the humoral immune system's response only show modest and short-lived effects on viral loads (Poignard *et al.*, 1999; Wei *et al.*, 2003). Due to the external pressure and a high innate mutation rate, viral escape to cellular as well as humoral immune responses is common. Escape mechanisms include mutations in the *env* gene and therefore altered gp160 epitopes, down-regulation of MHC-I molecules, as well as the evolvement of N-linked glycolisation of immunogenic spikes on the virus' surface (s. 1.1.3). In total, these factors lead to an initial upsurge in viral loads of up to 10^6 - 10^7 viral RNA copies / ml plasma within the first weeks of infection, followed by a mainly CTL-mediated decline to 10^3-10^4 copies/ml, before the establishment of an delicate equilibrium consisting of continuous viral replication on the one side and immunologic containment, including CTL-mediated cell-death on the other. In this phase, the initial increased CD-4^+ T-cell turnover rate can be temporarily compensated by an accelerated recruitment of hematopoietic stem cells and lymphoid progenitor cells from the bone marrow. However, low-level immune activation, infection of progenitor cells and the inability of the human immune system to eradicate the virus ultimately result in depletion of the CD-4^+ cell lineage, exhaustion of the adaptive immune system and the clinical presentation of AIDS.

1.1.6 Clinical course of infection

The course of an HIV infection can be separated into three stages: (i) the stage of acute infection (AHI), (ii) the latent stage, and (iii) the stage of AIDS. (i) broadly describes the first two to six weeks after infection, before the immune system actively combats the infection. It is characterised by an abundance of CD-4^+ targets, combined with a low immunologic defence, which accounts for a high viral replication rate of approximately ten hours, and viral loads of up to $44*10^7$ copies / ml within the first three weeks of exposure (Little *et al.*, 1999). Due to high viral loads, the absence of HIV-specific clinical symptoms as well as an undiagnosed HIV status, a disproportional amount of lateral transmissions take place at this stage of infection and effective diagnostic and therapeutic measure remain extremely important in early infection (Pilcher *et al.*, 2004; Brenner *et al.*, 2007). While older diagnostic tests solely relied on the presence of anti-HIV antibodies, modern fourth generation enzyme-linked immunosorbent assay (ELISA)-tests simultaneously test for anti-HIV antibodies, as well as the soluble p24 antigen (Hurt *et al.*, 2017). According to a country's national guidelines, positive ELISA tests are subsequently confirmed, mostly via western blot or a supplemental HIV-IgG-sensitive antibody differentiation assay (Kondo *et al.*, 2018).

Three to four weeks after transmission, the emergence of HIV-specific CTLs temporarily reconstitutes CD-4^+ T-cell counts and marks the transition to (ii) (Borrow *et al.*, 1994; Koup *et al.*, 1994). This latent phase is named after the long-term absence of clinical symptoms, however, on a cellular level, immune activation and viral replication continues despite vigorous immune responses, causing continuously detectable viral load levels of 10^3-10^6 copies per ml and hereby a gradual depletion of CD-4 T-cells (Perelson *et al.*, 1996; Finzi and Siliciano, 1998). Furthermore, pro-viral DNA has the ability to persist in viral reservoirs such as resting T-memory or progenitor cells for years and re-enter the replication cycle only upon activation of the host cell (Cavert *et al.*, 1997; Chomont *et al.*, 2009). The duration of this latent stage varies significantly between affected individuals and depends on many internal as well as external factors (Langford, Ananworanich and Cooper, 2007). The majority of PLWH, however, show a median progression rate of ten years before presenting symptoms due to immunosuppression. At this point the CD-4^+ T-cell count has declined to a level, where it can no longer sufficiently orchestrate the adaptive immune system, exposing the human organism to opportunistic infections but also leaving the immune system unable to further contain HIV itself. The rise in viral load is representative for the virus' systemic infiltration into the central and peripheral nervous system, muscles, skin, mucosae and internal organs. Clinical symptoms at this stage

include weight loss, chronic fatigue, night sweats, neurological disorders, diarrhoea as well as diverse oral mucosal and skin infections.

Progression to the most advanced stage three, equivalent to the clinical diagnosis of AIDS, is defined according to the CDC classification for various age groups. A diagnosis of AIDS for individuals above the age of six can be made by meeting either one of the following criteria: a total CD-4^+ T-lymphocyte count of <200/ µl, a CD-4^+-T-lymphocyte-to-total-lymphocytes ratio <14%, or the presence of any AIDS-defining clinical condition. The comprehensive list of AIDS-defining conditions includes HIV-related conditions, e.g. HIV encephalopathy or a general wasting syndrome, invasive opportunistic infections by other pathogens such as cerebral toxoplasmosis, PCP, internal candidiasis, cytomegalo- or JC-virus-infections, recurrent infections without immunity, as well as malignant conditions such as Kaposi sarcomas, Burkitt and other lymphomas, or invasive cervical cancer. Untreated, the median survival rate in this stage is three years. (Marcus *et al.*, 2016; Antiretroviral Therapy Cohort Collaboration *et al.*, 2017; Centers for Disease Control and Prevention, 2019; Cohen, 2019). However, nowadays, more and more studies hint, that in an ideal setting, early diagnosed and promptly treated individuals show a near-normal life expectancy (Marcus *et al.*, 2016; Antiretroviral Therapy Cohort Collaboration *et al.*, 2017).

1.1.7 Viruses, groups and subtypes

Shortly after the initial description in 1983, a rising number of reports from all over the world provided detailed information on a retrovirus, termed Human T-cell Lymphotropic Type III (HTLV-III) at that time. In 1986, however, genetic analyses on two isolates from young men from Guinea-Bissau and Cape Verde, respectively, who showed low CD-4^+ T-cell counts and clinically pathognomic signs of an underlying AIDS, yet repeatedly tested negative for serological HTLV-III-tests, revealed a second causative virus for the observed condition (Clavel *et al.*, 1986). While its morphology only showed subtle differences to previous isolates, antigenic sequences within the *env*-gene deviated substantially from HTLV-III's known sequence, expressing a new gp140 protein and thus explaining the lack of reactive antibodies to established ELISA-tests. These results allowed for a new taxonomy, labelling the viruses in accordance to their chronological description HIV-1 and HIV-2.

Both viruses originate from different zoonotic transfers of SIV strains to the human (Hirsch *et al.*, 1989; Keele *et al.*, 2006; Sharp and Hahn, 2010). In case of HIV-1, first transmission events

of SIV_{cpz} from chimpanzees (cpz) to humans most likely occurred in the first half of the 20th century, with the oldest sequenced isolates dating back to 1959 and 1960. Both viruses were isolated from individuals in Kinshasa, Democratic Republic of the Congo (DRC), rendering this area likely to be the birthplace of HIV-1 (Zhu *et al.*, 1998; Sharp and Hahn, 2008; Worobey *et al.*, 2008). HIV-2 arose from transmissions of SIV_{smm} from sooty mangabeys (smm) to the human in western Africa, most likely in Guniea-Bissau. Time of the first transmission also dates back well before HIV-2's epidemiological surfacing and isolation in 1986 to the first half of the 20th century, although presumably after transferal of HIV-1 to the human (Lemey *et al.*, 2003).

Both viruses, HIV-1 as well as HIV-2, share concepts in terms of transmission routes, cellular tropism, infection as well as replication cycle. HIV-2. However, has proven to be significantly less virulent and less pathogenic than HIV-1. Over the course of an infection with HIV-2, $CD\text{-}4^+$ T-cell counts are generally higher, viral loads lower and time of disease progression to AIDS significantly longer, with a high proportion of long-term-non-progressors (Marlink *et al.*, 1994; van der Loeff *et al.*, 2010; Nyamweya *et al.*, 2013). On the other hand, viral integration as well as observed transmission rates are similar to those of HIV-1, when adopted to the lower viremia, hence implying a lower grade of immune activation for HIV-2 infection as the underlying cause for better clinical outcomes (O'Donovan *et al.*, 2000; MacNeil *et al.*, 2007). Regardless of the reason for virological containment, lower absolute transmissions counts resulted in a lower spread of HIV-2 globally. Unlike HIV-1, HIV-2 has only been described sporadically across the world and is predominantly confined to West Africa (Soriano *et al.*, 2000; Barin *et al.*, 2007; Torian *et al.*, 2010).

Along with the discovery and the rapid expansion of symptomatic HIV-1 cases to pandemic dimensions in the mid 1980s came a high amount of detailed reports of viruses in the following years. Isolates from different geographical origins showed immense genetic diversity, warranting the need of a sufficient classification system as well as the generation of databases for genomic sequence information (Li, Tanimura and Sharp, 1988). While the Los Alamos National Laboratory (LANL) in the USA soon provided a platform for a sequence database, HIV's current classification system was proposed in 1999 (Robertson, 2000; Kuiken, Korber and Shafer, 2003). Herein, HIV-1 and HIV-2 sequences are classified according to their phylogenetic homology to one another. These homologies between HIV-1's and HIV-2's respective main clustering groups are starkly divergent, rendering a mutual ancestor unlikely. Furthermore, multiple groups show closer relations to SIV variants than to fellow HIV strains. Hence, these groups must originate

from separate transmission events (Keele *et al.*, 2006; Van Heuverswyn *et al.*, 2006). To date, four groups have been described for HIV-1: group M (main), O (outlier), N (non-M, non-O) and P. Group M and N derive from SIV_{cpz} lineages from chimpanzees, whereas group O and P have ancestor in SIV strains from gorillas (Keele *et al.*, 2006; Van Heuverswyn *et al.*, 2006; Plantier *et al.*, 2009). HIV-2, despite being the less virulent virus, experienced more transmission events to humans in the past, resulting in eight groups, A through H. Once transmitted to the human, the founder viruses adapt to the host's environment and external factors, marking the starting point of the virus' evolution within the new host-species. As for HIV-1 groups N and P, only a few representative viruses have been isolated from individuals in Cameroon, so far (Plantier *et al.*, 2009; Rodgers *et al.*, 2017). HIV-1 group O infections have been reported sporadically in European countries with close relations to Cameroon, but also almost exclusively remain endemic in Cameroon and neighbouring countries. Responsible for HIV's pandemic proportions are group M clades, accounting for >95% of all HIV infections worldwide (Santos and Soares, 2010). Tracing the genetic evolution pattern from its transferal, group M clades can be assigned to multiple subtypes. Currently, the ten subtypes A, B, C, D, F, G, H, J, K, and most recently L have been described (Yamaguchi *et al.*, 2019). Subtypes A and F can be further classified into sub-subgroups A1, A2, A3, A4 and A6, as well as F1 and F2, respectively. While fragments of a hypothesised subtype E can be observed in the circulating recombinant form (CRF) 01_AE, proof of a pure full-genome subtype E remains elusive (Carr *et al.*, 1996). Similar reasoning explains the absence of a subtype I (Gao *et al.*, 1998).

As infected cells continue to express the CD-4, as well as CCR-5 and/or CXCR-4 co-receptors, super-infections by different HIV strains can occur. Furthermore, HIV's genomic structure of harbouring two genomic copies within its nucleocapsid, coupled with events of template-strand switching during reverse transcription (s. 1.1.4), allow the synthesis of genetic hybrids, whose genomic RNA partially consist two and more subtypes (Hu and Hughes, 2012). Upon first description, a new recombinant is termed as a unique recombinant form (URF). Only after appearance in more than three epidemiologically unrelated individuals are these viruses termed CRFs (Robertson, 2000).

On a global scale, HIV-1 subtype C is the most prevalent, accounting for half of all HIV infections worldwide. It predominates the local epidemics in southern and eastern Africa and is also highly prevalent in parts of Asia and South America. Subtype A is mostly reported in East and West Africa. Of note is, that its recombinants CRF01_AE and CRF02_AG show the highest

prevalence rates across all CRFs and are highly endemic throughout parts of Asia as well as West and Central-West Africa, respectively. Subtype B isolates have been reported at lower rates in most parts of the world, however, are traditionally responsible for the majority of HIV infections in North America, Europe and Australia (Bbosa, Kaleebu and Ssemwanga, 2019).

Sequential polymorphisms within the genomic structures of different subtypes are not only important for tracing HIV's epidemiology and evolution, but also for diagnosis and treatment of an infection, as they may impact intrinsic drug efficacy, resistance pathways and primary diagnosis results (Taylor *et al.*, 2008; Aghokeng *et al.*, 2009; Peeters, Aghokeng and Delaporte, 2010; Tebit *et al.*, 2016). In times of increasing migration, and growing global tourism, tracking the local prevalence rates remains a necessary tool in order to be able to provide the best possible care

1.2 Antiretroviral Therapy

1.2.1 Overview

The search for effective treatment regimens against HIV began with its emergence and has since evolved from sole prophylaxis of opportunistic infections to an array of HIV specific drugs, targeting very distinct mechanisms within its replication cycle. Until 2018, seven targets could be successfully transformed into pharmaceutical drugs for ART. In chronological order according to their approval by the U.S. Food and Drug Administration (FDA), these are: Nucleoside reverse transcriptase inhibitors (NRTI), non-nucleoside reverse transcriptase inhibitors (NNRTI), protease inhibitors (PI), fusion inhibitors, CCR5-antagonists, Integrase strand-transfer inhibitors (INSTI) and post-attachment inhibitors. In addition, side and main effects of other drugs, that do not target the virus specifically, but enhance effectiveness of specific treatment options, are used in current regimens (Zeldin and Petruschke, 2003; Shah *et al.*, 2013).

Initiation of therapy is recommended for all patients, regardless of their respective CD-4^+ T-cell count, and early initiation, ideally within one week of diagnosis, is directly correlated to better virological responses (Group, 2015; Cohen *et al.*, 2016; Mateo-Urdiales *et al.*, 2019). While not being able to eradicate HIV, ART aims at long-term control of the infection, suppressing viral loads below the detection threshold of 50 copies/ml, preventing disease progression to AIDS, as well as its transmission to other individuals (Mahungu, Rodger and Johnson, 2009; Deeks, Lewin and Havlir, 2013). This benefit, however, requires life-long medication and with more

than 30 available FDA-approved compounds to choose from raises the question for an optimal treatment regimen (Aidsinfo, 2020).

Historically, the use of single and dual-drug ART regimens, with low genetic barriers has only shown short-lived success, as resistant variants quickly emerged under therapeutic pressure (Frost and McLean, 1994). Variants harbouring drug resistance mutations (DRM) may also pre-exist in patients before the initiation of ART, as AHI is associated with high HIV turnover rates and thus predestines for the existence of resistant viruses (Nowak *et al.*, 1997). Combination of available treatment options has proven superior to single-drug regimens and nowadays most guidelines recommend a combination of three drugs from at least two different classes (Hammer *et al.*, 1996; Bartlett *et al.*, 2006). Most commonly this entails two NRTIs, combined either with a NNRTI, PI or INSTI compound, respectively.

As of 2019, the WHO recommends a first-line treatment regimen for adults and adolescents, consisting of two NRTIs, namely Tenofovir disoproxil fumarate (TDF) and either Lamivudine (3TC) or Emtricitabine (FTC), combined with the INSTI Dolutegravir (DTG) ('WHO | Update of recommendations on first- and second-line antiretroviral regimens', 2019). National treatment guidelines are generally in line with the WHO's recommendations, however, local DRMs, drug-availability and cost-effectiveness have to be taken into account.

1.2.2 Nucleoside reverse-transcriptase inhibitors (NRTIs)

Due to good virologic suppression rates, affordability, as well as clinical experience, NRTIs are considered a backbone to most modern cART regimens (Riddler *et al.*, 2008; Hill and Sawyer, 2009; Paton *et al.*, 2014). Administered orally as prodrugs, they specifically target the reverse transcription of viral RNA to cDNA. Structurally, the inhibition of RT's DNA polymerisation step is explained by the activated drug's similarity to deoxy nucleoside triphosphates (dNTPs). As such, they compete with dNTPs for DNA-incorporation and thereby can inhibit the elongation process of the reverse transcription. As NRTIs lack a hydroxyl group relative to the 3'-OH group in a C2-deoxyribose, which is required for 3'-5'-phosphodiester-bonding the following dNTP, their incorporation leads to strand termination. (Arts and Hazuda, 2012).

Today, six FDA-approved NRTI options exist for cART, which can be categorised according to their structural resemblance to the four different dNTPs. T-analogues include Stavudine (d4T) and Zidovudine (ZDV/AZT), the first drug to be used in ART in 1987. C-analogues comprise FTC as well as 3TC, while G- and A-analogues consist of Abacavir (ABC) and Didanosine

(ddI), respectively. A seventh option of competitive RT inhibition is given with Tenofovir (TFV), a nucleotide reverse-transcriptase inhibitor (NtRTI) expressing resemblance with A-monophosphate (Fung, Stone and Piacenti, 2002). Specificity for relevant RT inhibition derives from the uniqueness of the catalysed reaction, as well as its lack of proofreading capabilities. Unlike RT, eukaryotic DNA polymerases express 3'-5'-exonucleolytic activity and may therefore recognise and excise incorrectly incorporated N(t)RTIs, resulting in lower grades of inhibition (Roberts, Bebenek and Kunkel, 1988; McCulloch and Kunkel, 2008). Nonetheless, adverse effects occur under treatment with NRTIs and NtRTIs. These range from milder symptoms, such as nausea, headaches and gastro-intestinal symptoms to drug specific interactions, for instance TDF's association with chronic kidney disease or hypersensitivity reactions to ABC in selective individuals (Martin *et al.*, 2004; Scherzer *et al.*, 2012). Additionally, all N(t)RTIs share a methodological side effect, with higher prevalence in first-generation NRTIs. This side effect originates in partial inhibition of eukaryotic DNA-polymerases, most importantly of mitochondrial DNA polymerase γ, and can lead to depletion of mitochondrial DNA and oxidative stress therein (Lewis, Day and Copeland, 2003). Clinical presentation include progressive polymyo- and neuropathy, hepatic failure and lactic acidosis (Margolis *et al.*, 2014). Resistance to N(t)RTIs under treatment can mostly be attributed to either pyrophosphorolysis, an excision of the respective drug from the primer strand, or an altered RT enzyme expressing better differentiation between NRTIs and dNTPs. Pyrophosphorolysis is generally observed following treatment with thymidine analogues, whereas other analogues can select for individual mutations, predominantly Met184Val/Ile, or Lys65Arg, which both confer resistance to the selecting drug(s) (Boyer *et al.*, 2001; García-Lerma *et al.*, 2003; Liu and Shafer, 2006).

1.2.3 Non-nucleoside reverse-transcriptase inhibitors (NNRTIs)

NNRTIs are non-competitive, HIV-1 specific, RT inhibitors. HIV-2, HIV-1 group O and primate lentiviruses harbour intrinsic resistance to this class of ART. NNRTI's binding point to HIV-1 RT is situated in proximity to the active site in a hydrophobic NNRTI-binding pocket (NNIBP) (Tantillo *et al.*, 1994; Stefan G. Sarafianos *et al.*, 2009). In absence of NNRTIs, this pocket does not exist and the enzyme undergoes conformational changes upon its formation (Kohlstaedt *et*

al., 1992). Although the exact mechanisms of inhibition have not yet been established, collective data suggest that the steric implications of binding NNRTIs impair the correct alignment of the primer strand in relation to incoming dNTPs. In that, the enzyme is no longer able to catalyse the translocation that frees the nucleoside-binding site, renders the lastly bound nucleotide the primer for the next dNTP and thereby enables elongation (Ding *et al.*, 1998; Das *et al.*, 2007; Stefan G Sarafianos *et al.*, 2009). Currently, five NNRTIs are available for cART: Nevirapine (NVP), Efavirenz (EFV), Etravirine (ETR), Rilpivirine (RPV) and Doravirine (DOR). NNRTIs high selectivity for HIV-1's RT results in fewer on-target adverse effects, when compared to NRTI, however distinctive and sever off-target side-effects can still be observed (Montessori *et al.*, 2004; Arendt *et al.*, 2007).

Clinically, cART regimens consisting of two NRTIs, plus one NNRTI showed promising results in treatment-naïve patients, with trials reporting 89% of patients showing viral suppression of <50 RNA copies/ml plasma at a 96-weeks follow-up in 2008 (Riddler *et al.*, 2008). However, owing to a relatively low genetic barrier to resistance, widespread use of NNRTIs resulted in a slow accumulation of resistance-associated mutations (RAM) that not only confer resistance to individual NNRTIs but also expressed significant levels of cross-resistance to other NNRTI (Delaugerre *et al.*, 2001; Gupta *et al.*, 2018). Furthermore, pre-existence of low-level resistant variants within a population bears a two-times increased risk of virological failure, when treated with a NNRTI containing cART regimen (Cozzi-Lepri *et al.*, 2015). Nowadays, prevalence rates of RAMs against NNRTIs exceeding 10% in NNRTI-naïve populations have been reported in many low- and middle-income countries and can be explained by transmitted drug-resistances (TDRs). The almost unimpaired viral fitness of NNRTI-resistant mutants allows for their assertion against wild-type strains during reversion, and thereby the transmission and spreading of NNRTI-resistant HIV-1 viruses (Deeks, 2001; Zazzi, Hu and Prosperi, 2018).

RAMs against NNRTI generally involve amino-acid substitutions surrounding the NNIBP, sterically blocking the pocket, or hindering hydrophobic interactions between the enzyme and its inhibitor (Stefan G. Sarafianos *et al.*, 2009). Cross-resistances, conferred by single point mutations, for instance at position Lys103 or Tyr181, have been described throughout the class of NNRTIs (Shafer, 2006; Wheeler *et al.*, 2010; Melikian *et al.*, 2014).

1.2.4 Protease inhibitors (PIs)

PIs act late in HIV's replicative cycle, inhibiting an enzyme that is greatly responsible for virion maturation, rather than primary infection of a cell. Yet, PR is indispensible and its inhibition causes effective viral suppression (Hughes *et al.*, 2011). The mechanism of action lies in a competitive inhibition of PR's highly conserved active site (s. 1.1.4). Herein, PIs interact with the enzyme's catalytic carboxyl groups of Asp25 and Asp125 via hydrogen bonds to the PI's hydroxyl-group. Binding affinity is further increased by the drug's design to interact with an array of conserved, and well-described residues within the enzyme's pocket (Weber *et al.*, 1989; Lv, Chu and Wang, 2015). While this conservation is responsible for PI's high efficiency, as well as their relatively high genetic barrier to resistance, it also accounts for a strong chemical conservation within PIs themselves (Tang and Shafer, 2012; Aoki *et al.*, 2018). Consequently, this leads to structural and physiological similarities between the individual compounds. These include a low bioavailability coupled with a high potential for drug-drug interactions, adverse effect and broad cross-resistances to other PIs (Williams and Sinko, 1999; Wensing, van Maarseveen and Nijhuis, 2010; Arts and Hazuda, 2012; Pau and George, 2014). In order to reduce the first-pass effect and increase bioavailability, PIs are nowadays administered with a booster, consisting of a sub-therapeutic dose of either Ritonavir (/r), or Cobicistat. The second-generation PI Darunavir (DRV) allows a once-daily formulation while proving to be more efficient in terms of viral suppression and maintaining a better side-effect and resistance profile (Pau and George, 2014; Eron *et al.*, 2019). Previously, hypercholesterolemia, hypertriglyceridemia, hyperglycemia, impaired insulin sensitivity, as well as permanent lipodystrophy and other metabolic changes were frequently observed in patients treated with a PI-containing regimen. However, recent clinical studies and meta-analyses revealed an improved metabolic profile, with discontinuation rates due to adverse effects at 4.7% at 192 weeks and reported cardiac events rates at 0.18% (9/4992) for patients treated with DRV/r (Orkin *et al.*, 2013; Martinez *et al.*, 2014; Antinori *et al.*, 2018). Of note is that second-generation PIs continue to show a low bioavailability below 40% and rely on co-administration of CYP3A4 (Cytochrom P450 3A4)-inhibitors in order to constantly surpass efficient blood-levels and not exceed toxic dosages (Rittweger and Arasteh, 2007; Kakuda *et al.*, 2014). CYP3A4, however, is a key component in an abundance of drug-metabolising pathways, and its expression levels are greatly influenced not only by pharmaceutical drugs, but also by other molecules in commonly used consumables, such as citrus fruits or herbal products (Eagling, Profit and Back, 1999; Bailey and Dresser, 2004). CYP3A4's inhibition or induction may result in unpredictably high or low PI

blood-levels, and can thereby cause severe adverse effects, or alternatively failure of PI based regimens (Zanger and Schwab, 2013). Despite these limitations for the use of PIs in the general population, selective patients may benefit greatly from their incorporation into cART regimens. Especially the high genetic barrier renders DRV a viable option for second-line and salvage therapy.

1.2.5 Integrase inhibitors (INSTIs)

Integrase inhibitors were the latest class of ART to be approved by the FDA. Owing to their characteristic to specifically target the IN-catalysed strand-transfer reaction, opposed to IN's 3'-end processing, they will be referred to as Integrase strand-transfer inhibitors (INSTIs) (1.1.3). Efficient long-term inhibition of viral cDNA integration is crucial, as delayed integration, up to 72 hours post-infection, may result in an unimpaired viral replication cycle (Thierry *et al.*, 2015). This is achieved by stable and functionally irreversible incorporation of INSTIs into the PIC. Here, all INSTIs inhibit the nucleoproteocomplex's active site via two distinct mechanisms: On the one hand, three planar oxygen atoms chelate IN's two necessary divalent metal ions. On the other hand, the INSTI's halobenzyl group interacts with the viral DNA's processed 3'-CA dinucleotide, the opposing strand's guanine base, as well as the enzyme's residues 145 and 146 in the catalytic pocket. These drug-enzyme and drug-DNA interactions lead to conformational changes within IN's active site that ultimately displace the open 3'-OH terminus by a few Å. Consecutively, the intasome complex remains a stable unit albeit IN no longer being able to perform the nucleophilic substitution reaction necessary for viral integration.

The two functional groups are separated by a linker sequence of varying length and rigidity (see Fig. 1.3) (Nowak *et al.*, 2009; Hare *et al.*, 2010; Xue, Liu and Yao, 2012; Thierry, Deprez and Delelis, 2017).

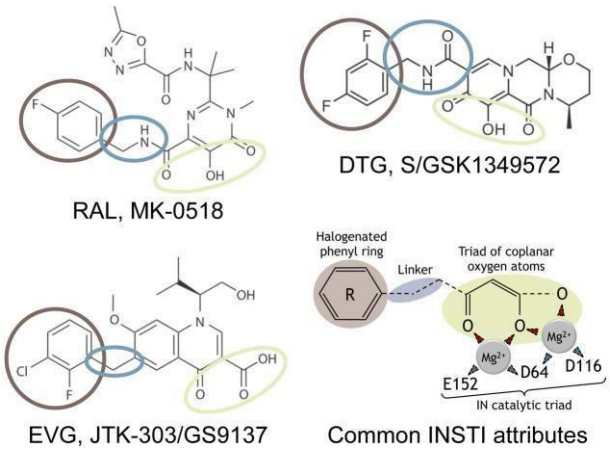

Fig. 1.3 Chemical structure of the INSTIs RAL, EVG and DTG, according to Thierry et al. (Thierry, Deprez and Delelis, 2017)**. This figure is used under the CC BY 4.0 license (https://creativecommons.org/licenses/by/4.0/) and no changes were made to the original figure. Mutual INSTI characteristics are colour-coded. Yellow marks the triad of co-planar oxygen atoms, brown the halogenated benzyl-ring and blue the linker region connection the two functional units.**

Available InSTIs comprise first-generation compounds Raltegravir (RAL) and Elvitegravir (EVG), as well as second-generation InSTIs Dolutegravir (DTG) and the most recently approved ART-drug Bictegravir (BIC). Since first approval in 2007 and 2012, respectively, RAL and EVG have shown non-inferior, and in case of RAL superior virological efficacy in treatment-naïve patients, when compared to NNRTI based cART (Rockstroh *et al.*, 2013; Zolopa *et al.*, 2013). Especially in the early phases of treatment, InSTIs are characteristically associated with a swift decline in viral loads, and throughout the course of treatment express few side effects or drug-drug interactions. Additionally, due to the unique mechanism of action, they have proven to be a viable option in second-line or salvage therapy for individuals, failing PI, and/or NNRTI based regimens (Grinsztejn *et al.*, 2007; Molina *et al.*, 2012; Viani *et al.*, 2015). However, the use of RAL and EVG also bears limitations in the form of unfavourable pharmacokinetics and a suboptimal resistance profile. While RAL requires a twice-daily formulation for optimal efficacy, EVG is mainly metabolized via CYP3A4 pathways and therefore relies on the co-administration of boosting agents, rendering EVG susceptible to a similar complication potential as seen in PIs (s. 1.2.4) (Klibanov, 2009; Eron *et al.*, 2011). Furthermore, RAL and EVG express modest barriers to resistance with one or two point mutations being able to confer high-level resistance to either of the drugs. Two of the most commonly observed high-level RAMs, namely Q148H/K/R/N and N155H/S/T/D, are selected for by both drugs, which explains high levels of cross-resistance between the two.

With DTG, approved in 2013, the first second generation INSTI became available, followed by BIC in 2018. Both drugs aim at overcoming RAL's and EVG relatively low genetic barrier to resistance, while maintaining their favourable side effect's profile and improving their pharmacokinetic interaction potential. Structurally, second-generation INSTIs display similarities to first-generation INSTIs, having three planar oxygen atoms, as well as a halogenated benzyl group. Of note is the slightly prolonged linker region that offers these molecules more internal flexibility and allows its two active groups to be positioned individually (s. fig 1.3). Thereby, the aromatic side chains may penetrate deeper into the vacated pocket, previously occupied by the viral 3'-A, and form a more stable complex (Smith *et al.*, 2018). This results in significantly slower dissociation rates between INSTIs and the PIC. Depending on the applied model, BIC and DTG show median dissociation half-life rates of 135h and 79h, respectively, compared to 14h for RAL and 13.6h for EVG (Hightower *et al.*, 2011; White *et al.*, 2017). Hence, it has been hypothesized that second-generation INSTIs are less prone to be affected by single-nucleotide replacement and express a higher genetic barrier to resistance (Engelman, 2019). Indeed, this effect could also be observed clinically, where DTG showed higher virological response rates with fewer emergences of RAMs than RAL and proved to be effective in patients, harbouring otherwise first-generation INSTI resistant viruses (Cahn, Anton L Pozniak, *et al.*, 2013; Castagna *et al.*, 2014). In treatment-naïve settings, DTG-containing regimens also showed superior viral suppression rates than other regimens (Molina *et al.*, 2015; Snedecor *et al.*, 2019). If virological failure occurred under treatment with a DTG containing regimen, in the majority of cases this was due to resistance to one of the co-administered drugs (Fourati *et al.*, 2015). Nevertheless, resistances against DTG and BIC have been confirmed in *in-vitro* drug-resistance selection assays, however no clear emergence-pattern of RAMs, following a distinct pathway to mutational resistance could be identified so far (Quashie *et al.*, 2012; Oliveira *et al.*, 2014). Among the more commonly observed mutations in DTG-resistant strains are substitutions at positions Arginine (Arg)263, Glu148 and Glycine (Gly)118, possibly isolating these positions as pivotal in conferring resistance to DTG (Castagna *et al.*, 2014; Munir *et al.*, 2015; Cardoso *et al.*, 2018). Most of them are extremely scarce in formerly treatment-naïve patients, as they are associated with a significant reduction in the enzyme's activity and viral replication rates (Mesplède *et al.*, 2013; Anstett *et al.*, 2015). The existence of compensatory, secondary RAMs following the emergence of primary RAMs has been

hypothesized, however, to date a direct path to restoring all enzymatic functionalities remains elusive.

The combination of superior efficacy in terms of viral suppression, good tolerability with few drug-drug interactions and a high genetic barrier to resistance, together with the availability of once-daily fixed-dose formulations, render DTG containing cART regimens a viable and preferred option in most modern first-line cART guidelines. Questions that need to be answered in the future include INSTI's long-term side-effects with regards to weight-gain, neuropsychiatric adverse effects, and renal implications, as well as the feasibility of NRTI-, specifically TDF/Tenofovir alafenamide (TAF)-sparing dual-therapy regimens containing DTG/DRV (Hoffmann *et al.*, 2017; Norwood *et al.*, 2017; Jabłonowska *et al.*, 2019; Venter *et al.*, 2019).

1.3 HIV in South Africa

1.3.1 Epidemiology

South Africa is home to the largest single-country population of PLWH. In 2018, 7.7 million people were reported to be living with the virus in South Africa, accounting for more than 20% of HIV's global burden of 37.9 million PLWH for the same year. Prevalence rates remain high at 20.4% in the general adult population aged 15-49 with higher prevalence rates in high-risk populations. While 240000 people were newly infected with the virus, 71000 people died from AIDS-related illnesses (UNAIDS, 2019).

In 2014, UNAIDS issued a list of targets that, if achieved, could potentially form a pathway to end the AIDS epidemic by 2030. These targets foresee, that by 2020 every country reaches the goal of having 90% of PLWH knowing their HIV status, 90% of people who know their HIV status receiving cART and 90% of the people on cART having suppressed viral loads (WHO, 2014).

South Africa adopted the 90-90-90 targets in its forth version of the national strategic plan (NSP). Herein, eight goals and strategies in order to reduce morbidity and mortality associated with HIV, tuberculosis (TB), and sexually transmitted diseases are formulated (The South African National AIDS Council, 2017). Strategies to meet UNAID's fast-track goals are multi-facetted and revolve around educational, preventive and therapeutic measures. Foremost, the

implementation of decentralised HIV-testing approaches, such as work place, community- and home-based testing facilities as well as a high rate of screening of new-borns exposed to HIV has shown high impact. Coupled with the introduction of various international initiatives, such as the Unitaid 'Self-Testing Africa (STAR)' initiative, offering HIV self-test kits, or USAID's 'Determined, Resilient, Empowered, AIDS-free, mentored and safe (DREAMS)' partnership that supports HIV response in adolescent girls and young women, a lot of progress has been made in raising awareness and testing for HIV. This allowed South Africa to reach the first goal of having 90% of PLWH knowing their HIV status by 2017 (Unitaid, 2018; Ingold *et al.*, 2019; UNAIDS, 2019). However, while individual projects haven shown that the targeted goals and their timeline remain feasible on a smaller scale, nation wide results for linking the target cascade remain unsatisfactory (Kharsany *et al.*, 2018; Médecins Sans Frontières (MSF), 2019). With 88% of people who receive cART showing viral suppression, but only 69% of people, who know their HIV status, receiving cART, currently a total of 55% of all PLWH show sufficient viral suppression, compared to the aim of 72% in UNAIDS' fast-track goals (UNAIDS, 2019).

1.3.2 National treatment guidelines

South Africa's national treatment programme, initiated in 2004, has since grown to become the largest cART treatment programme in the world, currently treating 4.8 million PLWH yearly (UNAIDS, 2019). Its treatment guidelines have been updated regularly and the latest iteration was issued in October 2019. Prior to this version, the preferred treatment regimen consisted of a backbone of TDF/FTC or TDF/3TC, combined with a core compound of EFV, DTG or RPV as third drug (Southern African HIV Clinicians Society *et al.*, 2017). However, the extensive use of NNRTI in the past has lead to rising number of transmitted NNRTI resistant variants. National drug resistance surveillance in 2017 reported resistance rates against NNRTIs of 19% in treatment-naïve patients and 38% in patients with prior exposure to cART (National Institute for Communicable Diseases, 17AD). These data, along with the wider availability of second-generation INSTIs in South Africa, have led to the latest change of South Africa's consolidated national HIV treatment guidelines. As of October 2019 these guidelines solely recommend a fixed-dose TDF, 3TC and DTG containing first-line regimen for treatment-naive adult men and adolescent boys $\geq$ 35 kg and $\geq$ 10 years of age as well as for treatment-naïve women and adolescent girls $\geq$ 35 kg and $\geq$ 10 years of age, who are not of childbearing potential (Southern African HIV Clinicians Society *et al.*, 2018; National Department of Health, 2019). As DTG, if

initiated peri-conceptively, has been associated with a minimal increase of neural tube defects, pregnant women and women wanting to conceive should receive counselling, on the risks and benefits of DTG and the alternative EFV based TEE therapy (TDF, FTC, EFV) (Zash *et al.*, 2018; Vannappagari and Thorne, 2019). Thereafter the client should be able to make an informed decision on the individually preferred treatment regimen. DTG based therapy initiated after the closure of the neural tube by the end of the sixth week of pregnancy (fourth week after conception) is not associated with birth defects and remains recommend throughout in this group.

Second and third-line therapy is dependant on the regimen used for first-line therapy and requires the confirmation of virological failure on two consecutive occasions, at least three months apart, as well as the optimization of adherence. Of note is, that due to its higher genetic barrier, virological failure because of DRMs is not considered to occur early in a DTG based regimen. Therefore a patient is required to be on DTG based therapy for at least two years, before a switch to second-line therapy should be considered (National Department of Health, 2019).

1.4 Aim of this study

In the absence of a cure for human HIV infections, cART and viral suppression remains a cornerstone in our response to combating the virus. With UNAID's 90-90-90 goals, a unified and global approach has been offered to realistically end the AIDS epidemic by 2030. South Africa, being home to the largest population of PLWH worldwide, has achieved to meet UNIAD's first 90-goal, however, with only 69% of people, who know their HIV status receiving cART, linkage to HIV care has proven unsatisfactory. The latest implementation of DTG-based cART in the national treatment guidelines may help overcome this obstacle, in that TDF/3TC/DTG is associated with fewer side-effects, when compared to previous treatment options, and therefore may lead to fewer discontinuation rates due to adverse effects, while having shown to be more effective in short- and long-term viral suppression. Furthermore, DTG's high genetic barrier to resistance renders treatment failure, due to emergent RAMs, unlikely. Lastly, a simplification of national ART guidelines, offering the same treatment option for the vast majority of the population, may lead to a demystification and therefore destigmatisation of cART in the eye of the public and could broaden the societal acceptance of HIV.

These outlooks, however, rely on a stable and high efficacy of DTG and while other parts of HIV-1's genome have been extensively studied in the past, the IN region only became of interest recently. Similar to previous studies that found genetic inter-subtype diversity to influence previous cART regimens, the presence of naturally occurring polymorphisms within the IN gene in different clades, CRFs and URFs, may influence their susceptibility to DTG, as well as their genetic barrier to individual non-polymorphic mutational resistance pathways (Wainberg and Brenner, 2012; Doyle, Dunn, Ceccherini-Silberstein, De Mendoza, Garcia, Smit, Fearnhill, A. G. Marcelin, *et al.*, 2015; Rogers *et al.*, 2018). With this study we aim to provide further data on INSTI-susceptibility, the diversity of HIV's IN gen in South Africa as well as a baseline mutational profile for South Africa's HIV-1 subtype C predominated context.

2 Material

2.1 Equipment

Piece of Equipment	Supplying company
GeneAmp PCR System 9700 thermal cycler	Applied BioSystems, USA
Nanodrop™ ND 1000	Nanodrop Technologies Inc., USA
GeneAmp® 9700 PCR system thermal cycler	Applied Biosystems, USA
UV-ITEC Prochem Gel Dock System	Whitehead Scientific, South Africa
Eppendorf Centrifuge 5424R	Eppendorf, Germany
ABI prism® 3130XL automated DNA genetic analyzer	Applied Biosystems, USA
Digital Vortex-Genie 2	Applied Biosystems, USA

2.2 Additional materials

Material	Supplying company
Gloves	Lasec, South Africa
Collection tubes (2ml)	Qiagen, Germany
Disposable transfer pipets (3.2ml)	Samco, Japan
Eppendorf safe-lock tubes (1.5ml, 2ml)	Eppendorf, Germany
Erlenmeyer flask (250ml, 1000ml)	Schott
Measuring glass cylinder (100ml, 250ml, 1000ml)	Schott
Pipet tips, filtered (1000µl, 200µl, 100µl, 10µl)	Quality scientific plastic, USA
Pipet tips, unfiltered (1000µl, 200µl, 100µl, 10µl)	Quality scientific plastic, USA
Parafilm® M All-Purpose	Bemis, USA
PCR- reaction tubes (0.2ml)	Applied Biosystems, USA
PCR-Microplate plate (96 wells)	Thermo Fisher Scientific, USA
MicroAmp® 8-Cap Strips	Thermo Fisher Scientific, USA

2.3 Reaction Kits

Reaction Kit	Supplying Company	Catalogue number
QIAamp DNA Blood Mini Kit	Qiagen, Germany	51106
QIAquick PCR Purification Kit	Qiagen, , Germany	28 106
BigDye™ Terminator cycle sequence ready Kit	Applied BioSystems, USA	4 337 035
BigDye XTerminator Purification Kit	Applied BioSystem, USA	4 374 408

2.4 Reagents

Reagent	Supplying Company
Nuclease free water	Qiagen, Germany
Agarose	Lonza, USA
Gel Loading Dye, Purple 6X	New England Biolabs®, USA
Novel Juice gel stain	GeneDirex, USA
5x Sequencing Buffer	Applied BioSystems, USA
Quick-Load® 1kb DNA Ladder	New England Biolabs, USA
Ethanol 100%	Sigma-Alrich, USA

2.5 Enzymes

Enzyme	Supplying company
SuperScript™ III One-Step RT-PCR enzyme mix	Invitrogen, USA
GoTaq® Flexi DNA Polymerase	Promega, USA
BigDye™ Terminator v3.1	Applied Biosystems, USA

2.6 Primer

All primer in this study were ordered from Ingaba Biotech, South Africa, and were received as lyophilisates. After reception, the lyophilised primers were treated with nuclease-free water according to the manufacturer's instructions in order to obtain a 10 µM concentration. The solutions were further diluted to obtain final concentrations of 5nM for RT-PCR and 10nM for nested PCR. Diluted aliquots were stored at -20°C until use. All primer's sequences were adopted from a previous publication on an IN amplification and sequencing assay that had yielded satisfactory results for HIV-1 subtypes A, B, C, D, F, G and J, as well as CRFs CRF01-AE, CRF02-AG, CRFF03-AB, CRF12-BF and CRF13-cpx (Laethem *et al.*, 2008).

Primer for RT-PCR

Primer	Sequence	HXB2 positions
Poli 5 (forward)	5’-CACACAAAGGRATTGGAGGAAATG-3’	4162 - 4185
Poli 8 (reverse)	5’-TAGTGGGATGTGTACTTCTGAAC-3’	5217 - 5195

Primer for nested PCR

Primer	Sequence	HXB2 positions
Poli 7 (forward)	5’-AACAAGTAGATAAATTAGTCAGT -3’	4186 - 4208
Poli 6 (reverse)	5’-ATACATATGRTGTTTTACTAARCT-3’	5130 - 5107

Additional inner primer for sequencing reactions

Primer	Sequence	HXB2 positions
Poli 2 (forward)	5’-TAAARACARYAGTACWAATGGCA-3’	4745–4766
KVLO 83 (reverse)	5’-GAATACTGCCATTTGTACTGCTG-3’	4772–4750

2.7 Computer software and online tools

Computer programs

Software-Name	Developer
Sequencher Version 5.0	Gene Codes Corporation, USA
Molecular Evolutionary Genetics Analysis (MEGA) version 7.0 and version X	Pennsylvania State University

Software-Name	Developer
Microsoft Excel 2011	Microsoft
Microsoft Word 2011	Microsoft

Websites, algorithms and onlinetools

Name	Website
Los Alamos National Library sequence database	https://www.hiv.lanl.gov/content/sequence/HIV/mainpage.html
Multiple Sequence Comparison by Log- Expectation (MUSCLE) algorithm (implemented in MEGA)	https://www.ncbi.nlm.nih.gov/pmc/articles/PMC390337/
REGA HIV- 1 & 2 Automated Subtyping Tool (Version 3.41)	http://www.bioafrica.net/manuscripts/Pineda_SubtypingTools2013.pdf
jumping profile Hidden Markov Model	http://jphmm.gobics.de/
Stanford University HIV Drug resistance database	https://hivdb.stanford.edu/
Consensus Maker	https://www.hiv.lanl.gov/content/sequence/CONSENSUS/AdvCon.html
Basic Local Alignment Search (BLAST) tool	https://blast.ncbi.nlm.nih.gov/Blast.cgi
GraphPad QuickCalcs tool	https://www.graphpad.com/quickcalcs/contingency1/

3 Methodology

3.1 Sample collection and storage

Samples used for analyses were collected and stored as part of the larger Tygerberg Virology study cohort. This cohort's samples were sampled between 1998 and 2002 and stored at -80 degree Celsius (°C) at the division of medical virology at the Tygerberg university hospital, South Africa (s. 4.1). After reception of individual EDTA containing whole-blood samples, these were spun to separate blood components, and blood plasma was consecutively frozen at -80°C for long-term storage. In order to prevent RNA degradation as well as the formation of cryoprecipitates from repetitive thawing and freezing, plasma samples were stored in aliquots containing 200µl each. Thawing was carried out gradually at room temperature. During all pipetting steps and reaction mix preparations, the samples and amplicons were continuously stored on ice.

3.2 Laboratory analyses

3.2.1 Extraction of viral RNA

Given the potential risk of contamination and infection, all steps involving the handling of intact virions were carried out in a separated laboratory, dedicated for nucleic acid extraction reactions, under a workbench hood with laminar airflow. Pipet tips were changed after every transfer of liquids and gloves were changed in between steps of the protocol.

Starting the laboratory analyses of HIV-1's genome, viral RNA had to be extracted from the patients' plasma samples. For this step the QIAamp Viral RNA Mini Extraction Kit's (Qiagen, Germany) spin protocol was used according to the manufacturer's instructions. In short, specific RNA binding properties of a silica-based membrane combined with the washout of the remaining plasma components result in purified RNA products.

In detail, the provided reagents were firstly prepared according to the manufacturer's handbook: lyophilized carrier RNA was dissolved in Buffer AVE and diluted to 1µg/ml, pure ethanol (96-100%) was added to Buffer AW1 and AW2 and a varying amount of carrier-RNA-containing Buffer AVE and Buffer AVL was mixed, depending on the amount of samples to be extracted, right before starting the extractions. Following the recommendations, I used 5.6µg of carrier-RNA per sample. Simultaneously, frozen EDTA-plasma samples were thawed at room temperature.

For viral lysis, 560µl of prepared, carrier RNA containing Buffer AVL was pipetted into a 1,5ml micro-centrifuge tube, followed by 140µl of EDTA-plasma sample. Thorough mixing was ensured by pulse-vortexing for 15 seconds and brief centrifuging for 30seconds thereafter

removed possible drops from the inside of the lids. Buffer AVL effectively lyses viral particles and also inactivates present RNases, however the addition of carrier RNA can bind instable, single-stranded viral RNA and increase its resilience to residual RNase activity. Furthermore it increases the yield of RNA-binding to the QIAmp membrane (s. below). After a ten-minute incubation period, during which the lysing-process is completed, 560µl of ethanol (96-100%) was added, again, followed by pulse-vortexing and centrifuging, to obtain a 1260µl reaction mix. Half of this mix (630µl) was then carefully pipetted onto a QIAmp Mini column placed in a 2ml collection tube and centrifuged at 6000 x g (8000 rpm) for one minute. Thereafter the filtrate was discarded and the QIAmp Mini column was loaded with the second half (630µl) of the reaction mix, followed by centrifugation under the same conditions. During this step viral and carrier RNA selectively bind to the silica-based membrane, whereas the remaining mixture is spun through the membrane into the filtrate. However, contaminants that are retained in the membrane during a single centrifugation can severely impair downstream reactions and need to be removed. Therefore two consecutive washing steps with 500µl of Buffer AW1 and Buffer AW2, respectively, are included in this protocol. After application of Buffer AW1, centrifugation is set at 6000 x g (8000 rpm) for one minute, whereas centrifugation after application of Buffer AW2 is carried out at 20 000 x g (14 000 rpm) for three minutes. An additional centrifugation after discarding the filtrate, eliminating the possibility of contamination of the membrane with AW2 flow-through during deceleration of the centrifuge was performed at 20 000 x g (14 000 rpm) for one minute. To yield a higher concentration of viral RNA, the QIAmp Mini columns were placed in clean 1.5ml micro-centrifuge tubes and 50µl of eluting Buffer AVE was applied to the spin column, followed by incubation at room temperature for one minute. A last step of centrifugation at 6000 x g (8000 rpm) for one minute eluted the purified and concentrated viral RNA into 50µl of Buffer AVE. Purified RNA products were either used directly for RT-PCR (s. 3.3.3), or stored at -80°C.

3.2.2 Photometric measurement of RNA concentration

Purity and concentration of a DNA or RNA sample can be assessed by spectrophotometry. This is based on the observation, that different molecules show a distinctive absorption pattern, based on the wavelength of the emitted light. DNA, RNA and single nucleotides show an absorption peak at a wavelength of 260nm, whereas contaminants show a different absorption spectrum. Proteins, for instance, generally absorb light at a wavelength of 280nm, whereas EDTA absorbs near 230nm. Knowing these absorption spectrums, the measured

260nm/280nm as well as the 260nm/230nm ratios can be used for estimating the nucleotides/contaminants content in a sample. 260nm/280nm ratios ~1.8 are generally considered to be pure for DNA, whereas, due to the higher absorption of Uracil in RNA compared to Thymidine in DNA, ratios ~2.0 are considered pure for RNA. 260nm/230nm ratios normally range between 2.0-2.2 for pure samples, however composition of the respective nucleotide-strand can substantially alter these results. Additionally, when handling pure samples, knowledge of the substrate's known absorption and the sample's measured absorption value at a given wavelength can be used to convert the sample's absorption values into its concentration using Beer's law. For DNA, this conversion can be made manually using the following formula, however most spectrophotometers calculate the concentration internally:

$$\text{Concentration of single-stranded DNA} = A_{260} \times 33 \text{ mg/}\mu\text{L}$$
$$\text{Concentration of double-stranded DNA} = A_{260} \times 50 \text{ mg/}\mu\text{L}$$

In our study, spectrophotometry was performed on a ™ND1000 Spectrophotometer (Thermo Fisher Scientific, USA). Spectrophotometric assessment of samples was used at two points in our workflow. Once, after extraction of viral RNA (3.3.1), and a second time after purification of PCR results before sequencing reactions (3.3.6). Each time the spectrophotometer was calibrated using the buffer, in which the RNA/DNA was dissolved in. In case of newly extracted RNA, calibration was done with Buffer AVE. For these measurements, 260nm/280nm absorption results were >3.3 throughout, whereas 260nm/230nm spectrophotometry showed results <1.0. No clear correlation between NanoDrop results and successful PCR or sequencing reactions could be made at this stage. Possible explanations could be a high content of Adenine and Thymidine in the HIV-1 IN fragment, showing a >2 fold increase absorption pattern at 260nm compared to Guanidine and Cytosine, combined with the presence of EDTA, absorbing strongly near 230nm (Geuther, 2007). On the other hand, Buffer AVE used in extracting viral RNA (s. 3.3.1) contains sodium azide to prevent microbial growth and the expression of RNases. Sodium azide is known to affect spectrophotometric measurements between 220-280 nm. Additionally, the amount of dissolved carrier-RNA, used to protect viral RNA from RNases, as well as increase yield of isolation, exceeds the amount of viral RNA and will vastly determine spectrophotometric results. Therefore NanoDrop results after extraction of viral RNA were not taken into account when planning future experiments.

Spectrophotometric purity assessment of purified PCR products later in the workflow yielded expected results of 260nm/280nm ratios between 1.9-2.2 and 260nm/230nm ratios between 1.9-2.6 (3.3.6).

3.2.3 Reverse-Transkriptase Polymerase Chain Reaction (RT-PCR)

In order to analyse specific genomic regions, such as the IN segment, multiple copies of this part of the genome need to be present in a pure, highly concentrated form (s. 3.3.7). Amplification of genomic segments is performed by using the concepts of polymerase chain reactions. Herein, the natural course of replication via RNA- and/or DNA-dependent DNA polymerases is imitated and the scale of replication is greatly increased. By executing multiple iterations of a reaction cascade consisting of DNA-denaturation, primer annealing and strand elongation, the yield of the desired DNA fragment is increased exponentially, theoretically by 2^{n+1} for DNA based PCRs and 2^n for RT-PCR with n being the number of cycles. Requirements for this reaction are the presence of a template strand, containing the desired fragment, a RNA/DNA-dependent DNA polymerases catalysing the polymerisation, forward and reverse primers, flanking the desired segment and thereby marking start and end point of the reaction, a reaction mix, providing dNTPs, Mg^{2+} ions and buffers for optimal biochemical conditions and an inert reaction medium, merging all the necessary components. Working with RNA-based organisms, such as HIV-1, the relatively unstable ssRNA-strand in a first step needs to be reversely transcribed into cDNA before commencing the replication cycles. This process was performed using the SuperScript™ III One-Step RT-PCR System with Platinum™ *Taq* High Fidelity DNA Polymerase (Thermo Fisher Scientific, USA). Containing two enzyme mixtures, a reverse transcriptase (SuperScript™ III) as well as a DNA-dependent DNA polymerase (Platinum™ Taq High Fidelity DNA Polymerase), reverse transcription into cDNA and amplification thereof can be carried out in one reaction batch. SuperScript™ III is a retroviral RT that has been synthetically altered to express increased thermo-stability with an optimal reaction temperature of 45-60°C as well as a reduced RNase H activity. This results in RNA/cDNA hybrids, as RNase H will not degrade the template strand after assembling the corresponding DNA strand (s. 1.1.4).

Platinum™ Taq High Fidelity DNA Polymerase contains a mixture of recombinant Taq DNA polymerase, Pyrococcus species GB–D polymerase, and Platinum™ Taq antibodies. Taq DNA polymerase is a heat-stable DNA polymerase, being able to sustain temperature exceeding 90°C. However, lacking a proofreading mechanism, Taq DNA polymerase relies

on the activity of GB-D polymerase to express 3' to 5' exonuclease activity in order to increase fidelity of the polymerisation reaction. Platinum™ Taq antibodies inhibit Taq DNA polymerase's enzymatic activity at room temperature, which allows for a so-called "hot start". In that, the Taq DNA polymerase can selectively be activated during the replication process after successful RT by increasing the reaction temperature to 94°C leading to denaturation of the antibodies and a restored enzymatic activity. Furthermore, temperatures of 94°C result in denaturation of SuperScript™ III as well as the RNA/cDNA hybrids. These information lead to the following cycling conditions:

Step	Temperature	Duration	
Reverse transcription	50° C	20 min	
Hot start	94° C	2 min	
Denaturation	94 °C	15 sec	40 cycles
Primer Annealing	50 °C	30 sec	
Elongation	68 °C	90 sec	
Final Elongation	68 °C	10 min	
Storing	4 °C	Indefinite	

Set temperatures for denaturation and elongation were adopted from the enzyme's manufacturer's recommendation. Ideal primer annealing temperature for RT-PCR was recommended to be 10°C below their respective melting temperatures (T_M), which depends on the primers' length, as well as their ACTG content, as AT pairs form two hydrogen bonds, whereas CG pairs exhibit three. Additional factors such as base-stacking effects or the exact concentration of Mg^{2+} ions were not taken into account in the following calculations. For gene-specific primers exceeding 15 bases, the melting temperature can be roughly estimated by:

$$T_M[°C] = 69.3°C + [(41(n_G + n_C) / s) - (650 / s)]°C$$

$n_{G/C}$ = number of G or C nucleosides

s = number of all nucleosides per sequence

With primers Poli5 and Poli8 used for RT-PCR (s. 2.5), T_M were estimated at 59.3°C and 58.3°C, respectively. Annealing temperature below 50°C have not shown better yields, therefore the annealing temperature was set at 50°C.

Reaction mixes for RT-PCR were composed as follows:

Reagents	Volume	Concentration
Nuclease free water	8.5 µl	-
2xReaction Mix	12.5 µl	-
Forward Primer (Poli 5)	0.5 µl	5 nM
Reverse Primer (Poli 8)	0.5 µl	5 nM
Enzyme Mix	0.5 µl	-
RNA Template	2.5 µl	Not available
Total	25 µl	

Reaction mixtures were always pipetted on ice into 0.2ml PCR reaction tubes, and thoroughly mixed. SuperScript™ III One-Step RT-PCR System with Platinum™ *Taq* High Fidelity DNA Polymerase (Thermo Fisher Scientific, USA) was added second to last, before adding the RNA template in a different laboratory. Thermocycling was carried out on a GeneAmp PCR System 9700 thermal cycler (Perkin Elmer, Boston, USA). After RT-PCR, the amplicons were either used directly for Nested PCR or stored at 4°C until further processing.

3.2.4 Nested Polymerase Chain Reaction

To further increase sensitivity and specificity of gene amplification, all amplicons from RT-PCR were subject to a second round of PCR. Herein, the already amplified estimated 1056bp long DNA fragment containing the IN-gene was used as a template to specifically increase the yield of a shorter, 945bp long sequence within this fragment. As the template has previously been transformed into cDNA, a single DNA-dependent DNA polymerase is sufficient for the reaction and a "hot start" is not required. I used a GoTaq® Flexi DNA Polymerase Kit with colourless 5X GoTaq® Flexi Buffer (Promega, USA) for nested PCRs. The GoTaq® DNA polymerase contains intrinsic exonuclease activity and does not need additional proofreading mechanism for a high-fidelity reaction. Unlike other reaction buffers, however, the 5X GoTaq® Flexi Buffer does neither contain Mg^{2+} ions, which are a necessary co-factor for all DNA polymerases, nor dNTPs that a required for successful elongation. Therefore these

components were added to the reaction mix separately. The final reaction mix for nested PCR was composed as follows:

Reagents	Volume	Final Concentration
Nuclease free water	28.5 µl	-
5xReaction Mix	10 µl	-
dNTP Mix	1 µl	0.2 mM each
Forward Primer (Poli 7)	2 µl	10 nM
Reverse Primer (Poli 6)	2 µl	10 nM
$MgCl_2$	3 µl	1.5 mM
Enzyme Mix	0.5 µl	2.5 u
DNA Template	3 µl	Not available
Total	50 µl	

Thermocycling was carried out on a GeneAmp PCR System 9700 thermal cycler (Perkin Elmer, Boston, USA) under following parameters:

Step	Temperature	Duration	
Initial denaturation	95 °C	2 min	
Denaturation	95 °C	20 sec	40 cycles
Primer Annealing	55 °C	30 sec	
Elongation	72 °C	90 sec	
Final Elongation	72 °C	10 min	
Storing	4 °C	Indefinite	

Denaturation and elongation temperatures were adopted from the enzyme's manufacturer's recommendation. Primer annealing temperatures were calculated for primers Poli 6 and Poli 7 (s. 2.2) using Promega's T_M Calculator available on their BioMath website (www.promega.com/biomath/). Depending on the interpretation of ambiguity code in Poli 6, T_M for Poli 6 and Poli 7 ranged between 52°C and 55°C. Following the reference article for these primers we decided to adopt the 55°C used in the original experiments (Swanson, Devare and Hackett, 2003).

After nested PCR, all amplicons were either directly screened for successful IN amplification via agarose gel electrophoresis, or stored at 4°C until further processing.

3.2.5 Agarose Gel electrophoresis

Agarose gel electrophoresis is commonly used to separate molecules according to their electric charge and size. An electric field is applied to a gel matrix, where the molecules to be studied are loaded onto. The direction of travel for a given molecule is mainly determined by its resulting electric charge. Since most molecules of biochemical interest carry a negative charge, they are applied onto the gel in wells near the cathode. The speed at which a molecule travels through the gel matrix is dependant on its charge as well as the its size. Small molecules travel faster through the gel matrix than larger ones. Visualisation is achieved by the addition of a fluorescent gel stain to the matrix that interacts with the travelled molecules and becomes fluorescent after exposure to UV-light.

In DNA strands, uncharged units of 5-carbon-sugars and purine or pyrimidine bases are connected by phosphodiester bonds. At physiological pH-values, these bonds carry a negative charge, which results in an overall negative charge of the entire molecule. As a DNA strand's charge is directly correlated to its length (plus one per incremented bp), gel electrophoresis involving DNA fragments are often referred to as separating only by size, negating the invariable factor of the molecule's electric charge.

Expecting a fragment size of 945bp (s. 2.2), I used 1.5% agarose gel matrices for electrophoresis. These were prepared by fully dissolving 1.5g of molecular grade agarose (Invitrogen, USA) in 100ml of 1 x Tris-Acetate-EDTA (TAE) buffer while heating the mixture in a microwave oven for four minutes. After a first phase of cooling, 10µl of gel stain (Novel Juice, GeneDirex, USA) was added and gradually mixed. Thereafter, the mixture was poured into a plastic electrophoresis gel tank, a comb creating wells for the appliance of PCR products was inserted and the gel was allowed to set. Larger gels were prepared with 2.25g of agarose, dissolved in 150ml of 1 x TAE buffer and otherwise identical preparation method. After hardening of the gel, the comb was removed and 1 x TAE buffer was added until fully covering the gel. Then, 5µl of every nested PCR product were mixed with 1µl of 6 x loading dye (Thermo fisher scientific, USA) and subsequently pipetted into the gel's wells. For verification and estimation of the amplified fragments, 6µl of a Quick-Load® 1kb DNA Ladder (New England BioLabs, United Kingdom) was loaded alongside the PCR products. Quality control was achieved by also adding the PCR's positive and negative control to the gel. Electrophoresis was carried out at 80mV for 30 minutes. Potential bands were visualised with the Alliance Chemilluminscence and Fluorescence gel imaging system (UVItec, UK).

Samples expressing a signal at an estimated fragment size of 945bp were considered successfully amplified on the IN gene and were used for analyses.

3.2.6 PCR product purification

All nested PCR products, expressing visually positive amplification of the expected fragment-length of 945 bases in the agarose gel electrophoresis (s. 3.3.5), were subsequently purified using the QIAquick PCR Purification Kit (Qiagen, Germany). Aim of this reaction is to remove primers, excessive nucleotides, polymerases and salts from the amplicons, which could possible impair the ensuing sequencing reaction.

In a first step, Buffer PE and Buffer PB were prepared by adding ethanol (96-100%) and pH indicator I, respectively, according to the manufacturer's instructions. Thereafter, five parts (225µl) of prepared Buffer PB were mixed with one part (45µl) of the remaining amplicon in a 1.5ml centrifuge tube, after 5µl were used for gel electrophoresis. In case its colour was brown, the solution was titrated with 3 M sodium acetate, until turning yellow, indicating a pH of $\leq$ 7.5. The mixture was then transferred onto a QIAquick spin column, placed in a 2 ml collection tube and centrifuged in a microcentrifuge (Eppendorf Centrifuge 5424R, Germany) at 13 000 rpm for one minute. In this step, similar to the process of extracting the viral RNA from plasma samples (s. 3.3.1), the viral DNA binds to the columns matrix. After discarding the flow-through, 750µl of prepared washing Buffer PE were applied to the QIAquick spin column, followed by two separate centrifugation steps in order to remove residual ethanol from the matrix. Elution of the matrix-bound DNA into a 1.5ml microcentrifuge tube was done by the appliance of 30µl of Buffer EB to the centre of the QIAquick spin column's membrane, followed by an incubation period of one minute at room temperature and centrifugation at 13 000 rpm for one minute.

After successful elution of the DNA, 1µl was used to spectrophotometrically measure the samples purity, as well as its concentration (s. 3.3.2). Satisfactory purity results permitted the concentration values to be used to dilute the samples to 12.5 ng/µl, which we have found to be optimal for our sequencing protocol (s. 3.3.7).

3.2.7 Sanger-sequencing reaction

All purified PCR products, presumably containing the HIV-1 IN coding DNA fragment, were subsequently sequenced via conventional Sanger sequencing technique. The cycling reactions, as well as the purification of the products was carried out using the BigDye™ Terminator v3.1 Cycle Sequencing Kit (Applied Biosystems, USA) and the BigDye® XTerminator™

Purification Kit (Applied Biosystems, USA), respectively, while the capillary electrophoresis, laser stimulation and signal interpretation was run on an ABI 3130XL genetic analyser (Applied Biosystems, USA).

The Sanger sequencing technique relies on a PCR-like reaction, in which a polymerase repeatedly synthesises a new DNA strand by inserting dNTPs corresponding to the template strand. For Sanger sequencing, however, the reaction mix not only contains dNTPs, but to a smaller degree also di-deoxynucleoside-triphosphates (ddNTPs). This leads to strand termination, as no 3'-hydroxyl-group is available for strand elongation. Linking the four respective ddNTPs to four different distinct fluorescents, this reaction creates corresponding DNA strand fragments of various lengths that end in one of the four ddNTPs. Consecutively using capillary electrophoresis to separate the fragments by length and laser stimulation of the strand-terminating fluorescent can directly determine the dNTP intended to be incorporated at that position by the ddNTP's emitted signal's wavelength. Important to this is that only one primer can be used per sequencing PCR reaction, as different fragments, from different parts of the double-stranded template would impair the position-matching of the emitted signals. Furthermore, four primers are necessary to sequence all four possible double-stranded, nested-PCR products on both strands. Two of these can be reused from nested-PCR, two inner primers, however, are sequencing specific. The primers we used in our assay were nested-PCR primers Poli6, and Poli7, as well as sequencing primers Poli 2F and KVLO83 (s. 2.5). Poli6 and Poli7 share a mutual annealing temperature of 55°C, but Poli 2F shows an optimal annealing temperature of 50°C, while KVLO83 primers anneal best at 60°C. Implications resulting from these different annealing temperatures had to be taken into account, when planning sequencing PCRs (s. below). For sequencing reactions, I used optical 96-well reaction plates (Thermo Fisher Scientific, USA) sealed with MicroAmp® 8-Cap Strips (Thermo Scientific, USA). Each well contained a reaction volume of 5µl, with ~25ng of purified PCR-product as template, each.

In detail, the reaction mixes consisted of the following:

Reagents	**Quantity**	**Concentration**
BigDye™ Terminator 3.1 Ready Reaction Mix	0.5 µl	-

Reagents	Quantity	Concentration
BigDye™ Terminator v3.1 5X Sequencing Buffer	1.5 µl	-
Primer	0.5 µl	5 µM
Nuclease-free water	0.5 µl	-
DNA template	2 µl	12.5 ng/µl
Total	5 µl	

Thermocycling was carried out on a GeneAmp PCR System 9600 thermal cycler (Perkin Elmer, Boston, MA, USA), when only one primer, or primers with similar annealing temperatures (Poli6, Poli7) were used on one reaction plate. Reaction plates, containing primers with different annealing temperatures were run on a Veriti™ 96-Well Fast Thermal Cycler (Thermo Fisher Scientific, USA). Cycling parameters were set as follows:

Step	Temperature	Duration	
Initial Denaturation	96 °C	1 min	
Denaturation	96 °C	10 sec	25 cycles
Annealing	Depending on primer used	7 sec	
Elongation	60 °C	4 sec	
Storing	4 °C	Indefinite	

Since this thermocycling reaction again leaves excess dNTPs, ddNTPs and other contaminants in suspension, which could impair interpretation of the DNA fragments during sequencing. Therefore an additional purification step is necessary. For this step we used BigDye® XTerminator™ Purification Kit (Thermo Fisher Scientific, USA). XTerminator Solution was vortexed, premixed with SAM solution, and 60.5µl of the mixture were added to every well of the reaction plate. Regularly vortexing the premix solution for ~5 seconds in between pipetting steps is important to keeping the particles of the XTerminator Solution in a homogenous suspension and ensuring their equal distribution in the wells. After adding the XTerminator/SAM solution mixture to the wells, the entire plates were vortexed on a Digital Vortex-Genie 2 (Applied Biosystems) at 1800rpm for 30 minutes, followed by a final centrifugation at 1000 x g for two minutes on a swinging bucket Eppendorf Centrifuge 5810 R (Eppendorf, Germany). Final plates were then run overnight on an ABI 3130XL genetic analyser (Applied Biosystems, USA).

3.2.8 Sequence-Assembly

Raw data trace files containing sequence chromatograms were retrieved from the ABI 3130XL genetic analyser (Applied Biosystems, USA), imported into Sequencher version 5.0 (Gene Codes Corporation, USA), and subsequently analysed. Sequence ends with poor quality were removed, and overlapping parts of four trace files from the four primer-reactions per sample were assembled into one single continuous sequence strand. This strand was then consecutively screened for ambiguities in the form of double- or other multi-peaks, exported from Sequencher version 5.0 (Gene Codes Corporation, USA) and saved in a fasta format.

3.3 Software analyses

3.3.1 Alignment of sequences

All nucleotide as well as protein alignments were conducted in MEGA v7 software using the MUSCLE alignment algorithm with a gap penalty of -400 and 0 for opening and extending a gap for nucleotide alignments, and -2.9 and 0 for protein alignments, respectively (Edgar, 2004; Kumar, Stecher and Tamura, 2016). Hydrophobicity multiplier here was set at 1.2. The number of iterations was limited to eight, while using the UPGMB clustering method (combination of Unweighted Pair Group Method with Arithmetic Mean and neighbour-joining) for all eight iterations. The minimal diagonal length was 24.

In-house sequenced files were loaded into MEGA v7 and nucleotide-aligned to each other before being manually edited to all be in the first reading frame. The sequence set was then translated into their corresponding amino acid sequences and manually aligned to the HXB2 IN amino acid sequence (correspondent to nucleotide positions 4230-5096). Overhanging parts exceeding this fragment were removed from analyses to not influence phylogenetic interference of the IN genome. Subsequently, all sequences were screened for the presence of stop codons or larger insertions or deletions and in the absence thereof saved in two separate, either nucleotide-, or amino acid-aligned fasta files. Pre-aligned sequences, retrieved from online databases, were screened for alignment and sequence quality in MEGA v7 according to the same criteria.

3.3.2 Phylogenetic analyses

Phylogenetic interference was conducted in MEGA v7 as well as in MEGA X. Phylogenetic trees based on the neighbour-joining (NJ) tree-building model, were computed in MEGA v7

(Saitou and Nei, 1987). Phylogenetic trees based on the Maximum-Likelihood model (ML) were inferred in MEGA X. An IN reference set was retrieved from the Los Alamos Database (https://www.hiv.lanl.gov/content/sequence/HIV/mainpage.html), in form of the 2018 filtered web curated DNA alignment, selective for the HXB2 positions 4230-5096 and only including HIV-1 group M subtypes A-K without recombinants. From this set of 2322 sequences, 92 South African sequences and one sequence from the DRC were manually chosen as reference sequences for the ML phylogenetic tree. Additional sequences from individual BLAST-search results were added for specific analysis of two individual sequences (s. 4.4.2). The best fitted nucleotide substitution model for inferring the evolutionary distances between the sequences for the total set of all sequences used in the ML-tree was determined by a model-test, run on MEGA v7. Here, the model with the lowest Bayesian Information Criterion (BIC) was considered to be the model that described the substitution pattern the most precisely. In case of the ML-tree displayed in this study, the General Time Reversible (GTR) model was calculated to be best fitted with a proportion of 0.4429 of invariable sites (+I) and a presumed discrete gamma-shape heterogeneity parameter (+G) of 0.3086. As a ML-tree always requires to be rooted to an ideally equally distant relative to all taxons, one HIV-1 Group O sequence was chosen as out-group.

3.3.3 Consensus generation

The generation of consensi was computed using the Consensus Maker online tool (https://www.hiv.lanl.gov/content/sequence/CONSENSUS/AdvCon.html). The unanimity threshold was set at 0.95, the majority threshold at 0.51.

3.4 Database derived sequences

3.4.1 HIV-1_{ZA} sequences

All additional sequences were obtained from the Los-Alamos sequence database (http://www.hiv.lanl.gov/). Search criteria are summarized in tbl. 3.1. The query for HIV-1_{ZA} resulted in 347 results (07.11.2019). Since the aim of this study was to characterize INSTI DRMs and NOPs in an INSTI-naïve setting, only sequences sampled before the introduction of the first INSTI, RAL, in 2007 were of interest in order to eliminate the possibility of transmitted drug-resistance mutations and altered polymorphism expression. Therefore, 32 sequences that either did not state a sampling year, or were sampled in the year 2007 or later were excluded from analysis. Another eleven sequences were excluded, as they belonged to our TV cohort and multiple entries could not be ruled out. Furthermore, inclusion of TV

sequences could impact the comparative value between our study cohort and HIV-1_{ZA} sequences. Fragments and problematic sequences were excluded. After selecting one sequence per patient, 292 IN sequences were retrieved from the database. Reassessment of the sequence-quality revealed the presence of stop codons in five sequences, three more showed severe insertions. Ultimately, 284 sequences were used for further analysis. Thereof, 32 were partially sequenced, targeting the IN gen, whereas the remaining 252 sequences were part of Near-Full-Length-Sequences (NFLG).

	HIV-1_{ZA}	**HIV-$1C_{ZA}$**	**HIV-$1C_{non\text{-}ZA}$**	**HIV-1B**
Virus	HIV-1	HIV-1	HIV-1	HIV-1
Subtype	-	C	C	B
Sampling country	ZA	ZA	-	-
More sequence information	Drug-naïve	Drug-naïve	Drug-naïve	Drug-naïve
Defined start	4230	4230	4230	4230
Defined end	5096	5096	5096	5096

Table 3.1: Search criteria for sequence queries in the Los Alamos HIV Sequence Database (https://www.hiv.lanl.gov/components/sequence/HIV/search/search.html). Categories and criteria are directly adopted form the search interface. Defined start and end positions are in relation to the HXB2 reference strain, as implemented in the database.

3.4.2 HIV-$1C_{za}$ sequences

After subtype analysis of the HIV-1_{ZA} sequence set (s. 4.3.2), only nine sequences were listed as non-subtype C. Exclusion of these sequences resulted in a set of 275 HIV-$1C_{ZA}$ sequences used for the generation of the consensus strand (s. 4.6).

3.4.3 HIV-$1C_{non\text{-}ZA}$ sequences

The Los Alamos sequence database query, using the search criteria shown in tbl. 3.1 returned 1268 sequences. Selecting one sequence per patient resulted in a set of 868 sequences. Hereof, 314 were sampled in South Africa and consequently were removed from this cohort. The remaining 554 sequences were screened for sequence quality and six samples were subsequently excluded due to the presence of stop codons (s. 3.5.1). Therefore, the search criteria led to a set of 548 HIV-1 subtype C sequences from 14 different countries, other than South Africa. Country distribution were stated as followed: Zambia (455/548), United States of America (1/548), Tanzania (37/548), Sweden (9/548), Saudi Arabia (10/548), Pakistan (1/548), Nepal (2/548), Mozambique (1/548), Kenya (3/548), India (17/548), Georgia (1/548), Spain (5/548), China (2/548), Canada (3/548).

3.4.4 HIV-1B sequences

Search criteria for drug-naïve HIV-1 subtype B IN sequences from tbl. 3.1 returned 2628 sequences, however, after selecting one sequence per patient, 1558 sequences were retrieved. Quality control excluded twelve sequences due to the presence of stop codons. 1546 sequences were used as a drug-naïve HIV-1 subtype B IN reference set.

3.5 Statistical and graphic analyses

Graphic visualization of data regarding molecular distances between HIV-1 sequences and subtyping thereof was done using either the software, or online tools, indicated at the respective analysis. Visualisation of numeric data, such as the mutational profiles of inter- and intra-subtype diversity, was conducted on Excel. For assessing the dependency of categorial data, such as the comparison of mutational profiles of different viral strains, the two-tailed Fisher's exact test was used. Calculation of this test was conducted on the free GraphPad's QuickCalcs online tool for the analysis of 2x2 contingency tables (https://www.graphpad.com/quickcalcs/contingency1/). The level of significance was set at $p<0.05$. Owing to the methodological bias of an increased α-level in multi-hypotheses testing, the Bonferroni correction has been applied as a post-hoc test to all test-panels. The corrected level of significance remained at $p<0.05$.

4 Results

4.1 Study design and patient demographics

The samples used for analyses ($n = 169$) were part of a larger cohort of 1222 samples, collected in the Western Cape province between the years of 1998 and 2002 (TV-cohort). Collection sites included an academic hospital clinic, private clinics, state clinics, the Western Province Blood Transfusion Service (WPBTS), and a sex worker cohort in Cape Town. This cohort represents a collective of individuals from multiple ethnic backgrounds, as well as different sexual orientations. Transmission of HIV occurred via MTCT or hetero- and/or homosexual contacts. As all samples were collected before the initiation of the national ART treatment programme, these samples are considered ART-naïve. The 169 samples used in this study were selected randomly. The study cohort's demographics are summarized in table 4.1.

	Positive IN amplification	No IN amplification	Total
Amount (n)	96	73	169
Age (years)	30.6 ± 12.3	33.1 ± 9.9	31.3 ± 11.8
Female (%)	68.8	78.8	73.8

Table 4.1: Description of the study cohort's demographics. No significant difference in age could be observed between positively und unsuccessfully amplified HIV-1 samples. A high percentage of female participants was present in all subsets of this study cohort.

4.2 Laboratory results

Subtyping of HIV-1 Integrase requires the isolation of viral RNA from the randomly selected patient's EDTA-plasma samples (3.1). Thereafter, RNA was transcribed into cDNA and amplified via PCR (3.2.3, 3.2.4). Positive RNA extraction and its amplification were confirmed by agarose gel electrophoresis (3.2.5).

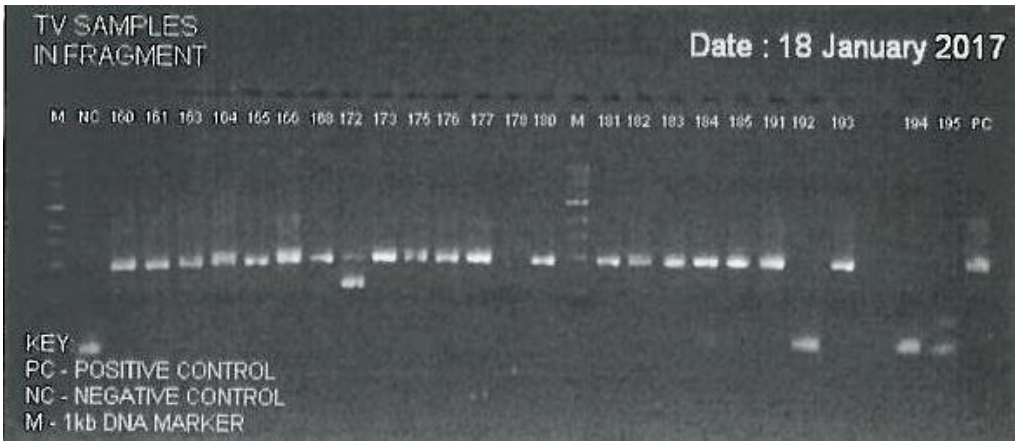

Fig. 4.1: Separation of the nested-PCR products on a 1.5% agarose gel to confirm the successful amplification of the IN-segment. As DNA ladder the Quick-Load® 1kb DNA Ladder (New England BioLabs, United Kingdom) was used. For TV samples 160, 161, 163, 164, 165, 166, 168, 173, 175, 176, 177, 180, 181, 182, 183, 184, 185, 191, 193, and the positive control, a sole signal just below 1000 base pairs was visible, indicating successful amplification of the IN-segment. TV-samples 178, 192, 194, 195 and the negative control showed no signal, indicating unsuccessful amplification. TV 172 shows two signals, a weak one below 1000 base pairs, and a stronger one relative to a shorter fragment. This indicates the amplification of multiple fragments.

Positive amplification of an estimated 945bp long fragment was confirmed for 96/169 (56.2%) samples. These amplicons were subsequently purified, photometrically measured, diluted and sequenced (3.2.6, 3.2.7). Sequencing results from the four respective primers per RNA copy were then aligned to a HXB2 reference strand and formed into one consensus strain per sample. Successful alignment of all four primers and satisfactory sequencing quality results, was obtained for 95/96 (99%) samples. One sample (TV172) did not align to the reference strand, despite showing good quality results for all primer strands. In retrospect, its gel electrophoresis showed at least a two-signal spectroscopy: One expected signal relative to

a ~1000bp long DNA fragment and a stronger signal relative to a significantly shorter fragment. One possible explanation could be alternative primer-binding during (RT-)PCR or sequencing reactions, which may have resulted in sequences from different genomic fragments, explaining their inability to align to the HXB2 reference strain. Low amounts of stored plasma sample did not allow for a second extraction cycle, however, repeated purification and sequencing reactions did not result in better alignment. Potential solutions for this samples could have been a repeated amplification cycle, or alternatively gel extraction and purification of the visible band at ~1kb, negating the alternative primer binding and the shorter amplified segment. Neither of these methods were used in this study.

One sample (TV413) contained a 62 base gap and another one (TV346) harboured a deletion relative to position 4998 on HXB2 (Fig. 4., Fig. 4.). However, both sequences aligned to the remainder of the sequences on the entire segment and were therefore included in further analyses.

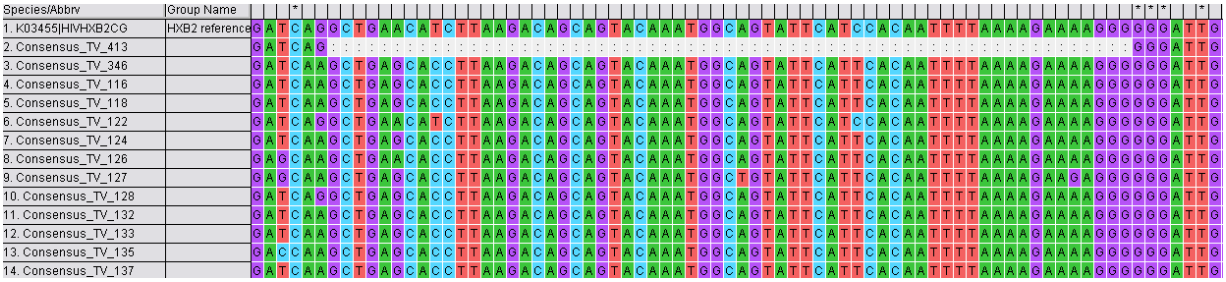

Fig. 4.1: A section of the TV-cohort alignment in MEGA v7. The sample TV413 in second position harbours a 62 base gap, relative to HXB2 position 4734-4796. Short adjacent sequence parts align well with the other sequences.

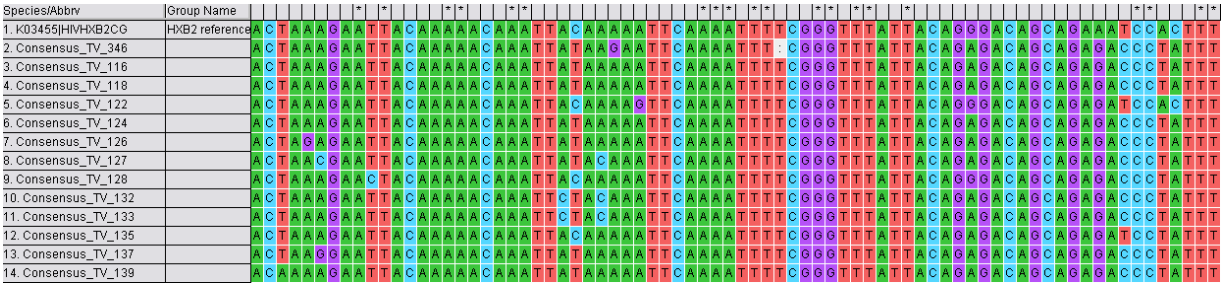

Fig. 4.2: A section of the TV-cohort alignment in MEGA v7. The samples TV 346 is shown to contain a deletion of one base, relative to HXB2 position 4998. The remainder of this sequence aligns well to the other sequences.

4.3 Subtyping of HIV-1 Integrase

4.3.1 Cohort sequences

Results from HIV online subtyping tools REGA v.3.41 and jpHMM are summarized in fig. 4.3. In total, REGA v.3.41 identified 84/92 (91.3%) HIV-1 IN fragments to solely belong to subtype C. 5/92 sequences (5.4%) (TV 122, TV 356, TV 404, TV 420 and TV 431) were classified as subtype B, and 1/92 (1.1%) (TV 412) as subtype A. Two sequences (TV406 and

TV439) were labelled subtype C-like, with bootstrap values of 81% and 99%, respectively. REGA's bootscan analyses are presented in Fig. 4.4 and Fig. 4.5. The posterior probability based algorithm, implemented in jpHMM subtyping tool on the other hand classified both samples as pure subtype C (Fig. 4.6, Fig. 4.7).

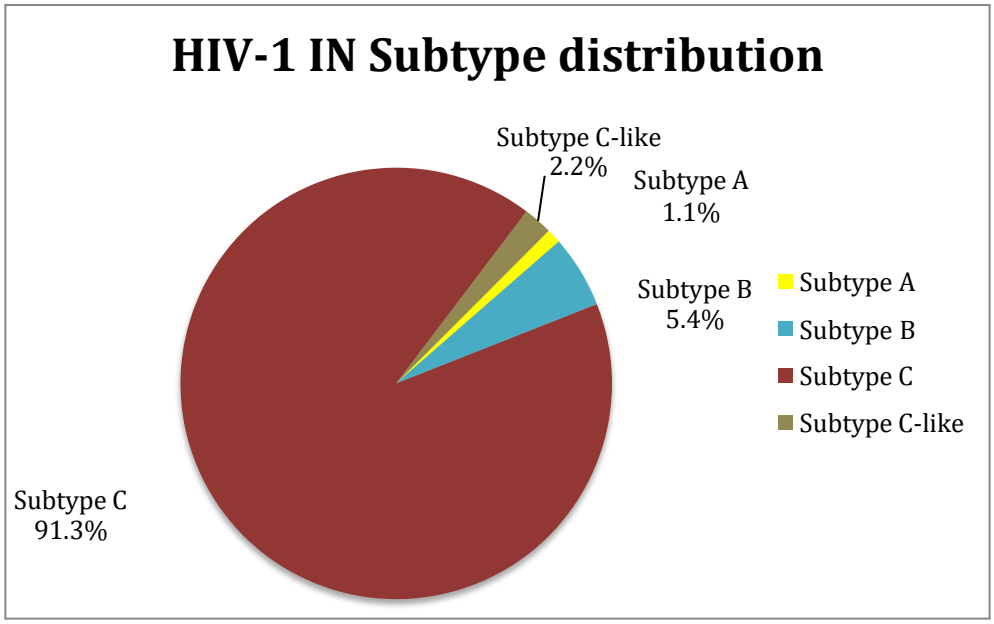

Fig. 4.3: HIV-1 Integrase subtype distribution within our study cohort. Subtyping was performed using the REGA HIV 1&2 online subtyping tool v. 3.41. 91.3% (84/92) samples were classified as subtype C, 5.4% (5/92) as subtype B and 1.1% (1/92) as subtype A. Two samples were assigned as Subtype C-like.

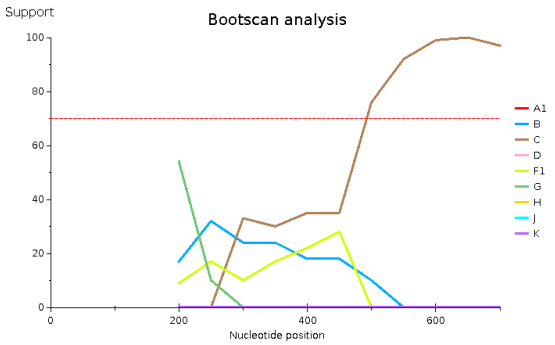

Fig. 4.4: HIV subtype recombination bootscan analysis of TV406 conducted in REGA HIV-1&2 online subtyping tool v.3.41. Bootscan was performed with a window size of 400 and a stepwise increment of 50. The dotted red line indicates a bootstrap confidence of 70%. Significance of classifying TV406 as subtype C is given only from nucleotide position ~500 onwards (relative to HXB2 position 4700). Prior segments cannot be assigned to any subtype with certainty.

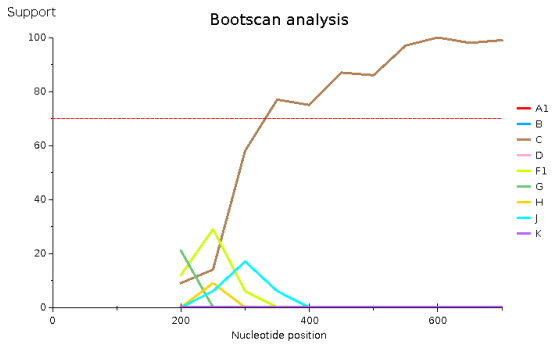

Fig. 4.5: HIV subtype recombination bootscan analysis of TV439 conducted in REGA HIV-1&2 online subtyping tool v.3.41. Bootscan was performed with a window size of 400 and a stepwise increment 50. The dotted red line indicates a bootstrap confidence of 70%. Only from nucleotide positions ~300 onwards (relative to HXB2 position 4530) can TV 439 be assigned to Subtype C with significance. Prior segments cannot be assigned to any subtype with certainty.

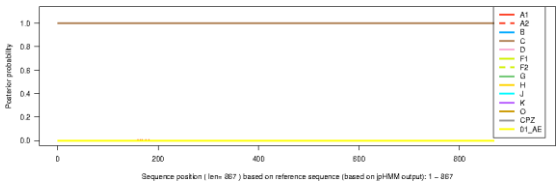

Fig. 4.6: HIV subtype recombination analysis of TV 406 using the jumping profile Hiden Markov Model online tool. Through all 880 nucleotides, TV 406 is classified as subtype C with a posterior probability of ~1.

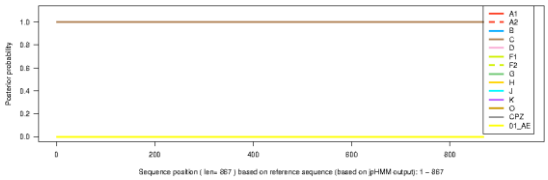

Fig. 4.7: HIV subtype recombination analysis of TV 439 using the jumping profile Hiden Markov Model online tool. Through all 880 nucleotides, TV 439 is classified as subtype C with a posterior probability of ~1

4.3.2 Database-derived sequences

Database-derived sequences included a selection of South African IN fragments, collected from drug-naïve PLWH before 2007. Detailed search criteria are explained in 3.4.1. After quality control, 284 sequences were used for drug-mutation analysis. Of these sequences, 275/284 (96.8%) were listed as subtype C, 6/284 as subtype B (2.1%) and 1/284 as subtype BD (0.35%) and 2/284 (0.7%) as subtype A. A reassessment of these ratios was omitted.

4.4 Phylogenetic analyses

4.4.1 Neighbour-Joining Tree

In a first phylogenetic analysis, a Neighbour-Joining tree consisting of all 95 cohort sequences was constructed to screen for possible contamination (Fig. 4.8). Here, high homology can be observed between samples TV176 and TV425, TV345 and TV432 as well as between TV403 and TV407. Based on indifferent entries of TV176 and TV425, as well as TV345 and TV432

in our database, these sample-pairs were most likely isolated from the same patient. In this case I removed one sample each (TV425 and TV432) from further analysis.

Despite being sampled at the same collection site, TV403 and TV407 do not share demographics and therefore were not considered to being isolated from the same patient. However, these samples were also never processed in direct succession during the laboratory work, rendering isolated in-lab contamination that does not affect additional samples unlikely. One source of contamination could be the processing of whole-blood samples upon initial reception at our laboratory, however, we do not own protocols thereof. Alternatively, close, but not identical homology could indicate close relations between HIV-hosting individuals and presumably a shared transmission route. Since both isolates were sampled at the same collection site and the patients' sexual orientation matched their respective gender, sexual transmission and collection in early infection could possibly explain a high degree of genetic homology. Regardless of origin, the close homology warranted exclusion of one sequence from downstream analysis (TV407).

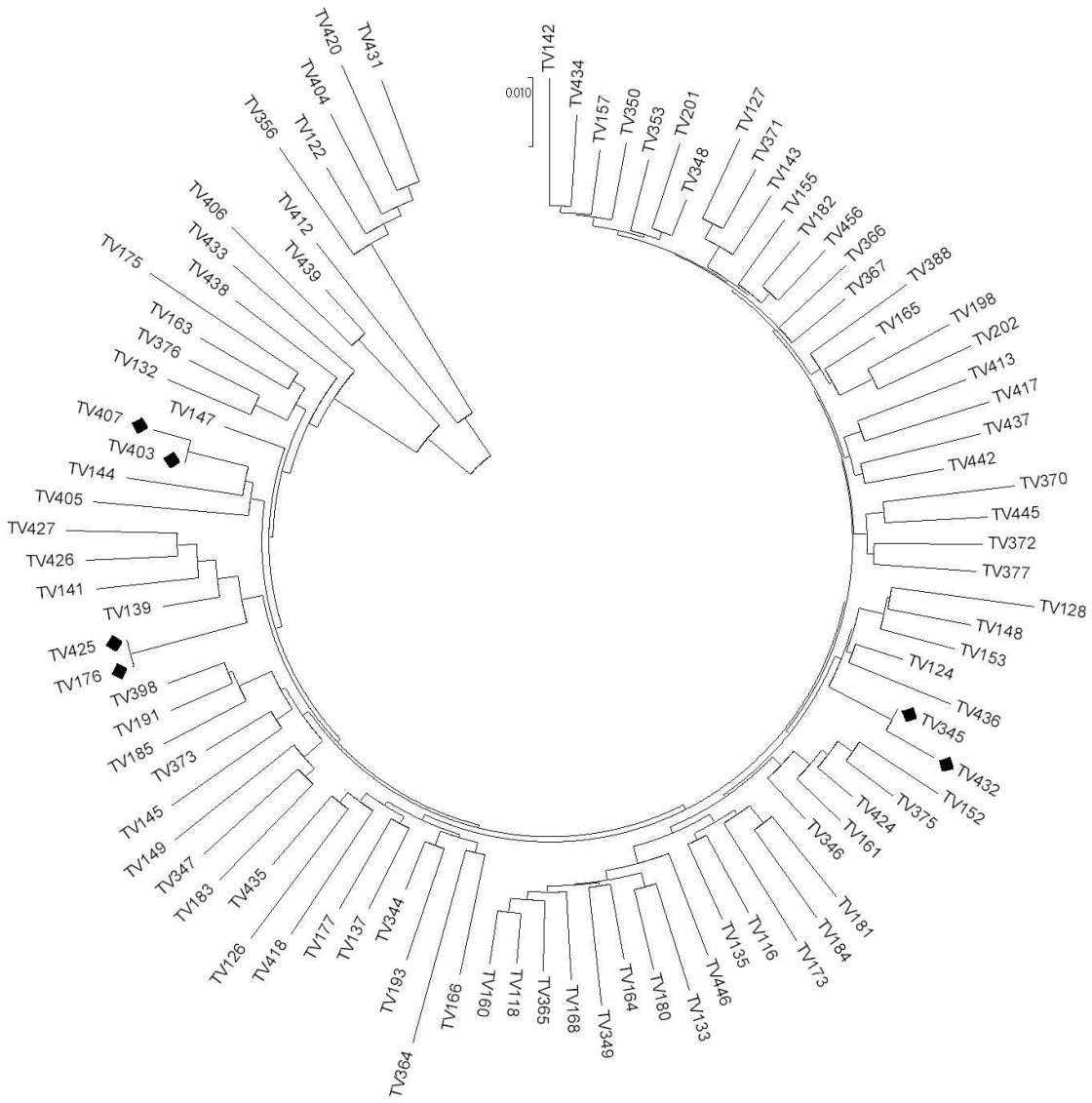

Fig. 4.8: Evolutionary relationships of the cohort's sequences. The evolutionary history was inferred using the Neighbour-Joining method. The optimal tree with the sum of branch length = 1.44453524 is shown. The tree is drawn to scale, with branch lengths in the same units as those of the evolutionary distances used to infer the phylogenetic tree. The evolutionary distances were computed using the Maximum Composite Likelihood method and are in the units of the number of base substitutions per site. The analysis involved 95 nucleotide sequences. Codon positions included were 1st+2nd+3rd. Evolutionary analyses were conducted in MEGA v7.

4.4.2 Maximium-Likelihood Tree

With the remaining 92 samples I computed a maximum-likelihood tree (Fig. 4.9, fig 4.10). Alignment sequences were retrieved from the HIV LANL database and consisted mostly of a selection of the curated 2018 HIV-1 filtered web alignment, including only group M, subtypes A-K sequences without the inclusion of recombinants. For greater clustering accuracy only subtype A, B, C, F1, F2 and G sequences were included. F1, F2 and G sequences were added, due to the ambiguous results of TV406 and TV439 in REGA's bootscan analysis (s. 4.3.1). Three additional sequences from BLAST-search results for these two samples were also added to the reference set.

In the phylogenetic tree, 86 samples clustered with subtype C reference strains, five with subtype B and one with subtype A. The two samples, TV406 and TV439, which could not be unanimously subtyped by online subtyping tools, can be seen to cluster with some genetic distance to the remaining subtype C samples, however their closest relatives are AY772691 and MG693909, two samples classified as subtype C. Sample JQ670835 was isolated from a cohort in Mozambique in 2012, and is listed as CG recombinant on the IN gene in the Los Alamos database. In the original work it was described to express similar homology to subtype C and G sequences in a Recombinant Identification Program (RIP) 3.0 analysis, however, our analyses with REGA v3.41 and jpHMM subtyping tools as well as RIP 3.0 analysis and the shown phylogenetic inference could not confirm these results (Oliveira *et al.*, 2012). Instead, JQ670835 can be seen to cluster with subtype C samples, which is in accordance with both online subtyping tools. Neither sample from this study cohort expressed close enough homology to clade G or F1 sequences, in order to be considered recombinant on the IN gene.

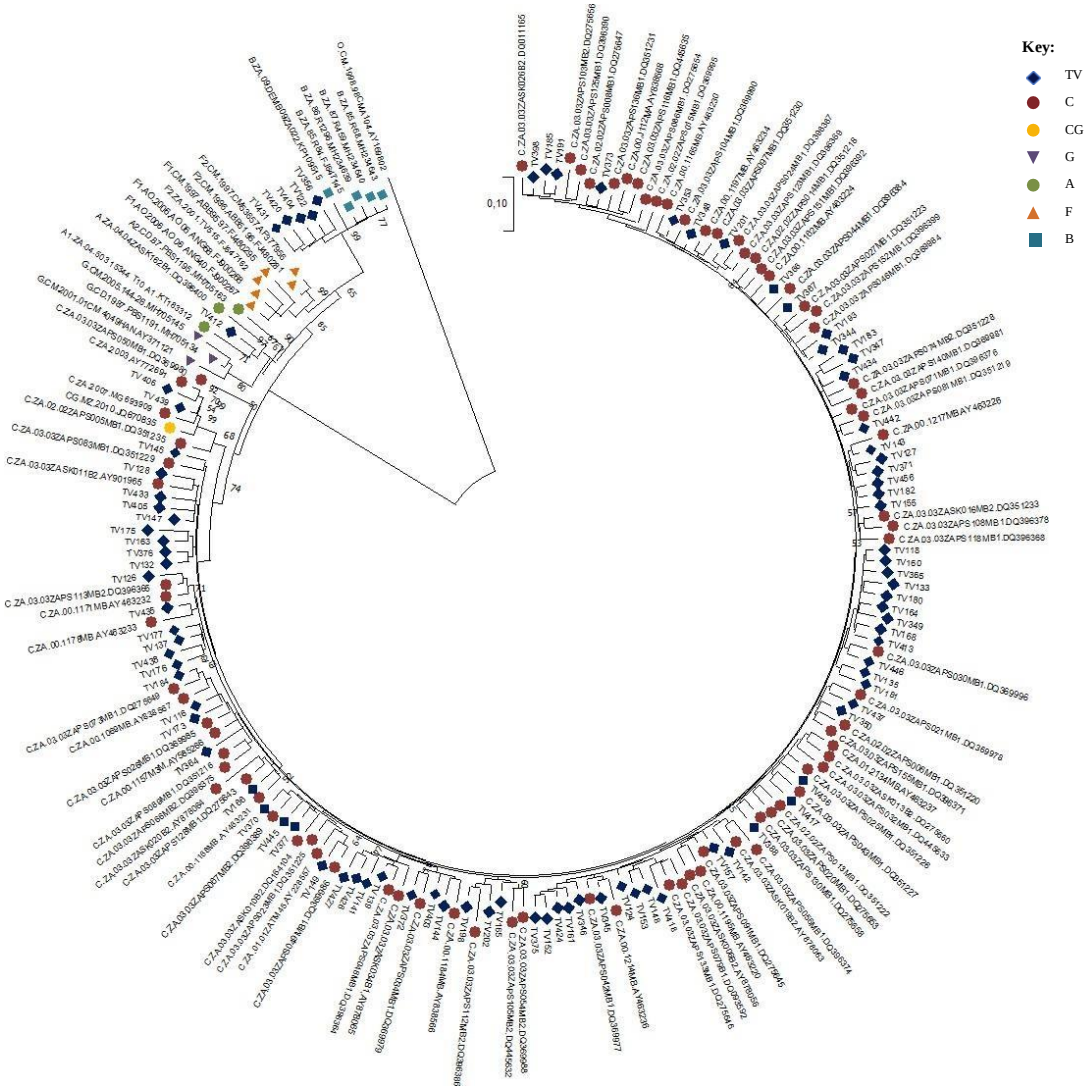

Fig. 4.9: Phylogenetic analysis of the HIV-1 Integrase segment inferred by the Maximum-Likelihood method and General Time Reversible model. The evolutionary history was inferred in MEGA X. The tree with the highest log likelihood (-17353.64) is shown. A discrete Gamma distribution was used to model evolutionary rate differences among sites (4 categories (+G, parameter = 0.3086)). The rate variation model allowed for some sites to be evolutionarily invariable ([+I], 44.29% sites). The tree is drawn to scale, with branch lengths measured in the number of substitutions per site. This tree was constructed with 1000 bootstrap replicates. Bootstrap values > 0.5 are indicated on the respective nodes in the tree. This analysis involved 92 sequences from this study cohort, as well as 100 sequences from the HIV LANL database and a manual selection of subtype F1, F2, G and CG recombinant sequences. One group O sample, also retrieved from the LANL database, was used as root for the ML-tree. Codon positions included were 1st+2nd+3rd+Noncoding. There were a total of 867 positions in the final dataset. Of the study cohort's samples, 86 clustered with subtype C sequences, five with subtype B and one with subtype A.

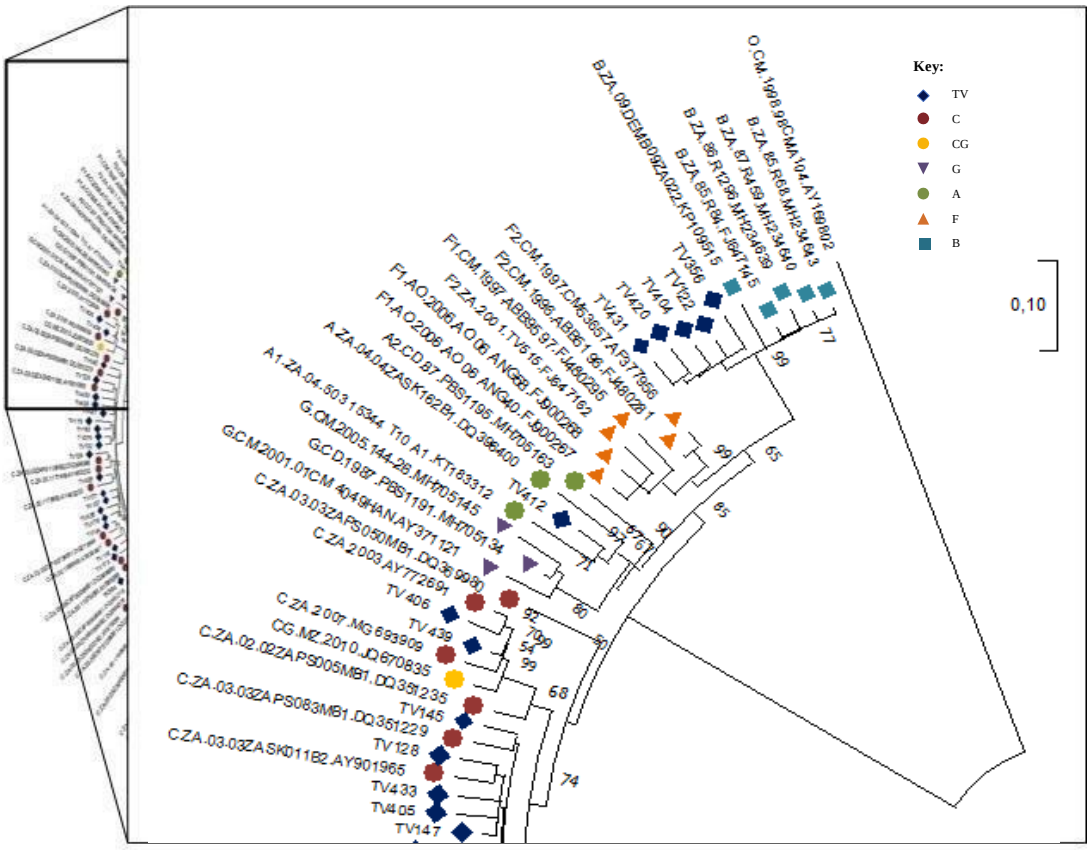

Fig. 4.10: Detailed enlargement of Fig 4.12, depicting evolutionary relationships of samples TV 356, 122, 404, 420, 431, 412, 406, 439, 145, 128, 433 and 405. TV 406 and TV 439 can be seen to cluster with subtype C reference sequences with a bootstrap support >0.5. The CG recombinant strain also clusters with subtype C samples, with distinct distance to subtype G sequences, however bootstrap values remain <0.5.

4.5 Resistance mutation analyses

4.5.1 Cohort sequences

Drug resistance analyses, using the Stanford University HIV drug resistance database, showed that 0/92 sequences (0%) harboured major INSTI RAMs and no sequence was found to express a reduced INSTI susceptibility. Accessory RAMs were present in 1/92 samples (1.1%): TV367 was found to contain the accessory drug mutation Gly140Glu, a non-polymorphic mutation that has been selected *in-vitro* before, but alone does not seem to confer resistance to INSTIs. Seven polymorphic amino acid substitutions that could potentially contribute to a reduced INSTI susceptibility were present in a total of 59/92 (64.1%) IN sequences. Amino acid substitution occurred at four different sites, namely Methionine (Met)50, Leucine (Leu)74, Serine (Ser)119 and Ser230. The most prevalent of these mutations in this study's cohort was conferred by a G→A substitution at position 150 of the IN fragment, relative to position 4379 on the HXB2 reference strand. This accounted for a Met50Isoleucine (Ile) mutation in 56/92 (60.1%) sequences. Additional diversity in this codon included T→C, as well as T→G substitutions at position 149 (4378 on HXB2), leading to

Met50Threonine (Thr) and Met50Arg mutations in 4/91 (4.4%) and 1/91 (1.1%) samples, respectively. All 56 Met50Ile, as well as Met50Thr substitutions were observed in subtype C sequences, whereas Met50Arg was present in one subtype B sample (TV404). The three remaining potentially impactful polymorphisms were detectable at lower rates with Ser230Asparagine (Asn) in 1/91 (1.1%) (TV364), Ser119Arg in 2/91 (2.2%) (TV405, TV370), Leu74Met in 1/91 (1.1%) (TV366) and Leu74Ile in 4/91 (4.4%) (TV122, TV128, TV173, TV367) samples. With the exception of Leu74Ile in two cases (TV122, TV128), all other polymorphisms were observed in strains harbouring the Met50Ile polymorphism.

	Major RAMs	**Accessory RAMs**	**Other Mutations**	**Subtype**	**Reduced susceptibility**
TV 367	-	Gly140Glu	Met50Ile	C	-
TV 139	-	-	Met50Thr	C	-
TV 145	-	-	Met50Thr	C	-
TV 163	-	-	Met50Thr	C	-
TV 176	-	-	Met50Thr	C	-
TV 404	-	-	Met50Arg	B	-
TV 364	-	-	Ser230Asn + Met50Ile	C	-
TV 122	-	-	Leu74Ile	B	-
TV 128	-	-	Leu74Ile	C	-
TV 173	-	-	Leu74Ile + Met50Ile	C	-

Table 4.2: Summary of observed IN mutations in the sequenced TV cohort. The terms major and accessory RAMs, as well as "other mutations" were used according to the Stanford University HIV drug resistance database. With the exception of the Met50Ile mutation that was present in 56/92 samples (60.1%), all commented IN mutations are listed in this table.

4.5.2 Database-derived sequences

After quality control (3.4.1), 284 drug-naïve IN fragments were screened for INSTI RAMs. The results are summarized in tbl. 4.3. 8/284 (2.8%) sequences showed amino acid substitutions at the following five classified major resistance sites: Thr66, Glu92, Tyrosine (Tyr)143, Ser147 and Glutamine (Gln)148. The most prevalent major RAM involved the amino acid residue Thr at codon 66, being replaced by Alanine (Ala) in 2/276 and by Ser in 1/276 fragments. Tyr143His occurred in 2/276 sequences, whereas Glu92Gly, Ser147Gly and Gln148His were identified in one sequence, respectively.

Accessory RAMs were observed in 8/284 (2.8%) sequences, with Thr97Ala being the most prevalent at 1.8% (5/284), followed by Glu157Gln at 0.7% (2/284). Ambiguity at nucleotide 487 of the IN fragment, relative to position 4716 on the HXB2 strand, lead to a RGA-codon and thus to a Gly163GlyArg substitution in 1/284 samples. Co-existence of major and accessory resistance mutations did not occur.

Of note is the diversity of the Met50 codon in this set of sequences. 165/284 (58.1%) sequences expressed the G→A mutation at position 150 (4379 on HXB2), responsible for the Met50Ile polymorphism and an additional 5/284 (1.8%) sequences were listed with G/A ambiguity, accounting for a Met50MetIle mutation analysis. Furthermore, 19/284 (6.7%) sequences harboured a Met50Thr substitution and 2/284 (0.7%) Met50Arg. Other directly relevant polymorphisms in this set included Leu74Ile in 21/284 (7.4%) and Leu74Met in 2/284 sequences (0.7%), Ser119Arg in 2/284 (0.7%), and His51Gln, Phenylalanine (Phe)121Ser and Valine (Val)151Ile in in 1/284 sequences (0.4%), respectively. Noteworthy co-occurrence of major or accessory RAMs with relevant polymorphisms include the combination of Thr66Ala+Met50Ile in 2/284, Tyr143His+Met50Ile in 1/284, Ser147Gly+Met50Ile in 1/284, Gln148His+Met50Ile in 1/284, Thr97Ala+Met50Ile in 3/284, Tyr143His+Leu74Ile in 1/284 and Thr97Ala+Met50Ile+Leu74Ile in 1/284 sequences.

Comparing the observed RAM-prevalence rates in the sequenced TV cohort with the database-derived sequences (0/92 vs. 8/284 and 1/92 vs. 8/284 for major and minor RAMs, respectively) the differences do not reach significant levels, with $p = 0.2076$ for major RAMs and $p = 0.6944$ for accessory RAMs according to the Fisher's exact test.

Accession number	Major RAMs	Accessory RAMs	Other Mutations	Subtype	Reduced susceptibility
DQ396385	Thr66Ala	-	Met50Ile	C	EVG: 60 fold RAL: 15 fold
AY878061	Thr66Ala	-	Met50Ile	C	EVG: 60 fold RAL: 15 fold
DQ056418	Thr66Ser	-	Met50Thr	C	-
AY463234	Tyr143His	-	Leu74Ile	C	EVG: 15 fold RAL: 60 fold DTG: 10 fold BIC: 10 fold
DQ351237	Tyr143His	-	Met50Ile	C	EVG: 10 fold RAL: 60 fold DTG: 5 fold

Accession number	Major RAMs	Accessory RAMs	Other Mutations	Subtype	Reduced susceptibility
					BIC: 5 fold
DQ164115	Glu92Gly	-	-	C	EVG: 30 fold RAL: 15 fold
DQ164107	Ser147Gly	-	Met50Ile	C	EVG: 60 fold
GU253432 *	Gln148His	-	Met50Ile	C	EVG: 60 fold RAL: 60 fold DTG: 25 fold BIC: 25 fold
DQ275646	-	Thr97Ala	Met50Ile	C	EVG: 10 fold RAL: 10 fold
DQ093596	-	Thr97Ala	Met50Ile	C	EVG: 10 fold RAL: 10 fold
AY772698	-	Thr97Ala	-	C	EVG: 10 fold RAL: 10 fold
DQ369992	-	Thr97Ala	Met50Ile	C	EVG: 10 fold RAL: 10 fold
GQ872532	-	Thr97Ala	Met50Ile + Leu74Ile	C	EVG: 10 fold RAL: 10 fold
DQ351222	-	Glu157Gln	Met50Leu	C	EVG: 10 fold RAL: 10 fold
DQ396380	-	Glu157Gln	-	C	EVG: 10 fold RAL: 10 fold
GU253450	-	Gly163GlyArg	-	C	EVG: 15 fold RAL: 15 fold

Table 4.3: Summary of observed IN mutations in the database-derived IN sequences. The terms major and accessory RAMs, as well as "other mutations" were used according to the Stanford University drug resistance database. All major and accessory RAMs are displayed in this table. Sample GU253432 (marked *) expressed low-level DTG resistance. The Stanford University drug resistance database recommended the twice-daily administration of DTG, if DTG were to be used. For the purpose of clarity, the listing of singular commented mutations was omitted in this table.

4.6 Generation of a South African HIV-1 subtype C IN consensus (HIV-1 C_{ZA})

4.6.1 Identification of naturally occurring polymorphisms

All HIV-1 subtype C IN sequences from this study cohort (n=86), as well as from the database-derived South African IN fragments (n=275) were included in the generation of HIV-1 C_{ZA}. In comparison to the HXB2 reference strain, the HIV-1C_{ZA} consensus identified the following 19 naturally occurring polymorphisms: Asp10Glu, Asp25Glu, Val31Ile, Met50Ile, Phe100Tyr, Leu101Ile, Thr112Val, Gly123Ser, Thr125Ala, Arg127Lys, Lys136Gln, Val201Ile, Thr218Ile, Asparagine (Asn)232Asp, Leu234Ile, Ala265Val, Arg269Lys, Asp278Ala and Ser283Gly. Four polymorphic sites (Asp10Glu, Gly123Ser, Arg127Lys and Asn232Asp) were considered to be polymorphic on the HXB2 strain and could not be confirmed in any investigated cohort. Therefore these polymorphisms were not included in the following analyses. The remainder of the polymorphisms were used to subsequently investigate genetic inter and intra subtype diversity at these positions.

4.6.2 Naturally occurring polymorphisms HIV-1 C_{cohort} versus HIV-1$C_{ZA(LANL)}$

Comparative analysis between this study cohort's subtype C and the database-derived subtype C IN fragment (3.4.2) was carried out, in order to investigate, whether the studied cohort is representative for South African drug-naïve subtype C IN sequences, or whether the two groups needed to be analysed separately in future analyses. Seven polymorphisms (Met50Ile, Thr125Ala, Val201Ile, Thr218Ile, Ala265Val, Arg269Lys and Ser283Gly) were further increased in our study cohort (s. Fig. 4.11). However, none of these results were significant (s. tbl 4.4). This indicates, that the selected cohort can be seen as a representative group of South African IN sequences and permitted their mutual analysis.

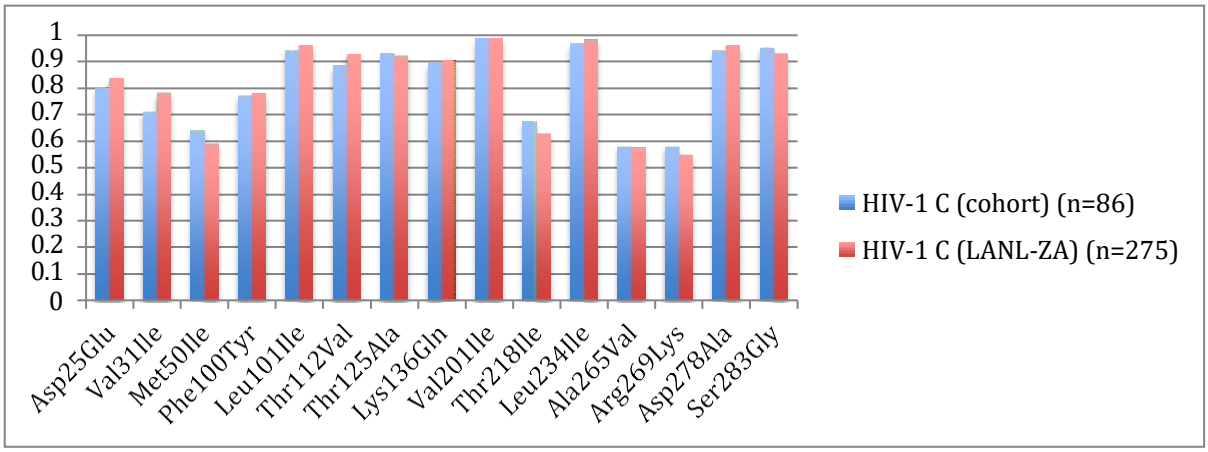

Fig. 4.11: Comparison between the mutational profiles of HIV-1 C (cohort) sequences and HIV-1 C (LANL-ZA) sequences for the Integrase region. Cohort sequences consist of all subtype C sequences from this study cohort (n=86). LANL-ZA sequences consist of all publicly available South African HIV-1 C Integrase sequences, sampled before 2007. Seven mutations (Met50Ile, Thr125Ala, Val201Ile, Thr218Ile, A265V, Arg269Lys, Ser283Gly) show slightly higher prevalence rates in the study cohort.

	Cohort sequences (n=86)	LANL-ZA sequences (n=275)	p-value	Corrected p-value
Met50Ile	55/86 (64.0%)	162/275 (58.7%)	0.4500	1
Thr125Ala	80/86 (93.0%)	254/275 (92.4%)	1	1
Val201Ile	85/86 (98.8%)	271/275 (98.2%)	1	1
Thr218Ile	58/86 (67.4%)	173/275 (62.7%)	0.5202	1
Ala265Val	50/86 (58.1%)	158/275 (57.5%)	1	1
Arg269Lys	50/86 (58.1%)	151/275 (54.9%)	0.6209	1
Ser283Gly	82/86 (95.3%)	256/275 (93.1%)	0.6149	1

Table 4.4: Prevalence rates for individual polymorphisms in the study cohort, compared to their respective prevalence rates in South African HIV-1 C Integrase sequences. Study cohort sequences consisted of all subtype C sequences from this study (n=86). LANL-ZA sequences consist of all publicly available South African HIV-1C Integrase sequences, sampled before 2007. Every comparison of prevalence rates between the two cohorts was tested for statistical significance via the Fisher Exact Test. Post-hoc testing for corrected p-values in multi-hypotheses testing array was carried out via the Bonferroni procedure.

4.6.3 Naturally occurring polymorphisms HIV-1C$_{ZA}$ versus HIV-1B$_{global}$

Comparing the combined drug-naïve South African HIV-1 subtype C IN fragments (n=361) to a set of drug-naïve HIV-1 subtype B sequences (n=1541) (s. 3.4.4) the presence of the naturally occurring polymorphisms in South African Integrase fragments was confirmed. All polymorphisms were significantly more prevalent in South African HIV-1 subtype C sequences, than in the reference set of global HIV-1 subtype B sequences (s. fig. 4.12, tbl 4.5).

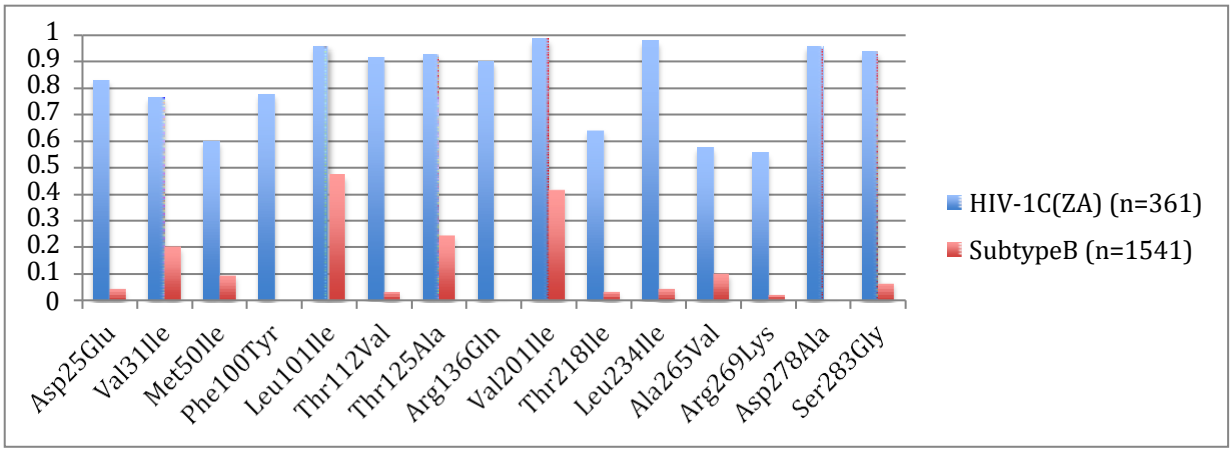

Fig. 4.12: Comparison between the mutation profiles of HIV-1 C_{ZA} and HIV-1 B sequences for the Integrase region. HIV-1 C_{ZA} (n=361) consists of 92 newly sequenced samples and 275 sequences retrieved from the LANL database. HIV-1B sequences (n=1541) consist of all online available drug-naïve subtype B IN fragments. All 15 polymorphisms can be seen to be enriched in the South African cohort.

	HIV-1C_{ZA} (n=361)	Global Subtype B (n=1541)	p-value	Corrected p-value
Asp25Glu	299/361 (82,8%)	69/1541 (4.48%)	<0.0001	<0.0015
Val31Ile	276/361 (76.5%)	313/1541 (20.31%)	<0.0001	<0.0015
Met50Ile	217/361 (60.1%)	147/1541 (9.54%)	<0.0001	<0.0015
Phe100Tyr	280/361 (77.8%)	1/1541 (0.06%)	<0.0001	<0.0015
Leu101Ile	345/361 (95.6%)	732/1541 (47.50%)	<0.0001	<0.0015
Thr112Val	331/361 (91.7%)	48/1541 (3.11%)	<0.0001	<0.0015
Thr125Ala	334/361 (92.5%)	377/1541 (24.5%)	<0.0001	<0.0015
Lys136Gln	326/361 (90.3%)	6/1541 (0.39%)	<0.0001	<0.0015
Val201Ile	356/361 (98,6%)	643/1541 (41.73%)	<0.0001	<0.0015
Thr218Ile	231/361 (64.0%)	47/1541 (2.21%)	<0.0001	<0.0015
Leu234Ile	353/361 (97.8%)	70/1541 (4.54%)	<0.0001	<0.0015
Ala265Val	208/361 (57.6%)	157/1541 (10.19%)	<0.0001	<0.0015
Arg269Lys	201/361 (55.7%)	31/1541 (2.01%)	<0.0001	<0.0015
Asp278Ala	345/361 (95.6%)	6/1541 (0.39%)	<0.0001	<0.0015
Ser283Gly	338/361 (94.29%)	105/1541 (6.81%)	<0.0001	<0.0015

Table 4.5: Prevalence rates for individual polymorphisms in the South African HIV-1 C consensus sequences compared to their respective prevalence rates in a drug-naïve HIV-1 subtype B Integrase reference set. Every comparison of prevalence rates between the two cohorts was tested for statistical significance via the Fisher Exact Test. Post-hoc testing for corrected p-values in multi-hypotheses testing array was carried out via the Bonferroni procedure.

4.6.4 *Naturally occurring polymorphisms HIV-1C$_{ZA}$ versus HIV-1C$_{non\text{-}ZA}$*

When comparing HIV-1C sequences from South Africa with HIV-1C sequences collected from countries other than South Africa, differences can be observed between their mutational profiles: 14/15 polymorphisms occurred more frequently in South African HIV-1C sequences (Fig. 4.13). However, statistical significance for these results at a post-hoc corrected level of significance ($p<0.05$) was only given for four polymorphisms: Met50Ile, Phe100Tyr, Thr112Val and Ala265Val (Table 4.6: Prevalence rates for individual polymorphisms in the South African HIV-1 C consensus sequences, compared to their respective prevalence rates in HIV-1 C Integrase sequences sampled in countries other than South Africa. Every comparison of prevalence rates between the two cohorts was tested for statistical significance via the Fisher Exact Test. Post-hoc testing for corrected p-values in multi-hypotheses testing array was carried out via the Bonferroni procedure.). The largest margin can be observed for the Met50Ile polymorphism, being present in 60.1% of drug-naïve South African HIV-1 subtype C samples, compared to 37.0% of non-South African subtype C sequences.

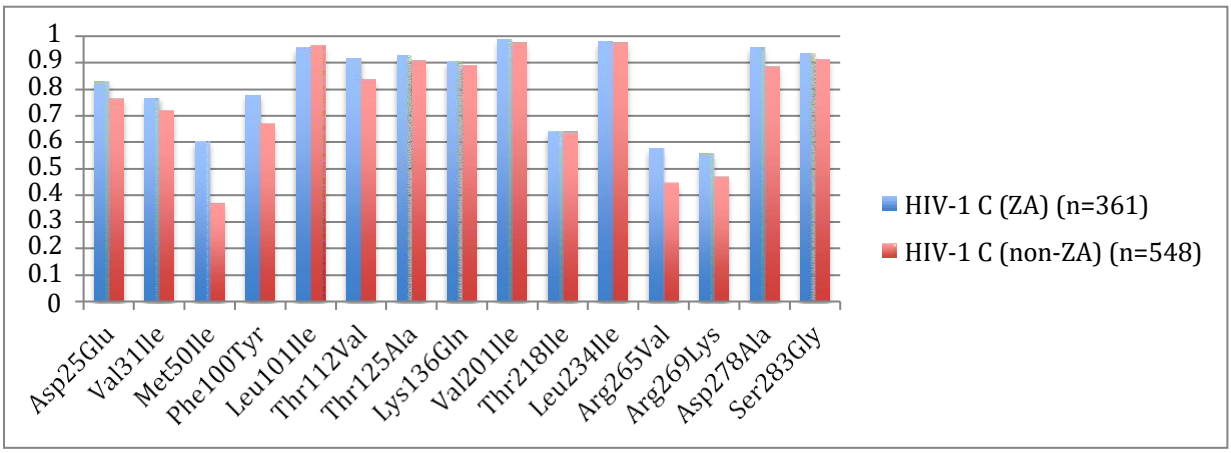

Fig. 4.13: Comparison between the mutation profiles of HIV-1 C$_{ZA}$ and HIV-1 C (non-ZA) sequences for the Integrase region. HIV-1 C$_{ZA}$ (n=361) consists of 92 newly sequenced samples and 275 sequences retrieved from the LANL database. HIV-1 C (non-ZA) sequences (n=548) consist of all online available subtype C IN fragments, sampled in countries other than South Africa. With the exception of Leu101Ile, every polymorphism can be seen to be more prevalent in South African HIV-1C Integrase sequences.

	HIV-1 C_{ZA} (n=361)	HIV-1 C_{non-ZA} (n=548)	p-value	Corrected p-value
Asp25Glu	299/361 (82,8%)	417/548 (76.1%)	0.0162	0.2268
Val31Ile	276/361 (76.5%)	394/548 (71.9%)	0.1435	1
Met50Ile	217/361 (60.1%)	203/548 (37.0%)	>0.0001	<0.0014
Phe100Tyr	280/361 (77.8%)	366/548 (66.8%)	0.0004	0.0056
Thr112Val	331/361 (91.7%)	459/548 (83.8%)	0.0004	0.0056
Thr125Ala	334/361 (92.5%)	496/548 (90.5%)	0.3302	1
Lys136Gln	326/361 (90.3%)	485/548 (88.5%)	0.4446	1
Val201Ile	356/361 (98,6%)	534/548 (97.4%)	0.3433	1
Thr218Ile	231/361 (64.0%)	349/548 (63.7%)	0.9439	1
Leu234Ile	353/361 (97.8%)	532/548 (97.1%)	0.6733	1
Ala265Val	208/361 (57.6%)	244/548 (44.5%)	< 0.0001	<0.0014
Arg269Lys	201/361 (55.7%)	257/548 (46.9%)	0.0101	0.1414
Asp278Ala	345/361 (95.6%)	513/548 (93.4%)	0.1913	1
Ser283Gly	338/361 (94.29%)	499/548 (91.1%)	0.1695	1

Table 4.6: Prevalence rates for individual polymorphisms in the South African HIV-1 C consensus sequences, compared to their respective prevalence rates in HIV-1 C Integrase sequences sampled in countries other than South Africa. Every comparison of prevalence rates between the two cohorts was tested for statistical significance via the Fisher Exact Test. Post-hoc testing for corrected p-values in multi-hypotheses testing array was carried out via the Bonferroni procedure.

5 Discussion

5.1 HIV-1 Integrase subtyping distribution in South Africa

Since the emergence of the retrovirus, termed HIV, South Africa's epidemic has been predominated by HIV-1 subtype C clades. Our study cohort consisted of archived blood samples, sampled between 1998 and 2002. Of 96/169 (56.8%) successfully sequenced samples, 92 IN sequences were subject to subtyping and further analyses. Thereof, 86/92 (93.5%) sequences were identified as subtype C, 5/92 as subtype B (5.4%) and 1/92 (1.1%) as subtype A. For two isolates, TV406 and TV 439, the genetic subtype could not be unambiguously established using online subtyping tools, however, the majority of the 967 base spanning sequences could be assigned to subtype C with significance, without proof of a clear breakpoint to another subtype in the bootscan analysis. Furthermore, phylogenetic inference with subtype C, B and A, as well as G, CG and F1 clades, which expressed homology to the TV samples at bootscan confidence levels <70%, showed both samples to cluster with HIV-1 subtype C samples, albeit expressing some genetic distance to all other studied subtype C samples in our cohort.

These results are in line with previously reports of subtype diversity in South Africa, both from the time of sampling of our cohort, as well as more recent publications. Van Harmelen et al. reported in 1999 similar prevalence rates of 92% (98/107) for subtype C, 7% (7/107) for subtype B and 1.9% (2/107) for subtype A (Harmelen *et al.*, 1999). Unlike mounting evidence for shifting epidemics in many countries, such as Canada, Portugal or France, where a substantial amount of new strains were found to become endemic in recent years, the subtype diversity in South Africa seems to be more stable on a larger scale (Chaix *et al.*, 2013; Siemieniuk, Beckthold and Gill, 2013; Carvalho *et al.*, 2015). In 2010, Papathanasopoulos et al. subtyped 42 South African isolates based on their *pol*-sequence and reported a HIV-1 subtype C prevalence rate of 97% (Papathanasopoulos *et al.*, 2010). Four years later, in 2014, Jacobs et al. characterised 65 HIV samples from the Western Province, South Africa, based on their *gag* and *pol* sequence and found 93.8% (61/65) to belong to subtype C (Jacobs *et al.*, 2014). These results can be interpreted as a stable epidemic, however, the emergence of reported recombinant variants is of concern. While no CRF or URF was present in our study cohort, both groups, Papathanasopoulos et al., as well as Jacobs et al., reported the presence of recombinant strains A1/C, C/A1 (n=2, 4.8%) and B/C (n=3, 4.6%) respectively. Additional recombinants have recently been described by Bessong et al. and Msimanga et al. (Msimanga, Vardas and Engelbrecht, 2015; Bessong and Iweriebor, 2016). These findings can be interpreted as an indicator for a development in the South African HIV epidemic that progresses slower than in other geographical regions, mostly due to the sheer amount of total HIV-1 C infections in the country, but in consequences could lead to immense genetic diversity in the future.

The subtyping results from our cohort, on the other hand, potentially marking a reference point to the shift in recombination events, cannot be considered absolute, as they have some limitations. Firstly, a study cohort with n=92 is too small to draw conclusions over a time span of almost two decades. Other studies, performed by Wilkinson et al., which were based on the same archived TV cohort, reported at least one unique AC and AD recombinant to be present in this cohort. Apart from showing the correct methodological approach by using near-full-length-genome (NFLG) sequences (s. below), this study points out the limitation of representing a 1222 samples large cohort by a randomly chosen subset of 169 samples. Secondly, our cohort was sampled at distinct collection sites, which were located in the Western Cape exclusively. By positioning collection sites strategically, for instance at popular long-distance truck routes, the representation of genetic variability outside the Western Province had been increased, however, the results of this relatively small cohort can still not

be considered representative for the entire country. In particular, rural and remote areas are most likely not represented accordingly in this study. Lastly and most importantly, subtyping analysis in our study was performed purely based on the IN fragment. Other genomic areas, possibly belonging to another subtype were not analysed. Therefore recombination could have only been observed, if the recombination site had been located within the IN gen. In order to obtain more reliable results, NFLG sequences, or at least fragments from different genomic sites at both ends of the genome, must be analysed. Thus, statements from our study regarding subtype diversity can only be considered valid for the IN region.

5.2 Resistance-associated mutations against INSTIs in South Africa

5.2.1 Major and minor RAMs in the study cohort's IN sequences

The presence of major INSTI-RAMs in cART- or INSTI-naïve settings would have immense impact, as DTG is quickly becoming the main therapeutic in the South African national ART program (National Department of Health, 2019). However, so far all studies from diverse geographical as well as genetic backgrounds on treatment-naïve patients have confirmed the absence or very low prevalence of RAMs against INSTIs (Passaes *et al.*, 2009; Mulu, Maier and Liebert, 2015; Alaoui *et al.*, 2018; Karade, Sen and Sashindran, 2019; Liu *et al.*, 2019).

In this study, we screened 169 archived cART-naïve plasma samples, sampled in South Africa's Western Cape region before the introduction of a national HIV treatment program, for the presence of major and accessory RAMs against INSTIS, as defined by the Stanford University Drug-Resistance Database (https://hivdb.stanford.edu/). Full-length IN sequences, spanning all 268 amino acids, were obtained for 95/169 (56.2%) plasma samples. No major RAM was found to be present in out cohort, and only one sample (TV367) was found to harbour the accessory resistance mutation Gly140Glu.

The yield in positive PCR amplification of 56.2% (95/135), however, is significantly lower than reported succession rates of >75% in comparable South African studies on treatment-naïve samples (Papathanasopoulos *et al.*, 2010; Bessong and Nwobegahay, 2013). A retrospective analysis of the randomly selected samples used in this study revealed that 46.6% (34/73) of the samples yielding no successful amplification were listed as HIV-1 negative in the corresponding database, and therefore could not contain viral RNA. Excluding these samples from the succession ratios gave corrected rates of 71.1% (96/135) and 70.4% (95/135) for positive PCR amplification and IN sequence obtainment, respectively. Another factor that could have affected the success of viral extraction and PCR-amplification is the long-term storage of the samples and potential subsequent RNA degradation. RNA

degradation could have occurred either by repetitive thawing and re-freezing cycles of individual samples, or by re-assortment of the freezing facilities over the years, involving the handling and transportation of the entire TV-cohort's samples between freezing units. Despite always handling frozen samples with care and continuously storing them on ice during transport, damage to fragile RNA-strands could not be ruled out. Whether other factors, such as low viral load titres, or an altered genetic code in PCR-primer binding sites, affected extraction and amplification reactions, was not established in this study.

Of the 95 obtained HIV-1 IN sequences, one accessory resistance mutation was present in our study: The subtype C classified TV367-sequence showed a G➔A mutation at position 419, which lead to a 5' -GAA codon, compared to a 5'-GGA codon at position 140, and thus accounted for the accessory mutation Gly140Glu. The Stanford University Drug Resistance Database lists this mutation as unusual, however, other substitutions at this site, such as Gly140Ser/Ala/Cys involving either G➔C mutations at position 419, and/or G➔U or G➔A mutations at position 418, are non-polymorphic and play an important role in the Gln148His/Arg/Lys+Gly140Ser/Aala resistance pathway, in that it restores the reduced viral replicative capacity of the Gln148His/Arg/Lys mutation. Gln148His and Gln148Arg/Lys alone reduce RAL and EVG susceptibility by 5-10 and 30-100 fold, respectively, however in combination with Gly140Ser/Ala/Cys, they reduce RAL and EVG susceptibility by >100 fold and in combination with additional mutations may also confer high-level resistance against DTG and BIC (Shafer, 2006). Generally, subtype C clades are considered more resistant to acquiring the Gly140Cys and Gly140Ser mutations and thereby to the entire Gln148His/Arg/Lys+Gly140Ser/Ala resistance pathway, as subtype C clades contain a GGA codon at position 140, instead of the GGC codon observed in subtype B clades (Brenner *et al.*, 2011; Doyle, Dunn, Ceccherini-Silberstein, De Mendoza, Garcia, Smit, Fearnhill, A.-G. Marcelin, *et al.*, 2015). This bears a higher genetic barrier, as a mutation to Gly140Cys (UGC) or Gly140Ser (AGC) would require a transition or a transversion of Guanine (418), respectively, combined with a transversion of Adenine (420), compared to a single transition of Guanine (418) in subtype B clades.

On the other hand, the subtype C specific GGA codon at position 140, despite granting a higher barrier to Gly140Ser/Cys mutations, could pose a target to human Apolipoprotein B mRNA Editing Catalytic Polypeptide-like (APOBEC) 3F enzymes. These enzymes specifically recognise 3'-TC di-nucleotides within the proviral DNA and catalyse a deamination of Cytosine, resulting in a 5'-GA➔5'-AA mutation of the corresponding strand during replication. In that, the GGA codon at position 140 could be altered to GAA and

thereby code for the Gly140Glu mutation. A bias towards di-nucleotides in glycine's and tryptophan's coding GGA and UGG codons has been established for the GG recognising APOBEC 3G proteins by Neogi et al., however, proof thereof for the close relative APOBEC 3F is still missing (Neogi *et al.*, 2013).

The impact of the Gly140Glu mutation on INSTI susceptibility has also not yet been fully determined. Despite involvement in the Gln148His/Arg/Lys pathway, the Gly140 position has also proven critical for a second mutational pathway, involving the major RAM Glu92Gln.

In a recent BIC selection assay, conducted by Andreatta et al., G140E has emerged as a secondary mutation in primary RAM Glu92Gln positive mutants after a passage of 117 days in 2.3% of the clones. After 215 days it was present in almost 60%, proving the escape advantage of this variant under BIC pressure. Over the course of the entire selection assay, testing four primary RAM mutants with all four available INSTIs each, the Glu92Gln/Gly140EGlu variant was the only strain to express high-level resistance to all available INSTIs, including second-generation INSTIs DTG and BIC, with >100 fold increased half-maximal effective concentrations (EC_{50}). At the same time this mutant showed a drastically reduced replicative capacity of 0.6% compared to the used HIV-1 NL4-3 reference strain, doubting its viability *in-vivo* (Andreatta *et al.*, 2018). A second *in-vitro* description of the Gly140Glu substitution had been provided previously in 2010, when Goethals et al. selected Gly140Glu in an assay pressuring with the experimental second-generation inhibitor MK-2048. In this case, however, no increased EC_{50} values were noted (Goethals *et al.*, 2010).

Although 5'-GA➔5'-AA hypermutations due to APOBEC 3F enzymes are commonly observed in HIV-1 infections, Gly140Glu has, to our knowledge, only been observed *in-vivo* once before: Rhee et al. performed a comparative analysis of all available HIV-1 group M IN sequences in 2008, including 1961 treatment-naïve IN sequences. This analysis found Gly140Glu to be present in only one subtype G sequence (1/1961) from Cuba (Liddament *et al.*, 2004; Rhee *et al.*, 2008).

Reason for the scarceness of this mutation *in-vivo* might lie in the nature of APOBEC proteins, as they introduce G➔A mutations in great numbers throughout the proviral genome. Neogi et al. could show in an Indian cohort that the G➔A hypermutation ultimately leads to the introduction of stop codons and otherwise defective virions in the majority of cases, which hinder the expansion of these variants *in-vivo*, explaining the lack thereof in population screenings (Neogi *et al.*, 2013). Notwithstanding the relative unlikeliness of APOBEC enzymes to contribute to drug resistances of entire viral strains on a larger scale, the

possibility of APOBEC-edited variants that do not contain stop-codons or an otherwise defective genome being viable and circulating in individuals cannot be ruled out on an individual level and may here severely impact virological outcomes (Mohammadzadeh *et al.*, 2019). Furthermore, G➔A mutations may also occur independent of APOBEC enzymes, as seen in *in-vitro* assays.

In case of our cohort, the time of sampling, before the global introduction of INSTIs, eliminates the possibility of Gly140Glu having emerged under non-reported INSTI pressure, or being transmitted as a TDR (s. below). Also, the patient's demographics as well as documented history were insignificant for prior drug-exposure or otherwise notable medical conditions. Therefore, in spite of the potential impact on replicative capacity, the Gly140Glu substitution has to be considered a natural variant in our cohort, either circulating in this patient *in-vivo*, or having mutated early in our laboratory workflow. Methods to either confirm or refute these findings could include repetitive extractions from the patient's plasma sample, follow-up isolations from subsequent blood samples, cloning assays, assessing the replicative fitness of this variant, as well as a review of the patient's history with regards to virological outcomes, particularly if initiated on an INSTI-based ART regimen. However, due to financial, informational and time constraints, these investigations were omitted in this study.

Apart from Gly140Glu, other RAMs, potentially impacting INSTI susceptibility were not observed in this study. These results are consistent with prior analyses of INSTI-naïve South African HIV-1 IN fragments, as well as larger clinical trials, recently conducted in South Africa: Fish et al. analysed 73 INSTI-naïve IN fragments in 2010, isolated from PLWH from the greater Johannesburg metropolitan area and detected no present mutation to directly impact INSTI susceptibility. However, INSTI-selected polymorphic mutations Thr97Ala and Glu157Gln, which in combination with other RAMs contribute to INSTI-resistance (Thr97Ala > Glu157Gln), were present at a rate of 1/73 (1.4%) and 3/73 (4.1%), respectively (total 4/73, 5.5%) (Fish *et al.*, 2010). Additional information was provided by Papathanasopoulos et al. and Bessong et al. in 2010 and 2013, who analysed a total of n=121 IN sequences, (32 and 89, respectively) from a similar geographical area. While both groups reported the absence of major RAMs, a total of six sequences were found to harbour one minor resistance mutation each. These were Gly163Arg and Val151Ile in the study conducted by Papathanasopoulos et al. (2/32, 6.3%) and Gln95Lys, Gln95Proline (Pro), Glu157Lys and Arg263Ser in the cohort analysed by Bessong et al. (4/89, 4.5%) (Papathanasopoulos *et al.*, 2010; Bessong and Nwobegahay, 2013). By the numbers, the differences in minor resistance

mutations between our cohort (1/91, 1.1%) and other South African cohorts (4/73, 5.5%, 2/32, 6.5%, and 4/89, 4.5%) are not statistically significant, and thus do not allow speculations on different RAM prevalence rates within the country, or the respective mutational patterns thereof. In addition, this study and the studies conducted by Fish et al., Papathanasopoulos et al. and Bessong et al. show differences in their inclusion criteria, warranting further caution, when comparing results:

In the present study, we included archived blood-samples, sampled between 1998 and 2002, a time period prior to the initiation of a national treatment program in South Africa in 2004 and the global introduction of INSTIs in 2007. While all 73 samples, analysed by Fish et al. were INSTI-naïve, 51/73 (70.0%) reported prior exposure to antiretroviral drugs, other than INSTIs. For the work, conducted by Papathanasopoulos et al., the period for sample-collection was stated as 2004-2007. Based on the provided GenBank accession number for the IN fragments screened in this study, 19/30 were listed as being sampled in 2007, giving rise to the possibility of DRMs and undocumented INSTI usage. Bessong et al. mentioned in their study that 1/3 (33.3%) of the participants estimated the year of infection to be 2008, whereas 2/3 (66.7%) assumed the date of infection to be before 2008. However, all samples were sampled in the year 2008.

The effects of theses differences in inclusion criteria on the studies results are debatable, and while the effect of non-INSTI based cART on the prevalence of INSTI-RAMs is considered minimal, the effect of TDR on INSTI resistance is not yet fully understood. However, case reports, such as reported by Young et al. or Boyd et al., who describe the cases of cART-naïve patients, harbouring multi-drug and INSTI resistant HIV-1 variants upon first diagnosis of infection, highlight the risk of TDRs against all available therapy options in treatment-naïve settings (Boyd *et al.*, 2011; Young *et al.*, 2011). As discussed in the report by Boyd et al., TDRs cannot exclusively be transmitted upon initial transmission of the virus, but may also occur in re- and super-infections of resistant HIV variants in patients, who were previously infected with a non-resistant HIV strain (Boyd *et al.*, 2011). Especially with regard to the time of occurrence in the two reports (2008 and 2009), TDRs must be considered a possibility from the time of introduction of new cART compounds. By only including samples before the widespread use of cART in South Africa and before the global introduction of INSTIs, we reduced the risk of ART-exposure as a cofounder to our study's results. Therefore we consider our results to represent a baseline, unbiased INSTI resistance rate.

In total, the absence of major RAMs in our cohort predicts good treatment efficiency for second-generation INSTIs in treatment-naïve settings in South Africa, although more research

on mutational resistance pathways, especially for the second-generation INSTIs DTG and BIC is needed in order to evaluate the significance of individual mutations, such as Gly140Glu. Questions that need to be answered with regards to this mutation in particular include, apart from its prevalence rates in larger cohort, the likelihood of selection by BIC and DTG *in-vivo* and in non-Glu92Gln-mutants, as well as its impact on the virus' replicative capacity.

The main limitation of this study arises from our inclusion criteria. In strictly including samples collected between 1998 and 2002, this study cannot represent the current state of ART resistance in South Africa, but rather a mutational baseline prior to ART exposure. Additionally, we only focused on the IN gen itself as origin of INSTI resistance, however genomic regions outside the IN gen have been described to contribute to INSTI resistance as well (Malet *et al.*, 2017). As these were not taken into account when assessing the level of INSTI-resistance, the true resistance may differ from the results presented in this study. Lastly, a study size of n=92 sequences may over- or underrepresent the true resistance rates, thus, larger cohorts are necessary to increase the study's significance and accuracy.

5.2.2 Major and minor RAMs in database-derived IN sequences

In order to better represent the drug-naïve HIV-1 IN fragment for the South African context, we included all South African publicly available drug-naïve HIV-1 IN sequences sampled before 2007 in order to rule out the effects of TDRs on our results (s. 5.2.1). Of the 284 samples retrieved from the Los Alamos sequence Database (https://www.hiv.lanl.gov/content/sequence/HIV/mainpage.html), 252 were part of NFLG sequences, whereas 32 were specifically sequenced on the IN gene alone. This set of 32 IN fragments consisted of sequences, previously described in studies by Fish et al. (n=21) and Papathanasopoulos et al. (n=11), however, in order to avoid a selection bias, these sequences remained in our analyses (Fish *et al.*, 2010; Papathanasopoulos *et al.*, 2010). Further sequences from these studies were not included, due to their sampling year or previous ART exposure.

Drug-resistance analysis revealed the presence of major RAMs in 8/284 (2.8%) sequences, with accessory RAMs being present at the same rate (8/284, 2.8%). Co-existence of >1 RAM was not observed. In the eight sequences harbouring major INSTI-RAMs, Thr66Ala and Tyr143His were identified in two sequences each, while Thr66Ser, Glu92Gly, Ser147Gly and Gln148His were present in one sequence, respectively. Thr66Ala, Glu92Gly and Ser147Gly have mostly been previously observed in patients receiving RAL or EVG based therapy and

are known to confer low-level resistance against EVG of 5-, 10- and 5-fold, respectively. Tyr143His is considered a transitional mutation in the two-base substitution pathway between the Tyr143 wild-type and the Tyr143Arg variant. While Tyr143His seemingly does not confer resistance, Tyr143Arg is associated with high-level resistance against RAL (60-fold) and low-intermediate-level resistance against EVG (10-fold), especially if combined with additional major or accessory RAMs (Liu and Shafer, 2006; Shafer, 2006). Of note is, that susceptibility to second-generation INSTIs DTG and BIC remains unaffected by the presence of either of these mutations. Similar to other major RAMs present in this cohort, Gln148His alone reduces RAL and EVG susceptibility by 5-10 fold. However, if co-existent with the compensating mutation Gly140Ser/Ala/Cys, Gln148His / Gly140Ser/Ala/Cys increases resistance to RAL and EVG dramatically and may also affect DTG and BIC susceptibility, especially if combined with additional mutations, such as Glu138Lys, Asn155His, Leu74Met or Thr97Ala (Shafer, 2006). Informational data on Thr66Ser is scarce, as it has been only observed once before by Rhee et al., however, to our knowledge, it has not been associated with INSTI resistance in previous studies (Rhee *et al.*, 2007).

Of the eight accessory mutations, observed in this study, Thr97Ala was present in 5/284 (1.8%), Glu157Gln in 2/284 (0.7%) and Gly163GlyArg in 1/284 (0.35%) sequences, respectively. Thr97Ala and Glu157Gln are both polymorphic and have been described at similar rates in other ART-naïve cohorts and are both selected for by RAL, EVG and DTG based cART (Varghese *et al.*, 2010; Abram *et al.*, 2017). Thr97Ala's impact on INSTI susceptibility remains elusive, as both, Thr97Ala mediated high-level resistance, as well as the lack thereof have been reported.

Similarly, studies on Glu157Gln indicate that Glu157Gln alone does not confer noteworthy resistance to INSTIs (Saladini *et al.*, 2017). At the same time, Glu157Gln showed compensating capabilities in the Arg263Lys resistance pathway, a well-documented pathway for DTG resistance in clinical settings (Charpentier and Descamps, 2018). Herein, the addition of Glu157Gln to Arg263Lys mutant variants increased their respective DTG-resistance by 10-fold (Anstett *et al.*, 2016).

The complexity of accessory resistance mutations is underlined by a case report of Danion et al. in 2015, in which a patient experienced virological failure under DTG based cART, after being initiated on a RAL containing first-line regimen. Switch of the INSTI compound was indicated due to Glu157Gln associated RAL resistance. Virological failure occurred despite appropriate adherence, with no other RAM than Glu157Gln being present (Danion *et al.*, 2015). This highlights the complexity of synergistic effects of mutations and polymorphisms

that are not classified as RAMs on their own, yet, in combination, may lead to reduced INSTI efficacy.

In total, the presence of a major INSTI-RAM prevalence rate of 2.8% (8/284) in a treatment-naïve population is in stark contrast to comparable reports from other geographical regions, such as Europe, Brazil, India, China and Morocco, all reporting their absence. (Passaes *et al.*, 2009; Doyle, Dunn, Ceccherini-Silberstein, De Mendoza, Garcia, Smit, Fearnhill, A. G. Marcelin, *et al.*, 2015; Alaoui *et al.*, 2018; De Francesco *et al.*, 2018; Karade, Sen and Sashindran, 2019; Liu *et al.*, 2019). For the African continent, however, similar results were reported by Inzaule et al. in 2018. In a transnational, Sub-Saharan case-control study, including 425 IN sequences from INSTI-naïve individuals from Kenya, Nigeria, South Africa, Uganda and Zambia, major INSTI-RAMs were present at a rate of 2.4%. Observed mutations and respective prevalence rates were stated as follows: Thr66Ala/Ile (0.7%), Glu92Gln (0.5%), Tyr143Cys/Ser (0.7%), Ser147Gly (0.2%) and Gln148Arg (0.5%) (Inzaule *et al.*, 2018). In addition, analysis of accessory RAMs revealed the mutation Gln95Lys, Thr97Ala, Glu157Gln, and Gly163Arg/Lys to be present at rates of 0.5%, 4%, 0.7% and 0.7%, respectively. These mutational sites, as well as their respective mutational prevalence are similar to results obtained in our study, indicating a greater genetic variability of the IN region in sequences from the African continent, despite not having been exposed to INSTIs in the past.

While INSTI-RAM rates of 2.8% or 2.4% are relatively low compared to pre-treatment NNRTI resistance rates of 11.5% in South Africa, as recently described by Chimukangara et al., a significantly higher variability in the IN region is still of concern, as INSTIs either already have been introduced, or are soon to be introduced into many African national HIV treatment programs, with the hope of low primary resistant rates, as observed in other parts of the world (UNAIDS, 2017; Chimukangara *et al.*, 2019).

Of note is, that a 2.8% RAM-prevalence rate of is not equal to a resistance rate of 2.8%, as the mainly used second-generation INSTI DTG as well as BIC show a higher genetic barrier to resistance and remain effective despite the presence of most RAL and EVG resistance pathway's mutation. Of the observed RAMs, only Gln148His/Arg and Glu157Gln have been described in direct connection to DTG resistances, although the accumulation of multiple RAMs, mostly observed in patients with RAL and/or EVG experience, can also cause resistance to DTG. On the other hand, long-term experience with second-generation INSTIs is still lacking, and new mutational pathways may still develop under increased DTG pressure. Furthermore, Anstett et al. showed, that the concurrence between laboratory analysis and viral

in-vivo resistance is insufficient for second-generation INSTIs (Anstett *et al.*, 2017). Combined with the findings from Danion et al., this points out the lack of understanding of accessory RAMs and their mechanism of action (Danion *et al.*, 2015).

With the knowledge we have today, our data suggest good efficacy of INSTI-based cART in South Africa, especially if constricted to DTG-based therapy, however, more, most importantly, clinical research is needed, in order to characterise DTG's and BIC's mutational pathways and the role of accessory RAMs therein. First clinical data on DTG efficacy on the South African context was obtained by the recently conducted ADVANCE trial. Herein, the 48-weeks results confirmed our prediction of high DTG efficacy, in that both TAF/FTC/DTG, as well as TDF/FTC/DTG reached their respective endpoint of non-inferiority when compared to a EFV based regimen and remained highly efficient throughout this study with no resistances emerging to the DTG-based regimen (Venter *et al.*, 2019).

Limitations of the analysis of the database-derived sequences are similar to the analysis of our TV cohort's limitations, in that a set of sequences, collected more than 12 years ago, is by default not able to represent the current state of resistance in a given population. Furthermore, our analysis focused on the IN-gene alone, despite NFLG sequences being available for 252 sequences, which left out mutations outside the IN gene that potentially influence the susceptibility to INSTIs (Malet *et al.*, 2017).

5.3 South African HIV-1 C IN inter and intra-subtype diversity

For the analysis of inter and intra-subtype diversity, all HIV-1 IN sequences that were identified as belonging to subtype C were used to generate a South African drug-naïve HIV-1 IN consensus sequence. This consensus was subsequently compared to a reference subtype B clade, in order to identify 15 natural polymorphisms occurring at a rate >50% in our cohort. In a first step, we showed that the identified polymorphisms were distributed equally in both, the retrieved, as well as the sequenced dataset, allowing us to use a combined dataset for comparison to other HIV-1 clades.

Secondly, comparative analysis between South African HIV-1 and global HIV-1 B sequences proved, that all 15 NOPs were indeed significantly enriched in South African subtype C sequences, marking an inter-subtype diversity of at least 6.6% (19/288) on an amino-acid and 7.7% (67/867) on a nucleotide level. Interestingly, 6/15 NOPs, namely Phe100Tyr, Leu101Ile, Thr112Val, Thr125Ala, Lys136Gln and Val201Ile were located in the IN's CCD domain, whereas only 2/15 (Asp25Glu, Val31Ile) and 6/15 (Thr218Ile, Leu234Ile, Ala265Val, Arg269Lys, Asp278Ala, Ser283Gly) were found in the enzyme's NTD and CTD domain,

respectively. Met50Ile is located in the linker region, connecting the NTD and the CCD (Rogers *et al.*, 2018). The CCD's metallic co-factor binding Asp-Asp-Glu motif, consisting of residues Asp64, Asp116 and Glu152, as well as the DNA binding residues Gln148, Lys156 and Lys159 or adjacent sites, however, remained highly conserved in both subtypes. A similar rate of 7% inter-subtype diversity for the IN region was reported by Brenner et al. in 2011 in a study cohort from Quebec, Canada (Brenner *et al.*, 2011). In this study, important implications of inter-subtype diversity were highlighted in that the Glycine coding codon at position 140 in subtype C clades consisted of GGA instead of GGC in subtype B (5.2.1). We could confirm these findings with GGA being the most prevalent codon at this position at a rate of 74% (267/361) compared to a GGC codon in 0.3% (1/361) of sequences. Of the remaining 92 sequences, 91/361 (25.2%) contained a GGG codon while one sample harboured a GAA (Gly140Glu) codon at this site.

Brenner et al. also conducted phylogenetic analyses of their cohort's samples and found subtype C clades to cluster in distinct sub-clusters, indicating significant intra-subtype diversity, depending on their respective geographical origin. With these findings in mind, we analysed the set of drug-naïve South African HIV-1 C IN sequences in comparison to all publicly available drug-naïve HIV-1 C IN sequences from outside South Africa, in order to determine the degree of intra-subtype diversity for subtype C clades and its impact on INSTI susceptibility. This analysis revealed the endemic presence of at least four NOPs (Met50Ile, Phe100Tyr, Thr112Val, Ala265Val) in South African HIV-1 C IN sequences. By using methods of molecular modelling, Rogers et al. recently assessed the impact of Met50Ile, Phe100Tyr and Ala265Val, amongst other non-subtype B specific polymorphisms, on the structural integrity of the intasome complex (Rogers *et al.*, 2018). The effect of Thr112Val was not investigated in this study. Of the three modelled NOPs, only residues at IN position 50 of the two inner IN molecules were observed to directly interact with viral cDNA. Phe100Tyr and Ala265Val were seen as parts of structural, hydrophobic clusters in the CCD and the CTD, respectively. By replacing the Met-residue at position 50 with Ile, the hydrophobic interactions between the enzyme's residue and the DNA-bases' moieties are enhanced, potentially leading to a higher affinity of the IN-Met50Ile variant to the viral cDNA. Amino acid substitutions Phe100Tyr and Ala265Val do not affect any binding sites directly, however could affect steric conformation of the core structure, including the active site, and the critical residue Arg263, involved in binding viral cDNA, respectively (Rogers *et al.*, 2018).

A second study conducted by Singh et al. investigated the subtype-specific binding affinities of DTG and RAL/EVG. Results included *in-vitro* and *in-silico* analyses (Singh *et al.*, 2019). Here, non-subtype B consensi, harbouring the polymorphisms Val31Ile, Thr112Val, Thr124Ala, Thr125Ala, Val201Ile, Leu234Ile and Ser283Gly, were observed to show greater binding affinities to DTG than subtype B consensi (1.8-fold increase). Additional analyses revealed, that the difference in binding energy between first- and second-generation INSTIs is significantly larger in non-subtype B clades: For non-stubtype B IN, they were reported at -22.6 kcal/mol versus -8.7 kcal/mol and 6.6 kcal/mol for DTG, RAL and EVG, compared to corresponding values of -13.7 kcal/mol, -11.2 kcal/mol and -10.4 kcal/mol for subtype B IN, respectively (Singh *et al.*, 2019). These results imply a greater efficacy of DTG in non-subtype B settings, than observed in subtype B predominated studies (Cahn, Anton L. Pozniak, *et al.*, 2013; Raffi *et al.*, 2013).

Of note is, that both studies, conducted by Rogers et al. and Singh et al., used global HIV-1 C IN consensi as subtype reference, and thereby analysed only for inter-subtype diversity. While these data provide great insights into IN's divers genetics impact on INSTI-binding affinity on a broader level, concrete conclusions for subtype C predominated settings, taking intra-subtype diversity into consideration, are still lacking. Discrepancies in DTG-binding affinity between subtype C IN and subtype C_{ZA} IN could result from subtype C_{ZA} specific polymorphisms, such as Met50Ile, Phe100Tyr and Ala265Val that, to our knowledge, were not incorporated into the study conducted by Singh et al.. However, data from Rogers et al. suggest, that Met50Ile leads to a greater IN affinity to viral cDNA and thereby stabilizes the intasome complex. While this possibly further increases DTG's binding affinity, it might also affect DTG's genetic barrier to resistance, as it has been observed in drug-resistance assays as a secondary mutation to Arg263Lys, one of DTG's main mutational resistance pathways (Quashie *et al.*, 2012; Tsiang *et al.*, 2016; Andreatta *et al.*, 2018). Herein, Met50Ile does not seem to restore the reduced viral replicative capacity, that is associated with Arg263Lys, due to lower IN-DNA and IN-RNA interactions, but increases the viral resistance of Arg263Lys mutant variants against INSTIs (Wares *et al.*, 2014; Kessl *et al.*, 2016). With Met50Ile already being present at a rate of 60%, a single G➔A point mutation at nucleotide position 788 would lead to Met50Ile/Arg263Lys double mutants, however, still at the cost of a reduced enzymatic activity. The loss in Arg263Lys-mutant's replicative capacity and IN activity has been quantified by Wainberg et al. at 40%, and many *in-vitro* studies suggest that this substitution is a mutational dead-end for HIV-1 (Wainberg *et al.*, 2014; Mesplède *et al.*, 2015). Nonetheless, Arg263Lys mutants have been observed in *in-vivo* studies in subtype B as well

as subtype C infected patients, mostly, however, under sub-optimal DTG blood-levels, yet proving their viability and simultaneously hinting at the existence of so far undiscovered compensatory mutations (Cahn, Anton L. Pozniak, *et al.*, 2013; Pena *et al.*, 2019; Rhee *et al.*, 2019). One potential compensatory mutation that also increases Arg263Lys mediated DTG-resistance could be Glu157Gln, as indicated by Anstett et al. in 2016 (Anstett *et al.*, 2016). In cell-free assays, they could show, that Arg263Lys/Glu157Gln double mutants had higher strand transfer activities than sole Arg263Lys variants, while expressing wild-type-level DNA binding activities and remaining infectious of the TZM-bl cell line *in-vitro*. Moreover, the Arg263Lys/Glu157Gln double mutant increased DTG resistance by 10-folds (Anstett *et al.*, 2016). Of note is, that strains, harbouring only Glu157Gln, showed increased susceptibility to DTG.

These findings confirm the hypothesis that compensatory mutations exist for Arg263Lys mutants, and that initially higher DTG susceptibility rates are not necessarily associated with better long-term cART outcomes, but may compensate for RAMs that are associated with a penalty in the virus' replicative capacity. The exact molecular mechanisms, however, in which Glu157Gln or other compensating mutations for Arg263Lys mutants restore the viral strand transfer and DNA-binding activities as well as its infectivity, remain elusive.

In light of these information, our data suggest, that the HIV-1 C_{ZA} IN enzyme expresses substantial genetic inter- as well as intra-subtype diversity, however, no present polymorphism was found to directly impact INSTIs susceptibility. Therefore, we expect a good initial treatment response to the widespread use of DTG in South Africa. Nonetheless may present polymorphisms in the IN gene shape HIV IN's future evolution under DTG pressure and thereby alter DTG's susceptibility. In particular the presence of polymorphisms in DTG's Arg263Lys mutational pathway needs to be monitored, as individual polymorphisms, such as Met50Ile that is known to increase DTG resistance, are currently present in up to 60% of South African IN sequences. With the rollout of DTG to the majority of patients in South Africa, Met50Ile may facilitate and accelerate Arg263Lys-mediated DTG resistance upon its emergence.

In addition, more research on drug-specific mutational pathways needs to be conducted in order to clarify resistance routes and determine the molecular interactions between RAMs and polymorphisms on a regional level. With significant intra-subtype diversity and increasing rates of recombinants, we may likely reach a point in near future, at which categorical data, such as subtype-assignation will no longer appropriately represent HIV-1's diversity and the complex interactions of polymorphic and non-polymorphic mutations in a singular virus.

Therefore, more regional, potentially even patient-specific approaches might be required to assess a viral strain's resistance profile and provide an individual patient with the optimal treatment regimen available. The best way to ensure optimal care and collect insights into emerging resistance-patterns under DTG rollout at the same time is the initiation of an institutionalised pre-treatment drug-resistance-testing program, analysing a patient's IN gene before commencing cART as well as upon signs of virological failure. In doing so, South Africa could be one step closer to achieving UNAIDS 90-90-90 goals and ending the AIDS epidemic by 2030.

Of interest in the future will be the change in IN diversity under INSTI pressure. Since this study only included South African IN sequences prior to 2007, our results have to be considered a baseline diversity rate. By including cART-experienced and more recently collected samples from the Los Alamos database, their comparison to the presented results could have provided a first notion of INSTI's influence, however, as INSTIs were only used sporadically before 2019, such analyses were not included in this work. Moreover, possible polymorphisms in HIV-1 subtypes other than subtype C were not assessed. In order to fully represent South Africa's HIV epidemic, polymorphisms for other subtypes and common CRFs need to be defined in this context. Lastly, the statistical methods, which were applied to test for observed polymorphism rates in this study for statistical significance, are widely considered too conservative, therefore underestimating the significance of results. Choosing these underlined the statistical significance of the reported differences in HIV-1 C IN mutations, however, the appliance less conservative methods could have resulted in altered rates of NOPs in our cohort.

6 Conclusion

With UNAIDS 90-90-90 goals in mind, we believe that the current time period and the long-term efficacy of DTG based cART is pivotal, if the goal of ending the AIDS epidemic by 2030 is to remain realistic. Especially in a country as South Africa, carrying more than 20% of the global HIV burden, every available method to help contain the viral spread and maintain an effective first-line cART regimen should be applied. To date, the South African HIV epidemic is still predominated by HIV-1 group M subtype C strains. Our data suggest that primary RAMs directly affecting the susceptibility of second-generation INSTIs are absent in South Africa's cART-naïve context. Secondary mutations therein were present at normal rates of <3%. Viruses belonging to subtype C, however, expressed stark inter- and intra-subtype diversity with a high proportion of potentially DTG affecting polymorphisms. In

face of the uncertainty of DTG's long-term resistance pathways *in-vivo*, pre-treatment drug-resistance testing for primary RAMs may hold key information for maintaining DTG's efficacy, as it might detect mutations that require dose-adaptation for efficient viral suppression and subsequent elimination of partially resistant variants. In addition, it could provide critical *in-vivo* insights into DTG's unique resistance pathways, which would otherwise lead to the emergence of resistant strains evading the scope of detection. Therefore we urge for the initiation of an institutionalised pre-treatment drug-resistance-testing program in South Africa, testing before commencement of cART, as well as upon signs of virological failure.

7 Bibliography

Abram, M. E. *et al.* (2017) 'Lack of impact of pre-existing T97A HIV-1 integrase mutation on integrase strand transfer inhibitor resistance and treatment outcome.', *PloS one*. Public Library of Science, 12(2), p. e0172206. doi: 10.1371/journal.pone.0172206.

Adachi, M. *et al.* (2009) 'Structure of HIV-1 protease in complex with potent inhibitor KNI-272 determined by high-resolution X-ray and neutron crystallography.', *Proceedings of the National Academy of Sciences of the United States of America*. National Academy of Sciences, 106(12), pp. 4641–6. doi: 10.1073/pnas.0809400106.

Aghokeng, A. F. *et al.* (2009) 'Inaccurate diagnosis of HIV-1 group M and O is a key challenge for ongoing universal access to antiretroviral treatment and HIV prevention in Cameroon.', *PloS one*. Public Library of Science, 4(11), p. e7702. doi: 10.1371/journal.pone.0007702.

Aidsinfo (2020) 'FDA-Approved HIV Medicines.' Available at: https://aidsinfo.nih.gov/understanding-hiv-aids/fact-sheets/21/58/fda-approved-hiv-medicines (Accessed: 02 March 2020).

Alaoui, N. *et al.* (2018) 'Prevalence of resistance to integrase strand-transfer inhibitors (INSTIs) among untreated HIV-1 infected patients in Morocco', *BMC Research Notes*. BioMed Central, 11(1), p. 369. doi: 10.1186/s13104-018-3492-5.

Alberts, B. *et al.* (2002) 'Helper T Cells and Lymphocyte Activation'. Garland Science. Available at: https://www.ncbi.nlm.nih.gov/books/NBK26827/ (Accessed: 02 March 2020).

Anderson-Daniels, J. *et al.* (2019) 'Dominant negative MA-CA fusion protein is incorporated into HIV-1 cores and inhibits nuclear entry of viral preintegration complexes.', *Journal of virology*. American Society for Microbiology Journals, p. JVI.01118-19. doi: 10.1128/JVI.01118-19.

Andreatta, K. *et al.* (2018) *[Bictegravir] Integrase Inhibitor Resistance Selections Initiated with Drug Resistant HIV-1, Integrase Inhibitor Resistance Selections Initiated with Drug Resistant HIV-1*. Available at: http://www.natap.org/2018/CROI/croi_134.htm (Accessed: 02 March 2020).

Anstett, K. *et al.* (2015) 'Dolutegravir resistance mutation R263K cannot coexist in combination with many classical integrase inhibitor resistance substitutions.', *Journal of virology*. American Society for Microbiology Journals, 89(8), pp. 4681–4. doi: 10.1128/JVI.03485-14.

Anstett, K. *et al.* (2016) 'Polymorphic substitution E157Q in HIV-1 integrase increases R263K-mediated dolutegravir resistance and decreases DNA binding activity', *Journal of*

Antimicrobial Chemotherapy, 71(8), pp. 2083–2088. doi: 10.1093/jac/dkw109.

Anstett, K. *et al.* (2017) 'HIV drug resistance against strand transfer integrase inhibitors.', *Retrovirology*. BioMed Central, 14(1), p. 36. doi: 10.1186/s12977-017-0360-7.

Antinori, A. *et al.* (2018) 'Efficacy and safety of boosted darunavir-based antiretroviral therapy in HIV-1-positive patients: results from a meta-analysis of clinical trials.', *Scientific reports*. Nature Publishing Group, 8(1), p. 5288. doi: 10.1038/s41598-018-23375-6.

Antiretroviral Therapy Cohort Collaboration, A. *et al.* (2017) 'Survival of HIV-positive patients starting antiretroviral therapy between 1996 and 2013: a collaborative analysis of cohort studies.', *The lancet. HIV*. Elsevier, 4(8), pp. e349–e356. doi: 10.1016/S2352-3018(17)30066-8.

Aoki, M. *et al.* (2018) 'Mechanism of Darunavir (DRV)'s High Genetic Barrier to HIV-1 Resistance: A Key V32I Substitution in Protease Rarely Occurs, but Once It Occurs, It Predisposes HIV-1 To Develop DRV Resistance', *mBio*. Edited by V. R. Prasad, 9(2). doi: 10.1128/mBio.02425-17.

Arendt, G. *et al.* (2007) 'Neuropsychiatric side effects of efavirenz therapy', *Expert Opinion on Drug Safety*. Taylor & Francis, 6(2), pp. 147–154. doi: 10.1517/14740338.6.2.147.

Arts, E. J. and Hazuda, D. J. (2012) 'HIV-1 antiretroviral drug therapy.', *Cold Spring Harbor perspectives in medicine*. Cold Spring Harbor Laboratory Press, 2(4), p. a007161. doi: 10.1101/cshperspect.a007161.

Bailey, D. G. and Dresser, G. K. (2004) 'Natural products and adverse drug interactions.', *CMAJ : Canadian Medical Association journal = journal de l'Association medicale canadienne*. CMAJ, 170(10), pp. 1531–2. doi: 10.1503/CMAJ.1031558.

BAMPI, C. *et al.* (2004) 'The chaperoning and assistance roles of the HIV-1 nucleocapsid protein in proviral DNA synthesis and maintenance', *The International Journal of Biochemistry & Cell Biology*, 36(9), pp. 1668–1686. doi: 10.1016/j.biocel.2004.02.024.

Barin, F. *et al.* (2007) 'Prevalence of HIV-2 and HIV-1 group O infections among new HIV diagnoses in France: 2003–2006', *AIDS*, 21(17), pp. 2351–2353. doi: 10.1097/QAD.0b013e3282f15637.

Barre-Sinoussi, F. *et al.* (1983) 'Isolation of a T-lymphotropic retrovirus from a patient at risk for acquired immune deficiency syndrome (AIDS)', *Science*, 220(4599), pp. 868–871. doi: 10.1126/science.6189183.

Bartlett, J. A. *et al.* (2006) 'An updated systematic overview of triple combination therapy in antiretroviral-naive HIV-infected adults', *AIDS*, 20(16), pp. 2051–2064. doi:

10.1097/01.aids.0000247578.08449.ff.

Bbosa, N., Kaleebu, P. and Ssemwanga, D. (2019) 'HIV subtype diversity worldwide', *Current Opinion in HIV and AIDS*, 14(3), pp. 153–160. doi: 10.1097/COH.0000000000000534.

Bessong, P. O. and Iweriebor, B. (2016) 'A putative HIV-1 subtype C/CRF11_cpx unique recombinant from South Africa', *SpringerPlus*. Springer International Publishing, 5(1), p. 285. doi: 10.1186/s40064-016-1924-z.

Bessong, P. O. and Nwobegahay, J. (2013) 'Genetic Analysis of HIV-1 Integrase Sequences from Treatment Naive Individuals in Northeastern South Africa.', *International journal of molecular sciences*. Multidisciplinary Digital Publishing Institute (MDPI), 14(3), pp. 5013–24. doi: 10.3390/ijms14035013.

Blumenthal, R., Durell, S. and Viard, M. (2012) 'HIV entry and envelope glycoprotein-mediated fusion.', *The Journal of biological chemistry*. American Society for Biochemistry and Molecular Biology, 287(49), pp. 40841–9. doi: 10.1074/jbc.R112.406272.

Borrow, P. *et al.* (1994) 'Virus-specific CD8+ cytotoxic T-lymphocyte activity associated with control of viremia in primary human immunodeficiency virus type 1 infection.', *Journal of virology*. American Society for Microbiology Journals, 68(9), pp. 6103–10. Available at: http://www.ncbi.nlm.nih.gov/pubmed/8057491 (Accessed: 02 March 2020).

Boyd, S. D. *et al.* (2011) 'Transmitted raltegravir resistance in an HIV-1 CRF_AG-infected patient.', *Antiviral therapy*. NIH Public Access, 16(2), pp. 257–61. doi: 10.3851/IMP1749.

Boyer, P. L. *et al.* (2001) 'Selective excision of AZTMP by drug-resistant human immunodeficiency virus reverse transcriptase.', *Journal of virology*. American Society for Microbiology (ASM), 75(10), pp. 4832–42. doi: 10.1128/JVI.75.10.4832-4842.2001.

Brenner, B. G. *et al.* (2007) 'High Rates of Forward Transmission Events after Acute/Early HIV-1 Infection', *The Journal of Infectious Diseases*. Narnia, 195(7), pp. 951–959. doi: 10.1086/512088.

Brenner, B. G. *et al.* (2011) 'Subtype diversity associated with the development of HIV-1 resistance to integrase inhibitors', *Journal of Medical Virology*. John Wiley & Sons, Ltd, 83(5), pp. 751–759. doi: 10.1002/jmv.22047.

Brik, A. and Wong, C.-H. (2003) 'HIV-1 protease: mechanism and drug discovery', *Organic & Biomolecular Chemistry*. Royal Society of Chemistry, 1(1), pp. 5–14. doi: 10.1039/b208248a.

Cahn, P., Pozniak, Anton L, *et al.* (2013) 'Dolutegravir versus raltegravir in antiretroviral-experienced, integrase-inhibitor-naive adults with HIV: week 48 results from the randomised, double-blind, non-inferiority SAILING study', *The Lancet*, 382(9893), pp. 700–708. doi: 10.1016/S0140-6736(13)61221-0.

Cahn, P., Pozniak, Anton L., *et al.* (2013) 'Dolutegravir versus raltegravir in antiretroviral-experienced, integrase-inhibitor-naive adults with HIV: Week 48 results from the randomised, double-blind, non-inferiority SAILING study', *The Lancet*, 382(9893), pp. 700–708. doi: 10.1016/S0140-6736(13)61221-0.

Cao, Y. *et al.* (1995) 'Virologic and Immunologic Characterization of Long-Term Survivors of Human Immunodeficiency Virus Type 1 Infection', *New England Journal of Medicine*, 332(4), pp. 201–208. doi: 10.1056/NEJM199501263320401.

Cardoso, M. *et al.* (2018) 'Two cases of dolutegravir failure with R263K mutation', *AIDS*, 32(17), pp. 2639–2640. doi: 10.1097/QAD.0000000000001978.

Carr, J. K. *et al.* (1996) 'Full-length sequence and mosaic structure of a human immunodeficiency virus type 1 isolate from Thailand.', *Journal of virology*, 70(9), pp. 5935–43. Available at: http://www.ncbi.nlm.nih.gov/pubmed/8709215 (Accessed: 02 March 2020).

Carvalho, A. *et al.* (2015) 'Analysis of a local HIV-1 epidemic in portugal highlights established transmission of non-B and non-G subtypes.', *Journal of clinical microbiology*. American Society for Microbiology (ASM), 53(5), pp. 1506–14. doi: 10.1128/JCM.03611-14.

Castagna, A. *et al.* (2014) 'Dolutegravir in Antiretroviral-Experienced Patients With Raltegravir- and/or Elvitegravir-Resistant HIV-1: 24-Week Results of the Phase III VIKING-3 Study', *Journal of Infectious Diseases*, 210(3), pp. 354–362. doi: 10.1093/infdis/jiu051.

Cavert, W. *et al.* (1997) 'Kinetics of Response in Lymphoid Tissues to Antiretroviral Therapy of HIV-1 Infection', *Science*. American Association for the Advancement of Science, 276(5314), pp. 960–964. doi: 10.1126/science.276.5314.960.

Centers for Disease Control and Prevention (1983) 'Current Trends Update: Acquired Immunodeficiency Syndrome (AIDS) -

- United States'. Available at: https://www.cdc.gov/mmwr/preview/mmwrhtml/00000137.htm (Accessed: 02 March 2020).

Centers for Disease Control and Prevention (CDC) (1981) 'Kaposi's sarcoma and Pneumocystis pneumonia among homosexual men--New York City and California.', *MMWR. Morbidity and mortality weekly report*, 30(25), pp. 305–8. Available at: http://www.ncbi.nlm.nih.gov/pubmed/6789108 (Accessed: 02 March 2020).

Centers for Disease Control and Prevention (1982) *Current Trends Update on Acquired Immune Deficiency Syndrome (AIDS) --United States*. Available at:

https://www.cdc.gov/mmwr/preview/mmwrhtml/00001163.htm (Accessed: 02 March 2020).

Centers for Disease Control and Prevention (2019) *Stages of HIV Infection | HIV Risk Reduction Tool | CDC*. Available at: https://wwwn.cdc.gov/hivrisk/what_is/stages_hiv_infection.html (Accessed: 02 March 2020).

Chaix, M.-L. *et al.* (2013) 'Increasing HIV-1 Non-B Subtype Primary Infections in Patients in France and Effect of HIV Subtypes on Virological and Immunological Responses to Combined Antiretroviral Therapy', *Clinical Infectious Diseases*. Narnia, 56(6), pp. 880–887. doi: 10.1093/cid/cis999.

Charpentier, C. and Descamps, D. (2018) 'Resistance to HIV Integrase Inhibitors: About R263K and E157Q Mutations', *Viruses*. Multidisciplinary Digital Publishing Institute, 10(1), p. 41. doi: 10.3390/v10010041.

Cherepanov, P. *et al.* (2003) 'HIV-1 Integrase Forms Stable Tetramers and Associates with LEDGF/p75 Protein in Human Cells', *Journal of Biological Chemistry*. American Society for Biochemistry and Molecular Biology, 278(1), pp. 372–381. doi: 10.1074/JBC.M209278200.

Chimukangara, B. *et al.* (2019) 'Trends in Pretreatment HIV-1 Drug Resistance in Antiretroviral Therapy-naive Adults in South Africa, 2000–2016: A Pooled Sequence Analysis', *EClinicalMedicine*. Elsevier, 9, pp. 26–34. doi: 10.1016/J.ECLINM.2019.03.006.

Choi, E. *et al.* (2018) 'Recent advances in the discovery of small-molecule inhibitors of HIV-1 integrase', *Future Science OA*, 4(9), p. FSO338. doi: 10.4155/fsoa-2018-0060.

Chomont, N. *et al.* (2009) 'HIV reservoir size and persistence are driven by T cell survival and homeostatic proliferation', *Nature Medicine*. Nature Publishing Group, 15(8), pp. 893–900. doi: 10.1038/nm.1972.

Chun, T.-W. *et al.* (1995) 'In vivo fate of HIV-1-infected T cells: Quantitative analysis of the transition to stable latency', *Nature Medicine*. Nature Publishing Group, 1(12), pp. 1284–1290. doi: 10.1038/nm1295-1284.

Ciuffi, A. *et al.* (2005) 'A role for LEDGF/p75 in targeting HIV DNA integration', *Nature Medicine*. Nature Publishing Group, 11(12), pp. 1287–1289. doi: 10.1038/nm1329.

Clavel, F. *et al.* (1986) 'Isolation of a new human retrovirus from West African patients with AIDS', *Science*, 233(4761), pp. 343–346. doi: 10.1126/science.2425430.

Cohen, M. S. *et al.* (2016) 'Antiretroviral Therapy for the Prevention of HIV-1 Transmission', *New England Journal of Medicine*. Massachusetts Medical Society, 375(9), pp. 830–839. doi: 10.1056/NEJMoa1600693.

Cohen, M. S. (2019) 'Successful treatment of HIV eliminates sexual transmission.',

Lancet (London, England). Elsevier, 393(10189), pp. 2366–2367. doi: 10.1016/S0140-6736(19)30701-9.

Cozzi-Lepri, A. *et al.* (2015) 'Low-frequency drug-resistant HIV-1 and risk of virological failure to first-line NNRTI-based ART: a multicohort European case–control study using centralized ultrasensitive 454 pyrosequencing', *Journal of Antimicrobial Chemotherapy*. Narnia, 70(3), pp. 930–940. doi: 10.1093/jac/dku426.

Craigie, R. (2001) 'HIV integrase, a brief overview from chemistry to therapeutics.', *The Journal of biological chemistry*. American Society for Biochemistry and Molecular Biology, 276(26), pp. 23213–6. doi: 10.1074/jbc.R100027200.

Craigie, R. and Bushman, F. D. (2012) 'HIV DNA integration.', *Cold Spring Harbor perspectives in medicine*. Cold Spring Harbor Laboratory Press, 2(7), p. a006890. doi: 10.1101/cshperspect.a006890.

Danion, F. *et al.* (2015) 'Non-virological response to a dolutegravir-containing regimen in a patient harbouring a E157Q-mutated virus in the integrase region', *Journal of Antimicrobial Chemotherapy*. Oxford University Press, 69(6), pp. 2118–22. doi: 10.1093/jac/dkv012.

Das, K. *et al.* (2007) 'Crystal Structures of Clinically Relevant Lys103Asn/Tyr181Cys Double Mutant HIV-1 Reverse Transcriptase in Complexes with ATP and Non-nucleoside Inhibitor HBY 097', *Journal of Molecular Biology*. Academic Press, 365(1), pp. 77–89. doi: 10.1016/J.JMB.2006.08.097.

Deeks, S. G. (2001) 'International perspectives on antiretroviral resistance. Nonnucleoside reverse transcriptase inhibitor resistance.', *Journal of acquired immune deficiency syndromes (1999)*, 26 Suppl 1, pp. S25-33. doi: 10.1097/00042560-200103011-00004.

Deeks, S. G. *et al.* (2004) 'Immune activation set point during early HIV infection predicts subsequent CD4+ T-cell changes independent of viral load', *Blood*. American Society of Hematology, 104(4), pp. 942–947. doi: 10.1182/blood-2003-09-3333.

Deeks, S. G., Lewin, S. R. and Havlir, D. V (2013) 'The end of AIDS: HIV infection as a chronic disease.', *Lancet (London, England)*. NIH Public Access, 382(9903), pp. 1525–33. doi: 10.1016/S0140-6736(13)61809-7.

Delaugerre, C. *et al.* (2001) 'Resistance profile and cross-resistance of HIV-1 among patients failing a non-nucleoside reverse transcriptase inhibitor-containing regimen', *Journal of Medical Virology*. John Wiley & Sons, Ltd, 65(3), pp. 445–448. doi: 10.1002/jmv.2055.

Ding, J. *et al.* (1998) 'Structure and functional implications of the polymerase active

site region in a complex of HIV-1 RT with a double-stranded DNA template-primer and an antibody fab fragment at 2.8 Å resolution', *Journal of Molecular Biology*, 284(4), pp. 1095–1111. doi: 10.1006/jmbi.1998.2208.

Doyle, T., Dunn, D. T., Ceccherini-Silberstein, F., De Mendoza, C., Garcia, F., Smit, E., Fearnhill, E., Marcelin, A. G., *et al.* (2015) 'Integrase inhibitor (INI) genotypic resistance in treatment-naive and raltegravir-experienced patients infected with diverse HIV-1 clades', *Journal of Antimicrobial Chemotherapy*, 70(11), pp. 3080–3086. doi: 10.1093/jac/dkv243.

Doyle, T., Dunn, D. T., Ceccherini-Silberstein, F., De Mendoza, C., Garcia, F., Smit, E., Fearnhill, E., Marcelin, A.-G., *et al.* (2015) 'Integrase inhibitor (INI) genotypic resistance in treatment-naive and raltegravir-experienced patients infected with diverse HIV-1 clades', *Journal of Antimicrobial Chemotherapy*. Oxford University Press, 70(11), pp. 3080–3086. doi: 10.1093/jac/dkv243.

Eagling, V. A., Profit, L. and Back, D. J. (1999) 'Inhibition of the CYP3A4-mediated metabolism and P-glycoprotein-mediated transport of the HIV-1 protease inhibitor saquinavir by grapefruit juice components.', *British journal of clinical pharmacology*. Wiley-Blackwell, 48(4), pp. 543–52. doi: 10.1046/j.1365-2125.1999.00052.x.

Edgar, R. C. (2004) 'MUSCLE: multiple sequence alignment with high accuracy and high throughput.', *Nucleic acids research*. Oxford University Press, 32(5), pp. 1792–7. doi: 10.1093/nar/gkh340.

Engelman, A. and Cherepanov, P. (2014) 'Retroviral Integrase Structure and DNA Recombination Mechanism.', *Microbiology spectrum*. NIH Public Access, 2(6), pp. 1–22. doi: 10.1128/microbiolspec.MDNA3-0024-2014.

Engelman, A., Mizuuchi, K. and Craigie, R. (1991) 'HIV-1 DNA integration: mechanism of viral DNA cleavage and DNA strand transfer.', *Cell*, 67(6), pp. 1211–21. doi: 10.1016/0092-8674(91)90297-c.

Engelman, A. N. (2019) 'Multifaceted HIV integrase functionalities and therapeutic strategies for their inhibition.', *The Journal of biological chemistry*. American Society for Biochemistry and Molecular Biology, 294(41), pp. 15137–15157. doi: 10.1074/jbc.REV119.006901.

Eron, J. J. *et al.* (2011) 'Raltegravir once daily or twice daily in previously untreated patients with HIV-1: a randomised, active-controlled, phase 3 non-inferiority trial', *The Lancet Infectious Diseases*, 11(12), pp. 907–915. doi: 10.1016/S1473-3099(11)70196-7.

Eron, J. J. *et al.* (2019) 'Week 96 efficacy and safety results of the phase 3, randomized EMERALD trial to evaluate switching from boosted-protease inhibitors plus

emtricitabine/tenofovir disoproxil fumarate regimens to the once daily, single-tablet regimen of darunavir/cobicistat/emtricitabine/tenofovir alafenamide (D/C/F/TAF) in treatment-experienced, virologically-suppressed adults living with HIV-1', *Antiviral Research*, 170, p. 104543. doi: 10.1016/j.antiviral.2019.104543.

Faure, A. *et al.* (2005) 'HIV-1 integrase crosslinked oligomers are active in vitro.', *Nucleic acids research*. Oxford University Press, 33(3), pp. 977–86. doi: 10.1093/nar/gki241.

Finzi, D. *et al.* (1999) 'Latent infection of CD4+ T cells provides a mechanism for lifelong persistence of HIV-1, even in patients on effective combination therapy', *Nature Medicine*. Nature Publishing Group, 5(5), pp. 512–517. doi: 10.1038/8394.

Finzi, D. and Siliciano, R. F. (1998) 'Viral dynamics in HIV-1 infection.', *Cell*. Elsevier, 93(5), pp. 665–71. doi: 10.1016/s0092-8674(00)81427-0.

Fiorentini, S. *et al.* (2006) 'Functions of the HIV-1 matrix protein p17.', *The new microbiologica*, 29(1), pp. 1–10. Available at: http://www.ncbi.nlm.nih.gov/pubmed/16608119 (Accessed: 02 March 2020).

Fish, M. Q. *et al.* (2010) 'Natural Polymorphisms of *integrase* Among HIV Type 1-Infected South African Patients', *AIDS Research and Human Retroviruses*. Mary Ann Liebert, Inc. 140 Huguenot Street, 3rd Floor New Rochelle, NY 10801 USA , 26(4), pp. 489–493. doi: 10.1089/aid.2009.0249.

Fourati, S. *et al.* (2015) 'Cross-resistance to elvitegravir and dolutegravir in 502 patients failing on raltegravir: a French national study of raltegravir-experienced HIV-1-infected patients', *Journal of Antimicrobial Chemotherapy*, 70(5), pp. 1507–1512. doi: 10.1093/jac/dku535.

De Francesco, M. A. *et al.* (2018) 'Prevalence of Integrase Strand Transfer Inhibitors Resistance Mutations in Integrase Strand Transfer Inhibitors-Naive and -Experienced HIV-1 Infected Patients: A Single Center Experience.', *AIDS research and human retroviruses*, 34(7), pp. 570–574. doi: 10.1089/AID.2018.0006.

Frankel, A. D. and Young, J. A. T. (1998) 'HIV-1: Fifteen Proteins and an RNA', *Annual Review of Biochemistry*, 67(1), pp. 1–25. doi: 10.1146/annurev.biochem.67.1.1.

Frost, S. D. W. and McLean, A. R. (1994) 'Quasispecies dynamics and the emergence of drug resistance during zidovudine therapy of HIV infection', *AIDS*, 8(3), pp. 323–332. doi: 10.1097/00002030-199403000-00005.

Fung, H. B., Stone, E. A. and Piacenti, F. J. (2002) 'Tenofovir disoproxil fumarate: a nucleotide reverse transcriptase inhibitor for the treatment of HIV infection.', *Clinical therapeutics*, 24(10), pp. 1515–48. doi: 10.1016/s0149-2918(02)80058-3.

Gao, F. *et al.* (1998) 'An isolate of human immunodeficiency virus type 1 originally classified as subtype I represents a complex mosaic comprising three different group M subtypes (A, G, and I).', *Journal of virology*. American Society for Microbiology (ASM), 72(12), pp. 10234–41. Available at: http://www.ncbi.nlm.nih.gov/pubmed/9811767 (Accessed: 02 March 2020).

García-Lerma, J. G. *et al.* (2003) 'A novel genetic pathway of human immunodeficiency virus type 1 resistance to stavudine mediated by the K65R mutation.', *Journal of virology*, 77(10), pp. 5685–93. doi: 10.1128/jvi.77.10.5685-5693.2003.

Geuther, R. (2007) 'A. L. LEHNINGER, Biochemistry. The Molecular Basis of Cell Structure and Function (2nd Edition). 1104 S., zahlr. Abb., zahlr. Tab. New York 1975. Worth Publ. Inc.', *Zeitschrift für allgemeine Mikrobiologie*. John Wiley & Sons, Ltd, 17(1), pp. 86–87. doi: 10.1002/jobm.19770170116.

Goethals, O. *et al.* (2010) 'Primary mutations selected in vitro with raltegravir confer large fold changes in susceptibility to first-generation integrase inhibitors, but minor fold changes to inhibitors with second-generation resistance profiles', *Virology*. Academic Press, 402(2), pp. 338–346. doi: 10.1016/J.VIROL.2010.03.034.

Goldgur, Y. *et al.* (1998) 'Three new structures of the core domain of HIV-1 integrase: an active site that binds magnesium.', *Proceedings of the National Academy of Sciences of the United States of America*. National Academy of Sciences, 95(16), pp. 9150–4. doi: 10.1073/pnas.95.16.9150.

Gottlieb, M. S. *et al.* (1981) '*Pneumocystis carinii* Pneumonia and Mucosal Candidiasis in Previously Healthy Homosexual Men', *New England Journal of Medicine*, 305(24), pp. 1425–1431. doi: 10.1056/NEJM198112103052401.

Grinsztejn, B. *et al.* (2007) 'Safety and efficacy of the HIV-1 integrase inhibitor raltegravir (MK-0518) in treatment-experienced patients with multidrug-resistant virus: a phase II randomised controlled trial', *The Lancet*. Elsevier, 369(9569), pp. 1261–1269. doi: 10.1016/S0140-6736(07)60597-2.

Grohmann, D. *et al.* (2008) 'HIV-1 Nucleocapsid Traps Reverse Transcriptase on Nucleic Acid Substrates †', *Biochemistry*. American Chemical Society, 47(46), pp. 12230–12240. doi: 10.1021/bi801386r.

Group, T. I. S. S. (2015) 'Initiation of Antiretroviral Therapy in Early Asymptomatic HIV Infection', *New England Journal of Medicine*. Massachusetts Medical Society, 373(9), pp. 795–807. doi: 10.1056/NEJMoa1506816.

Gupta, R. K. *et al.* (2018) 'HIV-1 drug resistance before initiation or re-initiation of

first-line antiretroviral therapy in low-income and middle-income countries: a systematic review and meta-regression analysis.', *The Lancet. Infectious diseases*. Elsevier, 18(3), pp. 346–355. doi: 10.1016/S1473-3099(17)30702-8.

Hammer, S. M. *et al.* (1996) 'A Trial Comparing Nucleoside Monotherapy with Combination Therapy in HIV-Infected Adults with CD4 Cell Counts from 200 to 500 per Cubic Millimeter', *New England Journal of Medicine*. Massachusetts Medical Society , 335(15), pp. 1081–1090. doi: 10.1056/NEJM199610103351501.

Hare, S. *et al.* (2009) 'Structural Basis for Functional Tetramerization of Lentiviral Integrase', *PLoS Pathogens*. Edited by J. Luban. Public Library of Science, 5(7), p. e1000515. doi: 10.1371/journal.ppat.1000515.

Hare, S. *et al.* (2010) 'Retroviral intasome assembly and inhibition of DNA strand transfer', *Nature*. Nature Publishing Group, 464(7286), pp. 232–236. doi: 10.1038/nature08784.

Harmelen, J. H. Van *et al.* (1999) 'A Predominantly HIV Type 1 Subtype C-Restricted Epidemic in South African Urban Populations', *AIDS Research and Human Retroviruses*. Mary Ann Liebert, Inc. , 15(4), pp. 395–398. doi: 10.1089/088922299311376.

Van Heuverswyn, F. *et al.* (2006) 'SIV infection in wild gorillas', *Nature*, 444(7116), pp. 164–164. doi: 10.1038/444164a.

Hightower, K. E. *et al.* (2011) '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*. American Society for Microbiology (ASM), 55(10), pp. 4552–9. doi: 10.1128/AAC.00157-11.

Hill, A. and Sawyer, W. (2009) 'Effects of nucleoside reverse transcriptase inhibitor backbone on the efficacy of first-line boosted highly active antiretroviral therapy based on protease inhibitors: meta-regression analysis of 12 clinical trials in 5168 patients', *HIV Medicine*. John Wiley & Sons, Ltd (10.1111), 10(9), pp. 527–535. doi: 10.1111/j.1468-1293.2009.00724.x.

Hirsch, V. M. *et al.* (1989) 'An African primate lentivirus (SIVsmclosely related to HIV-2', *Nature*, 339(6223), pp. 389–392. doi: 10.1038/339389a0.

Hoffmann, C. *et al.* (2017) 'Higher rates of neuropsychiatric adverse events leading to dolutegravir discontinuation in women and older patients', *HIV Medicine*. John Wiley & Sons, Ltd (10.1111), 18(1), pp. 56–63. doi: 10.1111/hiv.12468.

Hottiger, M. and Hübscher, U. (1996) 'Human Immunodeficiency Virus type 1 reverse

transcriptase.', *Biological chemistry Hoppe-Seyler*, 377(2), pp. 97–120. Available at: http://www.ncbi.nlm.nih.gov/pubmed/8868066 (Accessed: 02 March 2020).

Hu, W.-S. and Hughes, S. H. (2012) 'HIV-1 Reverse Transcription', *Cold Spring Harbor Perspectives in Medicine*. Cold Spring Harbor Laboratory Press, 2(10). doi: 10.1101/CSHPERSPECT.A006882.

Hughes, P. J. *et al.* (2011) 'Protease Inhibitors for Patients With HIV-1 Infection: A Comparative Overview.', *P & T : a peer-reviewed journal for formulary management*. MediMedia, USA, 36(6), pp. 332–45. Available at: http://www.ncbi.nlm.nih.gov/pubmed/21785550 (Accessed: 02 March 2020).

Hurt, C. B. *et al.* (2017) 'Selecting an HIV Test: A Narrative Review for Clinicians and Researchers.', *Sexually transmitted diseases*. NIH Public Access, 44(12), pp. 739–746. doi: 10.1097/OLQ.0000000000000719.

Ingold, H. *et al.* (2019) 'The Self-Testing AfRica (STAR) Initiative: accelerating global access and scale-up of HIV self-testing', *Journal of the International AIDS Society*. Wiley-Blackwell, 22(Suppl Suppl 1), p. e25249. doi: 10.1002/JIA2.25249.

Inzaule, S. C. *et al.* (2018) 'Primary resistance to integrase strand transfer inhibitors in patients infected with diverse HIV-1 subtypes in sub-Saharan Africa', *Journal of Antimicrobial Chemotherapy*. Narnia, 73(5), pp. 1167–1172. doi: 10.1093/jac/dky005.

Jabłonowska, E. *et al.* (2019) 'Real-life study of dual therapy based on dolutegravir and ritonavir-boosted darunavir in HIV-1-infected treatment-experienced patients.', *PloS one*. Public Library of Science, 14(1), p. e0210476. doi: 10.1371/journal.pone.0210476.

Jacobs, G. B. *et al.* (2014) 'HIV-1 Subtypes B and C Unique Recombinant Forms (URFs) and Transmitted Drug Resistance Identified in the Western Cape Province, South Africa', *PLoS ONE*. Edited by M. Salemi. Public Library of Science, 9(3), p. e90845. doi: 10.1371/journal.pone.0090845.

Jaskolski, M. *et al.* (2009) 'Piecing together the structure of retroviral integrase, an important target in AIDS therapy', *FEBS Journal*. John Wiley & Sons, Ltd (10.1111), 276(11), pp. 2926–2946. doi: 10.1111/j.1742-4658.2009.07009.x.

Ji, J. and Loeb, L. A. (1992) 'Fidelity of HIV-1 reverse transcriptase copying RNA in vitro', *Biochemistry*, 31(4), pp. 954–958. doi: 10.1021/bi00119a002.

Joffre, O. P. *et al.* (2012) 'Cross-presentation by dendritic cells', *Nature Reviews Immunology*. Nature Publishing Group, 12(8), pp. 557–569. doi: 10.1038/nri3254.

Kakuda, T. N. *et al.* (2014) 'Pharmacokinetics and pharmacodynamics of boosted once-daily darunavir', *Journal of Antimicrobial Chemotherapy*. Narnia, 69(10), pp. 2591–

2605. doi: 10.1093/jac/dku193.

Karade, S., Sen, S. and Sashindran, V. K. (2019) 'Absence of Integrase Strand Transfer Inhibitor Associated Resistance in Antiretroviral Therapy Naïve and Experienced Individuals from Western India.', *AIDS research and human retroviruses*, 35(6), pp. 567–571. doi: 10.1089/AID.2018.0272.

Kati, W. M. *et al.* (1992) 'Mechanism and fidelity of HIV reverse transcriptase.', *The Journal of biological chemistry*, 267(36), pp. 25988–97. Available at: http://www.ncbi.nlm.nih.gov/pubmed/1281479 (Accessed: 02 March 2020).

Keele, B. F. *et al.* (2006) 'Chimpanzee reservoirs of pandemic and nonpandemic HIV-1.', *Science (New York, N.Y.)*. NIH Public Access, 313(5786), pp. 523–6. doi: 10.1126/science.1126531.

Kessl, J. J. *et al.* (2016) 'HIV-1 Integrase Binds the Viral RNA Genome and Is Essential during Virion Morphogenesis.', *Cell*. NIH Public Access, 166(5), pp. 1257-1268.e12. doi: 10.1016/j.cell.2016.07.044.

Kharsany, A. B. M. *et al.* (2018) 'Community-based HIV prevalence in KwaZulu-Natal, South Africa: results of a cross-sectional household survey', *The Lancet HIV*, 5(8), pp. e427–e437. doi: 10.1016/S2352-3018(18)30104-8.

Klibanov, O. M. (2009) 'Elvitegravir, an oral HIV integrase inhibitor, for the potential treatment of HIV infection.', *Current opinion in investigational drugs (London, England : 2000)*, 10(2), pp. 190–200. Available at: http://www.ncbi.nlm.nih.gov/pubmed/19197797 (Accessed: 02 March 2020).

Kohlstaedt, L. *et al.* (1992) 'Crystal structure at 3.5 A resolution of HIV-1 reverse transcriptase complexed with an inhibitor', *Science*, 256(5065), pp. 1783–1790. doi: 10.1126/science.1377403.

Kondo, M. *et al.* (2018) 'Comparative evaluation of the Geenius HIV 1/2 Confirmatory Assay and the HIV-1 and HIV-2 Western blots in the Japanese population.', *PloS one*. Public Library of Science, 13(10), p. e0198924. doi: 10.1371/journal.pone.0198924.

Kontijevskis, A., Wikberg, J. E. S. and Komorowski, J. (2007) 'Computational proteomics analysis of HIV-1 protease interactome', *Proteins: Structure, Function, and Bioinformatics*. John Wiley & Sons, Ltd, 68(1), pp. 305–312. doi: 10.1002/prot.21415.

Korber, B. T. *et al.* (1998) *Numbering Positions in HIV Relative to HXB2CG*. Available at: https://www.hiv.lanl.gov/content/sequence/HIV/REVIEWS/HXB2.html (Accessed: 02 March 2020).

Koup, R. A. *et al.* (1994) 'Temporal association of cellular immune responses with the

initial control of viremia in primary human immunodeficiency virus type 1 syndrome.', *Journal of virology*. American Society for Microbiology Journals, 68(7), pp. 4650–5. Available at: http://www.ncbi.nlm.nih.gov/pubmed/8207839 (Accessed: 02 March 2020).

Kuciak, M. *et al.* (2008) 'The HIV-1 transcriptional activator Tat has potent nucleic acid chaperoning activities in vitro.', *Nucleic acids research*. Oxford University Press, 36(10), pp. 3389–400. doi: 10.1093/nar/gkn177.

Kuiken, C., Korber, B. and Shafer, R. W. (2003) 'HIV sequence databases', *AIDS Reviews*. doi: 10.1017/cbo9781139164764.007.

Kumar, S., Stecher, G. and Tamura, K. (2016) 'MEGA7: Molecular Evolutionary Genetics Analysis Version 7.0 for Bigger Datasets', *Molecular Biology and Evolution*, 33(7), pp. 1870–1874. doi: 10.1093/molbev/msw054.

Kwong, P. D. *et al.* (1998) 'Structure of an HIV gp120 envelope glycoprotein in complex with the CD4 receptor and a neutralizing human antibody.', *Nature*. NIH Public Access, 393(6686), pp. 648–59. doi: 10.1038/31405.

Laethem, K. Van *et al.* (2008) 'A genotypic assay for the amplification and sequencing of integrase from diverse HIV-1 group M subtypes', *Journal of Virological Methods*, 153, pp. 176–181. doi: 10.1016/j.jviromet.2008.07.008.

Langford, S. E., Ananworanich, J. and Cooper, D. A. (2007) 'Predictors of disease progression in HIV infection: a review.', *AIDS research and therapy*. BioMed Central, 4, p. 11. doi: 10.1186/1742-6405-4-11.

Lemey, P. *et al.* (2003) 'Tracing the origin and history of the HIV-2 epidemic.', *Proceedings of the National Academy of Sciences of the United States of America*. National Academy of Sciences, 100(11), pp. 6588–92. doi: 10.1073/pnas.0936469100.

Lesbats, P., Engelman, A. N. and Cherepanov, P. (2016) 'Retroviral DNA Integration', *Chemical Reviews*. American Chemical Society, 116(20), pp. 12730–12757. doi: 10.1021/acs.chemrev.6b00125.

Lewis, W., Day, B. J. and Copeland, W. C. (2003) 'Mitochondrial toxicity of nrti antiviral drugs: an integrated cellular perspective', *Nature Reviews Drug Discovery*. Nature Publishing Group, 2(10), pp. 812–822. doi: 10.1038/nrd1201.

Li, M. *et al.* (2006) 'Retroviral DNA integration: reaction pathway and critical intermediates', *The EMBO Journal*, 25(6), pp. 1295–1304. doi: 10.1038/sj.emboj.7601005.

Li, W. H., Tanimura, M. and Sharp, P. M. (1988) 'Rates and dates of divergence between AIDS virus nucleotide sequences.', *Molecular Biology and Evolution*. Narnia, 5(4), pp. 313–330. doi: 10.1093/oxfordjournals.molbev.a040503.

Liddament, M. T. *et al.* (2004) 'APOBEC3F Properties and Hypermutation Preferences Indicate Activity against HIV-1 In Vivo', *Current Biology*, 14(15), pp. 1385–1391. doi: 10.1016/j.cub.2004.06.050.

Little, S. J. *et al.* (1999) 'Viral dynamics of acute HIV-1 infection.', *The Journal of experimental medicine*. Rockefeller University Press, 190(6), pp. 841–50. doi: 10.1084/jem.190.6.841.

Liu, L. *et al.* (2019) 'Lack of HIV-1 integrase inhibitor resistance among 392 antiretroviral-naïve individuals in a tertiary care hospital in Beijing, China', *AIDS*, 33(12), pp. 1945–1947. doi: 10.1097/QAD.0000000000002282.

Liu, T. F. and Shafer, R. W. (2006) 'Web resources for HIV type 1 genotypic-resistance test interpretation.', *Clinical infectious diseases : an official publication of the Infectious Diseases Society of America*. NIH Public Access, 42(11), pp. 1608–18. doi: 10.1086/503914.

van der Loeff, M. *et al.* (2010) 'Undetectable plasma viral load predicts normal survival in HIV-2-infected people in a West African village', *Retrovirology*. BioMed Central, 7(1), p. 46. doi: 10.1186/1742-4690-7-46.

Los Alamos HIV sequence databse (1998) *Landmarks of the HIV genome*. Available at: https://www.hiv.lanl.gov/content/sequence/HIV/MAP/landmark.html (Accessed: 02 March 2020).

Lv, Z., Chu, Y. and Wang, Y. (2015) 'HIV protease inhibitors: a review of molecular selectivity and toxicity.', *HIV/AIDS (Auckland, N.Z.)*. Dove Press, 7, pp. 95–104. doi: 10.2147/HIV.S79956.

MacNeil, A. *et al.* (2007) 'Direct Evidence of Lower Viral Replication Rates In Vivo in Human Immunodeficiency Virus Type 2 (HIV-2) Infection than in HIV-1 Infection', *Journal of Virology*, 81(10), pp. 5325–5330. doi: 10.1128/JVI.02625-06.

Mahungu, T. W., Rodger, A. J. and Johnson, M. A. (2009) 'HIV as a chronic disease.', *Clinical medicine (London, England)*. Royal College of Physicians, 9(2), pp. 125–8. doi: 10.7861/clinmedicine.9-2-125.

Malet, I. *et al.* (2017) 'Mutations Located outside the Integrase Gene Can Confer Resistance to HIV-1 Integrase Strand Transfer Inhibitors', *mBio*. Edited by D. Paraskevis, 8(5). doi: 10.1128/mBio.00922-17.

Malim, M. H. and Emerman, M. (2008) 'HIV-1 Accessory Proteins—Ensuring Viral Survival in a Hostile Environment', *Cell Host & Microbe*. Cell Press, 3(6), pp. 388–398. doi: 10.1016/J.CHOM.2008.04.008.

Marcus, J. L. *et al.* (2016) 'Narrowing the Gap in Life Expectancy Between HIV-

Infected and HIV-Uninfected Individuals With Access to Care.', *Journal of acquired immune deficiency syndromes (1999)*. NIH Public Access, 73(1), pp. 39–46. doi: 10.1097/QAI.0000000000001014.

Margolis, A. M. *et al.* (2014) 'A review of the toxicity of HIV medications.', *Journal of medical toxicology : official journal of the American College of Medical Toxicology*. Springer, 10(1), pp. 26–39. doi: 10.1007/s13181-013-0325-8.

Marlink, R. *et al.* (1994) 'Reduced rate of disease development after HIV-2 infection as compared to HIV-1.', *Science (New York, N.Y.)*. American Association for the Advancement of Science, 265(5178), pp. 1587–90. doi: 10.1126/science.7915856.

Martin, A. M. *et al.* (2004) 'Predisposition to abacavir hypersensitivity conferred by HLA-B*5701 and a haplotypic Hsp70-Hom variant', *Proceedings of the National Academy of Sciences*, 101(12), pp. 4180–4185. doi: 10.1073/pnas.0307067101.

Martinez, E. *et al.* (2014) 'Early lipid changes with atazanavir/ritonavir or darunavir/ritonavir', *HIV Medicine*. John Wiley & Sons, Ltd (10.1111), 15(6), pp. 330–338. doi: 10.1111/hiv.12121.

Mateo-Urdiales, A. *et al.* (2019) 'Rapid initiation of antiretroviral therapy for people living with HIV.', *The Cochrane database of systematic reviews*. John Wiley and Sons, Inc. and the Cochrane Library, 6(6), p. CD012962. doi: 10.1002/14651858.CD012962.pub2.

McCaffrey, R. A. *et al.* (2004) 'N-linked glycosylation of the V3 loop and the immunologically silent face of gp120 protects human immunodeficiency virus type 1 SF162 from neutralization by anti-gp120 and anti-gp41 antibodies.', *Journal of virology*. American Society for Microbiology Journals, 78(7), pp. 3279–95. doi: 10.1128/jvi.78.7.3279-3295.2004.

McCulloch, S. D. and Kunkel, T. A. (2008) 'The fidelity of DNA synthesis by eukaryotic replicative and translesion synthesis polymerases', *Cell Research*. Nature Publishing Group, 18(1), pp. 148–161. doi: 10.1038/cr.2008.4.

Médecins Sans Frontières (MSF) (2019) *HIV project in South Africa reaches 90-90-90 target ahead of deadline | MSF*. Available at: https://www.msf.org/hiv-project-south-africa-reaches-90-90-90-target-ahead-deadline (Accessed: 02 March 2020).

Meek, T. D. *et al.* (1989) 'Human immunodeficiency virus 1 protease expressed in Escherichia coli behaves as a dimeric aspartic protease.', *Proceedings of the National Academy of Sciences of the United States of America*. National Academy of Sciences, 86(6), pp. 1841–5. doi: 10.1073/pnas.86.6.1841.

Melikian, G. L. *et al.* (2014) 'Non-nucleoside reverse transcriptase inhibitor (NNRTI)

cross-resistance: implications for preclinical evaluation of novel NNRTIs and clinical genotypic resistance testing.', *The Journal of antimicrobial chemotherapy*. Oxford University Press, 69(1), pp. 12–20. doi: 10.1093/jac/dkt316.

Mesplède, T. *et al.* (2013) 'Viral fitness cost prevents HIV-1 from evading dolutegravir drug pressure', *Retrovirology*. BioMed Central, 10(1), p. 22. doi: 10.1186/1742-4690-10-22.

Mesplède, T. *et al.* (2015) 'The R263K substitution in HIV-1 subtype C is more deleterious for integrase enzymatic function and viral replication than in subtype B', *AIDS*, 29(12), pp. 1459–1466. doi: 10.1097/QAD.0000000000000752.

Mohammadzadeh, N. *et al.* (2019) 'Role of co-expressed APOBEC3F and APOBEC3G in inducing HIV-1 drug resistance.', *Heliyon*. Elsevier, 5(4), p. e01498. doi: 10.1016/j.heliyon.2019.e01498.

Molina, J.-M. *et al.* (2012) 'Efficacy and safety of once daily elvitegravir versus twice daily raltegravir in treatment-experienced patients with HIV-1 receiving a ritonavir-boosted protease inhibitor: randomised, double-blind, phase 3, non-inferiority study', *The Lancet Infectious Diseases*. Elsevier, 12(1), pp. 27–35. doi: 10.1016/S1473-3099(11)70249-3.

Molina, J.-M. *et al.* (2015) 'Once-daily dolutegravir versus darunavir plus ritonavir for treatment-naive adults with HIV-1 infection (FLAMINGO): 96 week results from a randomised, open-label, phase 3b study', *The Lancet HIV*. Elsevier, 2(4), pp. e127–e136. doi: 10.1016/S2352-3018(15)00027-2.

Montessori, V. *et al.* (2004) 'Adverse effects of antiretroviral therapy for HIV infection.', *CMAJ : Canadian Medical Association journal = journal de l'Association medicale canadienne*. Canadian Medical Association, 170(2), pp. 229–38. Available at: http://www.ncbi.nlm.nih.gov/pubmed/14734438 (Accessed: 02 March 2020).

Mouscadet, J.-F. and Tchertanov, L. (2009) 'Raltegravir: molecular basis of its mechanism of action', *European Journal of Medical Research*, 14(Suppl 3), p. 5. doi: 10.1186/2047-783X-14-S3-5.

Msimanga, P., Vardas, E. and Engelbrecht, S. (2015) 'HIV-1 diversity in an antiretroviral treatment naïve cohort from Bushbuckridge, Mpumalanga Province, South Africa', *Virology Journal*. BioMed Central, 12(1), p. 24. doi: 10.1186/s12985-015-0244-1.

Mulu, A., Maier, M. and Liebert, U. G. (2015) 'Lack of integrase inhibitors associated resistance mutations among HIV-1C isolates', *Journal of Translational Medicine*. BioMed Central, 13(1), p. 377. doi: 10.1186/s12967-015-0734-3.

Munir, S. *et al.* (2015) 'G118R and F121Y mutations identified in patients failing raltegravir treatment confer dolutegravir resistance', *Journal of Antimicrobial Chemotherapy*,

70(3), pp. 739–749. doi: 10.1093/jac/dku474.

National Department of Health (2019) *Southern African HIV Clinicians Society*. Available at: https://sahivsoc.org/SubHeader?slug=ndoh-and-who-guidelines (Accessed: 02 March 2020).

National Institute for Communicable Diseases (17AD) *Annual Review - NICD*. Available at: http://www.nicd.ac.za/wp-content/uploads/2017/03/NICD_AR_2016_17.pdf (Accessed: 02 March 2020).

Neogi, U. *et al.* (2013) 'Human APOBEC3G-mediated hypermutation is associated with antiretroviral therapy failure in HIV-1 subtype C-infected individuals.', *Journal of the International AIDS Society*. Wiley-Blackwell, 16(1), p. 18472. doi: 10.7448/IAS.16.1.18472.

Norwood, J. *et al.* (2017) 'Brief Report: Weight Gain in Persons With HIV Switched From Efavirenz-Based to Integrase Strand Transfer Inhibitor-Based Regimens.', *Journal of acquired immune deficiency syndromes (1999)*. NIH Public Access, 76(5), pp. 527–531. doi: 10.1097/QAI.0000000000001525.

Nowak, M. A. *et al.* (1997) 'Anti-viral Drug Treatment: Dynamics of Resistance in Free Virus and Infected Cell Populations', *Journal of Theoretical Biology*. Academic Press, 184(2), pp. 203–217. doi: 10.1006/JTBI.1996.0307.

Nowak, M. G. *et al.* (2009) 'Identifying amino acid residues that contribute to the cellular-DNA binding site on retroviral integrase', *Virology*. Academic Press, 389(1–2), pp. 141–148. doi: 10.1016/J.VIROL.2009.04.014.

Nowotny, M. (2009) 'Retroviral integrase superfamily: the structural perspective', *EMBO reports*, 10(2), pp. 144–151. doi: 10.1038/embor.2008.256.

Nyamweya, S. *et al.* (2013) 'Comparing HIV-1 and HIV-2 infection: Lessons for viral immunopathogenesis', *Reviews in Medical Virology*. John Wiley & Sons, Ltd, 23(4), pp. 221–240. doi: 10.1002/rmv.1739.

O'Donovan, D. *et al.* (2000) 'Maternal plasma viral RNA levels determine marked differences in mother-to-child transmission rates ofHIV-1 and HIV-2 in The Gambia', *AIDS*, 14(4), pp. 441–448. doi: 10.1097/00002030-200003100-00019.

Oliveira, M. *et al.* (2014) 'Resistance mutations against dolutegravir in HIV integrase impair the emergence of resistance against reverse transcriptase inhibitors.', *AIDS (London, England)*. doi: 10.1097/QAD.0000000000000199.

Oliveira, M. F. *et al.* (2012) 'Genetic diversity and naturally polymorphisms in HIV type 1 integrase isolates from Maputo, Mozambique: implications for integrase inhibitors.', *AIDS research and human retroviruses*. Mary Ann Liebert, Inc., 28(12), pp. 1788–92. doi:

10.1089/aid.2012.0058.

Orkin, C. *et al.* (2013) 'Final 192-week efficacy and safety of once-daily darunavir/ritonavir compared with lopinavir/ritonavir in HIV-1-infected treatment-naïve patients in the ARTEMIS trial', *HIV Medicine*, 14(1), pp. 49–59. doi: 10.1111/j.1468-1293.2012.01060.x.

Pancera, M. *et al.* (2014) 'Structure and immune recognition of trimeric pre-fusion HIV-1 Env', *Nature*. Nature Publishing Group, 514(7523), pp. 455–461. doi: 10.1038/nature13808.

Papathanasopoulos, M. A. *et al.* (2010) 'Characterization of HIV Type 1 Genetic Diversity Among South African Participants Enrolled in the AIDS Vaccine Integrated Project (AVIP) Study', *AIDS Research and Human Retroviruses*. Mary Ann Liebert, Inc. 140 Huguenot Street, 3rd Floor New Rochelle, NY 10801 USA , 26(6), pp. 705–709. doi: 10.1089/aid.2009.0281.

Passaes, C. B. *et al.* (2009) 'Lack of primary mutations associated with integrase inhibitors among HIV-1 subtypes B, C, and F circulating in Brazil.', *Journal of acquired immune deficiency syndromes (1999)*, 51(1), pp. 7–12. doi: 10.1097/QAI.0b013e31819df3b3.

Paton, N. I. *et al.* (2014) 'Assessment of Second-Line Antiretroviral Regimens for HIV Therapy in Africa', *New England Journal of Medicine*, 371(3), pp. 234–247. doi: 10.1056/NEJMoa1311274.

Pau, A. K. and George, J. M. (2014) 'Antiretroviral therapy: current drugs.', *Infectious disease clinics of North America*. NIH Public Access, 28(3), pp. 371–402. doi: 10.1016/j.idc.2014.06.001.

Peeters, M., Aghokeng, A. . and Delaporte, E. (2010) 'Genetic diversity among human immunodeficiency virus-1 non-B subtypes in viral load and drug resistance assays', *Clinical Microbiology and Infection*. Elsevier, 16(10), pp. 1525–1531. doi: 10.1111/J.1469-0691.2010.03300.X.

Pena, M. J. *et al.* (2019) 'Virological Failure in HIV to Triple Therapy With Dolutegravir-Based Firstline Treatment: Rare but Possible', *Open Forum Infectious Diseases*. Narnia, 6(1). doi: 10.1093/ofid/ofy332.

Perelson, A. S. *et al.* (1996) 'HIV-1 dynamics in vivo: virion clearance rate, infected cell life-span, and viral generation time.', *Science (New York, N.Y.)*. American Association for the Advancement of Science, 271(5255), pp. 1582–6. doi: 10.1126/science.271.5255.1582.

Pilcher, C. D. *et al.* (2004) 'Brief but Efficient: Acute HIV Infection and the Sexual Transmission of HIV', *The Journal of Infectious Diseases*. Narnia, 189(10), pp. 1785–1792.

doi: 10.1086/386333.

Plantier, J. C. *et al.* (2009) 'A new human immunodeficiency virus derived from gorillas', *Nature Medicine*. doi: 10.1038/nm.2016.

Poignard, P. *et al.* (1999) 'Neutralizing Antibodies Have Limited Effects on the Control of Established HIV-1 Infection In Vivo', *Immunity*. Cell Press, 10(4), pp. 431–438. doi: 10.1016/S1074-7613(00)80043-6.

Pommier, Y., Johnson, A. A. and Marchand, C. (2005) 'Integrase inhibitors to treat HIV/Aids', *Nature Reviews Drug Discovery*. Nature Publishing Group, 4(3), pp. 236–248. doi: 10.1038/nrd1660.

Preston, B., Poiesz, B. and Loeb, L. (1988) 'Fidelity of HIV-1 reverse transcriptase', *Science*, 242(4882), pp. 1168–1171. doi: 10.1126/science.2460924.

Quashie, P. K. *et al.* (2012) '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*, 86(5), pp. 2696–2705. doi: 10.1128/JVI.06591-11.

Raffi, F. *et al.* (2013) 'Once-daily dolutegravir versus raltegravir in antiretroviral-naive adults with HIV-1 infection: 48 week results from the randomised, double-blind, non-inferiority SPRING-2 study', *The Lancet*. Elsevier, 381(9868), pp. 735–743. doi: 10.1016/S0140-6736(12)61853-4.

Ren, G. *et al.* (2007) 'Single-particle image reconstruction of a tetramer of HIV integrase bound to DNA.', *Journal of molecular biology*. NIH Public Access, 366(1), pp. 286–94. doi: 10.1016/j.jmb.2006.11.029.

Rhee, S.-Y. *et al.* (2007) 'Natural variation of HIV-1 group M integrase: Implications for a new class of antiretroviral inhibitors', *Retrovirology*. BioMed Central, 4(1), p. 21. doi: 10.1186/1742-4690-4-21.

Rhee, S.-Y. *et al.* (2008) 'Natural variation of HIV-1 group M integrase: Implications for a new class of antiretroviral inhibitors', *Retrovirology*. BioMed Central, 5(1), p. 74. doi: 10.1186/1742-4690-5-74.

Rhee, S.-Y. *et al.* (2019) 'A systematic review of the genetic mechanisms of dolutegravir resistance', *Journal of Antimicrobial Chemotherapy*. Narnia, 74(11), pp. 3135–3149. doi: 10.1093/jac/dkz256.

Rice, P. and Kiyoshi, M. (1995) 'Structure of the bacteriophage Mu transposase core: A common structural motif for DNA transposition and retroviral integration', *Cell*. Cell Press, 82(2), pp. 209–220. doi: 10.1016/0092-8674(95)90308-9.

Riddler, S. A. *et al.* (2008) 'Class-Sparing Regimens for Initial Treatment of HIV-1 Infection', *New England Journal of Medicine*. Massachusetts Medical Society , 358(20), pp. 2095–2106. doi: 10.1056/NEJMoa074609.

Rittweger, M. and Arasteh, K. (2007) 'Clinical Pharmacokinetics of Darunavir', *Clinical Pharmacokinetics*, 46(9), pp. 739–756. doi: 10.2165/00003088-200746090-00002.

Roberts, J. D., Bebenek, K. and Kunkel, T. A. (1988) 'The accuracy of reverse transcriptase from HIV-1.', *Science (New York, N.Y.)*. American Association for the Advancement of Science, 242(4882), pp. 1171–3. doi: 10.1126/science.2460925.

Robertson, D. L. (2000) 'HIV-1 Nomenclature Proposal', *Science*. doi: 10.1126/science.288.5463.55d.

Rockstroh, J. K. *et al.* (2013) 'Durable Efficacy and Safety of Raltegravir Versus Efavirenz When Combined With Tenofovir/Emtricitabine in Treatment-Naive HIV-1–Infected Patients', *JAIDS Journal of Acquired Immune Deficiency Syndromes*, 63(1), pp. 77–85. doi: 10.1097/QAI.0b013e31828ace69.

Rodgers, M. A. *et al.* (2017) 'Identification of rare HIV-1 Group N, HBV AE, and HTLV-3 strains in rural South Cameroon', *Virology*. doi: 10.1016/j.virol.2017.01.008.

Rogers, L. *et al.* (2018) 'Structural Implications of Genotypic Variations in HIV-1 Integrase From Diverse Subtypes.', *Frontiers in microbiology*. Frontiers Media SA, 9, p. 1754. doi: 10.3389/fmicb.2018.01754.

Saitou, N. and Nei, M. (1987) 'The neighbor-joining method: a new method for reconstructing phylogenetic trees.', *Molecular Biology and Evolution*. Narnia, 4(4), pp. 406–425. doi: 10.1093/oxfordjournals.molbev.a040454.

Saladini, F. *et al.* (2017) 'The HIV-1 integrase E157Q polymorphism per se does not alter susceptibility to raltegravir and dolutegravir in vitro.', *AIDS (London, England)*, 31(16), pp. 2307–2309. doi: 10.1097/QAD.0000000000001616.

Santos, A. F. and Soares, M. A. (2010) 'HIV Genetic Diversity and Drug Resistance', *Viruses*. Multidisciplinary Digital Publishing Institute (MDPI), 2(2), p. 503. doi: 10.3390/V2020503.

Sarafianos, Stefan G *et al.* (2009) 'Structure and function of HIV-1 reverse transcriptase: molecular mechanisms of polymerization and inhibition.', *Journal of molecular biology*. NIH Public Access, 385(3), pp. 693–713. doi: 10.1016/j.jmb.2008.10.071.

Sarafianos, Stefan G. *et al.* (2009) 'Structure and Function of HIV-1 Reverse Transcriptase: Molecular Mechanisms of Polymerization and Inhibition', *Journal of Molecular Biology*. Academic Press, 385(3), pp. 693–713. doi: 10.1016/J.JMB.2008.10.071.

Scherzer, R. *et al.* (2012) 'Association of tenofovir exposure with kidney disease risk in HIV infection.', *AIDS (London, England)*. NIH Public Access, 26(7), pp. 867–75. doi: 10.1097/QAD.0b013e328351f68f.

Shafer, R. W. (2006) 'Rationale and Uses of a Public HIV Drug-Resistance Database', *The Journal of Infectious Diseases*, 194(s1), pp. S51–S58. doi: 10.1086/505356.

Shah, B. M. *et al.* (2013) 'Cobicistat: A New Boost for the Treatment of Human Immunodeficiency Virus Infection', *Pharmacotherapy: The Journal of Human Pharmacology and Drug Therapy*. John Wiley & Sons, Ltd, 33(10), pp. 1107–1116. doi: 10.1002/phar.1237.

Sharp, P. M. and Hahn, B. H. (2008) 'Prehistory of HIV-1', *Nature*. Nature Publishing Group, 455(7213), pp. 605–606. doi: 10.1038/455605a.

Sharp, P. M. and Hahn, B. H. (2010) 'The evolution of HIV-1 and the origin of AIDS', *Philosophical Transactions of the Royal Society B: Biological Sciences*. The Royal Society, 365(1552), p. 2487. doi: 10.1098/RSTB.2010.0031.

Siemieniuk, R. A., Beckthold, B. and Gill, M. J. (2013) 'Increasing HIV subtype diversity and its clinical implications in a sentinel North American population.', *The Canadian journal of infectious diseases & medical microbiology = Journal canadien des maladies infectieuses et de la microbiologie medicale*. Hindawi Limited, 24(2), pp. 69–73. doi: 10.1155/2013/230380.

Singh, K. *et al.* (2019) *HIV-1 integrase polymorphisms and the affinity of first- and second-generation integrase strand transfer inhibitors (INSTIs) to diverse subtypes, HIV Nordic Conference* . Available at: https://hivnordic.se/p11-17-hiv-1-integrase-polymorphisms-and-the-affinity-of-first-and-second-generation-integrase-strand-transfer-inhibitors-instis-to-diverse-subtypes-2/ (Accessed: 02 March 2020).

Singh, O. and Su, E. C.-Y. (2016) 'Prediction of HIV-1 protease cleavage site using a combination of sequence, structural, and physicochemical features', *BMC Bioinformatics*. BioMed Central, 17(S17), p. 478. doi: 10.1186/s12859-016-1337-6.

Sluis-Cremer, N. and Tachedjian, G. (2008) 'Mechanisms of inhibition of HIV replication by nonnucleoside reverse transcriptase inhibitors', *Virus research*. NIH Public Access, 134(1–2), pp. 147–156. doi: 10.1016/J.VIRUSRES.2008.01.002.

Smith, S. J. *et al.* (2018) 'Efficacies of Cabotegravir and Bictegravir against drug-resistant HIV-1 integrase mutants', *Retrovirology*. BioMed Central, 15(1), p. 37. doi: 10.1186/s12977-018-0420-7.

Snedecor, S. J. *et al.* (2019) 'Comparative efficacy and safety of dolutegravir relative to common core agents in treatment-naïve patients infected with HIV-1: a systematic review and

network meta-analysis.', *BMC infectious diseases*. BioMed Central, 19(1), p. 484. doi: 10.1186/s12879-019-3975-6.

Soriano, V. *et al.* (2000) 'Human immunodeficiency virus type 2 (HIV-2) in Portugal: clinical spectrum, circulating subtypes, virus isolation, and plasma viral load.', *Journal of medical virology*, 61(1), pp. 111–6. Available at: http://www.ncbi.nlm.nih.gov/pubmed/10745242 (Accessed: 02 March 2020).

Southern African HIV Clinicians Society, G. *et al.* (2017) *Southern African journal of HIV medicine., Southern African Journal of HIV Medicine*. Available at: https://sajhivmed.org.za/index.php/hivmed/article/view/776/969 (Accessed: 02 March 2020).

Southern African HIV Clinicians Society, M. A. *et al.* (2018) *Southern African journal of HIV medicine., Southern African Journal of HIV Medicine*. Available at: https://sajhivmed.org.za/index.php/hivmed/article/view/917/1276 (Accessed: 02 March 2020).

South African National AIDS Council (2017) *The National Strategic Plan | SANAC*. Available at: https://sanac.org.za/the-national-strategic-plan/ (Accessed: 02 March 2020).

Strebel, K. (2013) 'HIV accessory proteins versus host restriction factors.', *Current opinion in virology*. NIH Public Access, 3(6), pp. 692–9. doi: 10.1016/j.coviro.2013.08.004.

Swanson, P., Devare, S. G. and Hackett, J. (2003) 'Molecular Characterization of 39 HIV-1 Isolates Representing Group M (Subtypes A-G) and Group O: Sequence Analysis of *gag* p24, *pol* Integrase, and *env* gp41', *AIDS Research and Human Retroviruses*. Mary Ann Liebert, Inc. , 19(7), pp. 625–629. doi: 10.1089/088922203322231003.

Tang, M. W. and Shafer, R. W. (2012) 'HIV-1 antiretroviral resistance: scientific principles and clinical applications.', *Drugs*. Springer, 72(9), pp. e1-25. doi: 10.2165/11633630-000000000-00000.

Tantillo, C. *et al.* (1994) 'Locations of Anti-AIDS Drug Binding Sites and Resistance Mutations in the Three-dimensional Structure of HIV-1 Reverse Transcriptase', *Journal of Molecular Biology*, 243(3), pp. 369–387. doi: 10.1006/jmbi.1994.1665.

Taylor, B. S. *et al.* (2008) 'The challenge of HIV-1 subtype diversity.', *The New England journal of medicine*. NIH Public Access, 358(15), pp. 1590–602. doi: 10.1056/NEJMra0706737.

Tebit, D. M. *et al.* (2016) 'HIV-1 Group O Genotypes and Phenotypes: Relationship to Fitness and Susceptibility to Antiretroviral Drugs', *AIDS Research and Human Retroviruses*. Mary Ann Liebert, Inc., 32(7), p. 676. doi: 10.1089/AID.2015.0318.

Thierry, E., Deprez, E. and Delelis, O. (2017) 'Different Pathways Leading to Integrase Inhibitors Resistance', *Frontiers in Microbiology*. Frontiers, 7, p. 2165. doi:

10.3389/fmicb.2016.02165.

Thierry, S. *et al.* (2015) 'Integrase inhibitor reversal dynamics indicate unintegrated HIV-1 dna initiate de novo integration', *Retrovirology*, 12(1), p. 24. doi: 10.1186/s12977-015-0153-9.

Torian, L. V. *et al.* (2010) 'HIV Type 2 in New York City, 2000–2008', *Clinical Infectious Diseases*, 51(11), pp. 1334–1342. doi: 10.1086/657117.

Tsiang, M. *et al.* (2016) 'Antiviral Activity of Bictegravir (GS-9883), a Novel Potent HIV-1 Integrase Strand Transfer Inhibitor with an Improved Resistance Profile.', *Antimicrobial agents and chemotherapy*. American Society for Microbiology (ASM), 60(12), pp. 7086–7097. doi: 10.1128/AAC.01474-16.

UNAIDS (2017) *New high-quality antiretroviral therapy to be launched in South Africa, Kenya and over 90 low-and middle-income countries at reduced price | UNAIDS*. Available at: https://www.unaids.org/en/resources/presscentre/pressreleaseandstatementarchive/2017/september/20170921_TLD (Accessed: 02 March 2020).

UNAIDS (2019) *South Africa country report 2018 | UNAIDS*. Available at: https://www.unaids.org/en/regionscountries/countries/southafrica (Accessed: 02 March 2020).

UNAIDS and Aidsinfo (2019) *AIDSinfo|UNAIDS*. Available at: https://aidsinfo.unaids.org/ (Accessed: 02 March 2020).

Unitaid (2018) *Expanding HIV self-testing in Africa - Unitaid*. Available at: https://unitaid.org/project/self-testing-africa-star/#en (Accessed: 02 March 2020).

Vannappagari, V. and Thorne, C. (2019) 'Pregnancy and Neonatal Outcomes Following Prenatal Exposure to Dolutegravir', *JAIDS Journal of Acquired Immune Deficiency Syndromes*, 81(4), pp. 371–378. doi: 10.1097/QAI.0000000000002035.

Varghese, V. *et al.* (2010) 'HIV-1 integrase sequence variability in antiretroviral naïve patients and in triple-class experienced patients subsequently treated with raltegravir.', *AIDS research and human retroviruses*. Mary Ann Liebert, Inc., 26(12), pp. 1323–6. doi: 10.1089/aid.2010.0123.

Venter, W. D. F. *et al.* (2019) 'Dolutegravir plus Two Different Prodrugs of Tenofovir to Treat HIV', *New England Journal of Medicine*, 381(9), pp. 803–815. doi: 10.1056/NEJMoa1902824.

Viani, R. M. *et al.* (2015) 'Safety, Pharmacokinetics and Efficacy of Dolutegravir in Treatment-experienced HIV-1 Infected Adolescents: Forty-eight-week Results from IMPAACT P1093.', *The Pediatric infectious disease journal*. NIH Public Access, 34(11), pp.

1207–13. doi: 10.1097/INF.0000000000000848.

Wainberg, M. *et al.* (2014) 'The R263K mutation in HIV integrase that is selected by dolutegravir may actually prevent clinically relevant resistance to this compound', *Journal of the International AIDS Society*, 17(4 Suppl 3), p. 19518. doi: 10.7448/IAS.17.4.19518.

Wainberg, M. A. and Brenner, B. G. (2012) 'The Impact of HIV Genetic Polymorphisms and Subtype Differences on the Occurrence of Resistance to Antiretroviral Drugs', *Molecular Biology International*. Hindawi, 2012, pp. 1–10. doi: 10.1155/2012/256982.

Wang, J. Y. *et al.* (2001) 'Structure of a two-domain fragment of HIV-1 integrase: implications for domain organization in the intact protein.', *The EMBO journal*. European Molecular Biology Organization, 20(24), pp. 7333–43. doi: 10.1093/emboj/20.24.7333.

Wares, M. *et al.* (2014) '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*. BioMed Central, 11, p. 7. doi: 10.1186/1742-4690-11-7.

Weber, I. *et al.* (1989) 'Molecular modeling of the HIV-1 protease and its substrate binding site', *Science*. American Association for the Advancement of Science, 243(4893), pp. 928–931. doi: 10.1126/science.2537531.

Wei, X. *et al.* (2003) 'Antibody neutralization and escape by HIV-1', *Nature*. Nature Publishing Group, 422(6929), pp. 307–312. doi: 10.1038/nature01470.

Wensing, A. M. J., van Maarseveen, N. M. and Nijhuis, M. (2010) 'Fifteen years of HIV Protease Inhibitors: raising the barrier to resistance', *Antiviral Research*. Elsevier, 85(1), pp. 59–74. doi: 10.1016/J.ANTIVIRAL.2009.10.003.

Wheeler, W. H. *et al.* (2010) 'Prevalence of transmitted drug resistance associated mutations and HIV-1 subtypes in new HIV-1 diagnoses, U.S.–2006', *AIDS*, 24(8), pp. 1203–1212. doi: 10.1097/QAD.0b013e3283388742.

White, K. L. *et al.* (2017) 'BICTEGRAVIR DISSOCIATION HALF-LIFE FROM HIV-1 G140S+Q148H INTEGRASE/DNA COMPLEXES | CROI Conference', in *BICTEGRAVIR DISSOCIATION HALF-LIFE FROM HIV-1 G140S+Q148H INTEGRASE/DNA COMPLEXES*. Available at: http://www.croiconference.org/sessions/bictegravir-dissociation-half-life-hiv-1-g140sq148h-integrasedna-complexes (Accessed: 02 March 2020).

WHO (2014) *Fast-Track - Ending the AIDS epidemic by 2030 | UNAIDS*. Available at: https://www.unaids.org/en/resources/documents/2014/JC2686_WAD2014report (Accessed:

02 March 2020).

‘**WHO** | Update of recommendations on first- and second-line antiretroviral regimens’ (2019) *WHO*. World Health Organization. Available at: https://www.who.int/hiv/pub/arv/arv-update-2019-policy/en/ (Accessed: 02 March 2020).

Williams and Sinko (1999) ‘Oral absorption of the HIV protease inhibitors: a current update.’, *Advanced drug delivery reviews*, 39(1–3), pp. 211–238. Available at: http://www.ncbi.nlm.nih.gov/pubmed/10837775 (Accessed: 02 March 2020).

Worobey, M. *et al.* (2008) ‘Direct evidence of extensive diversity of HIV-1 in Kinshasa by 1960’, *Nature*, 455(7213), pp. 661–664. doi: 10.1038/nature07390.

Xue, W., Liu, H. and Yao, X. (2012) ‘Molecular mechanism of HIV-1 integrase-vDNA interactions and strand transfer inhibitor action: A molecular modeling perspective’, *Journal of Computational Chemistry*. John Wiley & Sons, Ltd, 33(5), pp. 527–536. doi: 10.1002/jcc.22887.

Yamaguchi, J. *et al.* (2019) ‘Complete genome sequence of CG-0018a-01 establishes HIV-1 subtype L’, *JAIDS Journal of Acquired Immune Deficiency Syndromes*, p. 1. doi: 10.1097/QAI.0000000000002246.

Yoder, K. E. and Bushman, F. D. (2000) ‘Repair of gaps in retroviral DNA integration intermediates.’, *Journal of virology*. American Society for Microbiology (ASM), 74(23), pp. 11191–200. doi: 10.1128/jvi.74.23.11191-11200.2000.

Young, B. *et al.* (2011) ‘Transmission of integrase strand-transfer inhibitor multidrug-resistant HIV-1: case report and response to raltegravir-containing antiretroviral therapy’, *Antiviral Therapy*, 16(2), pp. 253–256. doi: 10.3851/IMP1748.

Zanger, U. M. and Schwab, M. (2013) ‘Cytochrome P450 enzymes in drug metabolism: Regulation of gene expression, enzyme activities, and impact of genetic variation’, *Pharmacology & Therapeutics*. Pergamon, 138(1), pp. 103–141. doi: 10.1016/J.PHARMTHERA.2012.12.007.

Zash, R. *et al.* (2018) ‘Comparative safety of dolutegravir-based or efavirenz-based antiretroviral treatment started during pregnancy in Botswana: an observational study.’, *The Lancet. Global health*. Elsevier, 6(7), pp. e804–e810. doi: 10.1016/S2214-109X(18)30218-3.

Zazzi, M., Hu, H. and Prosperi, M. (2018) ‘The global burden of HIV-1 drug resistance in the past 20 years.’, *PeerJ*. PeerJ, Inc, 6, p. e4848. doi: 10.7717/peerj.4848.

Zeldin, R. K. and Petruschke, R. A. (2003) ‘Pharmacological and therapeutic properties of ritonavir-boosted protease inhibitor therapy in HIV-infected patients’, *Journal of Antimicrobial Chemotherapy*. Narnia, 53(1), pp. 4–9. doi: 10.1093/jac/dkh029.

Zhao, G. *et al.* (2013) ‘Mature HIV-1 capsid structure by cryo-electron microscopy and all-atom molecular dynamics’, *Nature*. Nature Publishing Group, 497(7451), pp. 643–646. doi: 10.1038/nature12162.

Zhu, J. and Paul, W. E. (2008) ‘CD4 T cells: fates, functions, and faults’, *Blood*. American Society of Hematology, 112(5), pp. 1557–1569. doi: 10.1182/blood-2008-05-078154.

Zhu, T. *et al.* (1998) ‘An African HIV-1 sequence from 1959 and implications for the origin of the epidemic’, *Nature*, 391(6667), pp. 594–597. doi: 10.1038/35400.

Zolopa, A. *et al.* (2013) ‘A Randomized Double-Blind Comparison of Coformulated Elvitegravir/Cobicistat/Emtricitabine/Tenofovir Disoproxil Fumarate Versus Efavirenz/Emtricitabine/Tenofovir Disoproxil Fumarate for Initial Treatment of HIV-1 Infection’, *JAIDS Journal of Acquired Immune Deficiency Syndromes*, 63(1), pp. 96–100. doi: 10.1097/QAI.0b013e318289545c.

8 Acknowledgments

I would like to express my deepest gratitude to my supervisors Dr. Graeme Brendon Jacobs and Prof. Dr. Carsten Scheller, for the patience, trust and encouragement I experienced throughout the course of this project and for granting me the opportunity to pursue my own ideas. Without their enthusiasm upon first contact, this project would not have been feasible. I would also like to thank Prof. Dr. Thomas Dandekar and Prof. Dr. Hartwig Klinker, for supervising this thesis and offering spontaneous help throughout the course of this thesis.

I would like to thank Prof. Dr. Wolfgang Preiser for allowing me to stay at the Division of Medical Virology at the Stellenbosch University, and to use their laboratory facilities to conduct my research. My special thanks goes to Dr. Adetayo Emmanuel Adegbenga Obasa, for his laboratory training and his willingness to help, whenever I needed assistance with my experiments. I would like to thank all past and present members of the Division of Medical Virology (research and diagnostics) for their hospitality, friendliness and their ability to effortlessly integrate and welcome me into their team. Such a stimulating and encouraging work environment is all new members can hope for.

I would like to appreciate the help of Dr. Gabriele Blum-Oehler and the Graduate School of Life Sciences (GSLS), who from the beginning encouraged me to pursue own ideas and whose cooperation and structural flexibility rendered this project possible.

For funding, I would like to thank the GSLS, the Stellenbosch University as well as the National Health Laboratory Service (NHLS) Research Trust.

I would like to express my sincere gratitude to my parents, my brother and Miss Anke Joubert, who supported me throughout all stages of this project and who at all times motivated me to believe in my work.