

**Investigating the development and evolution of
drug resistance in the HIV-1 *pol* gene**

By Anne Supang Martin

University College London

Thesis submitted for the degree of Doctor of Philosophy

I, Anne Supang Martin confirm that the work presented in this thesis is my own. Where information has been derived from other sources, I confirm that this had been indicated in the thesis.

Abstract

The prognosis of those infected with HIV-1 has improved significantly since the introduction of highly active antiretroviral therapy (HAART). This has led to complete suppression of HIV-1 replication and reduction of viraemia to undetectable levels. Initially HAART comprised of three classes of drugs, namely nucleoside reverse transcriptase inhibitors (NRTIs), non-nucleoside reverse transcriptase inhibitors (NNRTIs) and protease inhibitors (PIs), which target two important viral enzymes. Until recently, patients developing highly drug-resistant HIV-1 have had limited therapy options. This has changed in the last few years with the development and approval for use of second-generation NNRTIs and PIs, and three new classes of drugs including integrase strand transfer inhibitors (INSTIs). This means that patients undergoing INSTI-containing salvage therapy are taking drugs targeting all three HIV-1 *pol* genes; *protease (PR)*, *RT* and *IN*. However, little data is available on the evolution, interaction and linkage of drug resistance mutations throughout the *pol* gene. In this study, we developed a single genome sequencing assay of the full-length HIV-1 *pol* gene and used it to investigate the development and linkage of drug resistance mutations in sequential samples from two patients failing INSTI-containing salvage therapy. Different phylogenetic methods were used to explore the evolution and dynamics of drug resistance mutations in the full-length HIV-1 *pol* gene. Furthermore, we examined the effect of co-evolved *PR* and *RT* on the susceptibility of patient-derived viruses to INSTIs and viral replicative fitness. Our data indicate that the development of drug resistance mutations in *IN* is complex and is a fine balance between attaining high levels of drug resistance and decent replicative fitness. This is to a degree influenced by mutations in other regions of the HIV-1 *pol* gene. Taken together, the data suggests that larger regions of patient-derived HIV-1 genome should be examined in order to get a good understanding of HIV-1 drug susceptibility.

Dedicated to Prof. Peter W. Martin (1949-2009)

Dad, as a truly brilliant academic you have always been in my thoughts throughout my research; always missed, always loved and never forgotten.

Acknowledgments

Thank you to the Health Protection Agency (now Public Health England; PHE) for funding my PhD and the opportunity to travel to conferences. Thanks also go to Deenan Pillay (UCL Medical School) and Pat Cane (PHE) for their supervision and expertise.

I am particularly grateful for the supervision provided by Tamyo Mbisa (PHE). Your patience and relaxed attitude are something to be admired. Thank you so much for all the time you have put in to these past 4 years and thank you to your lovely wife and children for putting up with the extra work you must've taken home over the years!

Thanks to all of the AVU team, both past and present for all their technical support and friendship and for creating such a nice atmosphere to work in, in particular Chris Parry for our early morning chats and Katherine Sutherland for keeping me sane (a term I use subjectively). Kat, thanks for the sympathetic ear and all the distractions when things went wrong (and when they didn't). I will cherish our plasticine models and hazy memories of tequila nights!

A big thank you also goes to the phylo experts, Stephane Hue (UCL), Richard Myers (PHE) and Keiran Lythgow (PHE) who played an integral role in Chapter 4 of my thesis. Stephane, I appreciate all the time you spent teaching me, helping me generate and interpret data and providing your expert opinion. It can't have been easy dealing with the constant requests for data and answers!

Many thanks also to Vinnay Pathak and his team at the NCI in Fredrick for welcoming me into their lab for a month during my travel fellowship as well as providing me with the pHL(WT) vector. Eleanor Gray (UCL), thank you for providing me with technical advice and primers (including 5095-) to try and amplify my single genomes when they weren't playing nicely and Bridget Ferns (UCL) for providing me with the patient samples used for the single genome sequencing analysis. I must also thank John Poh (PHE) for his technical advice, which I admit may have fallen on deaf ears some of the time (sorry! Though I'm sure you knew), thank you also for your friendship.

Huge thanks to all my family for their support throughout the years. Michael, thank you for putting up with my doom and gloom tirades, you would've thought the end of the world was coming! I appreciate all the advice you gave and for helping me put things in perspective. The

Sankey clan and all my friends, thanks for the emotional support, sorry if I ranted a bit too often and for too long! Special mention to Hugh Kikuchi who had to put up with I don't know how many hours of whining during his days at the HPA and even now. Thanks for the trips to the -30 and for the entertaining stories post-tequila!

I suppose I ought to thank/blame Clare Soares for convincing me to undertake this PhD/ordeal (delete as appropriate!). Your pearls of wisdom and streams of emails have more than kept me entertained and on track for the past 4 years. Our pact; "no quitting unless the other does" seems to have done the trick...now it's your turn, don't think that now I've finished the pact is void! If I can do it, so can you. I believe in you (you are the wind beneath my wings and all that), if in doubt though, take a pigeon into the viva.

Last, but not least, Will, my long suffering boyfriend, who I'm sure will be just as glad as me to draw a line under this. I know I don't say this enough but I appreciate everything you have done for me, you have been my rock. Thank you for putting up with the multitude of mood swings- your patience and understanding was, is, unfaltering.

Contents

Abstract	3
Acknowledgments	5
Contents	7
Figures	16
Tables	19
CHAPTER 1	22
General introduction	22
1.1 Implications of HIV/AIDS	23
1.1.1 Global health and financial impact of the AIDS pandemic	23
1.1.2 Course of infection	23
1.2 The history, evolution and diversity of HIV	25
1.2.1 HIV-1 diversity	25
1.2.2 HIV-1 subtype distribution	26
1.3 HIV-1 genome structure	28
1.4 Viral proteins	29
1.4.1 Reverse transcriptase	29
1.4.2 Integrase	30
1.4.3 Protease	31
1.5 HIV-1 virion structure	34
1.6 The lifecycle of HIV	35
1.6.1 Receptor binding and entry	36

1.6.2	Reverse transcription	36
1.6.3	Nuclear transport and integration	38
1.6.4	Transcription, translation and transport of viral macromolecules	40
1.6.5	Virion release and maturation	41
1.7	Antiretroviral drugs	41
1.7.1	Fusion and entry inhibitors	42
1.7.2	Reverse transcriptase inhibitors	43
1.7.3	Integrase inhibitors	43
1.7.4	Protease inhibitors	44
1.8	Drug resistance	44
1.8.1	Resistance to reverse transcriptase inhibitors	45
1.8.1.1	NRTI resistance mechanisms	46
1.8.1.2	NNRTI resistance mechanisms	46
1.8.1.3	Resistance mechanisms of mutations in the C terminal region of reverse transcriptase	47
1.8.2	Resistance to raltegravir	52
1.8.2.1	The Q148R/H/K pathway	53
1.8.2.2	The N155H Pathway	55
1.8.2.3	The Y143R/C Pathway	57
1.8.3	Resistance to elvitegravir	58
1.8.4	Resistance to protease inhibitors	59
1.8.5	Resistance to other antiretrovirals	59
1.8.6	Transmitted drug resistance	60
1.9	The future of HIV-1 treatment	61
1.9.1	Problems of HIV eradication	62

1.9.2	Novel therapeutic strategies	63
1.9.3	Novel mechanisms of reverse transcriptase inhibition	64
1.9.4	Other inhibitors of integration	65
1.9.5	Maturation inhibitors	65
1.10	Measuring drug resistance and viral fitness	66
1.10.1	Phenotyping	66
1.10.2	Genotyping	68
1.10.3	Clinical significance of minority viral variants	70
1.10.4	Clinical significance of the linkage of drug resistance mutations	71
1.10.5	Clinical significance of HIV-1 diversity and fitness	72
1.11	Virus evolution	74
1.12	Project outline	78
CHAPTER 2		79
General materials and methods		79
2.1	Clinical samples	79
2.2	Molecular biology techniques	80
2.2.1	Retroviral expression vectors and plasmids	80
2.2.1.1	Molecular clones	80
2.2.1.2	p8.9NSX	80
2.2.1.3	pHL(WT)	80
2.2.1.4	pCSFLW	81
2.2.1.5	pMDG	81
2.2.2	Primer design	83
2.2.3	Luria-Bertani agar plates	83

2.2.4	Viral RNA extraction	84
2.2.5	Amplification of genomic viral RNA using polymerase chain reaction (PCR)	84
2.2.5.1	Denaturation of viral RNA	84
2.2.5.2	cDNA synthesis (reverse transcription-PCR) and first round PCR reactions	85
2.2.5.3	Nested PCR reactions	86
2.2.5.4	Addition of restriction enzyme site by PCR	87
2.2.6	Agarose gel electrophoresis and PCR product purification	88
2.2.7	Gel extraction and purification of DNA	88
2.2.8	DNA quantification	89
2.2.9	DNA Sequencing	89
2.2.10	Cloning	90
2.2.10.1	Restriction enzyme digestion	90
2.2.10.2	Dephosphorylation and ligation reactions	90
2.2.10.3	TOPO TA cloning [®]	91
2.2.10.4	Transformation of HB101 competent cells	91
2.2.10.5	Transformation of XL10-Gold ultracompetent cells	92
2.2.10.6	Transformation of One Shot [®] TOP10 chemically competent cells	92
2.2.10.7	Colony PCR	92
2.2.10.8	Mini-preps of plasmid DNA	93
2.2.10.9	Multi and single site-directed mutagenesis	93
2.2.11	Construction of <i>gag-pol</i> expression vectors for use in integrase inhibitor phenotypic susceptibility and replication capacity assays	94
2.2.11.1	Construction of p8.9NSXClal+vector	94
2.2.11.2	Construction of patient-derived <i>IN</i> expression vectors	95

2.2.11.3	Construction of patient-derived <i>PR+RT</i> expression vector (PR+RTp8.9NSXClal+)	96
2.2.11.4	Construction of patient-derived full-length <i>pol</i> expression vectors	96
2.2.12	Construction of wild-type subtype F and patient-derived subtype F <i>PR+RT</i> expression vectors	98
2.2.12.1	Construction of p8.9NSX-F	100
2.2.12.2	Construction of patient-derived subtype F <i>PR+RT</i> expression vectors in a subtype F specific vector	101
2.2.12.3	Construction of wild-type subtype F and patient-derived subtype F <i>PR+RT</i> expression vectors in a wild-type subtype B background	103
2.2.12.4	Construction of wild-type subtype F and patient-derived subtype F <i>RT</i> -containing retroviral vectors	103
2.2.13	Single genome sequencing	105
2.2.14	Long cDNA single genome sequencing	106
2.3	Cell culture	107
2.3.1	Cells	107
2.3.2	Virus Production	107
2.3.2.1	Transfection with PEI	108
2.3.2.2	Transfection with FuGENE6	108
2.3.3	Single-replication cycle drug susceptibility assay	108
2.3.3.1	Pseudovirus titration	110
2.3.3.2	Integrase strand transfer inhibitor phenotypic susceptibility assay	110
2.3.3.3	EC ₅₀ calculation	111
2.3.4	Replication capacity assay	112
2.3.4.1	p24 Enzyme-linked immunosorbent assay (ELISA)	113
2.4	Protein analysis	114
2.4.1	Antibodies	114

2.4.2	Western blotting	114
2.4.2.1	Band density quantification	115
2.5	Phylogenetics and bioinformatics	115
2.5.1	Sequence assembly	115
2.5.2	Sequence alignment	116
2.5.2.1	Single genome sequences	116
2.5.2.2	Subtype F sequences	116
2.5.3	Database analysis for C terminal mutations	116
2.5.4	Phylogenetic reconstruction of inpatient viral evolution	117
2.5.5	Inferring the genetic diversity of the HIV-1 <i>pol</i> gene of patients A and B	117
2.5.6	Analysis of inpatient viral evolution and phylodynamics using the coalescent theory	118
2.5.6.1	Bayesian Skyline reconstruction	118
2.5.7	Analysis of selection pressure on the HIV-1 <i>pol</i> gene of patients A and B	119
2.6	Statistics	120
2.6.1	Subtype F C terminal of <i>reverse transcriptase</i> analysis	120
2.6.2	Pairwise genetic distance	120
2.6.3	Phenotypic and replication capacity assays	120
CHAPTER 3		121
Investigation of the development and genetic linkage of drug resistance mutations in full-length <i>pol</i> gene of patients on RAL-containing salvage therapy using single genome sequencing		121
3.1	Introduction	121
3.2	Results	122

3.2.1	Development of a single genome sequencing assay encompassing full length <i>pol</i>	122
3.2.2	Clinical details of patients A and B	124
3.2.3	Single genome sequencing of patient A and B samples	125
3.2.4	Development and linkage of RAL resistance mutations in patient A	128
3.2.5	Development and linkage of RAL resistance mutations in patient B	129
3.2.6	Development of PI and RTI resistance mutations and genetic linkage to RAL resistance mutations in patient A and B	132
3.2.7	Presence of other polymorphism in <i>IN</i> , <i>PR</i> and <i>RT</i>	133
3.3	Discussion	135
CHAPTER 4		140
	Investigation of inpatient viral evolution during the development of resistance to RAL salvage therapy	140
4.1	Introduction	140
4.2	Results	141
4.2.1	Analysis of inpatient viral population dynamics during development of resistance to RAL salvage therapy	141
4.2.2	Analysis of inpatient viral genetic diversity during RAL salvage therapy	145
4.2.3	Analysis of drug selective pressure on the HIV-1 <i>pol</i> gene during development of resistance to RAL salvage therapy	155
4.3	Discussion	166
CHAPTER 5		170
	Inpatient relationship between viral drug susceptibility and replication fitness in the development of resistance to RAL therapy	170

5.1	Introduction	170
5.2	Results	171
5.2.1	Development of an integrase inhibitor phenotypic drug susceptibility assay	171
5.2.2	Construction of patient-derived viral vectors	176
5.2.3	Effects of patient-derived <i>IN</i> on RAL susceptibility	178
5.2.4	Effects of patient-derived full-length <i>pol</i> on RAL susceptibility	179
5.2.5	Effects of patient-derived <i>IN</i> on EVG susceptibility	182
5.2.6	Effects of patient-derived full-length <i>pol</i> on EVG susceptibility	185
5.2.7	Effects of patient-derived <i>pol</i> gene fragments on viral replication fitness	185
5.3	Discussion	189
CHAPTER 6		195
The role of the C terminal region of reverse transcriptase in drug resistance of HIV-1 Subtype F		195
6.1	Introduction	195
6.2	Results	196
6.2.1	Identification of mutations in the C terminal region of <i>RT</i> of HIV-1 subtype F associated with primary resistance mutations	196
6.2.2	Phenotypic characterization of C terminal <i>RT</i> mutations identified in subtype F to be associated with drug resistance	200
6.2.3	Analysis of viral protein expression and proteolytic processing of subtype F expressing recombinant viruses	202
6.3	Discussion	205
CHAPTER 7		209
General discussion and future work		209

7.1	Drug resistance in the era of HAART	209
7.1.1	The single genome sequencing assay	212
7.1.2	Analysis of inpatient viral evolution and the role of recombination	213
7.1.3	Phenotypic drug susceptibility cut offs	214
7.1.4	Single-cycle replication assay limitations	214
7.2	Drug resistance in the C terminal of <i>RT</i>	215
7.3	Future work	215
	REFERENCES	218

Figures

CHAPTER 1

Figure 1.1	Natural Course of HIV-1 Infection	24
Figure 1.2	a) HIV-1 subtype and contribution to the pandemic and b) Their distribution worldwide between 2004 and 2007	27
Figure 1.3	Schematic of HIV-1 genome RNA	28
Figure 1.4	Schematic representation of the HIV-1 virion	34
Figure 1.5	Schematic of the HIV-1 replication cycle	35
Figure 1.6	Schematic of the reverse transcriptase process	37
Figure 1.7	HIV-1 DNA integration	39
Figure 1.8	Mutations in <i>RT</i> associated with RTI resistance	45
Figure 1.9	An RNase H-dependent C terminal NRTI resistance mechanism	48
Figure 1.10	An RNase H-dependent C terminal NNRTI resistance mechanism	49

CHAPTER 2

Figure 2.1	a) The p8.9NSX <i>gag-pol</i> expression vector b) Unique restriction sites present in p8.9NSX <i>gag-pol</i> region that were used for cloning	81
Figure 2.2	a) The pHL(WT) HIV vector b) Unique restriction sites present in the pHL(WT) <i>pol</i> gene that can be used for cloning	82
Figure 2.3	<i>gag-pol</i> expression vectors expressing patient-derived <i>IN</i> , <i>PR+RT</i> or full-length <i>pol</i> a) map of p8.9NSXClal+ b) schematic diagram of <i>gag-pol</i> region of vectors expressing patient-derived <i>IN</i> only c) schematic diagram of <i>gag-pol</i> region of vector expressing patient-derived <i>PR+RT</i> (PR+RTp8.9NSXClal+)	

	d) schematic diagram of vector expressing patient-derived full-length <i>pol</i> gene	97
Figure 2.4	a) p8.9NSX-F b) <i>gag-pol</i> expression vectors containing patient-derived subtype F <i>PR+RT</i> in a wild-type subtype F background c) <i>gag-pol</i> expression vectors containing either wild-type subtype F or patient-derived subtype F <i>PR+RT</i> in a wild-type subtype B background	99
Figure 2.5	a) Schematic diagram of <i>gag-pol</i> region of p8.9NSX expressing wild-type subtype F and patient-derived <i>RT</i> only b) schematic diagram of <i>gag-pol</i> region of pHL(WT) expressing wild-type subtype F and patient-derived <i>RT</i> only	104
Figure 2.6	Single cycle replication assay	109
Figure 2.7	Schematic of the set up of pseudovirus titration	111
Figure 2.8	Schematic of the set up of the INSTI phenotypic susceptibility assay	112

CHAPTER 3

Figure 3.1	Patient viral load, CD4 count and antiretroviral therapy data	126
Figure 3.2	Viral variants present in a) patient A and b) patient B and the percent of single genomes they represent at each time point	130

CHAPTER 4

Figure 4.1	Maximum likelihood tree of viral single genome data from patient A	143
Figure 4.2	Maximum likelihood tree of viral single genome data from patient B	144
Figure 4.3	Pairwise genetic distances of the whole <i>pol</i> gene	146
Figure 4.4	Pairwise genetic distances of individual <i>pol</i> genes	147
Figure 4.5	Linear regression analysis; correlation between pairwise genetic distance and number of sequences amplified at each time point	149
Figure 4.6	Bayesian skyline reconstructio b of patient A viral <i>pol</i> genes	152
Figure 4.7	Alignment showing the positively selected position in patient A	157

Figure 4.8	Alignment showing the positively selected position in patient A	161
------------	---	-----

CHAPTER 5

Figure 5.1	Restriction sites present in the p8.9NSXClal+ <i>gag-pol</i> region that were used for cloning	172
Figure 5.2	Dose-response curves obtained using different top concentrations of INSTI	174
Figure 5.3	Patient-derived <i>IN</i> only and full-length <i>pol</i> control vectors studied in the single-cycle replication assay for phenotypic and replicative fitness determination	177
Figure 5.4	RAL susceptibilities of patient-derived <i>IN</i> only viruses	180
Figure 5.5	Comparison of RAL susceptibilities of patient-derived <i>IN</i> only and full-length <i>pol</i> viruses	181
Figure 5.6	EVG susceptibilities of patient-derived <i>IN</i> only viruses	183
Figure 5.7	Comparison of EVG susceptibilities of patient-derived <i>IN</i> only and full-length <i>pol</i> viruses	184
Figure 5.8	Effects of patient-derived <i>IN</i> only and full-length <i>pol</i> viruses on replicative fitness	188

CHAPTER 6

Figure 6.1	Amount of p24 present 48 hours post transfection of retroviral vectors	201
Figure 6.2	The pHL(WT) HIV expression vector	202
Figure 6.3	Western blot analysis to a) confirm the presence of viral proteins in cell lysates 12 and 24 hours post-transfection and b) determine the cleavage efficiency of pseudoviruses 24 hours post-transfection	203

CHAPTER 7

Figure 7.1	Graphical representation of the relationship between the replicative fitness and susceptibility of recombinant viruses from patient A	211
------------	---	-----

Tables

CHAPTER 1

Table 1.1	HIV-1 genes and their products	32
Table 1.2	Commonly used antiretrovirals in HIV-infected patients	42
Table 1.3	C terminal RTI resistance mutations and their biochemical mechanisms	50
Table 1.4	Secondary resistance mutation pattern associated with Q148H/R/K	54
Table 1.5	Secondary resistance mutation pattern associated with N155H	56
Table 1.6	Secondary resistance mutation pattern associated with Y143C/H/R/K	58
Table 1.7	Mutations currently used for transmitted drug resistance surveillance	61

CHAPTER 2

Table 2.1	Reverse primers used for cDNA synthesis	85
Table 2.2	Annealing temperatures for different first round reverse primers used for the amplification of full-length <i>pol</i> gene	86
Table 2.3	Reverse primers used for nested PCR of full-length <i>pol</i> gene	86
Table 2.4	Annealing temperatures for different second round reverse primers used for the amplification of full-length <i>pol</i> gene	87
Table 2.5	Primers used for sequencing of HIV-1 <i>gag-pol</i>	90
Table 2.6	Primers used for the introduction of restriction enzyme sites in the p8.9NSX <i>gag-pol</i> expression vectors	95
Table 2.7	Primers used for the introduction and knockout of restriction enzyme sites in p8.9NSX-F	98
Table 2.8	Primers used for the introduction and knockout of restriction enzyme sites in subtype F patient-derived retroviral vectors	102
Table 2.9	Dilution series used in the INSTI phenotypic susceptibility assay	110

CHAPTER 3

Table 3.1	First round and nested reverse primer combinations tested for full-length <i>pol</i> genotyping	124
Table 3.2	Primer combinations used in the amplification of each patient samples	127
Table 3.3	Drug resistance mutations found in each patient	127
Table 3.4	Variations in the numbers of major and minor PI and RTI resistance mutations occurring in patient A during the sampling period	134
Table 3.5	Variations in the numbers of major and minor PI and RTI resistance mutations occurring in patient B during the sampling period	134

CHAPTER 4

Table 4.1	Bayes Factors between different nucleotide substitution, demographic and molecular clock models to determine best fit models for patient A viral <i>PR</i> , <i>RT</i> and <i>IN</i> sequences	150
Table 4.2	Positively selected codons by algorithm in viral <i>PR</i> , <i>RT</i> and <i>IN</i> sequences from patient A	156
Table 4.3	Positively selected codons by algorithm in viral <i>PR</i> , <i>RT</i> and <i>IN</i> sequences from patient B	156
Table 4.4	Negatively selected codons by algorithm in viral <i>PR</i> , <i>RT</i> and <i>IN</i> sequences from patient A	163
Table 4.5	Negatively selected codons by algorithm in viral <i>PR</i> , <i>RT</i> and <i>IN</i> sequences from patient B	164

CHAPTER 5

Table 5.1	Reproducibility of the single-cycle assay in determining the RAL and EVG EC ₅₀ of the p8.9NSX wild-type B control	173
Table 5.2	RAL EC ₅₀ of patient-derived <i>IN</i> only viruses	179
Table 5.3	EVG EC ₅₀ of patient-derived <i>IN</i> only viruses	185
Table 5.4	Infectivity levels of patient0derived full-length <i>pol</i> viruses	187

CHAPTER 6

Table 6.1	Mutations present in the C terminal of <i>RT</i> of HIV-1 subtype F viruses isolated from patients in the PENTA-EPPIC network	198
Table 6.2	Analysis of the frequency of C terminal mutations in RTI-naive and -experienced patients infected with Subtype F from the Stanford University Database	199

CHAPTER 1

General introduction

The first clinical descriptions of acquired immunodeficiency syndrome (AIDS) occurred in 1981 during an epidemic that first emerged in men who have sex with men (MSM) in North America. The causative agent was later discovered in 1983 to be the human immunodeficiency virus (HIV). HIV is classified as a retrovirus with a single-stranded positive sense RNA genome, with its replication occurring via a DNA intermediate as a distinguishing feature. Specifically, HIV belongs to the lentivirus genus of the Retroviridae family which are characterised by a long incubation period ("*lente*" is Latin for slow). The development of nucleos(t)ide reverse transcriptase inhibitors (NRTIs), targeting the viral reverse transcriptase (RT) enzyme, in the 1990s as therapy for HIV infection gave only slight improvement to those infected as the development of drug resistance was rapid and those on monotherapy subsequently failed. However, the prognosis of infected individuals began to improve following the introduction of other drug classes targeting the viral protease (PR) and the non-nucleoside reverse transcriptase inhibitors (NNRTIs) in the late 1990s. This was the beginning of the era of highly active antiretroviral therapy (HAART) which utilised multiple drug classes in treatment regimens. Since then there has been an explosion of drugs targeting different HIV genes and steps of the viral life cycle. Sadly, resistance development remains a major barrier in the treatment of HIV infection and work is still ongoing to produce drugs that target novel steps of the viral life cycle for use as salvage therapy. There are multiple reasons why drug resistance arises, the main one being the inherent error-prone nature of RT. RT is renowned for its capacity to make mistakes during replication causing mutations to arise and the rapid turnover of viruses leads to the generation of more viral variants. In addition, recombination during reverse transcription can cause the acquisition of mutations from other HIV genomes. The development of drug resistance in the era of HAART needs to be understood as multiple genes are targeted yet little is known about the interplay between resistance mutations on different linked HIV genes. This is the primary goal of this study. In addition, the majority of research to date has been done on subtype B viruses; therefore a second goal of this study is to investigate the development of resistance in other HIV-1 subtypes.

1.1 Implications of HIV/AIDS

1.1.1 Global health and financial impact of the AIDS pandemic

Globally, it is estimated that there were 34.2 million people living with HIV in 2011, the majority of whom live in Sub-Saharan Africa. Furthermore, in 2011 an average of 6800 individuals worldwide were newly infected daily equating to 2.5 million new infections and there were 1.7 million deaths due to HIV-1 infection and AIDS that year alone. However, there has been a decrease in the overall number of AIDS-related deaths since the mid-2000s and a decline in incidence of HIV-1 infections since the late 1990s mostly due to the introduction of HAART (UNAIDS, 2010). Unfortunately, an increase in the number of people living with HIV infections has inevitably resulted in an increase in the financial burden and resources required for patient treatment and care. This pandemic has an approximate cost of \$1.2 billion annually in developing countries and it is estimated that this may reach \$35 billion by 2031, the 50th year of the pandemic (Hecht, 2009). It is obvious from these data that the impact of the HIV/AIDS pandemic on global health, as well as the global economy, is extensive.

1.1.2 Course of infection

The natural course of HIV infection usually follows a distinct pattern (Pantaleo et al., 1993) and is illustrated in Figure 1.1. After infection with the virus the patient normally undergoes a period of acute infection, lasting on average about 3 weeks. During this period symptoms are analogous to other viral infections and as such are not usually picked up in the clinic. These symptoms are broad and can include fever, rash, pharyngitis, lymphadenopathy and other “flu-like” symptoms. In addition diarrhea, nausea, vomiting, headaches, weight loss, thrush and neurological symptoms may be experienced. The symptoms of acute infection can be attributed to the associated rapid decline in CD4+ T cell count and sharp increase in viral load. Towards the end of the acute infection period the emergence of HIV-specific immune responses results in the decline of viral load and a partial recovery in the CD4+ T cell count resulting in the second phase of infection; clinical latency. The level of sustained viral load attained towards the end of acute infection (viral load set point) may be an indicator of the length of clinical latency; the lower the viral load set point the slower the progression to AIDS.

Clinical latency can last on average 10 years however, the patient may experience fatigue and generalised lymphadenopathy. In this clinically latent period the CD4+ T cell count declines

slowly whilst the viral load may remain relatively low. As the CD4+ T cell count falls below 200 cells/ μ L and the viral load rapidly increases, the risk of infection or reactivation of opportunistic pathogens increases and the patient enters late-stage disease/AIDS, the third phase of HIV infection. The opportunistic infections include: candidiasis of the oesophagus and respiratory tract, extrapulmonary coccidioidomycosis/ cryptococcosis, cryptosporidiosis, cytomegalovirus retinitis, Kaposi's sarcoma, tuberculosis, pneumocystis carinii pneumonia and toxoplasmosis which are often referred to as AIDS-defining illnesses.

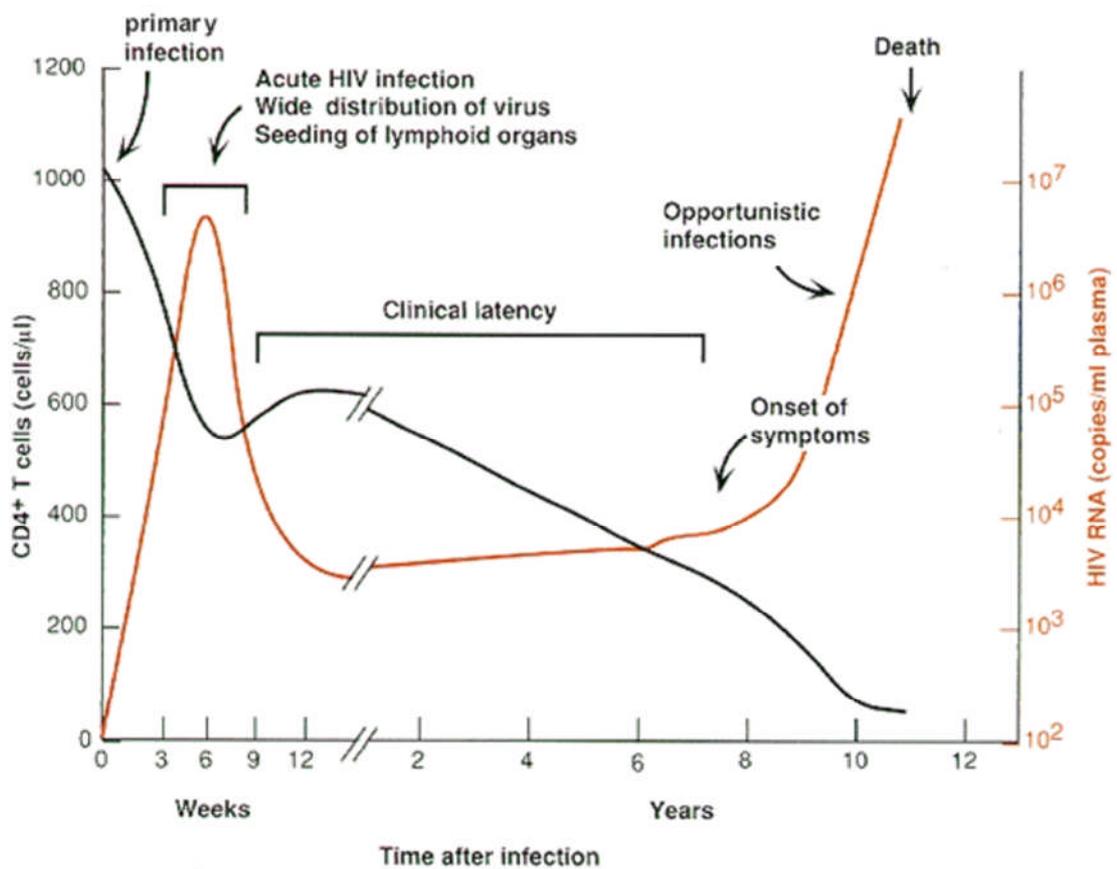


Figure 1.1 Natural course of HIV-1 infection. Natural HIV infection follows a characteristic course. During the acute phase of infection there is widespread distribution of the virus resulting in a sharp decrease in CD4+ T cells. Subsequently an immune response is initiated causing a reduction in detectable HIV RNA levels followed by a period of clinical latency. CD4+ T cell counts continue to decline in this period until it reaches a level below which there is an increased risk of opportunistic or reactivation of infections (Coffin et al., 1997).

1.2 The history, evolution and diversity of HIV

The origin of HIV is believed to be zoonotic, through multiple independent cross-species transmissions of simian immunodeficiency virus (SIV) from non-human primates into humans in West and Central Africa (Hemelaar, 2012; Peeters and Delaporte, 2012). These zoonotic transmissions gave rise to several HIV lineages: HIV-1 groups M, N, O and P and HIV-2 groups A-H. HIV-1 Group M is responsible for most of the current global pandemic whilst group O infections seem to be limited to the West-Central African region and groups N and P have only been identified in individuals residing in or originating from Cameroon. HIV-1 groups M and N are thought to have originated from SIVcpz found in *Pan troglodytes troglodytes* chimpanzees in West-Central Africa sometime between 1853 and the early 1900s and around 1963, respectively. It is thought that group M spread from South-Eastern Cameroon via the Congo River to the Democratic Republic of the Congo and from there initiated the global pandemic, whilst group N remained in South-Central Cameroon. Groups O and P viruses are closely related to the SIVgor virus which has been identified in Western lowland gorillas in Cameroon. It is unclear yet as to whether the cross-species event occurred between gorillas and humans or through chimpanzees via a closely related divergent ancestor of SIVgor. However, for HIV-1 group O, this event is thought to have occurred during the 1920s.

The origin of HIV-2 was proposed in 1989 to have come from circulating SIVsmm infections in sooty mangabey monkeys in West Africa (Hirsch et al., 1989). HIV-2 infections have mainly been restricted to this area although these infections are largely being replaced by HIV-1 infections (van der Loeff et al., 2006; Hamel et al., 2007). This thesis focuses on HIV-1 group M as it is the most clinically relevant form.

1.2.1 HIV-1 diversity

The diversity of the pandemic strain, HIV-1 group M, is substantial. Within this group there are 9 different subtypes, A-D, F-H, J and K, varying between 17-35% at the amino acid level (Korber et al., 2001; Peeters et al., 2010; Hemelaar et al., 2011). Subtypes A and F can be further divided into sub-subtypes, A1-4 and F1-2. Genetic variation within a subtype is thought to be between 8-17% at the amino acid level (Korber et al., 2001). In addition, there are over 50 known inter-subtype circulating recombinant forms (CRFs) that are defined as viral recombinants found in three or more epidemiologically unlinked individuals. Furthermore, there are also unique recombinant forms (URFs) which do not meet the criteria for CRFs and

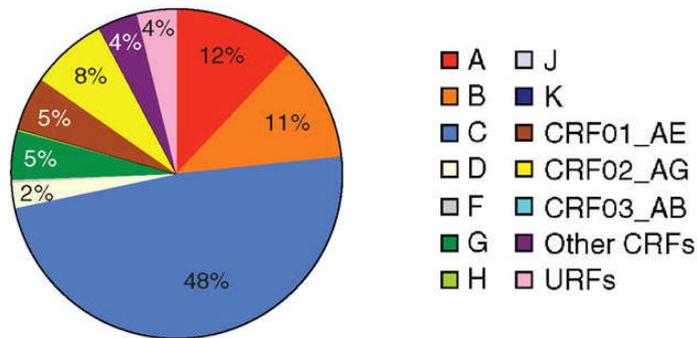
second generation recombinants (SRGs) that are generated through further recombination of CRFs with other subtypes or CRFs (Peeters et al., 2010; Hemelaar et al., 2011). Not surprisingly, this creates problems around vaccine design.

Historically, subtypes were determined by sequencing and phylogenetic analysis of the *env* and *gag* genes. However, nowadays any gene sequence can be used to determine the subtype of a virus, although sequencing only partial fragments of the genome may mean that recombinant forms can be missed.

1.2.2 HIV-1 subtype distribution

The distribution of subtypes varies worldwide as shown in Figure 1.2. The greatest diversity is found in Central Africa where all subtypes and many CRFs are found. Subtype B viruses are mainly found in the Western world i.e. North America, the Caribbean, Latin America, Western and Central Europe and Australia. In Asia, subtype A and various CRFs dominate. Subtype C which is mainly found in Southern Africa, Ethiopia and India is the major contributor to the pandemic responsible for ~48% of infections worldwide. Other subtype contributions vary between <1%-16% (Hemelaar et al., 2011).

a)



b)

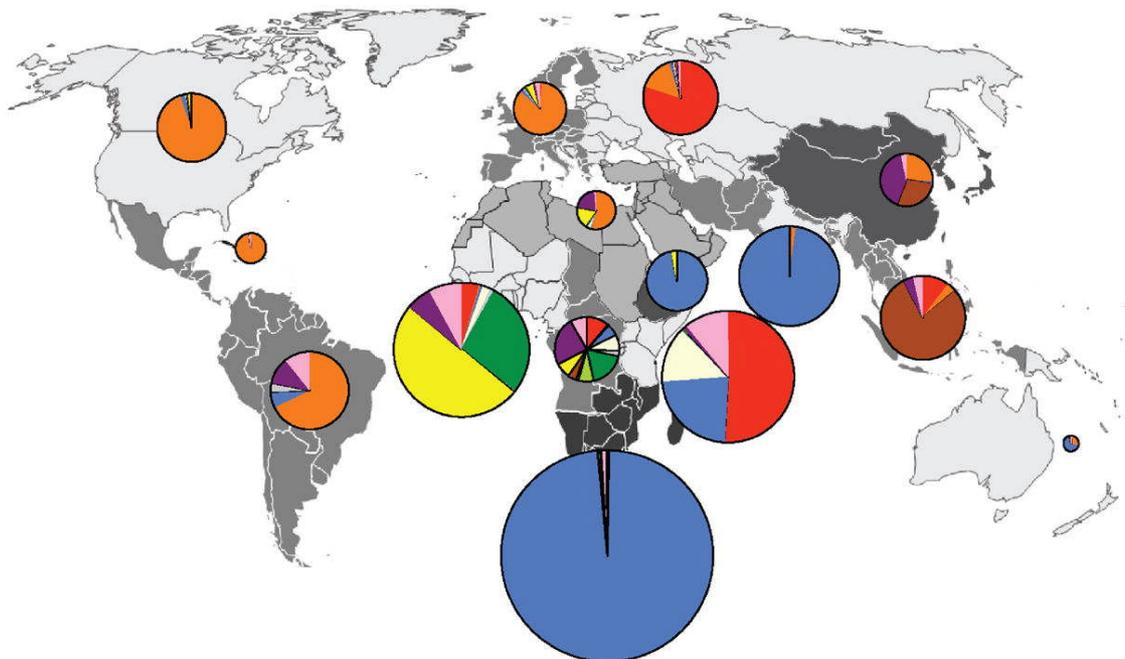


Figure 1.2 a) HIV-1 subtype contribution to the pandemic and b) their distribution worldwide between 2004 and 2007. Graphical distribution of HIV-1 subtypes and their contribution to the HIV pandemic. Subtype distribution is remarkably different across the world with the majority of infections occurring in Southern Africa, caused by HIV-1 subtype C viruses (Hemelaar et al., 2011).

1.3 HIV-1 genome structure

The HIV-1 genome is a single-stranded positive sense RNA of approximately 9 kb in length. In the virion HIV-1 genomic RNA is present as a dimer which is linked together by sequences near the 5' end of each RNA molecule. The HIV-1 RNA genome encodes the three major structural genes common to all retroviruses namely *gag*, *pol* and *env* as well as other regulatory and accessory proteins (reviewed by Coffin et al., 1997). The three structural genes are translated as polyprotein precursors which are subsequently processed by viral and cellular proteases to produce 9 viral proteins. The *gag* polyprotein precursor (Pr55^{Gag}) is produced from unspliced viral mRNA and subsequently cleaved to release the structural proteins: matrix (MA or p17), capsid (CA or p24), nucleocapsid (NC or p7) and p6. In addition, two spacer peptides, SP2 and SP1, are located within Pr55^{Gag} separating the CA and NC and the NC and p6, respectively (Freed, 2001).

The *gag-pol* polyprotein precursor (Pr160^{Gag-pol}) is the product of a frame shift event during Pr55^{Gag} translation which occurs 5% of the time and is responsible for the production of the viral enzymes PR, RT, and IN necessary for replication and virion (Jacks et al., 1988). The *env* polyprotein precursor (gp160) is produced from singly spliced viral mRNA and upon processing gives rise to two envelope subunit proteins, gp41 and gp120 which are involved in entry into the host cell. In addition, HIV-1 has two regulatory proteins (*tat* and *rev*) and four accessory proteins (*vif*, *vpu*, *vpr*, and *nef*) which are involved in the regulation and coordination of viral gene expression as well as playing other ancillary roles in the life cycle of the virus. These proteins are produced from alternatively spliced viral mRNA transcripts (reviewed by Freed, 2001 and Coffin et al., 1997). Figure 1.3 shows the organisation of the HIV-1 genome and its open reading frames.

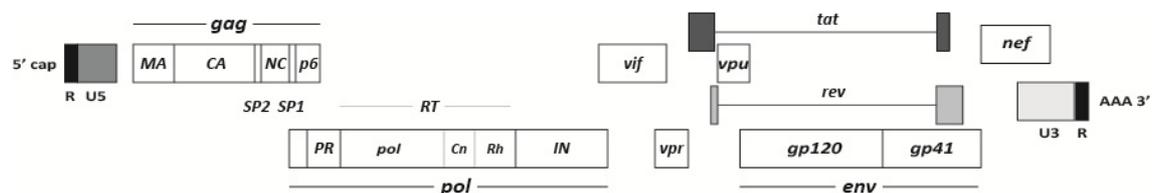


Figure 1.3 Schematic of HIV-1 genome RNA. Each RNA strand contains three principle genes, *gag*, *pol* and *env*, and six genes encoding two regulatory proteins (*tat* and *rev*) and four accessory proteins (*vif*, *vpu*, *vpr* and *nef*). These are found between regulatory sequences R (direct repeat sequences) and unique 5' and 3' sequences (U5 and U3) which are vital for viral

replication. These are duplicated during reverse transcriptase to generate the 5' and 3' long terminal repeats (LTRs) of the provirus.

1.4 Viral proteins

The main focus of this study concerns the HIV-1 *pol* gene. As such the *pol* gene products PR, RT and IN are described in more detail in this section. A brief description of all other viral proteins and their respective roles in the virus life cycle can be found in Table 1.1.

1.4.1 Reverse transcriptase

The RT enzyme is a heterodimer, consisting of related p66 and p51 subunits which combine to form a stable asymmetric complex (Divita et al., 1995). Each subunit contains a polymerase (*pol*) domain (amino acids 1-321), which is further separated into finger, thumb and palm subdomains, the structure of which resembles a right hand (Kohlstaedt et al., 1992). The C terminal regions of each subunit contain a connection (*cn*) subdomain (amino acids 322-440) whilst the p66 subunit also contains an RNase H (*rh*) domain (amino acids 441-560). The p51 subunit lacks this *rh* domain and therefore has no RNase H activity (Le Grice et al., 1991).

RT catalyses the synthesis of viral double-stranded (ds) DNA, which is required for integration into the host cell genome, by reverse transcription of the single-stranded genomic RNA molecule. As such, the enzyme has RNA-dependent and DNA-dependent DNA polymerase activity as well as RNase H activity (Coffin et al., 1997). In order for RT to catalyse each reaction the enzyme attains different substrate binding orientations, these being an RNase H cleavage competent orientation and the polymerase competent orientation. The switching between each orientation is a dynamic process which is achieved without dissociating from the nucleic acid (Abbondanzieri et al., 2008). The combining of both the DNA polymerization and RNA cleavage functions greatly increases the replication efficiency of the enzyme (Esposito et al., 2012).

The nucleic acid binding cleft of RT is primarily made up of residues of the p66 subunit and the nucleic acid makes contacts with both the *pol* and *rh* active sites, which are about 17-18 base pairs apart on the nucleic acid (Kohlstaedt et al., 1992). Highly conserved regions near the *pol* active site of the p66 subunit, together with two alpha helices of the p66 thumb subdomain act as a clamp, or “primer grip” to position the template-primer at the *pol* active site. The 3'OH

of the nucleic acid is close to the catalytic residues 110, 185 and 186 in the p66 palm subdomain (Larder et al., 1987), responsible for the chelation of two divalent ions (Patel et al., 1995), allowing for nucleophilic attack on an incoming nucleotide (Jacobo-Molina et al., 1993).

The fidelity of HIV-1 RT has been investigated in great detail and it is thought that much of the sequence diversity seen in HIV is due to the high error rates of RT. HIV-1 RT has no 3' or 5' exonuclease activity (Roberts et al., 1988) and therefore is unable to excise mis-paired nucleotides and perform nick-translation reactions. In addition, studies have shown that *in vitro* errors rates of RT translate to about 5 to 10 errors per HIV-1 genomes per round of replication (Preston et al., 1988). Recombination between different RNA molecules is also thought to occur during the reverse transcription process due to the multiple strand transfer reactions generating novel viral recombinants (Clavel et al., 1989).

1.4.2 Integrase

Integrase (IN) catalyses the insertion - integration - of newly synthesised double-stranded viral DNA into the genome of the infected cell. This is carried out by two enzymatic functions of IN; 3' end processing and DNA strand transfer (Bukrinsky et al., 1993a; Miller et al., 1997; Bushman, 1999; Lin and Engelman, 2003; Raghavendra et al., 2010). IN is a 32 kDa protein consisting of 288 amino acids. It is thought that the active form of the enzyme is a tetramer assembled from two dimers each bound to one of the viral DNA ends (Krishnan et al., 2010). Each monomer consists of three structural domains; the N terminal domain (NTD), the catalytic core domain (CCD) and the C terminal domain (CTD). The structure of the complete HIV-1 IN has so far proved to be elusive due to its poor solubility. However, structures of the three domains have been determined individually as well as that of two-domain fragments (Dyda et al., 1994; Chen et al., 2000; Wang et al., 2001).

These studies have revealed that the NTD consists of amino acids 1-49, and is made up of a triplet of α -helices containing a double histidine/cysteine (H₁₂-H₁₆-C₄₀-C₄₃) zinc binding motif which plays a role in the dimerization of IN monomers and the binding of cellular factors (McColl and Chen, 2010). The CCD consists of amino acids 50-212 and as the name suggests, contains the active site of the enzyme. This is represented by the D64, D116 and E152 (DDE) motif. This is necessary for the coordination of divalent metal ions (Mg²⁺ or Mn²⁺) that are essential for IN function (McColl and Chen, 2010). The remaining amino acids, 213-288, make up the CTD and contain SH3 domains that non-specifically bind to DNA (Engelman et al., 1994).

The recent elucidation of the crystal structure of prototype foamy virus (PFV) IN tetramer in complex with 3' processed viral DNA ends has greatly improved the understanding of the structure of full-length HIV-1 IN and its mode of action (Krishnan et al., 2010; Hare et al., 2010). This DNA-IN complex is called the intasome and is the minimal structure required for integration. The structural model of the HIV-1 intasome based on the PFV intasome reveals that the two viral DNA ends engaged by the IN active site are approximately five nucleotides apart which coincides with the number of nucleotides apart the target integration DNA sites are, further validating the intasome as the integration functional unit (Krishnan et al., 2010). Importantly, the intasome is the preferred target of integrase inhibitors, rather than IN itself. This suggests a direct involvement of the viral DNA ends in inhibitor binding.

1.4.3 Protease

The HIV protease is responsible for the cleavage of viral polyproteins into their functional proteins and as a result, the maturation of new viral particles. Protease is related to cellular aspartic proteases which use two apposed aspartate residues in their active site for catalysis. Functionality of the HIV protease comes through dimerisation, each 99 amino acid monomer contributing one Asp₂₅-Thr₂₆-Gly₂₇ triad to create a symmetric active site. This is essential for the enzyme's activity (Navia et al., 1989). The active site is gated by two flexible β hairpin loops (amino acids 46 to 56) which are known as flaps. These are thought to position the substrate in the active site enabling catalysis (Wlodawer and Erickson, 1993).

Protease recognises and cleaves at ten different cleavage sites, between seven and eight amino acids in length, within Pr55^{Gag} and Pr160^{Gag-pol}. Each site consists of different amino acid sequences (hydrophobic in nature), but secondary structures at each motif have proved to be very similar (Prabu-Jeyabalan et al., 2002). As protease dimerisation is required for function it is thought that upon oligomerisation of Pr160^{Gag-pol} during virion assembly, the protease domains of two polyproteins come together to form an active site enabling self-cleavage from the polyproteins and causing a cascade of stepwise cleavage of other functional units from Pr55^{Gag} and Pr160^{Gag-pol} (Sadiq et al., 2012).

Table 1.1 HIV-1 genes and their products

Gene	Protein	Functions
<i>gag</i>	Matrix (MA/ p17) Peripheral virion membrane protein	Viral entry (Yu et al., 1992) Gag intracellular transport (Rhee and Hunter, 1987; Weaver and Panganiban, 1990) Viral assembly (Henderson et al., 1983; Hansen et al., 1990; Bryant and Ratner, 1990) Regulates nuclear import of the pre-integration complex (Haffar et al., 2000)
	Capsid (CA/ p24) Core protein	Packaging of cyclophilin A (Gamble et al., 1996) and tRNA ^{lys3} into the virion (Kovaleski et al., 2007) Gag oligomerisation (Franke et al., 1994) Important for the nuclear import of the pre-integration complex (Yamashita and Emerman, 2004)
	Nucleocapsid (NC/ p7)	Viral RNA and DNA binding (Lapadat-Tapolsky et al., 1993; Schmalzbauer et al., 1996) Gag oligomerisation and viral assembly (Dawson and Yu, 1998) Interacts with PTAP domain of p6 to recruit cellular proteins required for budding (Dussupt et al., 2009) Increases incorporation of nucleotides by reverse transcriptase and its fidelity (Kim et al., 2012)
	p6	Incorporation of vpr into the virion (Checroune et al., 1995) Contains a PTAP domain which plays a role in virion budding (Huang et al., 1995; Gottlinger et al., 1991)
<i>vif</i>	vif	Protection from hypermutation of viral cDNA by host APOBEC3 restriction factors (Kao et al., 2003)
<i>vpr</i>	vpr	Inhibits cell division (Rogel et al., 1995) Nuclear transport of the preintegration complex (Heinzinger et al., 1994) Coactivator of HIV LTR (Vanitharani et al., 2001)
<i>tat</i>	tat	Viral transcription factor (Fisher et al., 1986; Dayton et al., 1986) Modulates cellular gene expression (for example Li et al., 2005a and Ju et al., 2009)
<i>rev</i>	rev	Regulates splicing and nuclear export of unspliced and singly spliced HIV mRNA (Pollard and Malim, 1998) Together with LEDGF regulates integrase (Benyamini et al., 2011)
<i>vpu</i>	vpu	Promotes degradation of CD4 receptors in the endoplasmic reticulum (Margottin et al., 1998)

		Promotes virion release by counteracting tetherin (Gottlinger et al., 1993; Neil et al., 2008) Directs interferon regulatory factor 3 for degradation (Doehle et al., 2012) Downregulates NTB-A at cell surface to evade lysis by natural killer cells (Bolduan et al., 2013)
<i>env</i>	gp120 (Surface glycoprotein)	Attachment to host cell membranes (Dalglish et al., 1984)
	gp41 (Transmembrane glycoprotein)	Membrane fusion (Kliger et al., 2000)
<i>nef</i>	nef	Down regulation of CD4 (Aiken et al., 1994) and MHC1 receptors (Collins et al., 1998) Targets internalised CD4 for degradation (daSilva et al., 2009)

A brief description of the functions of the HIV-1 *gag* and *env* structural genes and other accessory HIV-1 gene protein products are described here. In depth descriptions of the *pol* gene products can be found in the text.

1.6 The lifecycle of HIV

The lifecycle of HIV-1 consists of seven main stages as represented in Figure 1.5 and these are described in detail on the following pages.

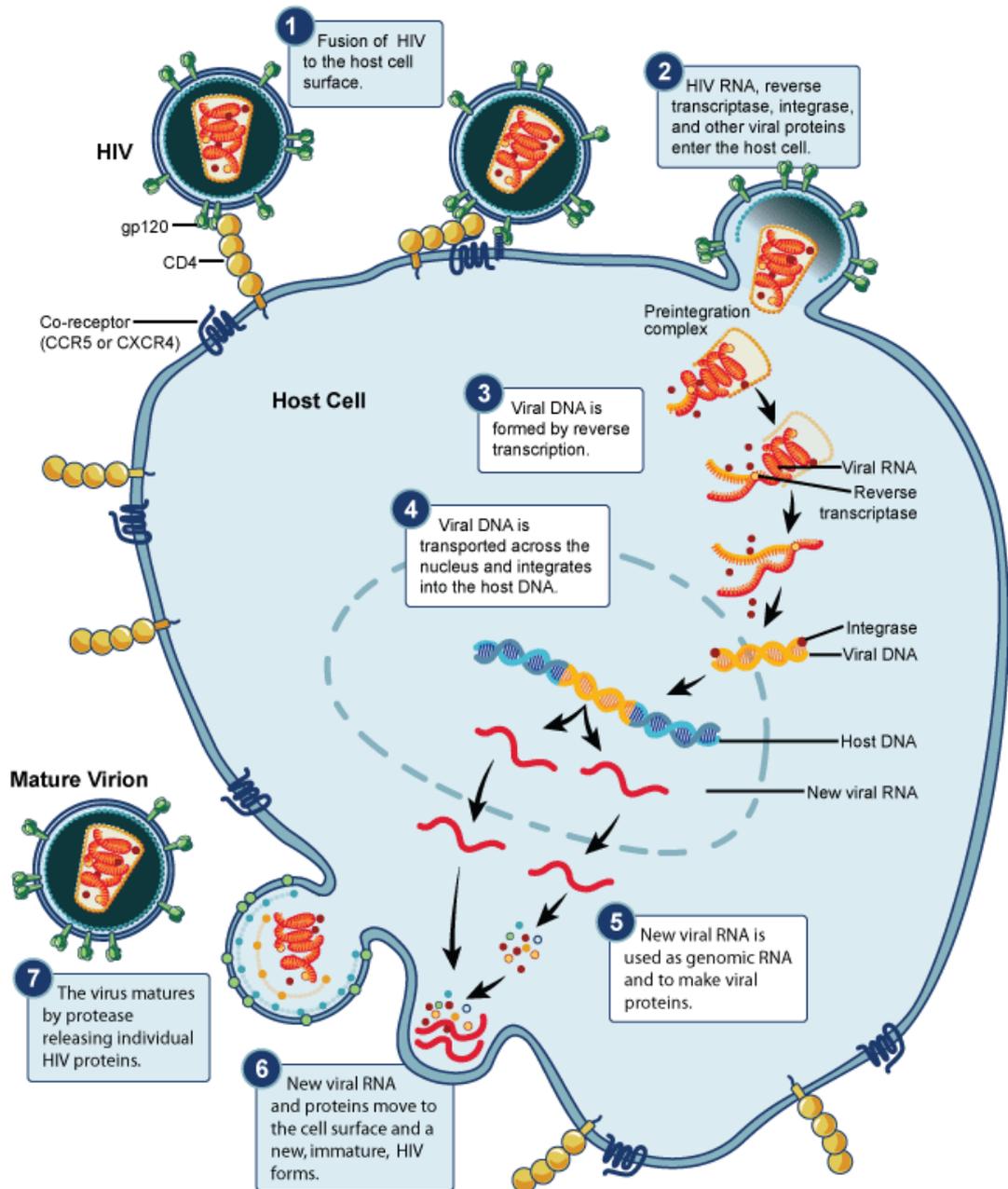


Figure 1.5 Schematic of the HIV-1 replication cycle. HIV-1 replication can be split into seven main steps: 1) Binding and fusion with host cell receptors, 2) entry, 3) reverse transcription, 4) nuclear transport and integration, 5) transcription of viral mRNAs and translation of the transcripts into viral proteins, 6) viral protein recruitment and assembly at the cell membrane and 7) virion release and maturation. Viral entry, reverse transcription, integration and maturation are major targets of HAART (NIH, 2012b).

1.6.1 Receptor binding and entry

Entry of the virus into a host cell involves the interaction of the viral surface glycoprotein gp120 firstly with a CD4 cell receptor (Dalglish et al., 1984) and subsequently with a chemokine co-receptor, CXCR4 or CCR5, found on macrophages and CD4+ T cell surfaces (Deng et al., 1996; Kozak et al., 1997). This binding induces a conformational change which exposes the fusion domain present in the N terminal region of the transmembrane glycoprotein gp41 (Kliger et al., 2000). This allows the fusion of viral and cellular membranes, enabling the viral capsid to enter the cell cytoplasm. Following entry into the host cell partial uncoating of the viral core occurs to release the ribonucleoprotein and in the process initiates reverse transcription of the single-stranded, positive-sense RNA into dsDNA (reviewed by Fassati, 2012).

1.6.2 Reverse transcription

The process of reverse transcription (Figure 1.6) starts when tRNA^{Lys3} bound to a primer binding site (PBS) upstream of the gene coding region acts as a primer providing a free 3'OH for the initiation of negative strand DNA synthesis. Complementary negative strand DNA (cDNA) is made by the RNA-dependent DNA polymerase activity of RT until the 5' end of the genomic RNA is reached. Subsequently, the RNA moiety of the DNA:RNA duplex is degraded by RT RNase H activity producing a negative strand strong-stop DNA. This strong-stop DNA is then able to anneal to the 3' end of the genomic RNA which contains repeated (R) sequences (which are present at both the 3' and 5' end of the viral genome) in a process termed first strand DNA transfer. Negative strand synthesis is then resumed and accompanied by RNase H degradation of the template RNA strand until the whole genome has been copied into negative sense DNA. Contained within the RNA template is a polypurine tract (PPT) which is relatively resistant to RNase H degradation, this portion of the RNA template, which has not been degraded, now serves as a primer for initiation of positive strand DNA synthesis, using the newly synthesised negative strand DNA as the template.

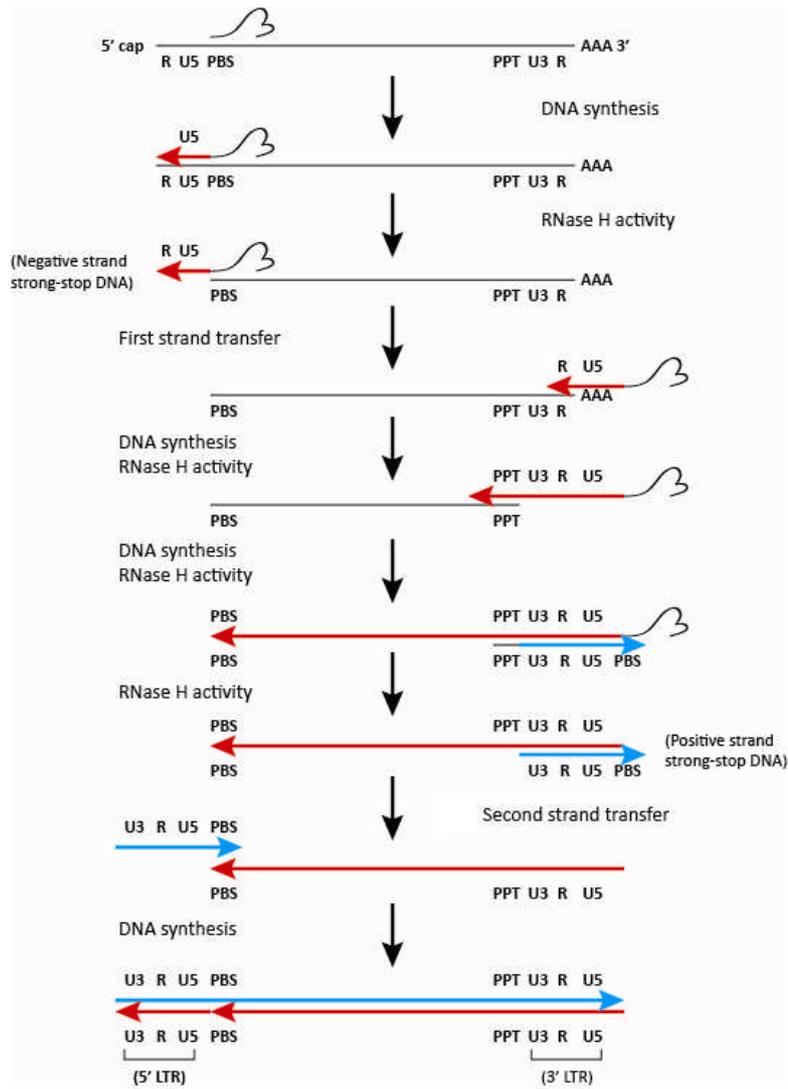


Figure 1.6 Schematic of the reverse transcriptase process. tRNA bound to the polymerase binding site (PBS) at the 5' end of the genome initiates reverse transcription. Negative strand DNA synthesis (**red**) occurs until the 5' end of the genome is reached whilst RNase H activity degrades the RNA (**black**) portion of the RNA:DNA duplex. Negative strand strong-stop DNA acts as a primer for synthesis at the 3' end after first strand DNA transfer to the complementary direct repeat sequence (R). DNA synthesis and RNase H degradation occurs until the 5' end (PBS region). The polypurine tract (PPT) region at the 3' end however is relatively resistant to RNase H degradation, and is therefore able to act as a primer for initiation of positive strand DNA synthesis (**blue**). The PPT is eventually degraded to generate the positive strand strong-stop DNA primer. This acts as a primer for positive strand DNA synthesis at the 5' end of the minus strand DNA after second strand DNA transfer to the PBS sequence. Negative and positive DNA synthesis proceeds to create the full-length genome and viral long terminal repeats (LTRs).

Transcription of the positive strand DNA is stopped after a portion of the tRNA^{Lys3} is transcribed (probably at a 1-methyl-adenosine residue that cannot act as a template for reverse transcription), creating a positive strand strong-stop DNA. Further RNase H activity eventually degrades the PPT in the RNA template primer and the tRNA exposing the PBS sequence in the positive strand strong-stop DNA enabling second strand DNA transfer and annealing with the PBS sequence in the negative strand DNA. Completion of positive strand DNA synthesis then occurs generating a dsDNA with identical sequences at both ends, termed long terminal repeats (LTRs). This dsDNA is then imported into the cell nucleus and integrated into the host cell genome (Sarafianos et al., 2009).

1.6.3 Nuclear transport and integration

It was first thought that the sites of integration were randomly selected; however, recent evidence suggests that integration into actively expressed host genes is favoured (Schroder et al., 2002; Mitchell et al., 2004; Wu and Burgess, 2004; Maxfield et al., 2005). Integration is speculated to involve four IN molecules, each dimer binding to either the 5' or 3' LTR of the dsDNA (Krishnan et al., 2010). The first step of viral DNA integration called 3' processing occurs in the cytoplasm prior to nuclear transport and involves cleavage of two nucleotides (G and T) from a conserved CA dinucleotide at both 3' ends of the DNA, creating 5' end overhangs and 3' OH groups as illustrated in Figure 1.7 (Brown et al., 1989; Vink et al., 1991; Leavitt et al., 1992).

This intermediate is then actively transported into the nucleus as part of the preintegration complex (PIC) via interactions with nucleopore complex proteins (Woodward et al., 2009). Studies have shown that viral IN, RT and MA proteins are present in the PICs, alongside host-derived proteins (Miller et al., 1997; Lee and Craigie, 1994; Farnet and Haseltine, 1991; Bukrinsky et al., 1993b; Raghavendra et al., 2010). In addition, there is evidence that viral vpr is required for the targeting of the viral dsDNA to the nucleus in non-dividing cells suggesting its presence in the PICs (Heinzinger et al., 1994).

Once in the nucleus the host factor, LEDGF/p75, binds to the host cell genomic DNA and the viral PIC facilitating the approach of the two IN dimers towards each other, thereby forming a tetramer and targeting viral dsDNA to host genomic DNA (Cherepanov et al., 2003; Ciuffi et al., 2005). In a one-step transesterification reaction known as strand transfer, the two 3' OH groups attack the host DNA in a staggered fashion (approximately 4-6 base pairs apart) separating the host DNA bonds and joining the viral 3' OH groups with the host 5' phosphate

groups (Bushman and Craigie, 1991). This intermediate is mismatched and gapped which induces a host DNA damage repair response resulting in the removal of the 5' mismatched dinucleotides, filling in of the gaps on the host DNA and finally the ligation of the viral dsDNA and host genomic DNA (Yoder and Bushman, 2000).

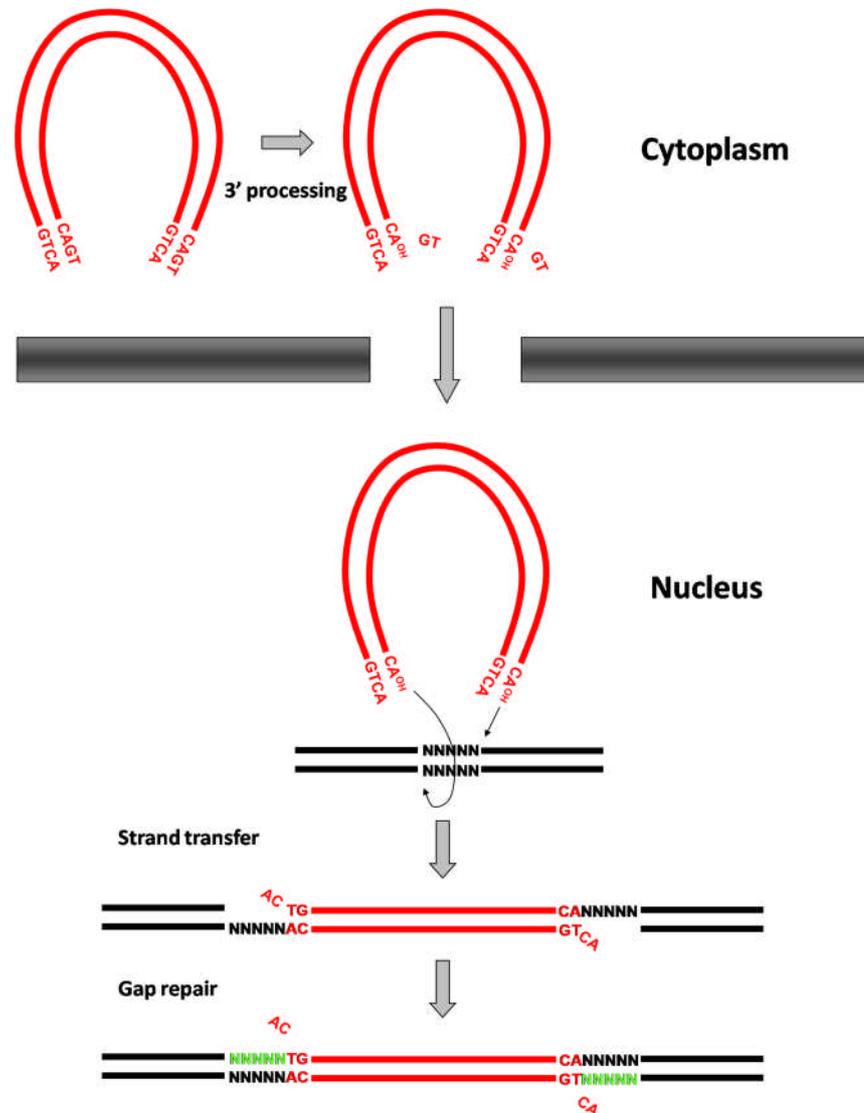


Figure 1.7 HIV-1 DNA integration. Following reverse transcription viral dsDNA is integrated into the host cell genome. This occurs in three steps 1) 3' processing, 2) strand transfer and 3) gap repair. HIV-1 IN is responsible for the first two steps. HIV-1 IN cleaves the 3' end GT dinucleotide from a conserved CAGT motif on each viral DNA strand (red) exposing a 3' hydroxyl group on the terminal adenosine. The 3' processed viral DNA is subsequently imported into the nucleus where strand transfer occurs resulting in the integration of viral DNA at positions approximately five base pairs apart into host DNA (black). Host DNA repair enzymes remove unpaired viral CA dinucleotides, fill in the five base pair gaps (green) and ligate the DNA ends (Mbisa et al., 2011a).

1.6.4 Transcription, translation and transport of viral macromolecules

Following the integration of viral dsDNA into the host chromosome and formation of a provirus, the virus can then take full advantage of the host transcription and translation machinery (Freed, 2001). The HIV LTR acts as the site for transcription initiation and contains multiple sequence motifs required for the recruitment of host RNA polymerase II (RNAP II) and additional host transcription factors. A specific role of the viral *tat* protein has been identified in this process. *Tat* is thought to activate transcription in trans, binding to the transactivating response sequence (TAR) in the viral DNA and recruiting positive-transcriptional elongation factor b complex (P-TEFb) which results in the phosphorylation of the C terminal domain of RNAP II and an increase in transcription.

Viral transcripts are then subject to the same processing events (5' cap addition, 3' end polyadenylation and splicing) as cellular mRNAs. These viral mRNAs are subsequently exported out of the nucleus where they are translated into Pr55^{Gag} and Pr160^{Gag-pol} (made from unspliced mRNA), *env*, *vif*, *vpu* and *vpr* (made from singly spliced mRNAs) and *rev*, *tat* and *nef* (made from multiply spliced mRNAs). The viral *rev* protein plays a role in the export of these viral mRNAs as well as in the regulation of viral RNA splicing. This is carried out via interactions with the Rev response element (RRE) present in the *env* gene of all unspliced and singly spliced mRNAs enabling the export of these mRNAs through the cellular nuclear export machinery before multiple splicing can occur (Pollard and Malim, 1998).

In the cytoplasm multiple regions in the Pr55^{Gag} play a central role in the recruitment of viral and host proteins to the cell membrane. Pr55^{Gag} is also essential in conducting virion assembly and budding (Gamble et al., 1996; Franke et al., 1994; Kovaleski et al., 2007; Dawson and Yu, 1998). In addition, the NC region of Pr55^{Gag} specifically is thought to be involved in recruiting the two viral genomic RNA molecules for packaging into the new virion (Lapadat-Tapolsky et al., 1993; Schmalzbauer et al., 1996). This is the crucial step which can lead to heterozygous recombination between different RNA molecules in the next round of infection in a dual infected cell (Hu and Temin, 1990; Vogt, 1971; Weiss et al., 1973). Transportation of the *env* polyproteins to the membrane however, occurs through the Golgi apparatus where glycosylation and cleavage into gp120 and gp41 occurs (Freed, 2001).

1.6.5 Virion release and maturation

The multimerisation of Pr55^{Gag} at the cell membrane forces membrane curvature whilst two “late domain” motifs in p6 appear to play a role in the recruitment of endosomal sorting complex required for transport (ESCRT) factors which facilitate virion budding from the host cell plasma membrane. After budding from the cell membrane the virus particle undergoes maturation which is carried out by the viral protease. This involves the stepwise cleavage of the viral polyproteins Pr55^{Gag} and Pr160^{Gag-pol} into their functional units. This causes the condensation of viral structural proteins into the characteristic conical core seen in mature virions by electron microscopy. Only after this process is the virion able to infect another cell (Sundquist and Krausslich, 2012).

1.7 Antiretroviral drugs

At present there are six classes of antiretrovirals (ARVs) licensed for use against HIV-1 infection each targeting different viral processes namely, fusion inhibitors and entry inhibitors, nucleos(t)ide reverse transcriptase inhibitors (NRTIs), non-nucleoside reverse transcriptase inhibitors (NNRTI), integrase strand transfer inhibitors (INSTIs) and protease inhibitors (PIs). A list of antiretroviral drugs that have been approved for clinical use against HIV infection can be found in Table 1.2. Currently, the British HIV Association (BHIVA) recommends that the first line of HAART should include two NRTIs, either emtricitabine (FTC) and tenofovir (TDF) or abacavir (ABC) and lamivudine (3TC), together with a third agent, either an NNRTI or a ritonavir-boosted PI or an INSTI. BHIVA guidelines recommend the start of HAART in a newly infected individual when their CD4+ cell count falls below 350 cells/ μ L or if there is any neurological involvement or any evidence of an AIDS-defining illness (BHIVA writing group., 2012).

Various combinations of these four different classes of drugs have been shown to be successful in achieving the aim of HAART which is the complete suppression of viral replication and reduction of viraemia to <50 copies/mL (Hammer et al., 2008). However, due to the development of drug resistance patients can fail on first line therapy and subsequent salvage therapies.

Table 1.2 Commonly used antiretrovirals in HIV-infected patients

Drug class	Drug or drug combination	Abbreviation
Fusion inhibitor (FI)	Enfuvirtide	ENF/T-20
Entry inhibitor	Maraviroc	MVC
Nucleos(t)ide reverse transcriptase inhibitors (NRTIs)	Abacavir	ABC
	Lamivudine	3TC
	Zidovudine	ZDV/AZT
	Didanosine	ddI
	Emtricitabine	FTC
	Stavudine	d4T
Non-nucleoside reverse transcriptase inhibitors (NNRTIs)	Tenofovir	TDF
	Efavirenz	EFV
	Etravirine	ETR
	Nevirapine	NVP
Integrase strand transfer inhibitor (INSTI)	Rilpivirine	RPV
	Raltegravir	RAL
*Protease inhibitors (PIs)	Atazanavir/ritonavir	ATV/r
	Darunavir/ritonavir	DRV/r
	Fosamprenavir/ritonavir	FPV/r
	Indinavir/ritonavir	IDV/r
	Lopinavir/ritonavir	LPV/r
	Nelfinavir	NFV
	Saquinavir/ritonavir	SQV/r
Tipranavir/ritonavir	TPV/r	

*PIs are generally boosted with ritonavir to inhibit the host enzyme responsible for PI metabolism (Kempf et al., 1997)

1.7.1 Fusion and entry inhibitors

Enfuvirtide, also known as T-20, is an inhibitor of viral entry specifically fusion of the viral and host cell membrane. Two hydrophobic alpha-helical regions of gp41 have been implicated in the conformational changes required for membrane fusion; enfuvirtide is a synthetic peptide that mimics one of these regions (Wild et al., 1994). It has been proposed that enfuvirtide binds to the highly conserved hydrophobic regions that normally induce the conformational change thereby inhibiting membrane fusion (Croteau et al., 2010; Kilby et al., 1998).

A second entry inhibitor was discovered during a screen of the Pfizer compound library to identify a small CCR5 ligand. A thousand analogues were synthesized and profiled from which maraviroc was subsequently selected. Maraviroc inhibits gp120 binding to the CCR5 co-receptor, preventing fusion with host cell membranes and viral entry into the cell. It has been

suggested that maraviroc causes the promotion of an inactive form of CCR5 which prevents CCR5-tropic viruses from utilising the co-receptor for entry (Dorr et al., 2005).

1.7.2 Reverse transcriptase inhibitors

Reverse transcriptase inhibitors (RTIs) were the first antiretroviral drugs to be developed. As there is no human equivalent to RT it was the obvious choice to target for inhibition. Currently there are 17 licensed RTIs which consist of two different drug classes; the NRTIs and the NNRTIs (BHIVA writing group., 2012). Though both classes target the RT enzyme each class has a different mechanism of action. NRTIs are nucleoside or nucleotide analogues of natural, cellular deoxynucleotides that are required for DNA replication. Upon entering the cell NRTIs are phosphorylated into their active form which can then be incorporated into the growing DNA chain by the viral RT during reverse transcription. However, as they lack the terminal 3' hydroxyl group needed for continual extension of the DNA chain, DNA replication is terminated prematurely (Hughes et al., 2008; Goody et al., 1991).

NNRTIs also target the viral RT; however, instead of binding to the active site of RT they bind to a hydrophobic binding pocket located away from the active site (Kohlstaedt et al., 1992). This interaction causes a conformational change in the enzyme which displaces catalytic residues thereby rendering RT inactive and unable to function (Shafer et al., 2000).

1.7.3 Integrase inhibitors

Raltegravir (RAL) was the first, and to date the only, integrase strand transfer inhibitor (INSTI) to be licensed for use in HIV-1 treatment. RAL which was approved for use in 2007 is a diketo acid (DKA) analog, its key feature being a β -hydroxy ketone moiety. DKA compounds were the first molecules to demonstrate selective and potent IN strand transfer inhibition with IC_{50} s in the nanomolar range (80 to 150 nM) in *in vitro* strand transfer assays (Hazuda et al., 2000). Further studies have shown that DKA compounds exert their antiviral activity through the chelation of the two metal ions that are coordinated by the DDE motif in the IN active site, thereby blocking their involvement in the strand transfer reaction (Grobler et al., 2002; Marchand et al., 2003).

Elvitegravir (EVG), inhibits integrase strand transfer through the same mechanism as RAL; however, it is not affected by one of the major resistance pathways that develops during RAL

treatment suggesting minor differences of action (Metifiot et al., 2011). Importantly, the blood concentration of EVG can be boosted by ritonavir or cobicistat enabling it to be administered once daily with the same efficacy and safety to that of RAL twice daily (Molina et al., 2012). EVG has recently been licensed for use in the USA in a once-daily, dual target, anti-HIV pill (the QUAD pill, Gilead Sciences) (Gilead, 2012) containing EVG, cobicistat and two NRTIs, tenofovir and emtricitabine (Cohen et al., 2011).

1.7.4 Protease inhibitors

Protease inhibitors (PIs) target the cleavage of the gag-pol polyprotein and subsequent maturation step of the virus life cycle and were a crucial development in HIV treatment as it provided a second class of drug targeting a different enzyme enabling HAART. To date, there are 8 licensed PIs used in the treatment of HIV-infected patients. PIs are designed to imitate the polyprotein cleavage sites, acting as competitive inhibitors, allowing them to bind to the protease active site. However, they cannot subsequently be cleaved by the protease. This therefore inhibits viral polyprotein cleavage and prevents viral maturation (Fun et al., 2012). PIs are usually boosted with ritonavir as it inhibits the host enzyme responsible for the metabolism of PIs, increasing plasma drug concentrations and therefore allowing for lower dosage of PIs (Kempf et al., 1997).

1.8 Drug resistance

Drug resistance is caused by mutations that occur in the viral genes that are targets for antiretroviral therapy and in some cases, other areas of the viral genome (Menendez-Arias, 2010). These mutations are a consequence of the error prone nature of the viral RT during genome replication, its tendency to recombine during reverse transcription and the rapid turnover rate of viral replication. The mutations can arise during suboptimal therapy or drug monotherapy which allow the continual replication and positive selection of resistant minority variants. Suboptimal therapy can result from patient non-adherence to the drug regimen, for example due to the toxic side effects of some of the drugs and the high pill burden. New developments have addressed the issue of patient adherence by simplifying dosage schedules, for example combining triple therapy into one pill (e.g., Atripla, which contains EFV, FTC and TDF and the QUAD pill) (Cohen et al., 2011) and boosting single drugs with cobicistat or

1.8.1.1 NRTI resistance mechanisms

Resistance to NRTIs can occur through two different mechanisms; discrimination or excision of the inhibitor. In the former, the viral RT is able to discriminate between the natural dNTP required for replication and the NRTI (Sarafianos et al., 1999; Gao et al., 2000; Feng et al., 2006). Some mutations for example, Q151M (in association with other accessory/secondary mutations), can enhance the interaction of RT with the 3' hydroxyl group of the natural dNTP which is not found in their drug counterparts, therefore allowing preferential binding to natural dNTPs (Deval et al., 2002). Other discriminatory resistance mutations include M184V/I, K65R, L74V/I and Y115F, the most common being M184V (Menendez-Arias, 2010).

The second NRTI resistance mechanism, excision, occurs when a pyrophosphate donor (thought to be ATP *in vivo*) attacks and cleaves the ultimate 3' phosphodiester bond of the incorporated NRTI, removing it from the chain thereby allowing elongation to continue (Meyer et al., 1998). Thymidine-associated mutations (TAMs) at amino acids 41, 67, 70, 210, 215 and 219, have been shown to enhance the binding of ATP in the correct orientation allowing an attack on the 3' phosphodiester bond of the chain terminating residue (Ehteshami and Gotte, 2008). TAMs are selected only by the thymidine analogues AZT and d4T; however, multiple TAMs cause significant resistance to other NRTIs. There are two overlapping mutational patterns associated with TAMs; Type 1 (M41L, L210W and T215Y), which are associated with higher levels of cross-resistance and Type 2 (D67N, K70R, T215F and K219QE) (Menendez-Arias, 2010). Clinically, TAMs are becoming less frequent due to the reduced use of AZT and d4T.

1.8.1.2 NNRTI resistance mechanisms

NNRTI resistance mutations occur in the hydrophobic NNRTI binding pocket and prevent binding of the drug to RT. Certain RT residues, including the eight known positions of primary drug resistance, 100, 101, 103, 106, 138, 181, 188 and 190, according to the Stanford HIV database (Shafer, 2006), are known to be involved in the stabilisation of NNRTI binding and when these are altered it disrupts the interactions between RT and the NNRTI (Das et al., 2007; Ren et al., 2001; Ren et al., 2004; Ren et al., 2006). For example, mutation of the wild-type tyrosine at 181 to cysteine is thought to weaken the contact between the pyridine groups of nevirapine and the aromatic side chains of tyrosine (Ren et al., 1995).

1.8.1.3 Resistance mechanisms of mutations in the C terminal region of reverse transcriptase

Recent studies have implicated mutations in the C terminal region of RT, which is not included in genotypic or phenotypic resistance testing, in resistance to RTIs (Hirsch et al., 2008). This includes the connection (cn) subdomain and the RNase H (rh) domain, encompassing amino acids 322-440 and 441-560, respectively. This raises the issue of whether the C terminal region of RT should be included in standard resistance tests.

Cn and rh mutations are located away from the binding sites of RTIs and therefore exclude any direct effects on their binding to RT. It has been shown that these mutations act in synergy with TAMs in the pol domain of RT, and indeed a positive correlation between TAMs and C terminal region mutations is seen in treatment-experienced patients (Yap et al., 2007; Delviks-Frankenberry et al., 2008; Nikolenko et al., 2005). It is hypothesised that there are two main resistance mechanisms to AZT attributed to the C terminal region of RT. Firstly, an RNase H-dependent mechanism (Figure 1.9), proposed by Nikolenko *et al.* in which C terminal mutations that reduce RNase H activity allow RT more time to excise the AZT-MP (Nikolenko et al., 2007). This enables the re-initiation of DNA synthesis before the RNA template has been degraded which would lead to the dissociation of the RT-template-primer complex (Delviks-Frankenberry et al., 2008).

The second mechanism is independent of RNase H activity. It is thought that cn subdomain mutations can directly affect the nucleotide excision activity of RT (Delviks-Frankenberry et al., 2008; Ehteshami et al., 2008). It is also thought that C terminal mutations can change the structure of the C terminal region of RT which in turn affects the pol active site via long-range interactions (Zelina et al., 2008). This could occur by the mutations influencing the binding of the RNA:DNA substrate to RT in a polymerase competent mode that favours nucleotide excision over RNase H cleavage (Brehm et al., 2007). A third mechanism that cannot be discounted is that C terminal region mutations may also counteract enzymatic defects in the pol domain (Delviks-Frankenberry et al., 2010).

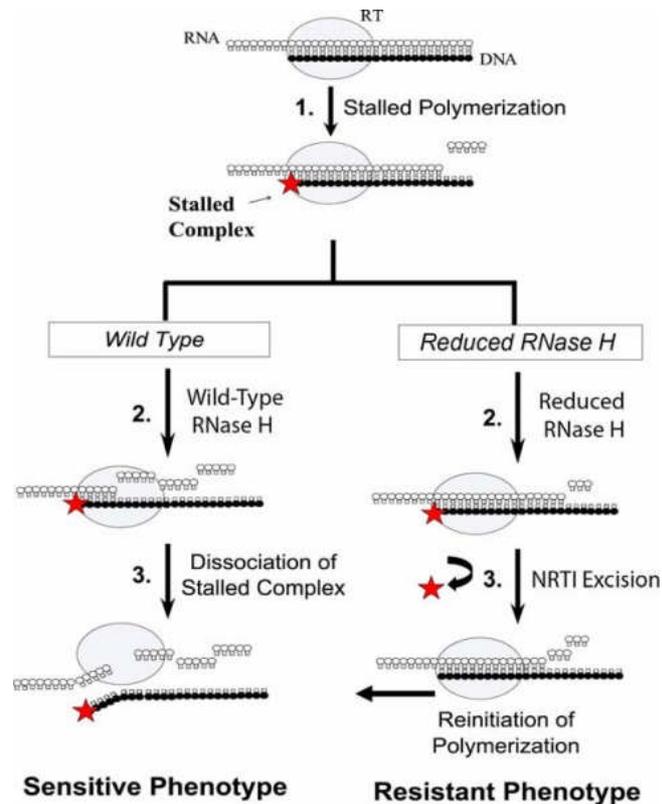


Figure 1.9 An RNase H-dependent C terminal NRTI resistance mechanism. Incorporation of AZT during reverse transcription leads to a stalled polymerisation complex. RNase H activity of RT in a sensitive virus causes the dissociation of the stalled complex. A resistant virus with reduced RNase H activity however, allows more time for the NRTI to be excised allowing resumption of DNA synthesis. Grey oval: reverse transcriptase; red star: AZT; white circles: RNA; black circles: DNA (Delviks-Frankenberry et al., 2010).

C terminal NNRTIs resistance mechanisms can also be grouped into RNase H-dependent and -independent mechanisms (Figure 1.10). The RNase H-dependent mechanism, described by Nikolenko and colleagues depends on the affinity of the NNRTI to RT. During DNA synthesis, the bound NNRTI can dissociate from RT thereby alleviating the constraint on polymerisation. If RNase H activity is reduced it allows the NNRTI more time to dissociate. However, if the NNRTI binding affinity for RT is high the dissociation will take longer and it is likely that the RNase H will degrade the RNA template resulting in the dissociation of the DNA primer from the RNA template. Therefore, in the presence of C terminal mutations that reduce RNase H activity it is expected that NNRTI resistance would increase further (Nikolenko et al., 2010). It

has been suggested that an RNase H-independent resistance mechanism is caused by long range effects of the C terminal region on the NNRTI binding pocket resulting in reduced NNRTI binding (Delviks-Frankenberry et al., 2010).

One of the most characterised of the C terminal region resistance mutations is the N348I, located in the *cn* subdomain. This resistance mutation causes high levels of resistance to both NRTIs and NNRTIs. It is highly associated with TAMs and M184V in the N terminal region but has actually been shown to appear first during RTI therapy (Yap et al., 2007; Gotte, 2007). It has been shown to decrease RNase H activity and increase ATP- P_i-mediated nucleotide excision (Ehteshami et al., 2008; Gotte, 2007). Examples of C terminal resistance mutations and their biochemical mechanism of action are shown in Table 1.3.

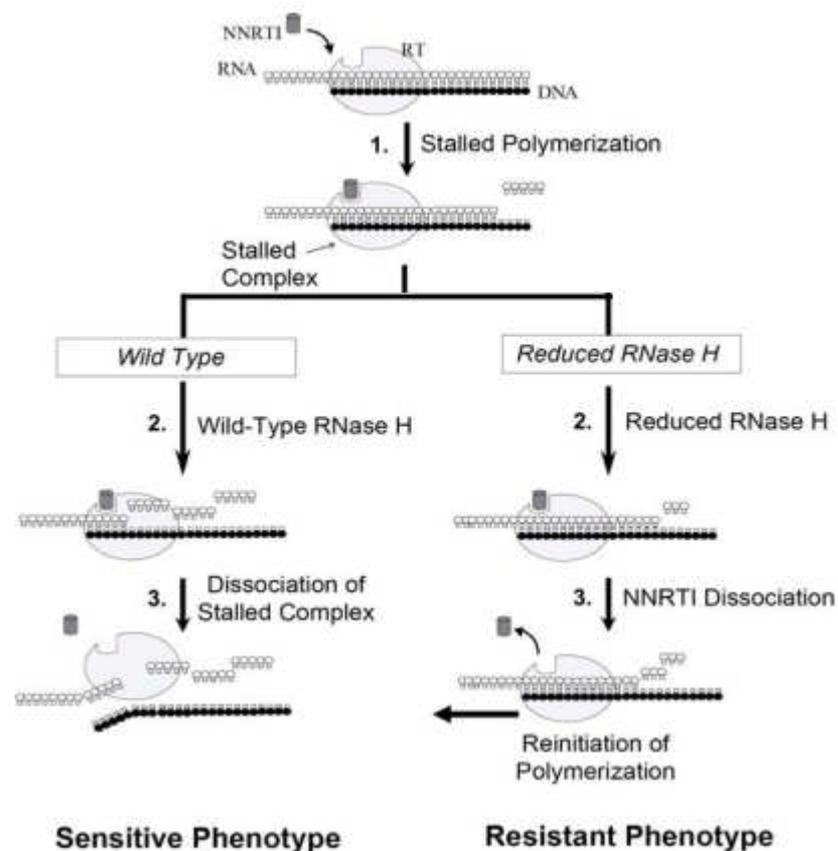


Figure 1.10 An RNase H-dependent C terminal NNRTI resistance mechanism. Incorporation of an NNRTI during reverse transcription leads to a stalled polymerisation complex. RNase H activity of reverse transcriptase in a sensitive virus causes the dissociation of the stalled complex. A resistant virus with reduced RNase H activity however, allows more time for the NNRTI to be dissociated allowing resumption of DNA synthesis. Grey oval: reverse transcriptase; grey cylinder: NNRTI; white circles: RNA; black circles: DNA (Delviks-Frankenberry et al., 2010).

Table 1.3 C terminal RTI resistance mutations and their biochemical mechanisms

C terminal resistance and/or associated mutations	RTI resistance	Biochemical mechanism
G333D/E (Kemp et al., 1998; Zelina et al., 2008)	Resistance to zidovudine and lamivudine in combination with M184V and TAMs	Selected <i>in vitro</i> Increases the ability of RT to bind AZTMP template/primer terminated complexes, allowing ATP-mediated excision Increased discrimination between dCTP and 3TC-triphosphate
G335C/D and E312Q (Nikolenko et al., 2007)	Increase AZT resistance	Selected <i>in vivo</i> Reduces RT template switching thereby increasing time available for nucleotide excision (May alter affinity of RT for template primer)
N348I (Sluis-Cremer et al., 2010; Yap et al., 2007; Delviks-Frankenberry et al., 2008; Nikolenko et al., 2007; Ehteshami et al., 2008; Hachiya et al., 2008)	Increases NVP and AZT resistance Low level resistance to ETR Increased ETR resistance when combined with Y181C TDF resistance when combined with TAMs	Selected <i>in vivo</i> by NVP and AZT Reduces RNase H activity Increased DNA synthesis in combination with TAMs (reduced RNase H activity and increased processive DNA synthesis facilitates the recruitment of PPI as a substrate for the excision reaction) May indirectly affect primer binding in p66 subunit (may alter the flexibility and mobility of the p66 thumb subdomain and reposition amino acids present in the NNRTI-binding pocket)
A360I/V (Nikolenko et al., 2007; Santos et al., 2008; Ehteshami et al., 2008)	Increases AZT resistance	Selected <i>in vivo</i> Part of the rh primer grip (may influence proper positioning of the primer to the RNase H active site) Increases processive DNA synthesis in combination with TAMs
V365I and A376S (Nikolenko et al., 2007)	Increase AZT resistance	Selected <i>in vivo</i> Reduce RNase H activity

T369I/A/V (Gupta et al., 2010b)	In combination with N348I resistance to NNRTI and AZT increases	Selected <i>in vivo</i> Located close to RT heterodimer interface of p66 and p51 and may affect NNRTI resistance by altering RT dimerisation
A371V and Q509L (Brehm et al., 2007; Dau et al., 2010)	Increase AZT resistance when combined with TAMs Increase 3TC, ABC and TDF resistance	Selected <i>in vitro</i> and <i>in vivo</i> Enhance RT-mediated AZT-MP excision

1.8.2 Resistance to raltegravir

Three different major mutation pathways have been described that cause resistance to RAL *in vitro* and *in vivo* from viruses isolated from patients failing on RAL treatment. These involve amino acid substitutions primarily in the IN catalytic core, at positions 148, 155 and less frequently, 143 (Cooper et al., 2008). These primary mutations are often accompanied by one or more secondary, accessory mutations which usually increase resistance and/or compensate for a replication deficit (Malet et al., 2008; Fransen et al., 2009a; Goethals et al., 2010; Baldanti et al., 2010; Canducci et al., 2010; Reigadas et al., 2010; Cooper et al., 2008). Some of these secondary mutations, for example L74I, T97A, E138K, V151I, G163R, V165I, V201I, I293M and T206S are natural polymorphisms which may influence the rate of resistance development in viruses that contain these mutations prior to RAL therapy (Ceccherini-Silberstein et al., 2008; Canducci et al., 2010; Myers and Pillay, 2008; Fransen et al., 2009a; Rhee et al., 2008). One primary mutation, however, is sufficient to cause a high level of resistance to RAL (Malet et al., 2008; Ferns et al., 2009). The three primary resistance pathways are mutually exclusive, in that they never occur on the same genome, (Ferns et al., 2009; Charpentier et al., 2008; Malet et al., 2009; Reigadas et al., 2010; Fransen et al., 2009a) but studies have shown that different pathways can be present in the same viral population within patients undergoing virological failure (Canducci et al., 2010; Malet et al., 2009).

It has been suggested that the evolution of RAL resistance is driven by the need to increase levels of resistance during drug selective pressure (Canducci et al., 2010; Fransen et al., 2009a). Indeed, switches from the 155 to the 148 pathway have been observed during treatment failure, and is thought to be possibly linked to higher resistance levels seen with the 148 pathway (Malet et al., 2009). Switches from both the 148 and 155 pathways to the 143 pathway have also been observed during treatment failure (Reigadas et al., 2010; Sichtig et al., 2009; Canducci et al., 2010). It is thought that the 143 pathway may have a replicative advantage over the other two pathways in the presence of RAL which may explain its eventual emergence during RAL therapy (Delelis et al., 2009). Interestingly, studies on the cessation of RAL treatment have shown that INSTI resistance mutations are quick to disappear, which demonstrates that these mutations may come at a fitness cost to the virus (Fun et al., 2010; Ferns et al., 2009). It is worthwhile noting that elvitegravir (EVG), the next most advanced INSTI, has been shown to have a common binding mechanism to RAL and therefore there may

be considerable cross resistance between RAL resistance mutations and EVG resistance mutations (Marchand et al., 2009; Hare et al., 2010).

1.8.2.1 The Q148R/H/K pathway

Arginine, histidine or lysine substitutions of the glutamine residue at position 148 have been shown to reduce the susceptibility to RAL and also cause a reduction in replication capacity of the virus (Delelis et al., 2009; Fransen et al., 2009a). This is perhaps why substitutions at this position often occur with the secondary mutations G140S/A or E138K, in order to compensate for this deficit in replication capacity (Fransen et al., 2009a; Malet et al., 2009; Metifiot et al., 2010; Delelis et al., 2009). These secondary mutations appear after the establishment of the 148 primary mutation and have been shown to increase the levels of RAL resistance and/or replication capacity (Fransen et al., 2009a; Malet et al., 2009; Nakahara et al., 2009), though interestingly, this is dependent on the nature of the amino acid substitution at position 148 (Fransen et al., 2009a). A study by Fransen *et al.* showed that on addition of G140S to a Q148R/H mutant, susceptibility to RAL was reduced compared to Q148R/H alone. However, addition of the G140S mutation to Q148K increased susceptibility to RAL. They also showed that the Q148R/H/K mutants had a reduced replication capacity compared to wild-type, but with the G140S mutation, replication capacity increased only for amino acid substitutions H and K (Fransen et al., 2009a).

A study by Metifiot *et al.* further revealed that the G140S/A mutants alone impaired integrase strand transfer activity without affecting the 3' processing reaction; however, the R, H and K mutants at 148 showed defective activities in both strand transfer and 3' processing. They hypothesised that substituting glutamine with arginine, histidine or lysine (all of which have longer and larger side chains) alters binding of the viral DNA to IN, therefore affecting both the strand transfer and 3' processing activities. However, through complementation studies they showed that only the G140S/Q148H double mutant was capable of restoring IN activities to near wild-type levels, and that this complementation only operates in *cis*. Metifiot *et al.* concluded that this double mutant was the only mutant to be able to restore the active site allowing 3' processing and strand transfer activities, whilst not re-establishing the RAL binding site (Metifiot et al., 2010). This may explain the evolution towards this replication efficient double mutant virus in patients failing on RAL treatment. Other secondary mutations associated with the Q148R/H/K resistance pathway can be found in Table 1.4.

Table 1.4 Secondary resistance mutation patterns associated with Q148H/R/K

Associated secondary mutations	Effects on drug resistance or viral fitness of Q148 mutants
N17S	
N17S + G163R	
V54I + E138K + G140A	Increases resistance to raltegravir and elvitegravir compared to addition of E138K + G140A; increases viral fitness
L63I + L74M + A128T + E138K + V151I	
L74M + G140A	
V79I + G140A + G163R	
E92Q + E138K + M154I	
T97A	
T112A + G140S + G163R	Increases raltegravir resistance compared to addition of G140S + G163R
H114Y + A128T	Increases elvitegravir resistance
T124A	
E138K/A	Depending on Q148 mutation, increases resistance to raltegravir and elvitegravir; increases viral fitness
E138K + G140A	Increases resistance to raltegravir and elvitegravir compared to addition of G140A alone; increases viral fitness
E138A + G140S + Y143H	Increases raltegravir resistance compared to addition of G140S alone
E138K + G140A + S230R	Increases resistance to raltegravir compared to addition of E138K + G140A
E138K + G163R	
G140A/S/C	Depending on Q148 mutation, increases resistance to raltegravir and elvitegravir; increases viral fitness
G140S + N155H	
G140S + K156N	
G140S/C + G163R/K	
G140S + E170A	
N155H	
N155H + E170A	
S147G	
G163R	

Taken from Mbisa *et al.* (Mbisa *et al.*, 2011a)

1.8.2.2 The N155H Pathway

This pathway is generally associated with lower RAL resistance than the 148 pathway, which may explain its eventual disappearance and replacement with either the 148 or 143 pathways during RAL treatment failure (Fransen et al., 2009a; Reigadas et al., 2010). This mutant has been shown to reduce the replication capacity of the virus; specifically, it has been shown to have an impaired strand transfer activity and to some extent impaired 3' processing activity (Malet et al., 2008; Fun et al., 2010).

A mechanism by which the N155H mutant causes resistance to RAL has been reported by Grobler *et al.* They proposed that the N155 residue in the IN active site interacts with the residues responsible for binding the magnesium cations required for IN activity. As RAL binds to IN through interactions with these metal ions, mutation of this residue may prevent RAL from binding by disrupting the metal ion-active site arrangement (Grobler et al., 2008).

Interestingly, studies have shown that the N155H pathway has a broad range of reductions in RAL susceptibility dependent on which secondary mutations have also been accumulated (Fransen et al., 2009a). However, it has been suggested that the 155 pathway does not confer a stable state of RAL resistance regardless of the accumulated secondary mutations. Malet *et al.* showed that the secondary mutations in quasispecies containing the N155H primary mutation in patient samples varied greatly over time (Malet et al., 2009).

The E92Q mutation has been shown to preferentially occur with N155H as well as L74I, T97A, E138K, V151I, G163R and I203M by clonal analysis (Fransen et al., 2009a; Sichtig et al., 2009; Malet et al., 2009; Fun et al., 2010). Fransen *et al.* also showed that with the E92Q mutation both the replication capacity and RAL susceptibility of the N155H mutant were reduced further. It was suggested that selection of E92Q after the establishment of N155H in the virus is due to the increase in RAL resistance rather than compensating for the replication capacity deficit (Fransen et al., 2009a). However, a recent study has shown that subsequent selection of the Q95K mutation together with the N155H mutation resulted in a virus with an increased RAL resistance as well as a partially restored replication capacity. Furthermore, upon addition of the V151I mutation, replication capacity was reduced but resistance was further increased, generating a virus with an overall higher level of fitness in the presence of RAL. Indeed, this triple mutant virus was shown to rapidly dominate the viral population in one patient during RAL treatment failure (Fun et al., 2010). Other secondary mutations associated with the N155H resistance pathway can be found in Table 1.5.

Table 1.5 Secondary resistance mutation patterns associated with N155H

Associated secondary mutations	Effects on drug resistance or viral fitness of N155 mutant
V72I	
V72I + E92G	
L74M + E92Q + V151I + E157Q	
E92Q/A/G	Increases raltegravir and elvitegravir resistance
E92Q + T97A	Increases raltegravir resistance compared to addition of E92Q alone
E92A + G163R	
Q95K	Increases raltegravir and elvitegravir resistance; increases viral fitness
Q95K + V151I	Increases raltegravir and elvitegravir resistance compared to addition of Q95K alone
T97A	Increases raltegravir and elvitegravir resistance; increases viral fitness
T97A + V125A + V151I	Increases raltegravir resistance compared to addition of V125A + V151I
T97A + V151I	Increases raltegravir resistance compared to addition of T97A alone
T124A + V151I	
V125A + V151I	
G140S	
G140S + Q148H	
Y143R/H	Increases raltegravir resistance
Y143R + E170A	
Q148R/H	
Q148H + E170A	
V151I	Increases raltegravir resistance
V151I + M154I	
V151I + G163R	
M154I	
G163R/K	Increases raltegravir and elvitegravir resistance; increases viral fitness
I204T	
D232N	

Taken from Mbisa *et al.* (Mbisa *et al.*, 2011a)

1.8.2.3 The Y143R/C Pathway

As with the other resistance pathways, mutations at position 143 in the viral *IN* gene have been shown to increase resistance to RAL whilst reducing the replication capacity of the virus (Canducci et al., 2010; Delelis et al., 2010; Reigadas et al., 2010). However, the level of RAL resistance exhibited by the 143 pathway has been shown to be higher than that of the Q148H/G140S double mutant and other primary mutations (Delelis et al., 2010). *In vitro* studies by Delelis *et al.* showed that Y143R/C mutants have impaired 3' processing and strand transfer activities. It was shown that 3' processing was more affected for the R mutant than the C mutant whereas the strand transfer activity was more efficient in the R mutant. Interestingly, both mutants had similar levels of resistance to RAL. This suggested that the Y143C mutant could be a transient form of the Y143R mutant, as substitution from a tyrosine to a cysteine involves 1 base pair (bp) change whilst substitution to an arginine requires 2 bp changes. Modelling analysis suggested that the Y143R/C mutation may allow for an alternative recognition site for DNA binding, in particular, the post-3' processing contact with the 5'AC overhang, whilst preventing the binding of RAL to the enzyme, similar to the observation with the Q148H/G140S double mutant (Delelis et al., 2010).

In another study the strand transfer activity of the Y143R/C mutants was shown to be severely impaired, but the 3' processing activity was only moderately affected. The study showed that these mutants were highly resistant to RAL but had a lower replication capacity than wild-type virus. Unpublished data from Fransen *et al.*, suggested that perhaps the PR and/or RT of the virus could be playing a role in rescuing the replication capacity deficit induced by the Y143R/C mutations (Reigadas et al., 2010).

Studies have shown that the Y143R/C mutations can be present either alone or in association with other secondary mutations at positions 72, 74, 92, 97, 163, 203 and 230 (Canducci et al., 2010; Malet et al., 2009; Reigadas et al., 2010) (Table 1.6). Little is yet known about the effects of these secondary mutations on the RAL resistance levels and replication capacity of the virus. One study however has shown that the addition of the T97A mutation with Y143R/C greatly reduces RAL susceptibility whilst also rescuing the catalytic defect of IN (Reigadas et al., 2011). Indeed, it is widely believed that these mutations, and other natural polymorphisms in the *IN* gene, play a positive role in IN activity and or sensitivity to RAL (Reigadas et al., 2010; Delelis et al., 2010; Canducci et al., 2010).

Table 1.6 Secondary resistance mutation patterns associated with Y143C/H/R/K

Associated secondary mutations	Effects on drug resistance or viral fitness of Y143CHRK mutants
L74M + T97A	
L74M + T97A + S119T + E138D	
L74M + T97A + E138A	Increases raltegravir resistance compared to addition of T97A alone
L74M + E138D	
T97A	Increases raltegravir resistance
T97A + E138A	
T97A + G140D + G163R	
G140S	Increases raltegravir resistance
N155H	
N155H + E170A	

Taken from Mbisa *et al.* (Mbisa *et al.*, 2011a)

1.8.3 Resistance to elvitegravir

EVG has a slightly improved resistance profile compared to RAL as HIV replication remains unaffected by substitutions at position 143 in the presence of EVG. There is still however a broad range of cross-resistance between the two INSTIs and high levels of resistance to EVG are observed with mutations at primary RAL resistance sites 148 and 155 (Metifiot *et al.*, 2011; Abram *et al.*, 2012).

In vivo and *in vitro* EVG selected mutations involve substitutions at amino acids 66, 92, 138, 140, 145, 147, 148 and 155 which includes RAL-associated primary and secondary mutations Q148R/H/K, N155H and E92Q (Shimura *et al.*, 2008; Waters *et al.*, 2009; McColl *et al.*, 2007). Similar to RAL resistant genotypes, it has been shown that EVG resistance patterns can switch during EVG therapy. In a study of 38 EVG-treated patients, 8 patients showed a genotypic switch from the E92Q to the Q148R/H resistance pathway which is thought to be associated with a higher level of resistance (Waters *et al.*, 2009).

Primary EVG resistance mutations consist of T66I, E92Q, S147G, Q148H/R/K and N155H, as determined by the Stanford University HIV drug resistance database (Shafer, 2006). As with RAL, mutations from glutamine to arginine or lysine at position 148 result in a higher resistance to EVG than the N155H resistance mutation. The E92Q and T66I EVG resistance mutations show a similar level of resistance to N155H whilst S147G confers the lowest level of resistance (Abram *et al.*, 2012).

Similar to RAL primary resistance mutations, EVG resistance mutations are associated with a loss of fitness (Shimura et al., 2008; Abram et al., 2012) and it appears that the replication capacities of these mutants are inversely correlated to levels of resistance (Marinello et al., 2008).

1.8.4 Resistance to protease inhibitors

There are 15 amino acid positions in PR that are known sites of primary drug resistance. Most PI primary resistance mutations lead to a conformational change in and around the PR active site at amino acids 30, 32, 46, 47, 48, 50, 54, 82 and 84, causing a reduced affinity for the inhibitor (Carrillo et al., 1998; Tie et al., 2005; Prabu-Jeyabalan et al., 2003; Ali et al., 2010; Shafer, 2006). In contrast, substitutions at amino acids 58, 74, 76, 83, 88 and 90 cause PI resistance indirectly in an as yet undefined mechanism (Ali et al., 2010). It is worth noting that most of the resistance mutations confer varying levels of resistance to multiple PIs and a major resistance mutation to one drug can be a minor resistance mutation to another (Johnson et al., 2011). In addition to causing reduced susceptibility to PIs these mutations also cause a decrease in the replication capacity of the virus. This can be compensated by the acquisition of secondary mutations elsewhere in PR as well as other regions of the HIV-1 genome. Unlike primary resistance mutations, some secondary mutations such as K20R, M36I, I62V, L63P, V77I and I93L, are common polymorphisms (Martinez-Cajas and Wainberg, 2007; Shafer, 2006).

Mutations in *gag* have also been found to contribute to resistance to multiple PIs. Most occur at *gag* cleavage sites and can either cause resistance to PIs directly or are strongly associated with resistance mutations in *PR* and act as compensatory or additive resistance mutations, for example Q430R, A431V and L449F in the NC/SP1 and SP1/p6 cleavage sites (Clavel and Mammano, 2010). Other mutations found in the *CA* and *MA* genes have also been found to contribute to PI resistance (Parry et al., 2009). The mechanisms by which mutations in *gag* cause resistance are yet to be fully understood.

1.8.5 Resistance to other antiretrovirals

Unfortunately, resistance also develops to the licensed ARVs targeting viral entry namely, maraviroc and enfuvirtide. Mechanisms of maraviroc resistance include the selection for intrinsic CXCR4-tropic viruses (Dorr et al., 2005) and the acquisition of mutations in the *env* gene which allow for CXCR4 co-receptor usage (Pastore et al., 2004). Alternatively, viruses can

accumulate mutations which allow CCR5 co-receptor use despite the block imposed by maraviroc and this is thought to be the main mechanism of resistance (Westby et al., 2007). A study by Westby *et al.* indicated that positions 316 and 323 in the V3 loop of *env* were key residues in resistance as reversion of both residues back to wild-type restored maraviroc susceptibility to wild-type levels (Westby et al., 2007).

Enfuvirtide resistance is conferred by mutations in the peptide interaction site, codons 36 to 38 of gp41 and proximal sites (Menendez-Arias, 2010). A study by Wei and colleagues found common resistance mutations in patients undergoing enfuvirtide monotherapy to be G36D, Q32R/H, I37V, V38M/A and Q39R (Wei et al., 2002), which have also been identified in other studies along with mutations Q40H, N42T and N43D/T/S (Marcelin et al., 2004; Menzo et al., 2004; Xu et al., 2005). As with resistance patterns to other ARVs, primary resistance mutations confer a fitness loss which can be compensated for by secondary mutations. For example, in a study of 17 patients receiving long-term enfuvirtide treatment the S138A mutation appeared after the acquisition of N43Q in almost all cases and resulted in a further increase in enfuvirtide resistance by three-fold (Xu et al., 2005). Further analysis demonstrated that the S138A mutation compensated for the impaired fusion caused by N43Q mutation (Izumi et al., 2009).

1.8.6 Transmitted drug resistance

The first report of transmitted drug resistance (TDR) was published in 1993 (Erice et al., 1993) when a patient with primary HIV-1 infection was identified to be infected with a virus resistant to zidovudine due to the T215Y NRTI resistance mutation. It appeared that the patient had a sexual partner who was HIV-positive and was receiving zidovudine-containing therapy. It was assumed that this was the mode of transmission as no other risk factors were found. Since then multiple reports have been published documenting both single- and multi-class TDR through sexual (Conlon et al., 1994; Conway et al., 1999; Hecht et al., 1998; Yerly et al., 1999), parenteral (Veenstra et al., 1995; de et al., 1996; Yerly et al., 1999) and vertical (Johnson et al., 2001; Colgrove et al., 1998) transmission. For this reason a list of TDR mutations was compiled to enable TDR surveillance and this is regularly updated (Bennett et al., 2009).

There are currently 93 drug resistance mutations that can be classified as TDR mutations for surveillance in newly infected, untreated persons (Table 1.7). These include 34 NRTI resistance

mutations at 15 RT positions, 19 NNRTI resistance mutations at 10 RT positions, and 40 PI resistance mutations at 18 PR positions (Bennett et al., 2009).

The transmission of INSTI drug resistance has not yet been widely reported, with only two reports to date. This could be because RAL is a relatively new drug. The first documented case of a transmitted INSTI resistant virus occurred in 2010. Genotypic analysis of a patient's pre-treatment sample identified major resistance mutations in the *PR* (L10I, V32I, M46I, A71V, V82A and L90M), *RT* (K70K/R, K103N and V106A) and *IN* (G140S and Q148H) genes (Fransen et al., 2010). The second case also involved the transmission of a multi-class drug resistant virus. Drug resistance mutations included L10I/V, V11I, I13V, G16E, K20I, L33F, M36I, M46L, Q58E, H69K, T47P and L89M in *PR*, T215Y and K103N in *RT* (at low levels) and N155H, E157Q and G163R in *IN* (Boyd et al., 2011).

Table 1.7 Mutations currently used for transmitted drug resistance surveillance

Drug class	Transmitted drug resistance mutations
NRTI	41L, 65R, 67N/G/E, 69D/ins, 70R/E, 74V/I, 75M/T/A/S, 77L, 115F, 116Y, 151M, 184V/I, 210W, 215Y/F/I/S/C/D/V/E, 219Q/E/N/R
NNRTI	100I, 101E/P, 103N/S, 106M/A, 179F, 181C/I/V, 188L/H/C, 190A/S/E, 225H, 230L
PI	23I, 24I, 30N, 32I, 46I/L, 47V/A, 48V/M, 50V/L, 53L/Y, 54V/L/M/A/T/S, 73S/T/C/A, 76V, 82A/T/F/S/C/M/L, 83D, 84V/A/C, 85V, 88D/S, 90M

Data from Bennett *et al.* (Bennett et al., 2009)

1.9 The future of HIV-1 treatment

Despite HIV-infected patients having a near normal life expectancy in the advent of HAART (Reid et al., 2013), there still remains the excess risk of heart disease and toxicity (reviewed by Flexner and Saag, 2013 and Boccarda et al., 2013). In addition, ARV therapy cannot eradicate or fully cure HIV infection. As the arena of ARV drug development becomes crowded and the financial implications of life-long treatment become apparent as well as the long term adverse effects of treatment and HIV infection itself, researchers are now looking towards strategies to cure HIV and develop novel therapeutic strategies. There are two potential strategies for cure; functional and sterilising. Strategies working towards a functional cure involve developing ways in which HIV infection can be controlled long-term, in the absence of HAART, mirroring

what is seen in long-term non-progressors, also called elite controllers, who are able to control HIV infection without the use of ARV therapy. Recently, claims of functionally curing a baby and fourteen people of HIV infection have come to light. The infected baby and the 14 infected adults were found to be able to control their infection upon treatment interruption for several years (in the case of the adults). The principle of “cure” appears to be centred on early treatment with HAART which is thought to reduce viral reservoirs and preserve immune responses thereby allowing successful infection control in the absence of ART (Roehr, 2013; Saez-Cirion et al., 2013).

The second strategy, developing a sterilising cure, involves eliminating all HIV-infected cells in a patient (Lewin and Rouzioux, 2011). This strategy was recently demonstrated in an HIV-infected patient suffering from leukemia in Germany who was “cured” of the infection after a bone marrow transplant from a donor with a 32 bp deletion in the CCR5 allele (Allers et al., 2011; Hutter et al., 2009). However, the validity of the cure is still inconclusive as recent evidence seems to indicate a detection of intermittent low levels of virus in the patient (Yukl et al., 2012). The CCR5 and CXCR4 co-receptors are crucial for HIV infection and it has been shown that a homozygous 32 bp deletion in the CCR5 allele is responsible for the absence of HIV infection in some multiply exposed individuals (Liu et al., 1996). A new report of a sterilising cure also as a consequence of bone marrow transplantation was recently described. In these cases, two men received stem cells which did not contain the protective mutation. This suggests that protection of the newly transplanted cells was through ART which both patients were receiving at the time (Mascolini, 2013). Nevertheless, this bone marrow transplant strategy appears unrealistic for use in the wider infected population due to the complexity and toxicity of the treatment.

1.9.1 Problems of HIV eradication

The major barrier defying attempts for a cure is the persistence of HIV in all patients on HAART. Viral persistence can take different forms; latently infected CD4+ T cells (Chun et al., 1997a; Chun et al., 1997b; Siliciano et al., 2003; Finzi et al., 1997), residual viral replication in cells (Chun et al., 2000; Zhu et al., 2002) and its presence in different anatomical reservoirs which have little penetration of HAART, for example the central nervous system, the retina and the testes (Pomerantz, 2002).

Particularly relevant to this thesis is the integration of HIV DNA into resting or activated CD4+ T cell (in the latency phase) which results in the establishment of stable and latently infected cell reservoirs (Chun et al., 1997a; Chun et al., 1997b; Siliciano et al., 2003; Finzi et al., 1997). Latency in these reservoirs may be maintained through various mechanisms including repressed transcription and negative regulation of T cell activation (Coiras et al., 2009). These integrated proviral reservoirs are problematic in cure strategies as they are concealed from the host immune system and in addition remain out of reach from current ARV drugs (Pomerantz, 2002).

1.9.2 Novel therapeutic strategies

Strategies have been devised and are being developed to overcome the problems of viral persistence with the aim of a functional or sterilising cure of HIV infection. One strategy to eliminate or reduce the latent HIV reservoir is by activating latently infected T cells causing the release of viral particles. The administering of HAART would then block further rounds of viral replication whilst the infected cell would die. In theory, the numbers of latently infected T cells would decline thereby reducing the latent reservoirs with the potential for complete elimination of the virus from the patient. This relies on HAART completely inhibiting viral replication. Compounds used for activating latently infected cells include IL-7, prostratin and histone deacetylase inhibitors (Lewin and Rouzioux, 2011).

Another proposed strategy to overcome viral persistence and cure patients of HIV is to eliminate or reduce residual viral replication. A number of studies in patients receiving HAART have looked at the effect of treatment intensification on residual viral replication (Buzon et al., 2010; Gandhi et al., 2012; Vallejo et al., 2012; Massanella et al., 2012), the results of each study however appear to be discordant. Treatment intensification involves the addition of ARV drugs, such as RAL, to an already suppressive regimen. In a study by Buzon *et al.*, intensification with RAL resulted in the increase of episomal 1- and 2-long terminal repeat (LTR) circles; a measure of unsuccessful HIV viral integration, however, total and integrated DNA levels remained stable. In addition, the study showed a decline in CD8+ T cell activation in patients with an increase in 1- and 2-LTR circles during RAL intensification, a specific indicator of reduced levels of HIV replication (Buzon et al., 2010). Conversely, a study by Gandhi *et al.* found no change in the levels of 2-LTRs and total viral DNA. They concluded that intensification with RAL had no effect on residual viral replication. Additionally, Gandhi *et al.* suggested the presence of a source of viral replication that is not reached by RAL to be responsible for the

ongoing residual replication or alternatively, that the presence of residual viraemia is not reliant on new rounds of replication for sustaining the low levels of viraemia seen in HAART-suppressed patients (Gandhi et al., 2012).

Strategies for making cells resistant to HIV infection, as per the strategy developed by doctors of the German patient cured of HIV, are also being developed as a novel way to potentially lead to the eradication of HIV from infected individuals. Different gene therapy approaches include eliminating CCR5 expression through engineered zinc-finger nucleases (Holt et al., 2010) and RNA-based gene therapy (DiGiusto et al., 2010; Li et al., 2005b).

Boosting immunity or vaccination against HIV is also an area which is being heavily researched. There are currently thirty-seven on-going clinical trials, the majority of which are still in the Phase I stage (www.IAVIreport.org/Trials-Database). So far there has been little promise of an effective vaccine due to the multiple challenges facing vaccine design including the sheer diversity of HIV (causing problems of antibody-antigen recognition) and the absolute correlates of protection from HIV infection have remained undefined despite extensive research into the host immune response to infection (Munier et al., 2011). Only three vaccines have made it to Phase III clinical trials (ALVAC-HIV [vCP1521], AIDSVAX B/E and ADISVAX B/B) and unfortunately the results have been somewhat disappointing despite showing good efficacy in preclinical animal model studies (Rerks-Ngarm et al., 2009; Pitisuttithum et al., 2006; Flynn et al., 2005).

Unfortunately, it is likely that no single strategy will enable the complete eradication of HIV from infected individuals but a combination of different strategies may prove to be the most effective path towards a functional or sterilising cure for HIV (Lewin and Rouzioux, 2011). In the absence of a broadly effective eradication strategy, researchers are finding novel ways to inhibit already targeted viral processes and also seeking ways in which to inhibit host interactions with viral proteins.

1.9.3 Novel mechanisms of reverse transcriptase inhibition

Researchers are now looking at novel ways to target the RT as resistance to older RTIs is a major limitation to their sustained use. New RTI compounds with novel modes of action currently being explored in pre-clinical studies include translocation defective RTIs (TDRTIs), delayed chain terminator RTIs (DCTRTIs), lethal mutagenesis RTIs and nucleotide competing RTIs (NcRTIs) as well as RNase H inhibitors (Esposito et al., 2012).

1.9.4 Other inhibitors of integration

Dolutegravir (DTG), a second generation INSTI, has recently undergone a phase 3 clinical trial to determine the efficacy of DTG as a first line therapy compared to RAL. The results showed that DTG had a non-inferior efficacy and a similar safety profile to RAL (Raffi et al., 2013). In addition DTG was recently licensed for use in the USA (FDA, 2013). Its mode of action is analogous to that of the first generation INSTIs RAL and EVG; chelation of divalent metal ions at the viral IN active site (Hightower et al., 2011; Kobayashi et al., 2011). However, DTG has an improved resistance profile (Johns, 2010) and as of yet no known drug resistance profiles have evolved in the clinic. This higher barrier for resistance is thought to be due to the tighter binding of DTG to both wild-type and resistant intasomes compared to RAL and EVG (Hightower et al., 2011). One study, using *in vitro* selection saw the development of R356K in IN together with the later appearance of H51Y. The authors determined that whilst R356K conferred low level resistance to DTG, the addition of H51Y dramatically reduced the replication of the virus whilst DTG resistance remained low (Quashie et al., 2012a; Mesplede et al., 2013). Mesplede *et al.* concluded that the lack of resistance evolution in patients was likely due to the production of severely replicative deficient virus, or a dead end virus.

Other IN inhibitors in pre-clinical development include allosteric inhibitors that prevent IN-host factor interactions, for example by inhibiting the LEDGF/p75-IN interaction (Quashie et al., 2012b). IN interaction with LEDGF/p75 is required for the tethering of the PIC to the host chromatin and the recruitment of other cellular factors which facilitate successful integration. Inhibition of this interaction can severely affect viral replication (Christ et al., 2010).

1.9.5 Maturation inhibitors

Maturation inhibitors target the viral maturation process by binding to Gag and blocking cleavage events as opposed to binding to the PR active site akin to PIs. The first in class of maturation inhibitors is bevirimat (BMV). Specifically, BVM inhibits the CA-SP1 cleavage and the formation of mature CA proteins causing the production of non-infectious virions (Adamson et al., 2009; Li et al., 2003). It is hypothesised that BVM binds to a pocket formed during Gag-Gag multimerisation. Phase I and II studies of safety and efficacy have shown maturation inhibition to be a potential target for treatment (Smith et al., 2007). However, it seems that the inhibition of maturation by BMV is affected by the viral sequence at the C

terminal end of the SP1 protein, in particular amino acids 6-8 which are relatively non-conserved among different HIV-1 subtypes (Van et al., 2009b).

1.10 Measuring drug resistance and viral fitness

Drug resistance testing of newly HIV-infected individuals and treated patients showing treatment failure is carried out in resource-rich countries as part of standard care and clinical management of HIV infection (NIH, 2012a; Hirsch et al., 2008). There are two major approaches that are used to measure viral drug resistance and fitness; phenotyping and genotyping. Phenotyping involves measuring viral activity in cell-based assays and tends to be more time consuming than current PCR-based genotyping methods.

1.10.1 Phenotyping

Historically, phenotypic drug susceptibility assays involved the cultivation of patient virus isolates; however, in addition to being very time consuming, the cultivation process can alter the original distribution of viral variants in the primary sample, therefore giving a susceptibility value that does not reflect the actual *in vivo* drug susceptibility (Japour et al., 1993; Kusumi et al., 1992). Nowadays, most phenotypic drug susceptibility assays involve the production of replication-competent or pseudotyped recombinant vectors containing patient-derived genes of interest that are subsequently used in multiple or single-cycle replication assays to determine levels of drug susceptibility as well as infectivity or viral replication capacity. This has the advantage of being faster and also usually provides a better reflection of drug susceptibility and viral fitness in the *in vivo* viral population (Hertogs et al., 1998; Kellam and Larder, 1994; Petropoulos et al., 2000).

The accurate interpretation of resistance assay results is important in the clinical management of patients to enable the determination of successful treatment regimens and to help minimize the development of drug resistance. With this in mind, cut offs have been established to determine when a drug no longer has antiviral activity with respect to a particular clinical isolate. Cut offs can be determined at technical (based on the reproducibility of an assay), biological and clinical levels (Perno and Bertoli, 2006). Biological cut offs are established by determining the mean fold change in susceptibility of a large cohort of isolates from treatment-naïve patients. If the fold change in susceptibility of a clinical isolate falls within the mean plus 2 standard deviations, the standard interpretation is that the isolate is susceptible.

Although biological cut offs provide a comparison between clinical isolates and those circulating in the treatment-naïve, HIV infected population, they do not provide information on the likely virological response to an antiviral. Clinical cut offs, on the other hand, try to establish the relationship between drug susceptibility and the virological response from treatment-experienced patients and often require two cut off values. Firstly, the fold change when antiviral activity starts to reduce, and secondly, the fold change when there is complete loss of antiviral activity (Haubrich, 2004).

The determination of clinical cut offs for individual antivirals has proved to be difficult due to the necessity of large data sets and the use of combination therapy as is the norm (Perno and Bertoli, 2006). As such, clinical cut offs for RAL and EVG have not yet been established.

Several methods have been developed to produce patient-derived recombinant viruses to allow the measurement of drug susceptibility and viral fitness. One of the first methods used to produce recombinant viruses was through homologous recombination in mammalian cells using regions of overlapping homology between patient-derived PCR amplicons and an expression vector in which the gene(s) of interest had been deleted. The PCR products and expression vector are delivered into mammalian cells by electroporation resulting in the recombination of the two DNA fragments and the production of viable virus that can be used to test for drug susceptibility (Kellam and Larder, 1994; Hertogs et al., 1998; Hu and Kuritzkes, 2010). Alternatively, cloning methods such as sequence-specific uracil deglycosylase-mediated cloning (Martinez-Picado et al., 1999) have also been developed to create recombinant vectors for use in drug susceptibility assays. However, the method in frequent use nowadays involves the restriction digestion and ligation of patient-derived PCR products directly into the expression vector utilising natural or engineered restriction sites within the vector (Frater et al., 2001; Parry et al., 2009; Buzon et al., 2008; Petropoulos et al., 2000). In addition, several groups have used site-directed mutagenesis to create singular drug class resistant or multi-class drug resistant viruses (Fransen et al., 2009a; Nakahara et al., 2009; Hu and Kuritzkes, 2010; Hu and Kuritzkes, 2012; Goethals et al., 2008; Gupta et al., 2009b).

The drug susceptibility of these recombinant or mutant viruses can then be measured in the presence of different amounts of drug using various methods to detect levels of infectivity. Single-cycle assays using pseudotyped recombinant viruses measure the number of cells infected in one round of virus replication whilst multiple-cycle assays use replication-competent virus vectors and involve the quantification of virus infectivity over a period of time. Commonly used as a method to detect viral infectivity is the inclusion of a reporter gene

in the viral genome, for example firefly or *Renilla* luciferase (Petropoulos et al., 2000; Parry et al., 2009; Buzon et al., 2008; Weber et al., 2011) or by inducing β -galactosidase production in the host cell via tat activation (Hu and Kuritzkes, 2010; Delelis et al., 2009). Others measure the production of virus by quantifying the amount of HIV-1 p24 capsid protein in the cell-free supernatant (Martinez-Picado et al., 1999; Fun et al., 2010) and or by measuring the appearance of cytopathic effects in cell culture (Kellam and Larder, 1994; Hertogs et al., 1998) compared to a control wild-type vector. The data are then used to determine the 50% or 90% inhibitory concentrations (EC_{50} or EC_{90}) using linear regression methods. The results are then expressed as a fold-change in drug susceptibility relative to wild-type control virus.

On the other hand, replication fitness can be determined as the level of infectivity per amount of p24 capsid in viral inoculum relative to that of wild-type control virus set at 100% (Parry et al., 2009). Other methods to determine viral replication fitness utilising multiple rounds of replication are growth competition assays (Buzon et al., 2008; Hu and Kuritzkes, 2010; Canducci et al., 2010). These usually involve the co-infection of cells with two or more vectors in which their rates of replication are directly compared. Outgrowth of one vector over the other allows a relative fitness rate to be determined. These assays may be performed with varying concentrations of each virus in multiple experiments and in the absence or presence of drug. The relative amounts of each vector can be quantified longitudinally using multiplex PCR with fluorescently labelled primers which are specific to each vector (Buzon et al., 2008) or real time PCRs of each specific sequence tag (Hu and Kuritzkes, 2010; Canducci et al., 2010). Alternatively, vectors labelled with different fluorescent dyes can be measured using Fluorescent Activated Cell Sorting (FACS) (Dykes et al., 2010).

1.10.2 Genotyping

Standard drug resistance assays have moved away from the more time consuming *in vitro* cell-based, phenotypic assays to the quicker, PCR-based genotypic assays. Standard genotyping assays usually consist of the amplification and sequencing of the viral genes of interest or parts thereof, usually *PR* and the N terminal of *RT* and the catalytic core domain of *IN*. These partial gene fragments typically encompass only the regions where known major resistance mutations develop. They are usually able to detect the genotype of viral variants which are present in a sample at 20% or more and give a consensus sequence based on the majority of viruses present. The sequence output is analysed using a drug resistance interpretation algorithm, for example the Stanford HIV Database (Shafer, 2006) which picks out both major and minor drug

resistance mutations present in the sequence. These algorithms translate the sequence mutational profile into levels of susceptibility to different antiretroviral drugs, information which is then used by clinicians to determine the best treatment options for the patient (Rhee et al., 2003).

These assays cannot always detect variants that are present in smaller numbers (Gega and Kozal, 2011; Palmer et al., 2005). Furthermore they are unable to detect the presence of genetic linkage of drug resistance mutations and/ or other mutations present as mutations that are detected are not necessarily found on the same genome. In addition as only partial fragments of the *pol* gene are sequenced, the investigation of the development of resistance across the whole *pol* gene (targets of the three main classes of ARV drugs) is limited. Few data is therefore available on the co-evolution and linkage of PI, RTI and INSTI resistance mutations found within the whole *pol* gene.

Other genotypic methods that can be used for the detection of drug resistance in a research situation include clonal analysis, single genome sequencing (SGS), point mutation assays and next generation or ultra-deep sequencing. For clonal analysis genes of interest are amplified by RT-PCR and inserted into a plasmid vector. DNA from individual colonies is then sequenced and drug resistance is interpreted from the sequence data, each colony representing one viral genome. However, as cloning occurs after the PCR step some genomes can be represented more than once, or not at all due to PCR sampling bias. This technique can be very sensitive depending on the number of colonies analysed and can also provide information on drug resistance mutations in the context of the other background mutations present in the virus as well as genetic linkage of mutations (Gega and Kozal, 2011; Winters et al., 2012).

Point mutation assays using allele-specific PCR are specifically designed to detect a few critical resistance mutations. These involve the comparison, in real-time, of amplifications of either wild-type or allele-specific mutations and the total sequence population. Primers are designed to incorporate either the wild-type residues or mutated sequence mismatch at or near the 3'end. The mismatched primer will not be able to initiate the amplification of wild-type sequences under stringent PCR conditions and so amplification using these primers signifies the presence of a mutated genotype (Johnson et al., 2007; Metzner et al., 2005). These point mutation assays are very sensitive and relatively quick to carry out. Unfortunately these assays are unable to detect all resistance mutations in one reaction. Additionally no other genetic information regarding linkage and mutational context can be obtained from this assay (Gega and Kozal, 2011).

SGS involves the dilution of cDNA to a concentration low enough to ensure that each subsequent PCR reaction will only amplify a single genome. The sequence obtained from each reaction is therefore derived from a single viral genome. A major advantage of this technique is that the assay-based error rate should be that of the RT step only (Palmer et al., 2005). A SGS assay developed by Palmer *et al.* found that their single genome sequencing assay had a background substitution rate of 1.1 assay-related errors/10,000 nucleotides amplified and an assay recombination rate of <1/66,000 nucleotides analysed. SGS assays can be very sensitive, depending on the number of sequences that are amplified. They also provide information on mutational context and genetic linkage (Palmer et al., 2005; Mbisa et al., 2011b; McKinnon et al., 2011).

Next generation or ultra-deep sequencing (NGS or UDS) is a relatively new technique and is only recently becoming widely commercially available. Its first reported use in the detection of HIV drug resistance was in 2005 (Simons et al., 2005). In basic terms NGS involves sequencing a dense array of DNA through multiple cycles of enzymatic reactions and imaging-based collection of data (Shendure and Ji, 2008). There are multiple NGS technologies that have been developed with differing sequencing chemistry and outputs. To date, the main NGS technologies which are available on different throughput platforms include Roche 454 (using emulsion PCR and pyrosequencing technology), Illumina (using in situ bridge amplification and reversible terminators for sequencing) and Ion Torrent (using emulsion PCR and release and detection of H⁺ ions for sequencing).

NGS is highly sensitive, and in some instances can detect viral variants present at as low as 1% of a sample. In addition, the throughput is very high and it is able to generate hundreds of thousands sequence reads at one time. However, these reads can be relatively short depending on the technology (from 35 to 700 bp) and so it has a limited use in the detection of genetic linkage across a long fragment of DNA. There are also issues regarding the storage and analysis of the masses of data produced as well as error rates with regards to homopolymeric regions (Gega and Kozal, 2011).

1.10.3 Clinical significance of minority viral variants

The clinical relevance of minority drug resistant variants (viral variants present as <20% of the viral population) and drug resistance mutation linkage has come under the spotlight in the advent of advancing technologies in drug resistance detection and clinical cut-offs are yet to be

established at which minority variants become clinically significant (Gega and Kozal, 2011). Some studies have shown that the presence of minority variants harbouring certain resistance mutations can cause treatment failure in some populations. In two studies, treatment-naive patients in which minority variants harbouring NNRTI resistance mutations (as detected by real-time allele specific PCR and UDS) at baseline were found to be more likely to fail NNRTI-containing treatment than those without NNRTI resistant minority variants at baseline (Johnson et al., 2008; Simen et al., 2009).

Conversely, a study concluded that PI resistance mutations present at low levels were not indicative of treatment failure. This study revealed that minority variants containing PI drug resistance mutations revealed by UDS at baseline were not indicative of treatment failure on a boosted PI-containing regimen, however this may be because the PI resistance mutations that were found had low genotypic resistance scores (Lataillade et al., 2010).

1.10.4 Clinical significance of the linkage of drug resistance mutations

The linkage of drug resistance mutations may enhance the drug resistance of a particular viral variant enabling it to successfully propagate in an environment of multiple drug therapy. Its clinical relevance comes into play when resistance mutations are not linked; drugs in a patient's regimen may still be active against individual viruses not harbouring the specific drug class resistance mutations even when the bulk genotype suggests otherwise. For example, a virus harbouring three genetically linked mutations conferring resistance to the three major drug classes, PIs, RTIs and INSTIs, may have a selective advantage under combination treatment pressure. If however the three PI, RTI and INSTI resistance mutations were not genetically linked then each drug class would still be active against viruses not harbouring the specific drug class resistance mutations. This is the rationale for using triple combination therapy; resistance mutations to each drug are likely to be present in the viral quasispecies at low levels but are unlikely to be present on the same genome. As of yet, there is not much evidence to determine the role of genetic linkage of drug resistance mutations in treatment failure (Gega and Kozal, 2011).

1.10.5 Clinical significance of HIV-1 diversity and fitness

It has been reported that HIV-1 subtype may influence susceptibility to antiviral drugs due to subtype-specific polymorphisms. Indeed, it is well documented that HIV-1 group O and HIV-2 viruses are inherently resistant to one of the major drug classes; NNRTIs (Isaka et al., 2001; Witvrouw et al., 2004; Descamps et al., 1997; Tuailon et al., 2004). Inherent HIV-2 NNRTI resistance has been attributed to the leucine residue at position 188 in the *RT* gene (Isaka et al., 2001).

What's more a study by Delviks-Frankenberry *et al.* found that a recombinant virus (CRF01_AE) containing TAMS displayed a 64-fold higher level of AZT resistance, whilst a subtype B virus harbouring the same TAMS exhibited a 13-fold higher level of resistance relative to a wild-type subtype B virus. Further analysis revealed that the T400 wild-type amino acid in CRF01_AE was responsible for this increased AZT resistance, in addition this residue increased AZTMP excision and reduced RNase H activity in the virus. This was reiterated when the A400 wild-type residue in subtype B was substituted with tyrosine. The converse occurred when the T400 residue in CRF01_AE was replaced with alanine. Delviks-Frankenberry *et al.* emphasise the need to develop and use subtype-specific genotypic and phenotypic assays to provide a true reflection of drug resistance (Delviks-Frankenberry et al., 2009). In addition, a systematic review of virologic and biochemical data investigating the role of subtypes in drug resistance also confirmed the potential role of subtype-specific polymorphisms in drug resistance; the type of resistance mutations to appear and the degree of resistance conferred may be affected (Martinez-Cajas et al., 2008).

It has also been shown that different mutations appear at different frequencies amongst treatment-experienced patients infected with different subtypes (Grossman et al., 2001; Pieniazek et al., 2000; Frater et al., 2001). Furthermore, there is evidence that the presence of some minor PI and RTI drug resistance mutations in treatment-naive patients influences the rate of development of major drug resistance mutations (Vergne et al., 2000).

In addition, many studies have examined the importance of subtype on rates of disease progression and the rates or routes of transmission (Kanki et al., 1999; Kiwanuka et al., 2008; Baeten et al., 2007; Yang et al., 2003; Nelson et al., 2007; Hu et al., 1999; Alaeus et al., 1999; Kunanusont et al., 1995). Studies have shown that infection with non-A subtypes caused a faster progression to AIDS (Kanki et al., 1999; Kiwanuka et al., 2008; Baeten et al., 2007). Conversely two studies found no difference in disease progression or virological response

(Alaeus et al., 1999; Pillay et al., 2002). However, there does appear to be a link between maternal viral subtype and mother to child transmission (MTCT). Several studies have suggested a higher rate of transmission of one subtype over another, though the results are contradictory (Yang et al., 2003; Blackard et al., 2001; Renjifo et al., 2004). A review of epidemiological and biological data suggested that differences in HIV-1 transmissibility and pathogenesis were likely to be influenced by differences in individual viruses rather than broader, subtype-specific differences. Other confounding factors such as the host and socio-behavioural factors may also play a role in transmissibility and pathogenesis. Interestingly, the review by Hu *et al.* concluded that subtype did not appear to be linked to any specific biological and phenotypic viral characteristics and that a single factor, such as viral subtype, is unlikely to be accountable for the observed differences. Other determinants including host factors, access to medical care and the presence of other infections may also contribute to the differences observed between subtypes (Hu et al., 1999).

Despite this evidence and the fact that subtype B only accounts for ~12% of HIV-1 infections worldwide (Geretti, 2006; Hemelaar et al., 2011) the focus of most HIV-1 research has remained on subtype B because it is the predominant subtype seen in resource-rich countries. An increase in the prevalence of non-B subtype infections in Western Europe and an increase in the roll-out of ARV therapies to resource-poor countries (UNAIDS, 2009; Hirsch et al., 2008) has however made it increasingly important to determine the full effects of subtype on treatment options and the evolution and mechanisms of HIV-1 drug resistance.

On the other hand, the clinical implications of a less fit virus to a patient are yet to be fully established, and it is unknown whether, or how much, viral fitness impacts on disease progression or treatment in a patient (Quinones-Mateu and Arts, 2002). Conceptually it may be thought that drug resistant viruses, as they are less fit may be beneficial to the patient as their replication is slowed down, which may in turn reduce the viral load and importantly, the risk of disease progression. Indeed, the intentional selection of less fit, drug resistant variants was proposed as a possible alternative use of ARV treatment in 1999 (Berkhout, 1999), however due to the rapid ability of HIV to acquire compensatory mutations, this may not prove to be successful. In addition, an *ex vivo* study on viral fitness performed by Quinones-Mateu *et al.* on an HIV infected cohort from Belgium found that viruses from long term elite suppressors were significantly less fit than viruses from patients with accelerated AIDS progression. They also found that there was a strong correlation between viral loads and relative fitness of viruses from both groups of patients. They suggested that viral fitness together with viral load

could be a predictor for the rate of disease progression (Quinones-Mateu et al., 2000). What's more a review published by Lobritz *et al.* gathered information on several studies which argued for the role of viral fitness in elite controllers of HIV. Two major conclusions were reached in the review. Firstly, compared to patients with a normal progression to disease, the *gag*, *pol* and *env* genes of elite controllers had a lower replication capacity and secondly, HIV targeted cell-mediated immune responses may further reduce replication capacity of an already less fit virus by pushing viral evolution towards the selection of an escape mutant with an even higher fitness cost (Lobritz et al., 2011). Regardless of the clinical implications, the determination of the replication capacity of patient-derived viruses is necessary to fully understand viral evolution, both at the population level and within a patient.

1.11 Virus evolution

The HIV-1 genome has a high degree of plasticity and is capable of rapid evolution due to inherently high mutation rates and selection. These high mutation rates are caused by reverse transcription errors, recombination (Rambaut et al., 2004) and the high rate of virus production, which is on average 10.3×10^9 virions per day (Perelson et al., 1996). In addition, the cellular polymerases involved in HIV replication can also be a source of mutation. Cellular DNA polymerases have a high fidelity and so are unlikely to significantly contribute to HIV mutation, however RNA polymerase II mediated transcription could potentially be a substantial source of viral mutation (Preston and Dougherty, 1996).

Errors occur during viral replication because of the lack of 3' to 5' exonuclease activity exhibited by viral RT (Roberts et al., 1988), reported error rates range between $3-10 \times 10^{-4}$ substitutions per nucleotide per round of replication (Preston et al., 1988; Roberts et al., 1988; Boyer et al., 1992; Hubner et al., 1992; Gao et al., 2004). Intrapatient error rates do not translate to evolutionary rates at the population level as one may expect, in fact, evolutionary rates in an epidemic tend to be lower. Reasons for this include the non-random transmission of slower-evolving viruses and transmission during later stages of disease progression when viral evolution has declined (Pybus and Rambaut, 2009).

Recombination is a form of genetic exchange between two or more genetically distinct parental genomes, generating hybrid nucleic acid sequences. In retroviruses, recombination occurs when a cell is co-infected with two or more different viruses; the resulting daughter virions may contain one RNA strand from each parental virus. During the next round of

infection virions containing heterozygous RNA strands can recombine through template switching during cDNA synthesis (Hu and Temin, 1990). In HIV-1 replication there are an estimated 3 recombination events occurring per genome, per round of replication (Zhuang et al., 2002).

Mutation and recombination generates viral variants in the viral population which are destined for one of two fates during evolution. The fate of a new viral variant is determined by its fitness in the current environment and the population size. Advantageous mutations are those that result in increased fitness of the virus, for example, those that result in immune escape or drug resistance. On the other hand unfavourable mutations are those that reduce viral fitness, for example, those that exert functional constraints on a gene or protein. As a result, viral variants with a fitness advantage increase in frequency until they become fixed in the population and are said to be positively selected, whilst unfavourable variants decrease in frequency until they have been eliminated from the population, also known as negative or purifying selection (Vandamme et al., 2009). Positive and negative selection can thus be distinguished by looking at the frequency of synonymous and non-synonymous mutations. Synonymous mutations result in neutral or silent changes which have no effect on the amino acid codon, whilst non-synonymous mutations result in an amino acid change. One can assume that synonymous mutations do not confer a selective advantage or disadvantage, if this is true then the differences observed between synonymous and non-synonymous mutations should then reflect selective adaptation. Consequently a high frequency of synonymous mutations compared to non-synonymous mutations is indicative of negative selection, whilst a high frequency of non-synonymous mutations compared to synonymous mutations is indicative of positive selection (Vandamme et al., 2009).

HIV-1 virus evolution is governed by selection, for example, selective adaptation to the host immune response or drug treatment (Zhang et al., 1997a; Chen et al., 2004; Ross and Rodrigo, 2002; Zanotto et al., 1999). In addition, random genetic drift in which mutations randomly become fixed or eliminated in the population also contributes to HIV-1 evolution (Gojobori et al., 1990; Holmes and de, 1998; Plikat et al., 1997; Frost et al., 2000; Shriner et al., 2004). However, it is thought that selective adaptation where mutations confer an advantage to the virus are fixed in the viral population plays the major role in inpatient viral evolution (Rambaut et al., 2004; Pybus and Rambaut, 2009). Whilst selective adaptation appears to be the major contributor to inpatient viral evolution, the same cannot be said for viral evolution at the epidemic/population level. This can be seen in the structure of phylogenetic

trees for each dataset. Trees inferring evolution at the epidemic level show little evidence of selective adaptation and show lineages co-existing for extended periods even though transmission of HIV-1 between hosts is usually accompanied by a bottleneck effect in which viral diversity is reduced causing the new viral population to evolve from a small donor subset (Cichutek et al., 1992; McNearney et al., 1992). In contrast, inpatient phylogenies reflect the successive fixation of advantageous mutations and the extinction of disadvantageous lineages (Rambaut et al., 2004).

It is the constant evolution and persistent selective pressures exerted by the host immune response and drug treatment that causes inpatient genetic diversity of HIV to fluctuate over time and in space. The virus is constantly evolving to reflect changes in immune (Liu et al., 2011; Allen et al., 2005; Liu et al., 2006) and drug (Harada et al., 2013; Kitrinos et al., 2005; Ibanez et al., 2000) pressures. Interestingly, due to different selective pressures exerted on different parts of the HIV genome, evolution across the whole viral genome can occur at different rates, and independently from each other (Leigh Brown and Cleland, 1996; Shi et al., 2010), thus diversity is not distributed evenly throughout the genome (Korber et al., 2001; Rossenkhani et al., 2012).

By observing evolutionary changes in individuals from a population the retrospective history and dynamics of that population can be explored, which forms the basic principle behind the coalescent theory. The discovery and development of the coalescent theory is credited to Kingman (Kingman, 1982a; Kingman, 1982b), Hudson (Hudson, 1983) and Tajima (Tajima, 1989). The theory states that in the absence of selection sampled sequences randomly “choose” their parent as we go back in time. Whenever two “pick” the same parent their lineages are said to coalesce. From this one can perform a retrospective inference of population dynamics from randomly selected individuals from a population, enabling the reconstruction of the dynamics of a population (Drummond et al., 2005). Whilst these methods have primarily been developed to analyse sequence data at the population level, they can be adapted for the analysis of data obtained from a single patient (Drummond et al., 2003; Grenfell et al., 2004).

Coalescent inference methods, as with other phylogenetic frameworks, require the implementation of different models such as molecular clock and nucleotide substitution models. Molecular clock models describe the rate of molecular evolution. It was first believed that evolution followed a constant rate or a molecular clock in which mutations were lost or fixed in a genome by random sampling or genetic drift (Kimura, 1968). This remained the

theory for evolution until studies began to provide evidence of evolution that departed from the normal clock-like rate; rates of evolution appeared to differ between different species and genes (Britten, 1986; Ayala, 1997; Hasegawa and Kishino, 1989). Molecular clock models addressing this issue have been developed allowing rates of evolution to vary over time, introducing the concept of relaxed molecular clock models (Thorne et al., 1998; Huelsenbeck et al., 2000).

Models of substitution are essential in the reconstruction of phylogenies and all other phylogenetic analysis. These models define the rates at which different substitutions occur along the branch of a tree. Substitution models are largely based on three parameters of sequence evolution; base frequency (the occurrence of a particular base at a particular sequence site), base exchangeability (the tendency of certain bases to be substituted for one another) and rate heterogeneity (the differences in substitution rates across the sequences). Each model of nucleotide substitution differs regarding the type and number of parameters taken into account (Page and Holmes, 1998).

A crucial requirement of studying the coalescent is the assumption of a demographic model. These models are a mathematical function that describes the change in the effective population size over time. There are currently six demographic models that can be used: constant population size, exponential growth, logistic growth (exponential growth followed by a constant population size), expansion growth (constant population size followed by exponential growth), piecewise con-exp-con (constant population sizes either side of exponential growth) and finally the Bayesian skyline plot (BSP) model. Unlike the other models, the BSP is a non-parametric model that uses Metropolis-Hastings Markov Chain Monte Carlo (MCMC) (Hastings, 1970) sampling to estimate the distribution of effective population size through time directly from a sample of sequences and can fit a wide range of demographic situations (Pybus and Rambaut, 2005; Drummond et al., 2005). MCMC sampling is often portrayed as walking through a tree landscape. During the walk a random tree is selected, T1, and compared to a second tree, T2. If the likelihood of T1 representing the data is smaller than T2, then T2 replaces T1 as the current tree. If the likelihood of T2 however is smaller than T1, T1 is kept as the current tree and the walker goes up the hill. This process is repeated a considerable number of times to find the best tree given the data, which represents the top of the highest hill.

The coalescent theory framework is commonly used for the estimation of parameters such as nucleotide substitution rate, effective population sizes (N_e) or viral diversity, and the date of

origin of the most common recent ancestor (MCRA). Frequently, groups have used coalescent-based approaches to unravel the origin and evolution of HIV-1 epidemics enabling the estimation of dates of HIV introduction into each population (Jung et al., 2012; Lemey et al., 2003; Hue et al., 2005; Gifford et al., 2007). Others have used coalescent methods to quantify genetic bottlenecks in cases of transmission (Edwards et al., 2006) and genetic diversity during infection and subsequent antiviral treatment (Tazi et al., 2011).

1.12 Project outline

This project aims to investigate the evolution and development of drug resistance across the whole HIV-1 *pol* gene. Specifically, we aim to determine the interaction of drug resistance mutations in the three main genes targeted by HAART in patients undergoing RAL-containing salvage therapy. A novel SGS assay was developed which allowed the amplification of HIV-1 full-length *pol* from longitudinal samples taken from two patients failing on RAL salvage therapy. This allowed us to explore the viral variants present and how they changed with respect to the patient regimen whilst also being able to look at drug resistance mutation linkage (Chapter 3). We investigated the inpatient viral evolution using multiple phylogenetic and bioinformatic techniques to determine the mechanisms at play during the development of drug resistance (Chapter 4). Furthermore, we investigate the effects of INSTI drug resistance mutations in the context of co-evolved highly drug-resistant *PR* and *RT* genes on resistance to RAL and EVG and viral replication fitness using a single-cycle replication assay (Chapter 5).

A secondary aim was to further our knowledge of drug resistance in less well defined subtypes whilst also determining the importance of the inclusion of the C terminal region of *RT* in routine drug resistance genotyping. Full-length viral *RT* sequences were amplified from patients infected with HIV-1 subtype F to try and determine the presence of any novel mutations occurring in the C terminal of *RT* that were associated with N terminal resistance mutations (Chapter 6).

These studies will provide an insight into drug resistance development in the era of multi-class ARV therapy targeting all three genes in HIV-1 *pol*. We hope that our findings may help devise better treatment regimens and help improve our understanding of the emergence of drug resistance, and consequently treatment failure, especially in patients undergoing RAL salvage therapy.

CHAPTER 2

General materials and methods

2.1 Clinical samples

Plasma samples from HIV-1 infected individuals used for the design and optimization of polymerase chain reactions (PCRs) and sequencing reactions were obtained from the Antiviral Unit (AVU) Clinical Services at Public Health England (PHE, Colindale) and the Mortimer Market Clinic, University College London Hospital (UCLH). The samples were previously genotyped for ARV resistance testing and the viral load and HIV-1 subtype data were available.

Plasma samples used in the study of the role of C terminal region of HIV-1 subtype F RT in the development of drug resistance were obtained from treatment-experienced children infected with HIV-1 subtype F1 enrolled in a EuroCoord-CHAIN study on HIV-1 infection in Europe. All samples were from Romania and were collected over a 6-year period from 2003 and 2008. EuroCoord-CHAIN is a large European collaboration of HIV observational cohorts (EuroCoord); including the PENTA-EPPICC network (Paediatric European Network for Treatment of AIDS-European Pregnancy and Paediatric HIV Cohort Collaboration, and the European Collaborative HIV and Anti-HIV Drug Resistance Network (CHAIN).

Plasma samples used for single genome sequence analysis of full-length HIV-1 *pol* gene were obtained from the Mortimer Market Clinic, UCLH, from patients undergoing salvage antiretroviral therapy (ART) containing RAL (600 mg daily) and other ARVs. Sequential samples taken before, during and after RAL therapy were obtained as was a single sample taken after a patient was re-initiated on RAL-containing therapy. Details of each patient are outlined in more depth in Chapter 3.

2.2 Molecular biology techniques

2.2.1 Retroviral expression vectors and plasmids

2.2.1.1 Molecular clones

The following HIV-1 molecular clones were obtained through the NIH AIDS Research and Reference Reagent Program (Germantown, Maryland, USA) and used for primer testing and PCR optimization: pHXB2D (subtype B), p97ZA012.1 (subtype C), p94UG114.1.6 (subtype D) and p93BR020.1 (subtype F).

2.2.1.2 p8.9NSX

p8.9NSX (a kind gift from Chris Parry, MRC, Uganda) is an HIV based *gag-pol* expression vector. It also expresses the HIV-1 accessory proteins *tat* and *rev*. However, the HIV-1 accessory protein genes *vif*, *vpu*, *vpr* and *nef*, as well as the *env* gene are deleted (Figure 2.1a) (Naldini et al., 1996). The vector contains multiple unique restriction enzyme sites flanking and within the *gag-pol* region which allow the cloning of various patient-derived *gag-pol* fragments (Figure 2.1b).

2.2.1.3 pHL(WT)

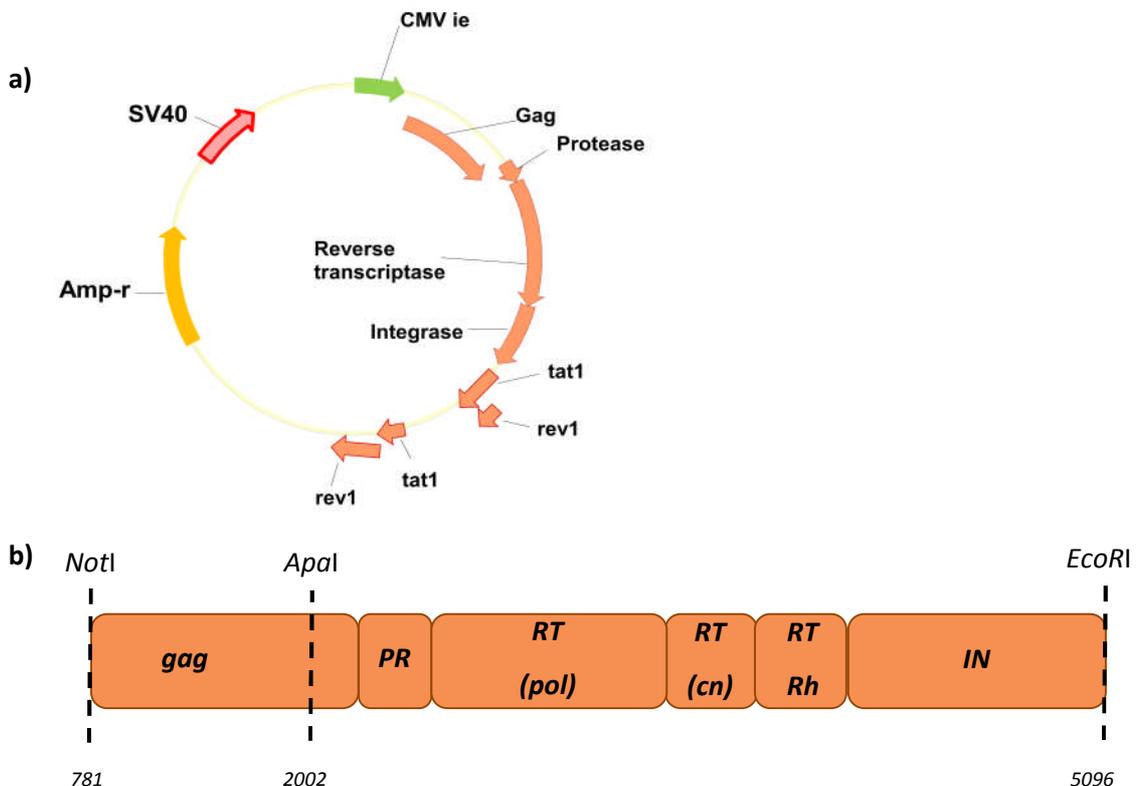
The pHL(WT) vector (a generous gift from Vinay Pathak, National Cancer Institute, Frederick) is an HIV-1 based vector expressing the *firefly luciferase* reporter gene in the *nef* open reading frame and all of the HIV-1 proteins except *env* (Figure 2.2a). The vector was generated by introducing two restriction enzyme sites, *Eco47III* (nucleotide position 3408 in HXB2, in the *pol* domain of *RT*) and *SpeI* (nucleotide position 3805 in HXB2, in the RNase H domain of *RT*) into pNLuc (Kiernan et al., 1998), an HIV-1 based vector expressing the *firefly luciferase* reporter gene. All *MscI*, *Clal* and *SpeI* restriction sites in *IN*, *luciferase* and *gag* genes were knocked out to ensure that these sites were unique only in the *RT* gene (Nikolenko et al., 2007). This allows the sub-cloning of various patient-derived *RT* fragments (Figure 2.2b).

2.2.1.4 pCSFLW

The pCSFLW is a reporter vector encoding firefly luciferase (kindly donated by Chris Parry, MRC, Uganda). It also encodes the HIV-1 packaging sequence and a spleen focus-forming virus (SFFV) promoter which is flanked by self inactivating HIV-1 long terminal repeats (LTRs) (Bainbridge et al., 2001; Wright et al., 2008).

2.2.1.5 pMDG

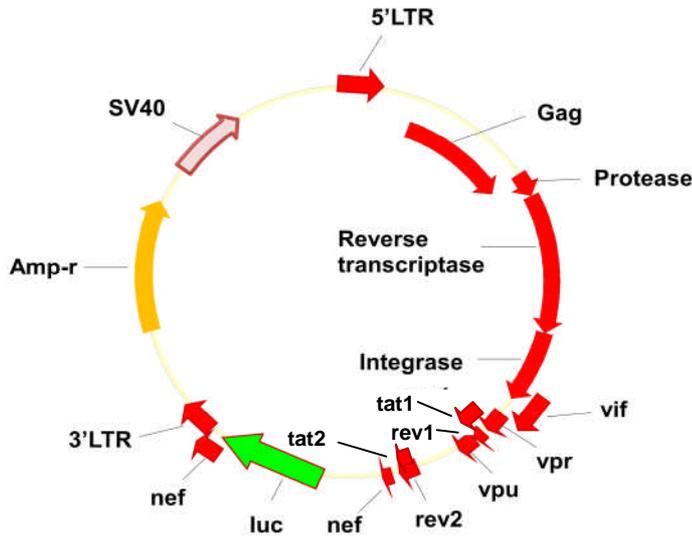
The pMDG is an envelope encoding vector expressing the vesicular stomatitis virus glycoprotein (VSV-g) as an envelope protein (Naldini et al., 1996).



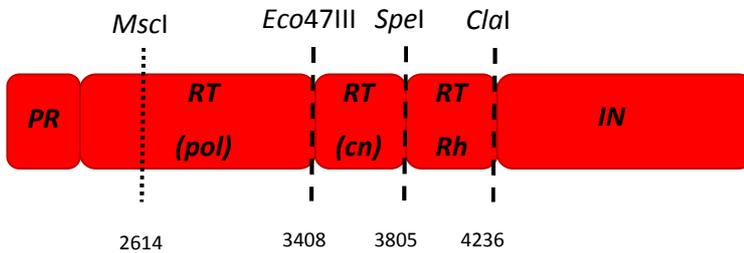
NB: Numbers in italics indicate the nucleotide position in HXB2

Figure 2.1 a) The p8.9NSX *gag-pol* expression vector and b) the unique restriction sites present in p8.9NSX *gag-pol* region that were used for cloning. The p8.9NSX plasmid vector expresses the HIV-1 *gag-pol* gene under the control of the Cytomegalovirus (CMV) promoter. The plasmid vector also contains a Simian Vacuolating Virus 40 (SV40) origin of replication (Ori) which allows replication of the plasmid in mammalian cells expressing the large T antigen such as 293T cells. In addition it contains an *ampicillin resistance* gene (*Amp-r*) for bacterial selection.

a)



b)



NB: Numbers in italics indicate the nucleotide position in HXB2; Indicates a natural enzyme restriction site

Figure 2.2 a) The pHL(WT) HIV vector and b) Unique restriction sites present in the pHL(WT) *pol* gene that can be used for cloning. The pHL(WT) vector expresses all HIV genes, (with the exception of *env*) and the *firefly luciferase* reporter gene under an HIV LTR promoter. Similar to p8.9NSX, it also contains the SV40 Ori and an *ampicillin resistance* gene (*Amp-r*).

2.2.2 Primer design

Oligonucleotide primers for use in PCRs were empirically designed by aligning representative sequences of HIV-1 subtypes A, B, C, D, and F in MEGA4.1 software (section 2.5.2) and looking for conserved areas upstream and downstream of the region of interest for amplification. Particular attention was paid to the 3' end of the primers which were designed to anneal to highly conserved regions and to end in one or more G or C nucleotides. Inter- or intra-subtype variations at nucleotide positions within the primer sequence were overcome by the use of ambiguous bases. Each primer ideally contained between 40-60% of G or C nucleotides, was 18-22 bp long and had a melting temperature (T_m) of 68-72°C which was theoretically determined using the simplified formula: $T_m = 4^\circ\text{C} \times (\text{number of G and C nucleotides in the primer}) + 2^\circ\text{C} \times (\text{number of A and T nucleotides in the primer})$.

Sequencing primers were designed in a similar way; however the GC content was not a critical factor in their design. For full coverage of the area desired for sequencing, primers were designed for every 500-600 nucleotides sequence stretch in both the forward and reverse orientation in order to provide overlapping sequencing reads.

Primers used for site-directed mutagenesis (SDM) were designed following the instructions of the QuikChange Site-Directed Mutagenesis Kit (Agilent Technologies). Briefly, the primers' optimal lengths were between 25 and 45 bp long and were individually designed for each target plasmid, unless the target region was conserved between plasmids. Each SDM primer was designed to incorporate the desired point mutation(s) close to the middle of the primer and primers used for the introduction or removal of a restriction site were designed to introduce the desired point mutation(s) without changing the amino acid coding of the plasmid.

The primers used in each PCR, sequencing and SDM reaction are described in the relevant sections.

2.2.3 Luria-Bertani agar plates

Luria-Bertani (LB) agar plates were made using in-house sterilized LB agar (in-house, Media Section, PHE, Colindale). This was melted in a microwave and left to cool before adding 100 µg/mL ampicillin (Sigma-Aldrich). Approximately 20 mL of liquid agar was added to 10 cm petri dishes and left to cool. LB agar plates were stored at -4°C until further use.

2.2.4 Viral RNA extraction

HIV-1 viral RNA was manually extracted from plasma using the QIAamp UltraSens Virus Kit (Qiagen) according to the manufacturer's protocol. The kit allows for the concentration of viral nucleic acids followed by purification through a silica-gel membrane spin column. Briefly, 200 μL of plasma, made up to 1 mL with phosphate buffered saline (PBS; in-house, Media Section, PHE, Colindale), was denatured and the nucleic acids precipitated by adding 800 μL of Buffer AC containing 5.6 μg of carrier RNA and incubated at room temperature for 10 minutes (min). The sample was then centrifuged and the pellet re-suspended in 300 μL of pre-warmed Buffer AR containing proteinase K and incubated at 40°C for 10 min to digest proteins. This was followed by the addition of 300 μL of binding Buffer AB. The lysate was then applied onto the QIAamp spin column followed by centrifugation allowing the nucleic acids to bind to the QIAamp membrane whilst contaminants pass through. The column was then sequentially washed twice with Buffer AW1 and AW2 followed by elution of purified RNA using 60 μL of elution Buffer AVE. The extracted RNA was either used immediately in cDNA synthesis or stored at -80°C for later use. Up to 1 mL of plasma was used for viral RNA extraction of samples with a viral load of less than 1,000 copies/mL.

2.2.5 Amplification of genomic viral RNA using polymerase chain reaction (PCR)

2.2.5.1 Denaturation of viral RNA

10 μL of extracted viral RNA was denatured at 65°C for 10 min together with 1.1 μM of reverse primer and 1.5 mM of dNTP. For the amplification of *PR+RT* (2.2 kb) the reverse primer used was 4368- whereas for the amplification of full-length *pol* gene (3.5 kb) one of 6 reverse primers (5317-, 5175-, 5254-, 5450-, KVL069 or 5254-deg) was used (Table 2.1). Following the RNA denaturation step, the denaturation sample mix was immediately incubated on ice and used in the RT-PCR reaction step.

2.2.5.2 cDNA synthesis (reverse transcription-PCR) and first round PCR reactions

cDNA synthesis and first round PCR reactions were performed in a single step using the QIAGEN® One Step RT-PCR Kit (Qiagen) according to the manufacturer's guidelines. The QIAGEN® One Step RT-PCR Kit contains all the components required for a one step cDNA synthesis and first round PCR amplification reaction. The QIAGEN OneStep RT-PCR Enzyme Mix includes Omniscript and Sensiscript RTs which provide highly specific and efficient reverse transcription of both larger and smaller RNA amounts. In addition, it includes a HotStarTaq DNA polymerase which allows a hot-start PCR for highly specific amplification after reverse transcription has occurred, temporally separating reverse transcription and PCR amplification.

The forward primer used in this step was 1849+ (5'-GATGACAGCATGTCAGGGAG-3'). A RT-PCR reaction mix was prepared using 1x QIAGEN OneStep RT-PCR Buffer, 2 µL of QIAGEN OneStep RT-PCR Enzyme Mix and 0.3 µM 1849+. This was added to the RNA denaturation sample mix described previously (section 2.2.5.1) and made up to 50 µL using nuclease-free H₂O (nH₂O).

The RT-PCR reaction mix was then run using a DNA Engine Dyad Peltier Thermal Cycler under the following cycling conditions; 50°C for 40 min, 95°C for 15 min, 94°C for 30 seconds, variable temperature for 30 seconds (for the amplification of *PR+RT* an annealing temperature of 50°C, was used whereas the annealing temperatures for the amplification of full-length *pol* gene varied depending on the primer used as indicated in Table 2.2.) and 72°C for variable time (for the amplification of *PR+RT* an extension time of 2.5 min was used whereas the extension time for the amplification of full-length *pol* gene was 3.5 min). Steps 3 to 5 were repeated an additional 44 times with a final extension step of 72°C for 3 min.

Table 2.1 Reverse primers used for cDNA synthesis

Primer name	Primer sequence (5'→3')
4368-	GCTAGCTACTATTTCTTTTGCTACT
5317-	AAATGCCAGTCTCTTTCTCCTG
5175-	TTCATAGTGATGTCTATAAAACC
5254-	CTAGTGGGATGTGTACTTCTG
5450-	TTATGTCCYGCTTGATAWTCACACCTAGG
KVL069 ^a	TTCTTCCTGCCATAGGARATGCCTAAG
5254-deg	CTAGTGGGATRTGTACTYCTG

^aprimer sequence obtained from Van *et al.* (Van *et al.*, 2008)

Table 2.2 Annealing temperatures for different first round reverse primers used for the amplification of full-length *pol* gene

Primer	Annealing temperature used (°C)
5317-	55.7
5175-	48.6
5254-	50.9
5450-	50
KVL069	57.5
5254-deg	57.2

2.2.5.3 Nested PCR reactions

The nested PCR reactions were performed using the Platinum® Taq DNA polymerase kit (Life Technologies) according to the manufacturer's protocol. Platinum® Taq DNA polymerase is a recombinant Taq DNA polymerase complexed with an antibody that blocks polymerase activity unless subject to denaturation providing a hot-start PCR for highly specific amplification.

All amplifications used the forward primer 1870+ (5'-GAGTTTTGGCTGAAGCAATGAG-3'). For the amplification of *PR+RT* only the reverse primer 4295- (5'-CTTTCATGCTCTTCTTGAGCCT-3') was used whereas for the amplification of full-length *pol* amplification with one of seven different reverse primers (including 5175-, 5254- or 5317- used in the cDNA synthesis step, see table 2.1) was used depending on the primer used in the RT-PCR reaction (Table 2.3).

Table 2.3 Reverse primers used for nested PCR of full-length *pol* gene

Primer	Primer sequence (5'→3')
KVL084 ^a	TCCTGTATGCARACCCCAATATG
5125-	ATGRTGTTTTACTAAACTKTTCC
5095-	TAATCCTCATCCTGTCTACYTGCCACAC
5222-deg	TGCTATAAAACCATCCTYTAGC

^aprimer sequence obtained from Van *et al.* (Van et al., 2008)

1 µL of the RT-PCR reaction was added to a PCR reaction mix containing 1x PCR buffer, 2 mM MgCl₂, 0.2 mM dNTP, 0.3 µM 1870+, 0.3 µM reverse primer, 2.5 Units (U) Platinum Taq and was made up to a volume of 50 µL using nfH₂O.

The nested PCR reaction was then run using a DNA Engine Dyad Peltier Thermal Cycler under the following cycling conditions; 94°C for 5 min, 94°C for 30 seconds, variable temperature for

30 seconds (for the amplification of *PR+RT* an annealing temperature of 55°C was used whereas the annealing temperatures for the amplification of full-length *pol* gene varied depending on the primer used as indicated in Table 2.4.) and 72°C for variable time (for the amplification of *PR+RT* an extension time of 1 min was used whereas the extension time for the amplification of full-length *pol* gene was 3.5 min). Steps 2 to 4 were repeated an additional 35 times with a final extension step of 72°C for 5 min.

Table 2.4 Annealing temperatures for different second round reverse primers used for the amplification of full-length *pol* gene

Primer	Annealing temperature (°C)
KVL084	58.6
5125-	51.9
5095-	50
5317-	56.7
5254-	50.9
5175-	48.6
5222-deg	52.2

2.2.5.4 Addition of restriction enzyme site by PCR

To facilitate the cloning of PCR products into the *gag-pol* expression vector unique restriction enzyme sites were introduced onto the flanking regions of the fragment to be cloned by PCR. The PCR reactions were carried out using the Platinum® Taq DNA polymerase kit (Life Technologies). 10 ng of PCR product was added to a PCR reaction mix containing 1x PCR buffer, 2 mM MgCl₂, 0.2 mM dNTP, 0.3 μM forward primer, 0.3 μM reverse primer, 2.5 U Platinum Taq and was made up to 50 μL with nfH₂O. The primers used for the addition of restriction enzyme sites contained the desired restriction site specific sequence at the 5' end and did not alter the amino acid coding. These are discussed under the relevant sections.

The PCR reaction was run using a DNA Engine Dyad Peltier Thermal Cycler under the following cycling conditions; 94°C for 5 min, 94°C for 30 seconds, variable temperature for 30 seconds (annealing temperatures differed depending on the primer combination used and these are discussed in each section where appropriate) and 72°C for variable time (extension times varied depending on the length of amplicon generated; in general an extension time of 30 second per kb was used. These are discussed in each section where appropriate). Steps 2 to 4 were repeated an addition 35 times with a final extension step of 72°C for 5 min.

2.2.6 Agarose gel electrophoresis and PCR product purification

PCR amplicons were identified by agarose gel electrophoresis using a 1% agarose gel prepared in Tris Borate EDTA (TBE; Life Technologies) buffer containing 1× RedSafe Nucleic Acid Staining Solution (Chembio). 6 µL of PCR reaction was mixed with 0.5 µL 10× BlueJuice™ Gel Loading Buffer (Life Technologies) and loaded onto each well of the gel. In addition, 6 µL of a 1 kb ladder (TrackIt™ 1 Kb Plus DNA Ladder, Life Technologies) was loaded onto one of the wells as a marker for product size determination. The gel was run at a constant voltage (125 V) for 25 min to separate the DNA fragments and was visualized using a Bio-Rad Molecular Imager® Gel Doc™.

Positive PCR products were purified using the Illustra GFX PCR DNA and Gel Band Purification Kit (GE Healthcare) following manufacturer's guidelines before use in downstream reactions. Briefly, the entire PCR reaction mix is added to 500 µL Capture buffer type 3 and any proteins in the sample are denatured. DNA binds to the silica membrane during centrifugation at 18626 x g for 1 min. Subsequent centrifugation with 500 µL Wash buffer type 1 at 18626 x g for 1 min in addition to a dry spin at 18626 x g for 1 min removes salts and other contaminants from the membrane bound DNA. The purified DNA is eluted from the membrane using 50 µL Elution buffer type 6 and quantified (section 2.2.8).

DNA products from restriction enzyme digests were identified using 0.8% agarose gel electrophoresis prepared in the same way as described above. 3 µL 10× BlueJuice™ Gel Loading Buffer (Life Technologies) was used to load the whole digest reaction (20 µL) into one of the wells. 6 µL of a 1kb ladder (TrackIt™ 1 Kb Plus DNA Ladder, Life Technologies) was also loaded onto one of the wells as a marker for product size determination. Running times varied between 30 and 60 min, depending on the restriction digest pattern expected. DNA fragments were extracted from the gel and purified as described in section 2.2.7.

2.2.7 Gel extraction and purification of DNA

DNA fragments were separated on agarose gel and the desired bands were excised before being purified using the Illustra GFX PCR DNA and Gel Band Purification Kit (GE Healthcare) following manufacturer's instructions. Briefly 10 µL Capture buffer type 3 was added for every

10 mg of gel slice and the agarose slice was dissolved by incubating at 60°C for 15 to 30 min. Subsequent steps follow those described in section 2.2.6.

The purified DNA fragments were then ethanol precipitated using 5 µL 3 M sodium acetate (pH 5.2) and 110 µL ethanol and incubated overnight at -20°C. After incubation, DNA was pelleted using a bench top centrifuge for 15 min at 18626 x g. The supernatant was then removed and the pellet rinsed using 100 µL of 70% ethanol and spun for a further 3 min. The ethanol was removed and the pellet left to air dry for approximately 10 min. The pellet was subsequently resuspended in 8 µL nfH_2O . 2 µL of the resuspended DNA was used for DNA quantification (described in section 2.2.8).

2.2.8 DNA quantification

DNA was quantified using the Thermo Scientific Nanodrop ND-1000 with 2 µL of purified PCR-derived or plasmid DNA following the manufacturer's instructions.

2.2.9 DNA Sequencing

Sequencing was carried out by the core sequencing service at PHE, Colindale, based within the Applied and Functional Genomic Unit (AFGU) using the ABI Genetic Analyser Capillary Platforms. Sequencing reactions were prepared containing 30 ng/µL of purified PCR product or 250-300 ng/µL of plasmid DNA and 0.83 µM of sequencing primer (Table 2.5) and made up to a volume of 6 µL with nfH_2O . Sequences were analyzed using Sequencher 4.9 software (Gene Codes Corporation). Resistance mutations and other deviations from consensus B were determined using the Stanford University HIV Drug Resistance Database tool available online (<http://sierra2.stanford.edu/sierra/servlet/JSierra?action=sequenceInput>).

Table 2.5 Primers used for sequencing of HIV-1 *gag-pol*

	Primer name	Primer sequence (5'→3')	Orientation
PR and RT primers	P7	CTTTARCTTCCCTCAGATCACTCT	Forward
	2600+	ATGGCCAAAAGTTAAACAATGGC	Forward
	2610-	TTCTTCTGTCAATGGCCATTGTTTAAC	Reverse
	3161+	GTATGTAGGATCTGACTTAGAAATAG	Forward
	3750+	GGAAAGACTCCTAAATTTAGACTA	Forward
	G	CCATCCCTGTGGAAGCACATTG	Reverse
	H	CTGTATTTCTGCTATTAAGTCTTTTGA	Reverse
	3957-	TGTCAGTAACATACCCTGCTTTTC	Reverse
	4227-	TAATTTATCTACTTGTTCATTTCTCC	Reverse
IN primers	4230+	TTGGAGGAAATGAACAAG	Forward
	4722+	AAAGTCAAGGAGTAGTAG	Forward
	4804-	GCCATTTGTAAGTCTGTC	Reverse
Gag primers	GAG1.5R	TCTATCCCATTCTGCAGC	Reverse
	GAGR2	ATGCTGACAGGGCTATACATTCTTAC	Reverse
	GAGF1	ATGGGTGCGAGAGCGTCAGTATT	Forward

2.2.10 Cloning

2.2.10.1 Restriction enzyme digestion

To prepare PCR products or plasmid DNA for cloning or to identify plasmids expressing insert of interest after cloning, restriction enzyme digestion was used. The required amount of DNA (typically 1-2 µg) was used for the digestion reaction using the appropriate restriction enzyme (New England BioLabs; NEB or Roche) and recommended buffer. This was incubated at the recommended temperature for 1 hour. Complete digestion and desired fragments were identified by agarose gel electrophoresis as described earlier (section 2.2.6).

2.2.10.2 Dephosphorylation and ligation reactions

The Rapid DNA Dephos and Ligation Kit (Roche) was used for the dephosphorylation of digested vector DNA arms and ligation of insert DNA following manufacturer's guidelines. The dephosphorylation reaction contained the following; 1 µg vector backbone, 1x rAPid Alkaline Phosphatase Buffer, 1 U of rAPid Alkaline Phosphatase and was made up to a volume of 20 µL with nfH_2O .

The dephosphorylation reaction was incubated at 37°C for 1 hour followed by a 10 min incubation at 75°C to inactivate the alkaline phosphatase. The mixture was either used immediately in the ligation reaction or stored at -20°C for later use.

The ligation reaction contained a total of 200 ng of DNA (molar ratio of vector DNA to insert DNA of 1:3), 1x DNA Dilution Buffer, 1x T4 DNA Ligation Buffer, 5 U of T4 DNA Ligase and was made up to 21 µL using nfH_2O .

The ligation reaction was incubated at room temperature for 1 hour. 2 µL of the ligation reaction was then used to transform HB101 competent cells (section 2.2.10.4).

2.2.10.3 TOPO TA cloning®

To facilitate TOPO TA cloning 3' A-overhangs were added to the PCR products using the Taq DNA polymerase kit (Life Technologies). 42.5 µL of purified DNA was added to a reaction containing 1x PCR Buffer, 1.5 mM MgCl_2 , 0.2 mM dATP and 1 U of Taq.

The reaction was incubated at 72°C for 15 min. The PCR products were then ligated into the pCR®2.1-TOPO® vector (3.9 kb) (Life Technologies) using 4 µL of the 3' A-overhang PCR reaction, 1 µL salt solution and 1 µL TOPO® vector.

The ligation reaction was incubated at room temperature for 30 min allowing the efficient ligation of the 3' A-overhangs of the PCR product with the 3' T-overhang of the TOPO vector. 2 µL of the ligation reaction was then used to transform One Shot® TOP10 chemically competent cells described in section 2.2.10.6. Colonies were screened by colony PCR (section 2.2.10.7) using the M13 forward (-20) (5'-GTAAAACGACGGCCAG-3') and the M13 reverse primers (5'-CAGGAAACAGCTATGAC-3') (Life Technologies).

2.2.10.4 Transformation of HB101 competent cells

HB101 competent cells (Promega) were thawed on ice and 50 µL of cells were aliquoted into pre-chilled flip cap tubes. 2 µL of ligated DNA was added to the cells and were left on ice for 30 min. Cells were heat shocked for 45 seconds in a 42°C water bath to allow uptake of DNA and immediately placed on ice for 2 min. 500 µL of S.O.C medium (Life Technologies) was added to the transformed cells and incubated for 1 hour at 37°C, with shaking at 200 rpm. Subsequently they were subject to centrifugation at 1520 x g for 3 min. 400 µL of the medium was removed

and the cell pellet was resuspended in the remaining 100 μ L medium. Transformants were recovered by spreading the 100 μ L of resuspended cells on an LB agar plate. These were left to incubate overnight at 37°C.

2.2.10.5 Transformation of XL10-Gold ultracompetent cells

XL-Gold ultracompetent cells (Agilent Technologies) were thawed on ice and 45 μ L of cells were aliquoted into pre-chilled 14 mL Falcon polypropylene round-bottom tubes. 2 μ L of the β -mercaptoethanol (β -ME) mix was added to each aliquot of cells and incubated on ice for 10 min, mixing gently every 2 min. 1.5 μ L of DpnI-treated DNA was added to an aliquot of cells and incubated on ice for 30 min. Cells are heat shocked in a 42°C water bath for 30 seconds and immediately placed on ice for 2 min. 500 μ L of pre-heated (42°C) LB broth (in-house, Media Section, PHE, Colindale) was added to the transformed cells and incubated at 37°C for 1 hour with shaking at 200 rpm. Transformants were recovered by spreading 75 μ L of cells onto LB agar plates. These were incubated overnight at 37°C.

2.2.10.6 Transformation of One Shot® TOP10 chemically competent cells

One Shot® TOP10 chemically competent cells (Life Technologies) were thawed on ice and 2 μ L of the TOPO® cloning reaction was added and gently mixed. The cells were left to incubate for 30 min on ice and were heat shocked for 30 seconds in a 42°C water bath and immediately transferred to ice. 250 μ L of S.O.C medium was added before incubation at 37°C with shaking at 200 rpm for 1 hour. Transformants were recovered by spreading 75 μ L of cells onto LB agar plates and incubated overnight at 37°C.

2.2.10.7 Colony PCR

Colony PCR was carried out using Taq DNA polymerase Kit (Life Technologies) to screen bacterial colonies that were positive for plasmids expressing the insert of interest. Single colonies were picked from a culture plate; half of the colony was streaked onto a master plate and the other half added to a PCR reaction mix containing 1x PCR Buffer, 1.5 mM MgCl₂, 0.2 mM dNTP, 1 μ M forward primer, 1 μ M reverse primer, 1 U Taq and made up to a volume of 20 μ L nH₂O. The primers that were used were dependent on the insert.

The PCR reaction was then run using an Applied Biosystems 2720 Thermal Cycler under the following cycling conditions; 95°C for 5 min, 94°C for 30 seconds, 50°C for 30 seconds and 72°C for 3-4 min (as a general rule, 30 seconds per kb). Steps 2 to 4 were repeated an additional 34 times with a final extension step of 72°C for 7 min.

Positive colonies were set up for mini-prep (as described in section 2.2.10.8) and subsequently sent for sequencing (section 2.2.9).

2.2.10.8 Mini-preps of plasmid DNA

DNA was extracted from a single *E.coli* colony, which was used to inoculate 5 mL LB-broth containing 100 µg/mL ampicillin (Sigma-Aldrich) and incubated overnight at 37°C with shaking at 200 rpm. The cultures were centrifuged at 4500 x g for 3 min and the supernatant was discarded. Mini preps of plasmid DNA were generated using the QIAprep Spin Miniprep Kit (Qiagen) following manufacturer's instructions. Firstly the cell pellet is resuspended in 250 µL Buffer PI and subsequently cells are lysed with 250 µL of Buffer P2. The reaction is neutralized with 350 µL of Buffer N3 and the cell lysates are cleared by centrifugation at 17,900 x g for 10 min. The supernatant is added to the QIAprep spin column where the plasmid DNA adsorbs to the silica gel membrane. The membrane is washed by centrifugation for 1 min at 17,900 x g with 750 µL Buffer PE to remove contaminants and then dried by centrifugation at 17,900 x g for 1 min. A second wash step of centrifugation with 500 µL Buffer PB was required by HB101 cells to remove trace nuclease activity. 50 µL Buffer EB was subsequently added to the membrane and incubated at room temperature for 1 min. Purified plasmid DNA was eluted by centrifugation at 17,900 x g for 1 min.

2.2.10.9 Multi and single site-directed mutagenesis

Multi and single site-directed mutagenesis was carried out using the QuikChange Lightning Multi Site-Directed Mutagenesis kit and QuikChange II XL Site-Directed Mutagenesis kit respectively (both Agilent Technologies).

The multi site-directed mutagenesis PCR reaction contained 1x QuikChange Lightning Multi reaction buffer, 100 ng of each mutagenic primer (if 1-3 primers are being used) or 50 ng of each mutagenic primer (if 4-5 primers are being used), 1 µL dNTP mix, 0.75 µL QuikSolution, 100 ng ds-DNA template and 1 µL QuikChange Lightning Multi enzyme blend, made up to a final volume of 25 µL with nfH_2O .

The PCR reaction was incubated in a DNA Engine Dyad Peltier Thermal Cycler under the following cycling conditions; 95°C for 2 min, 95°C for 20 seconds, 55°C for 30 seconds and 65°C for variable time (as a general rule; 30 second extension per kb of plasmid length). Steps 2 to 4 were repeated an addition 29 times with a final extension step of 65°C for 5 min.

The single site-directed mutagenesis PCR reaction contained 1x reaction buffer, 125 ng forward primer, 125 ng reverse primer, 1 µL dNTP mix, 3 µL QuikSolution, 10 ng ds-DNA template and 2.5 U Pfu Ultra HF DNA polymerase, made up to a final volume of 50 µL using H_2O .

The PCR reaction was incubated in a DNA Engine Dyad Peltier Thermal Cycler under the following cycling conditions; 95°C for 1 min, 95°C for 50 seconds, 60°C for 50 seconds and 68°C for variable time (as a general rule; 1 minute extension per kb of plasmid length). Steps 2 to 4 were repeated an addition 29 times with a final extension step of 68°C for 5 min.

After the PCR reaction was complete, DpnI digestion with 1 µL enzyme (10 U/µL) was then performed in order to digest the methylated, parental ds-DNA plasmid (the PCR generated plasmid is not methylated and therefore not digested). 1.5 µL of DpnI-treated DNA is then used to transform XL-10 Gold ultracompetent cells as outlined in section 2.2.10.5.

2.2.11 Construction of *gag-pol* expression vectors for use in integrase inhibitor phenotypic susceptibility and replication capacity assays

The p8.9NSX *gag-pol* expression vector was modified to enable the cloning of patient-derived *IN*, *PR+RT* or full-length *pol* genes. This was accomplished by the introduction of unique restriction enzyme sites *Apa*I, *Cl*aI and *Eco*RI in p8.9NSX and/or patient-derived fragments by site-directed mutagenesis or PCR. A schematic of the patient-derived vectors can be found in Figure 2.3. The primers used to introduce the restriction enzyme sites are shown in Table 2.6.

2.2.11.1 Construction of p8.9NSXClal+vector

To facilitate the cloning of patient-derived *PR+RT* and *IN* domains separately it was necessary to introduce a unique restriction enzyme site at the end of *RT* and beginning of *IN*. A restriction site search of p8.9NSX revealed that a *Cl*aI restriction enzyme site flanking *IN* amino acids 4/5 would be appropriate. Single site-directed mutagenesis (described in section 2.2.10.9) using

primers BRhINClal- and BRhINClal+ was carried out and mutants were transformed into XL10-Gold ultracompetent cells (section 2.2.10.5). DNA mini-preps were prepared from the colonies and sequenced using PR, RT and IN sequencing primers (Table 2.5) to confirm the presence of the *Clal* restriction site.

Table 2.6 Primers used for the introduction of restriction enzyme sites in the p8.9NSX *gag-pol* expression vector

Primer name	Primer sequence (5'→3')	Orientation
GagApaF	GCAGGGCCCCTAGGAAAAAGGGC	Forward
FEcoRI	TAAAGAATTCCATGTGTTAATCCTCATCC	Reverse
BRhINClal-	GCTGTGATATTCTCATGTTCTTCTTGGGCCTT ATCGATT CCATCC	Reverse
BRhINClal+	CTGGAATCAGGAAAGTACTATTTTTGGATGGA ATCGATA AAGGCC	Forward
2A4RhINClal+	CTGGAATTAGAAAAGTACTATTTTTGGATGGA ATCGATA AAGGCC	Forward

NB: bold type face in sequence indicates restriction enzyme site

2.2.11.2 Construction of patient-derived *IN* expression vectors

Restriction site PCR was used to introduce a *Clal* restriction enzyme site at the beginning of *IN* (flanking *IN* amino acids 4/5) and an *EcoRI* restriction enzyme site at the end of *IN* (flanking amino acids 3/4 upstream of *IN*) in the single genomes amplified through the SGS assay. The PCR reaction was set up as described in section 2.2.5.4 using primers BRhINClal+ and FEcoRI and an annealing temperature of 50°C and an extension time of 1.5 minutes to generate the 860 bp product.

Positive PCR products were sequenced with the *IN* sequencing primers (Table 2.5) to ensure that the *IN* sequences were identical to the original single genome sequences. The PCR products and p8.9NSXClal+ vector were digested using *Clal* and *EcoRI* restriction enzymes (NEB) and insert DNA purified using the Illustra GFX PCR DNA and Gel Band Purification Kit (GE Healthcare) as described in section 2.2.7. The PCR products were then directly cloned into the dephosphorylated p8.9NSXClal+ expression vector using the Rapid DNA Dephos and Ligation Kit (Roche) and 2 µL of the ligation reaction used to transform HB101 competent cells as described earlier (sections 2.2.10.2 and 2.2.10.4 respectively). The incubation steps after the addition of S.O.C medium however were carried out at 30°C instead of 37°C to minimize the number of mutations occurring in the plasmid DNA during bacterial growth. DNA mini-preps of the plasmid DNA from single colonies were sequenced using *IN* sequencing primers (Table 2.5)

to verify that the sequence of the cloned *IN* genes were identical to the original single genome sequences.

2.2.11.3 Construction of patient-derived *PR+RT* expression vector (PR+RTp8.9NSXClal+)

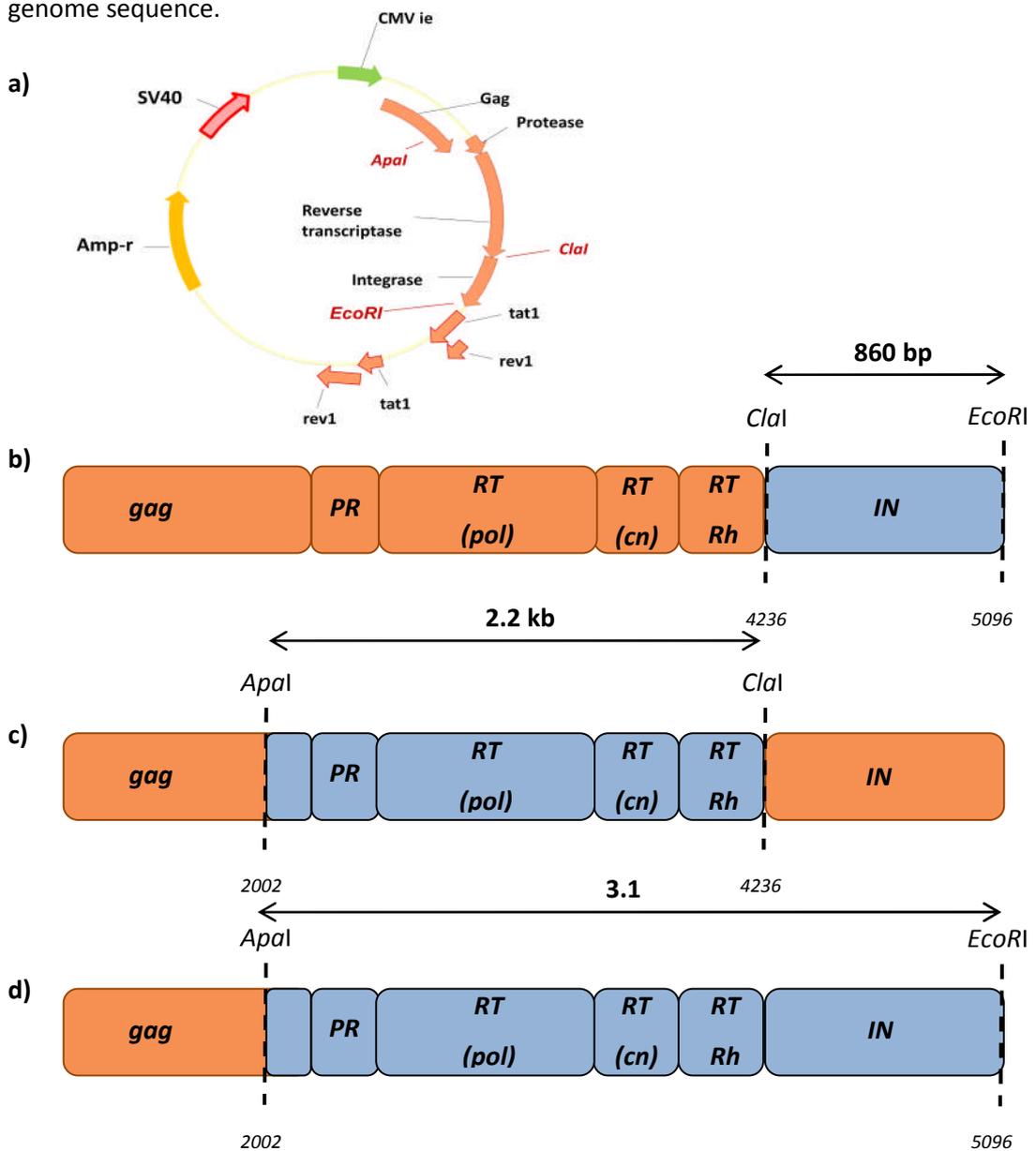
A patient-derived *PR+RT* expression vector was created by introducing an *ApaI* restriction enzyme site 242 bp downstream of *PR* and a *Clal* restriction site at the beginning of *IN* (flanking *IN* amino acids 4/5) in the 6B11 single genome. The primers used were GagApaF and BRHINClal- respectively. An annealing temperature of 50°C and an extension time of 2.5 minutes were considered to be adequate to amplify the 2.2kb *PR+RT* fragment.

The positive PCR product was sequenced with *PR* and *RT* primers (Table 2.5) to ensure that the sequence was identical to the original 6B11 *PR+RT* single genome. The PCR product was used for direct cloning into the p8.9NSXClal+ expression vector. Both the 6B11 PCR product and p8.9NSXClal+ were digested with the *ApaI* and *Clal* restriction enzymes (NEB). The 2.2 kb DNA fragment from the 6B11 PCR product was gel extracted, purified and ethanol precipitated as described in section 2.2.7 whilst the p8.9NSXClal+ was dephosphorylated (section 2.2.10.2). The purified 2.2 kb 6B11 insert and dephosphorylated vector were subsequently ligated and transformed into HB101 competent cells as described earlier (sections 2.2.10.2 and 2.2.10.4). Likewise, the bacterial culture incubation step during the transformation was carried out at 30°C instead of 37°C to minimize the number of mutations occurring in the plasmid DNA during bacterial growth.

2.2.11.4 Construction of patient-derived full-length *pol* expression vectors

The patient-derived *IN* expression vectors were digested with *Clal* and *EcoRI* restriction enzymes to generate the 860 bp *IN* insert. In addition the PR+RTp8.9NSXClal+ vector underwent digestion with the same enzymes to remove the *IN* fragment, creating the *PR+RT* backbone. As before, the insert was gel extracted, purified and ethanol precipitated whilst the vector backbone was dephosphorylated (sections 2.2.7 and 2.2.10.2 respectively). Subsequently ligation and transformation into HB101 competent cells was carried out (sections 2.2.10.2 and 2.2.10.4 respectively). Bacterial culture incubation steps during the transformation however were carried out at 30°C instead of 37°C to minimize the number of

mutations occurring in the plasmid DNA during bacterial growth. Sequencing with the PR, RT and IN primers (Table 2.5) ensured that extracted plasmid DNA was identical in PR and RT to the original 6B11 single genome sequence and identical in IN to the original cloned single genome sequence.



NB: Numbers in italics indicate the nucleotide position in HXB2; orange denotes sequences from the p8.9NSXClal+ vector; blue denotes sequences derived from patients

Figure 2.3 *gag-pol* expression vectors expressing patient-derived *IN*, *PR+RT* or full-length *pol* (a) map of p8.9NSXClal+, (b) schematic diagram of *gag-pol* region of vectors expressing patient-derived *IN* only, (c) schematic diagram of *gag-pol* region of vector expressing patient-derived *PR+RT* (*PR+RT*p8.9NSXClal+) and (d) schematic diagram of vector expressing patient-derived full-length *pol* gene. Different patient-derived fragments were sub-cloned into p8.9NSXClal+ using different restriction enzyme sites as indicated.

2.2.12 Construction of wild-type subtype F and patient-derived subtype F *PR+RT* expression vectors

The p8.9NSX *gag-pol* expression vector was modified to enable the expression of subtype F *gag-pol* sequences and the cloning of patient-derived *RT* domains. A schematic of the wild-type subtype F p8.9NSX vector backbone and wild-type subtype F- and patient-derived *PR+RT* and *RT* only expression vectors are shown in Figure 2.4. Primers used in this section are shown in Table 2.7.

Table 2.7: Primers used for the introduction and knockout of restriction enzyme sites in p8.9NSX-F

	Primer Name	Primer sequence (5'→3')	Orientation
Restriction PCR primers	A2+F NotI	AGCAGCGGCCGCAAGGAGAGAGATGGGTGC	Forward
	FEcoRI	TAAAGAATTCATGTGTTAATCCTCATCC	Reverse
SDM primers	FEcoRIKO-F	GTATCCAGCAGGAATTTGGTATTCCTTACAACCC CCAAAGTC	Forward
	FSpeIKO-F	GAAGTGATATAGCTGGAAC TACAAGCAC CCTTCA GGAAC	Forward
	FHpaIKO-F	ACACTAATGATGTAAAACAG TTGAC AGAAGCAGT GC	Forward
	FPolCnHpaI-F	CTTAGGGGAGCCAAGGCG TTAAC AGACATAGTG CCTACTGACT	Forward
	FChRhSpeI-F	TGGGAGTTTGTCAATACCCCC ACTAGT AAAAC TATGGTATCAGTTA	Forward
	FRhINClal-F3	GTTTCTAGATGGT ATCGATA AAGGCACAAGAGG	Forward
	SubFTins-F	GAACCCAGATTGTAAGACCATTTTAAAAGCATTG GGACCAGGG	Forward
	SubFTins-R	CCCTGGTCCCAATGCTTTTAAAATGGTCTTACAAT CTGGGTTC	Reverse

NB: bold type face in sequence indicates restriction enzyme site

2.2.12.1 Construction of p8.9NSX-F

A subtype F specific *gag-pol* expression vector was created by amplifying the *gag-pol* region from the HIV-1 subtype F molecular clone p93BR020.1 using primers that also introduced a *NotI* restriction enzyme site downstream of the *gag* gene initiation codon and an *EcoRI* restriction enzyme site upstream of the *IN* gene stop codon. The PCR reaction was set up as described in section 2.2.5.4 using restriction site PCR with primers A2+FNotI and FEcoRI to generate a 4.3kb product. An annealing temperature of 62°C and extension time of 4.5 minutes was deemed adequate.

The PCR product was then sub-cloned into a TA vector using the TOPO TA Cloning Kit (Life Technologies) as described in section 2.2.10.3.

To facilitate the cloning of different patient-derived *RT* domains, site-directed mutagenesis was carried out to introduce restriction enzyme sites *HpaI* (between the *pol* domain and *cn* subdomain in *RT*, flanking amino acids 289/290), *SpeI* (between the *cn* subdomain and RNase H domain in *RT*, flanking amino acids 421/422) and *Clal* (at the beginning of the *IN* gene, flanking amino acids 4/5) illustrated in Figure 2.4a. To make sure that the *HpaI*, *SpeI* and *EcoRI* restriction sites were unique, the naturally occurring sites in the *gag-pol* of p93BR020.1 were also eliminated by site-directed mutagenesis using either the QuikChange Lightning Multi Site-Directed Mutagenesis kit or the QuikChange II XL Site-Directed Mutagenesis kit (Agilent Technologies) as described in section 2.2.10.9.

DNA mini-preps of positive colonies were sequenced using *gag*, PR, RT and IN sequencing primers (Table 2.5) to verify the presence and absence of restriction enzyme sites and any other mutations.

The subtype F *gag-pol* fragment was then digested out of the TOPO TA vector using *NotI* and *EcoRI* restriction enzymes (NEB) to generate a 4.3 kb *gag-pol* insert. In parallel the p8.9NSX vector was also digested with *NotI* and *EcoRI* to generate the vector backbone (section 2.2.10.1). The insert was gel extracted, purified and ethanol precipitated whilst the vector backbone was dephosphorylated followed by ligation and the transformation of HB101 competent cells as described previously (sections 2.2.7, 2.2.10.2 and 2.2.10.4).

DNA mini-preps were prepared and sequenced using the PR, RT, IN and *gag* sequencing primers (Table 2.5) to verify successful ligation of the wild-type subtype F into the p8.9NSX vector.

2.2.12.2 Construction of patient-derived subtype F *PR+RT* expression vectors in a subtype F specific vector

PR+RT fragments were amplified from patients infected with HIV-1 subtype F using the GagApaF and R1SubtypeF primers to introduce an *Apal* site 242bp downstream of *PR* and a *Clal* restriction site at the beginning of the *IN* gene (flanking *IN* amino acids 4/5), generating a 2.2 kb fragment encompassing *PR+RT* (Figure 2.4b). The PCR reaction was prepared as previously described (section 2.2.5.4) an annealing temperature of 59.8°C and an extension time of 2.5 min were used.

The PCR products were then subcloned into a TOPO TA vector as previously described (section 2.2.10.3). Similar to the construction of p8.9NSX-F, internal restriction sites were introduced to separate the 3 domains of RT (Figure 2.4b). Some samples also required the knockout of internal *HpaI* and *SpeI* restriction sites. This was carried out using the QuikChange Lightning Multi Site-Directed Mutagenesis kit or the QuikChange II XL Site-Directed Mutagenesis kit (Agilent Technologies) as described in section 2.2.10.9. DNA mini-preps were prepared from positive colonies and sequenced using gag, PR, RT and IN sequencing primers (Table 2.5).

Both the p8.9NSX-F vector and patient-derived *PR+RT* TOPO vectors were subsequently digested with *Apal* and *Clal* restriction enzymes (NEB; section 2.2.10.1). The 2.2 kb patient-derived *PR+RT* fragment from the TOPO vector was gel extracted, purified and ethanol precipitated (section 2.2.7) whilst the p8.9NSX-F vector backbone was dephosphorylated followed by the ligation (section 2.2.10.2) and transformation of HB101 competent cells (section 2.2.10.4). DNA mini-preps of positive clones were sequenced using the PR and RT primers (Table 2.5) to verify that the clones were identical in amino acid sequence to the original PCR product.

A list of primers used for the construction of subtype F patient-derived retroviral vectors is shown in Table 2.8.

Table 2.8 Primers used for the introduction and knockout of restriction enzyme sites in subtype F patient-derived retroviral vectors

	Primer Name	Primer sequence (5'→3')	Orientation
Restriction PCR primers	GagApaF	GCAGGGCCCCTAGGAAAAAGGGC	Forward
	R1SubtypeF	CCTT ATCGAT CCCATCTAAAAACAGT	Reverse
SDM primers	195-1PolCnHpaI-F	TAGGGGAGCCAAGAC GTTAAC AGACATAG TGACATTGAC	Forward
	SubFPolCnHpaI-F	TAGGGGAGCCAAGGC GTTAAC RGAYATAG TGACRCTGAC	Forward
	183-2PolCnHpaI-F	TAGGGGAACCAAGGC GTTAAC AGACATAG TGACTGAC	Forward
	189-1PolCnHpaI-F	TAGGGGAACCAAGGC GTTAAC AGAAATAG TGCCACTGAC	Forward
	180-3PolCnHpaI-F	TAGGGGAGTCAAGGC GTTAAC AGACGTAG TGACTGAC	Forward
	197-1PolCnHpaI-F	TAGGGGAACCAAGGC GTTAAC AGACATAG TGACAATGAC	Forward
	SubFCnRhSpeI-F	TGTCAACACCCCC ACTAGT AAAACRTG GTATCAGTTAG	Forward
	180-3CnRhSpeI-F	TGCCAACCCCC ACTAGT AAAACATG GTATCAGTTAG	Forward
	183-2202-1CnRhS	TGTCAAYACYCCYCC ACTAGT AAAATTATGG TATCAGTTAG	Forward
	183-2HpaIKO-F	ACACTAATGATGTAAAAC GTTGAC AGAGG CAGTAC	Forward
	210-2HpaIKO-F	ACACTAATGATGTAGAAC GTTGAC AGATG CAGTAC	Forward
	SubFSpeIKO-F	CTGTCAATGATATACAGAA ACTTGT AGGAA AACTAAATTGGGC	Forward
	FPolCnHpaI-F	CTTAGGGGAGCCAAGGC GTTAAC AGACAT AGTGCCACTGACT	Forward
	FRhINClal-F	GTTTCTAGATGGG ATCGATA AAGGCACAAG AGG	Forward
	FRhINClal-F3	GTTTCTAGATGGT ATCGATA AAGGCACAAGA GG	Forward
	FRhINClal-F1	ATACTGTTTTTAGATGGT ATCGATA AAGGAA GGCGA	Forward
	FRhINClal-F2	GTA CTGTTTT AGATGGT ATCGATA AAGGAA GGCGA	Forward
	FRhINClal-R2	TCGCCCTTCCTT ATCGATA CCCATCTAAAAAC AGTAC	Reverse

NB: bold type face in sequence indicates restriction enzyme site

2.2.12.3 Construction of wild-type subtype F and patient-derived subtype F *PR+RT* expression vectors in a wild-type subtype B background

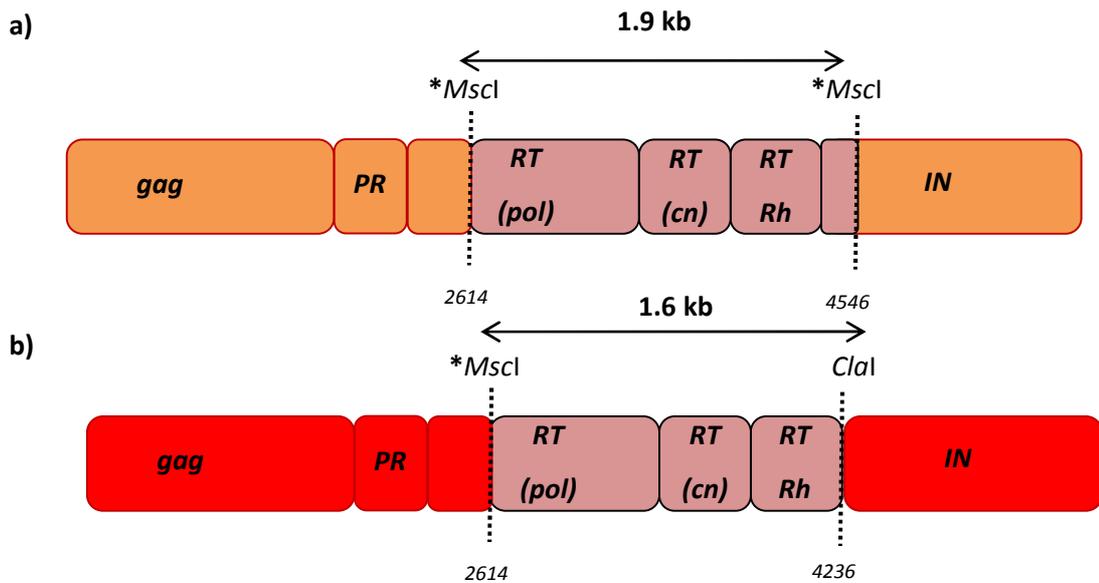
p8.9NSX-F and patient-derived *PR+RT* expression vectors were digested with *ApaI* and *Clal* restriction enzymes (NEB; 2.2.10.1). This generated a 2.2 kb fragment containing full length *PR+RT* (Figure 2.4c). The 2.2 kb *PR+RT*-containing fragment was gel extracted, purified and ethanol precipitated (section 2.2.7). In parallel the p8.9NSX was also digested with *ApaI* and *Clal* restriction enzymes and dephosphorylated. The 2.2 kb fragment was subsequently ligated with the dephosphorylated p8.9NSX vector (section 2.2.10.2) and transformed into HB101 competent cells as described in section 2.2.10.4. DNA mini-preps of positive clones were sequenced using the PR and RT primers (Table 2.5) to verify that the clones were identical in amino acid sequence to the original PCR product.

2.2.12.4 Construction of wild-type subtype F and patient-derived subtype F *RT*-containing retroviral vectors

p8.9NSX-F and patient-derived *PR+RT* expression vectors were digested with the *MscI* restriction enzyme (NEB; section 2.2.10.1). This generated a 1.9 kb fragment containing 1.6 kb of *RT* and 310 bp of *IN* through naturally occurring *MscI* restriction sites (Figure 2.5a). The 1.9 kb *RT*-containing fragment was gel extracted, purified and ethanol precipitated (section 2.2.7). In parallel the p8.9NSX (see section 2.2.1.2) was also digested with the *MscI* restriction enzyme and dephosphorylated (sections 2.2.10.1 and 2.2.10.2). The 1.9 kb fragment was subsequently ligated with the dephosphorylated p8.9NSX vector (section 2.2.10.2) and used to transform HB101 competent cells (section 2.2.10.4). DNA mini-preps of positive clones were sequenced using the PR and RT primers (Table 2.5) to verify that the clones were identical in amino acid sequence to the original PCR product.

An alternative vector was also used to construct wild-type subtype F and patient-derived subtype F *RT*-containing retroviral vectors (Figure 2.5b). The pHL(WT) (see section 2.2.1.3), p8.9NSX-F and patient-derived *PR+RT* expression vectors were digested with *MscI* and *Clal* restriction enzymes (NEB; section 2.2.10.1). The ~1.6 kb fragment from the p8.9NSX-F and patient-derived *PR+RT* enzyme reaction, encompassing *RT* only was gel extracted, purified and ethanol precipitated (section 2.2.7). The digested pHL(WT) vector was dephosphorylated and was ligated to purified insert as previously described in section 2.2.10.2. The ligated DNA was

transformed into HB101 competent cells (section 2.2.10.4) and mini-preps of positive clones were sequenced using the PR and RT primers (Table 2.5) to verify that the clones were identical in amino acid sequence to the original PCR product.



*Natural MscI restriction sites

NB: Numbers in italics indicate the nucleotide position in HXB2; **orange** denotes sequences from the p8.9NSX vector; **red** denotes sequences from pHL(WT); **maroon** denotes sequences derived from patients or p8.9NSX-F.

Figure 2.5 a) Schematic diagram of *gag-pol* region of p8.9NSX expressing wild-type subtype F and patient-derived *RT* only and b) schematic diagram of *gag-pol* region of pHL(WT) expressing wild-type subtype F and patient-derived *RT* only. Patient- and wild-type subtype F-derived *RT* fragments were sub-cloned into p8.9NSX and pHL(WT) using the *MscI* and *MscI* and *MscI* and *Clal* restriction enzyme sites.

2.2.13 Single genome sequencing

The single genome sequencing (SGS) assay described by Palmer *et al.* (Palmer *et al.*, 2005) was modified to amplify full length viral *PR*, *RT* and *IN* from patient samples. 20 μL of extracted viral RNA was denatured at 65°C for 10 min together with 0.08 μM of outer reverse primer KVL069 and 1 mM of dNTP.

First strand cDNA synthesis was carried out using the Superscript™ III Reverse Transcriptase kit (Life Technologies). The RNA denaturation mix was added to a reaction mix containing 1x First Strand Buffer, 1mM DTT, 20 U RNase-Out and 50 U Superscript III, made up to a final volume of 50 μL with nfH_2O . This mix was subsequently incubated in an Eppendorf Mastercycler for 50 min at 45°C followed by 10 min at 85°C.

Three-fold serial dilutions of the cDNA were performed in 2 mM Tris-HCl (pH 8.0) (Ambion) to a maximum of 1:2187. A cDNA reaction yielding 3 out of 10 positive PCR reactions was empirically determined and used to generate single genome products for each sample. According to Poisson's distribution, a cDNA dilution yielding 30% positive PCR reactions will contain a single genome 80% of the time. Subsequently, 1 μL of diluted cDNA was then used for first round PCR. The reaction was prepared per 96 well plate using 720 μL Platinum PCR SuperMix® High Fidelity (contains all reagents required for PCR amplification and includes the "hot-start" Platinum Taq previously described in section 2.2.5.3, as well as a polymerase that possesses a 3' exonuclease proofreading activity that increases fidelity; Life Technologies), 0.22 μM 1849+ and 0.22 μM KVL069 (sections 2.2.5.1 and 2.2.5.2).

9 μL of the reaction mix was used for each reaction and the cycling conditions used were as follows; 94°C for 2 min, 94°C for 30 seconds, 57.5°C for 30 seconds and 72°C for 3.5 min. Steps 2 to 4 were repeated an additional 44 times with a final extension step of 72°C for 5 min.

Nested PCR reactions were also carried out using Platinum® PCR SuperMix High Fidelity (Life Technologies). Again, the reaction mix was prepared per 96 well plate; 1440 μL Platinum PCR SuperMix High Fidelity®, 0.22 μM 1870+ and 0.22 μM inner reverse primer (either 5222-deg, 5095- or KVL084 (section 2.2.5.3)

19 μL of the reaction mix was mixed with 1 μL of first round PCR product for each reaction and the following cycling conditions were used; 94°C for 30 seconds, variable temperature for 30 seconds (different annealing temperatures were used depending on the primer set and can be

found in Table 2.4) and 72°C for 3.5 min. Step 1 to 3 were repeated an additional 40 times with a final extension step of 72°C for 5 min.

Positive PCR amplicons were identified using a 1% E-Gel® 96 Agarose (Life Technologies) and a 1 kb ladder (TrackIt™ 1Kb Plus DNA Ladder, Life Technologies) diluted 1:20 in nfH_2O . Positive PCR products were purified using Illustra GFX PCR DNA and Gel Band Purification Kit (GE Healthcare) (section 2.2.6) and sequenced using PR, RT and IN sequencing primers (Table 2.5). DNA sequences were analyzed as described in section 2.2.9. Any sequences that contained double peaks in both forward and reverse reads were classified as not being single genomes and therefore discarded.

2.2.14 Long cDNA single genome sequencing

For samples that proved difficult to amplify a long cDNA single genome sequencing protocol developed by Mary Kearney (NCI-Frederick) was used (Galli et al., 2010). The protocol uses modified cycling conditions however all reaction mixes remained as described previously (section 2.2.13). The cDNA synthesis cycling conditions were as follows, 50°C for 1 hour, 55°C for 1 hour and 70°C for 15 min.

Additionally, after cDNA synthesis, an RNase H step was carried out at 37°C for 20 min using 2 U of RNase H (Life Technologies). For first round PCR the following cycling conditions were used; 95°C for 2 min, 95°C for 30 seconds, 49.5°C for 30 seconds (the annealing temperature is 8°C lower than used previously for KVL069 to increase the likelihood of annealing to the template sequence) and 68°C for 4 min. Steps 2 to 4 were repeated an additional 44 times.

The nested PCR cycling conditions were as follows; 94°C for 2 min, 94°C for 15 seconds, variable temperature for 30 seconds (61.5°C for 5095- and 60.6°C for KVL084) and 68°C for 4 min. Steps 2 to 4 were repeated an addition 9 times followed by 94°C for 15 seconds, variable temperature for 30 seconds (61.5°C for 5095- and 60.6°C for KVL084) and 72°C for 3 min 30 seconds. The final 3 steps are repeated an additional 29 times, increasing the extension step by 5 seconds each cycle finishing with a final step of 72°C for 10 min.

The temperatures used in both annealing steps were 2°C higher than that used previously for 5095- and KVL084 to increase the sequence binding specificity. The extension step was also increased incrementally each cycle to ensure enough time for full extension of the template.

2.3 Cell culture

2.3.1 Cells

The human embryonic kidney 293T cell line which expresses the SV-40 large T antigen was obtained from the American Type Culture Collection (ATCC). Cells were maintained in growth media made up of Dulbecco modified Eagle medium (DMEM; Life Technologies) supplemented with 10% fetal bovine serum (FBS), 100 mM sodium pyruvate (Life Technologies), 50 U/mL of penicillin (Life Technologies) and 50 µg/mL streptomycin (Life Technologies). The cells were grown in a humidified 37°C incubator with 5% CO₂. The cells were subcultured every two to three days by rinsing the monolayer once with phosphate buffered saline (PBS) followed by incubation with Trypsin-EDTA (Life Technologies) at 37°C with 5% CO₂ for 1-2 min. Once the cells were detached the trypsin was neutralized by the addition of growth media, mixed by pipetting up and down to break up any cell clumps and split 1:4 or 1:6. Stocks of 293T cells were stored in liquid nitrogen at a density of 10⁶ cells/mL of freezing media (growth media plus 6% dimethyl sulfoxide; DMSO). To start a fresh passage of 293T cells a vial of frozen cells was removed from liquid nitrogen storage, thawed quickly in a 37°C water bath and seeded onto a 100 mm plate containing 10 mL of pre-warmed growth media. The plate was then incubated at 37°C with 5% CO₂ and the media was changed the following day.

2.3.2 Virus Production

A previously described method (Parry et al., 2009) was used to produce pseudoviruses containing fragments amplified from patient-derived viruses. To produce pseudoviruses 293T cells were seeded and co-transfected with the p8.9NSX vector expressing patient-derived *gag-pol* fragments together with two other plasmids; pCSFLW (section 2.2.1.4) and pMDG (section 2.2.1.5). The amounts were as follows: 300 ng of the p8.9NSX-derived *gag-pol* expression vector, 300 ng of the pMDG VSV-G envelope vector and 500 ng of pCSFLW. This was made up to a volume of 10 µL with nfH₂O. Transfection of pHL(WT)-derived vectors was carried out as with the p8.9NSX-derived vectors, however 750 ng of pHL(WT)-derived vector was used instead of 300 ng of p8.9NSX-derived vector and no pCSFLW was required. The amount of pMDG remained the same (300 ng) and the two vectors were also made up to a volume of 10 µL with nfH₂O.

Following transfection with p8.9NSX-derived vectors pseudovirions are produced that contain HIV-1 structural and enzymatic proteins encoded by the p8.9NSX *gag-pol* expression vector and an HIV-1 based genome encoding the *luciferase* reporter gene from the pCSFLW vector surrounded by a VSV-G envelope from the pMDG vector. However, transfection with pHL(WT)-derived vectors generated pseudovirions containing HIV-1 structural and enzymatic proteins in addition to the *luciferase* reporter gene encoded by pHL(WT) which are surrounded by the same VSV-G envelope.

2.3.2.1 Transfection with PEI

The day before transfection 293T cells were seeded into wells of a 6 well plate at 6.25×10^5 cells in 4 mL DMEM per well. The 10 μ L of vector mix (described in section 2.3.2) was added to 150 μ L Opti-MEM (Invitrogen). 10 μ L of linear polyethylenimine (PEI) (Alfa Aesar) was then added and the transfection mix was incubated at room temperature for 15 min. The mix was subsequently added drop-wise to the 293T cells in which the 4 mL DMEM was previously replaced with 2 mL fresh DMEM. Plates were incubated at 37°C with 5% CO₂. After 48 hours the pseudoviruses were harvested from the supernatant using 0.45 μ m filters and stored in two aliquots at -80°C.

2.3.2.2 Transfection with FuGENE6

Transfection with FuGENE 6 Transfection Reagent (Roche) was carried out in a similar fashion to PEI transfection. 293T cells were seeded into wells of a 6 well plate at 6.25×10^5 cells in 4 mL DMEM per well the day before transfection. 6 μ L of FuGENE 6 was added drop-wise to 70 μ L Opti-MEM (Life Technologies) and mixed by flicking. This was left for 3-5 min and subsequently the 10 μ L of the two or three vectors (section 2.3.2) was added drop-wise into the mix, this was once again mixed by flicking. The mix was then added drop-wise to the 293T cells in which the 4 mL DMEM was previously replaced with 2 mL fresh DMEM. Plates were incubated at 37°C with 5% CO₂. After 48 hours the pseudoviruses were harvested from the supernatant using 0.45 μ m filters and stored in two aliquots at -80°C.

2.3.3 Single-replication cycle drug susceptibility assay

The pseudovirions that are generated by co-transfection of two or three plasmids bud from the producer 293T cells and upon infection of fresh 293T cells the luciferase mRNA integrates

into the host cell genome. The expression of luciferase can be measured as an indicator of infection. Upon the addition of INSTIs during the infection stage of the single replication cycle the measure of subsequent luciferase expression is an indicator of resistance to the ARV drug and an EC_{50} can be calculated. Briefly the harvested pseudoviruses are titrated to determine the amount of pseudovirus needed to produce 10^6 Reflective Light Units (RLUs; measure of luciferase expression). This volume is then used to infect fresh 293T cells in the presence of serially diluted drug. Luciferase expression is measured and an EC_{50} value (half maximal inhibitory concentration; amount of drug required to inhibit a biological process by half) is calculated.

This is a single-cycle replication assay as the pseudovirions harvested after transfection do not contain gag-pol mRNA and so no further HIV virions can be produced. This permits the assay to be performed in a Category 2 laboratory rather than at Category 3. An outline of the procedure can be found in Figure 2.6.

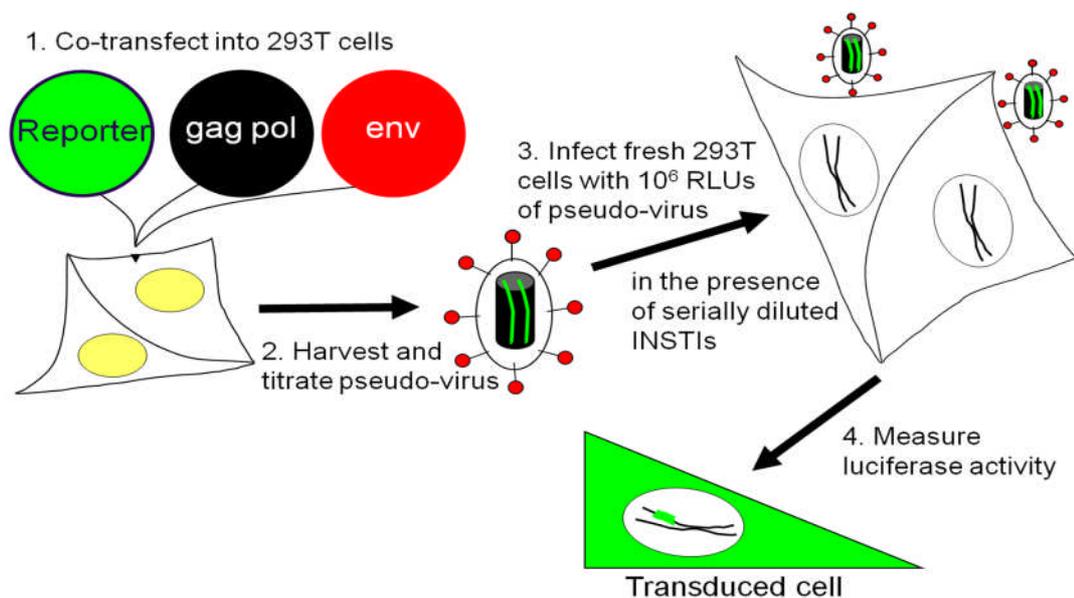


Figure 2.6 Single-cycle replication assay. (1) Co-transfection of the three plasmids namely, pCSFLW an HIV-based vector which expresses the *firefly luciferase* reporter gene (green), p8.9NSX expressing patient-derived *gag-pol* (black) and pMDG expressing VSV-G envelope glycoprotein (red) into 293T cells produces pseudovirions containing a luciferase expressing RNA genome (2). The pseudovirions are used to infect fresh 293T target cells resulting in the integration of proviral DNA encoding the *firefly luciferase* gene into the host cell genome (3). Infectivity is indicated by luciferase expression and drug susceptibility of the pseudoviruses is determined by infection in the presence of serial dilutions of antiretroviral drugs (4). Adapted and modified from Chris Parry (MRC, Uganda)

2.3.3.1 Pseudovirus titration

Harvested pseudoviruses were titrated on 293T cells to determine infectivity. Two-fold serial dilutions of each pseudovirus stock, up to 1:2084, were made in 50 μ L of DMEM growth media across a 96 well plate in duplicate. 2×10^6 293T cells in a volume of 5 mL per 96 well plate were harvested and plated out at 50 μ L per well. An illustration of this process can be found in Figure 2.7. Plates were left to incubate at 37°C with 5% CO₂ for 72 hours. Steady-glo substrate (Promega) was used to detect luciferase expression post-incubation. 50 μ L of substrate was added to each well and plates were incubated on a shaker for 10-15 min. Luciferase activity was measured using the GloMax-Multi Detection System (Promega).

2.3.3.2 Integrase strand transfer inhibitor phenotypic susceptibility assay

To measure pseudovirus susceptibility to INSTIs a phenotypic susceptibility assay was set up with RAL and EVG, a schematic of which can be found in Figure 2.8. Firstly, in a 96 well plate, 101 μ L DMEM was added into alternate rows of columns 2 to 12. 151 μ L of the doubled maximum drug concentration is added to column 1. For RAL (ARP980, NIBSC), 151 μ L of 6 μ M stock was added and for EVG (ARP991, NIBSC) a 200 nM stock was used. For the p8.9NSX control, patient-derived wild-type *IN* and the PR+RTp8.9NSXClal+ vectors however a maximum EVG concentration of 1 nM was used, i.e. 151 μ L of a 2 nM stock was added. 3-fold dilutions were performed across the plate, leaving column 12 drug free. The dilution series of each drug is depicted in Table 2.9.

Table 2.9 Dilution series used in the INSTI phenotypic susceptibility assay

	1	2	3	4	5	6	7	8	9	10	11
RAL (nM)	3000	1000	330	110	37	12	4.1	1.4	0.46	0.15	0.05
EVG (nM)	100	33.3	11.1	3.70	1.23	0.41	0.14	4.6 $\times 10^{-2}$	1.5 $\times 10^{-2}$	5.1 $\times 10^{-3}$	1.6 $\times 10^{-3}$
EVG (nM)	1	0.33	0.11	3.7 $\times 10^{-2}$	1.2 $\times 10^{-2}$	4.1 $\times 10^{-3}$	1.4 $\times 10^{-3}$	4.6 $\times 10^{-4}$	1.5 $\times 10^{-4}$	5.1 $\times 10^{-5}$	1.7 $\times 10^{-5}$

This dilution series is duplicated in the row below by transferring 50 μ L from the first row in to the second. For each pseudovirus to be tested a total 1.2×10^6 293T cells in 3 mL were harvested. The amount of virus required to give an approximate RLU of 10^6 was calculated from the previous titration results (section 2.3.3.1), this volume was added to the 3 mL of 293T

cells. 50 μL of the cell + pseudovirus mix was then added to all wells. Plates were incubated at 37°C with 5% CO_2 for 48 hours. Steady-glo substrate (Promega) was used to detect luciferase expression post-incubation. 50 μL of substrate was added to each well and plates were left to rock for 10 min. Plates were then read using the Glomax Multi Detection system (Promega).

2.3.3.3 EC_{50} calculation

Luciferase expression levels of pseudoviruses incubated in varying concentrations of INSTIs relative to that of no drug control were used to calculate the percent inhibition of virus replication in the presence of the drugs. The data was then used to determine the EC_{50} values for RAL and EVG by computing the best-fit line using the linear regression by the least squares method. Fold changes in EC_{50} of the patient-derived viruses were calculated relative to wild type p8.9NSX control.

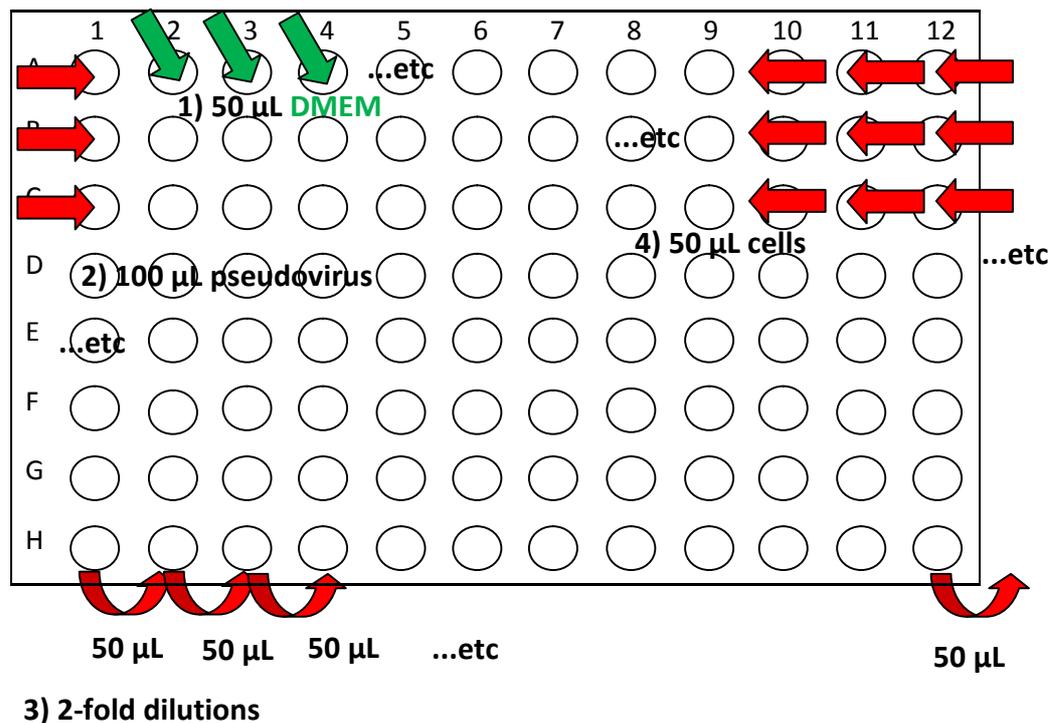


Figure 2.7 Schematic of the set up of pseudovirus titration. A schematic representation of pseudovirus titration set up. (1) Add 50 μL DMEM into columns 2-12 (2) Add 100 μL pseudovirus in column 1 (3) 2-fold dilutions across the plates mixing in between transfers (4) Add 50 μL cells at 2×10^4 cells per well to all wells. Plates are left to incubate at 37°C with 5% CO_2 for 72 hours and Steady-glo substrate was used to detect luciferase expression post-incubation.

2) 151 μL INSTI 1) 101 μL DMEM

5) 50 μL pseudovirus
+ cell mix

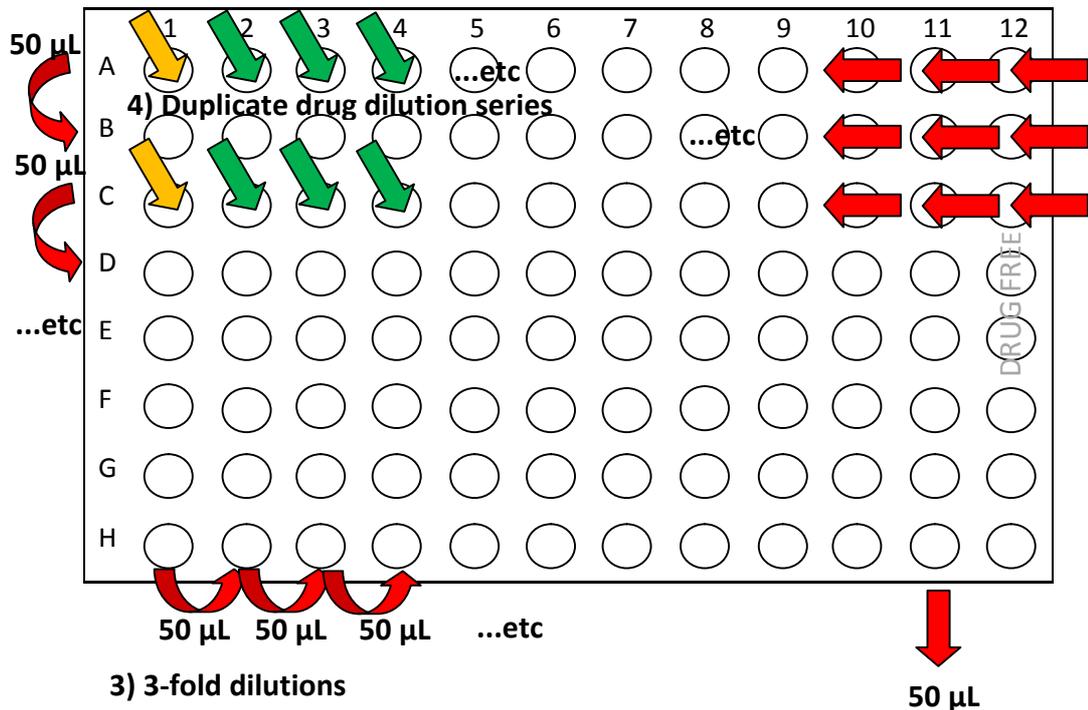


Figure 2.8 Schematic of the set up of the INSTI phenotypic susceptibility assay. A schematic representation of INSTI drug susceptibility assay set up. (1) Add 101 μL DMEM into columns 2-12 in alternate rows (for example A, C, E etc) (2) Add 151 μL of the doubled top drug concentration to column 1 (for RAL 151 μL of 6 μM stock and for EVG 151 μL of 200 nM stock) (3) 3-fold dilutions across the plates mixing in between transfers leaving column 12 as a drug free control (4) Duplicate the drug dilution series by adding 50 μL to the row below (for example add 50 μL from row A to row B, row C to row D etc) (5) Add 50 μL of pseudovirus + cell mix (volume of 10^6 RLU of pseudovirus in 3 mLs of cells at 2×10^4 cells) to all wells. Plates are left to incubate at 37°C with 5% CO_2 for 48 hours and Steady-glo substrate was used to detect luciferase expression post-incubation.

2.3.4 Replication capacity assay

To determine the replication capacity of pseudoviruses a serial dilution titration series of the virus samples on 293T cells was performed. Infectivity was determined at 48 hours post-infection using Steady-Glo (Promega). Infectivity data within the linear range were normalized to the amount of p24 capsid determined by ELISA and used to determine replication capacity

by expressing the values as a proportion of the patient-derived wild-type in *IN* virus control which was set to 100%. Results were calculated based on three independent experiments.

2.3.4.1 p24 Enzyme-linked immunosorbent assay (ELISA)

The Genscreen™ HIV-1 Ag Assay from Bio-Rad was used for the detection of HIV-1 p24 antigen in harvested pseudoviruses following the manufacturer's protocol. Briefly, 50 µL of specimen diluent was added to the wells, followed by 150 µL of 1:900 or 1:2700 DMEM- diluted pseudovirus. This was incubated for 60 min at 37°C in a dry-heat static incubator to allow the binding of any p24 antigen present in the sample to the antibodies coated on the well. Any bound antigen will remain attached during the subsequent washes. The wells were washed 5 times with a 30 second soak and blot dried on absorbent paper.

100 µL of working conjugate 1 was added to each well and incubated as previously but for 30 min, allowing the biotinylated sheep anti-p24 antibody in conjugate 1 to bind to the antigen that is attached to the wells. These antigen-antibody complexes remain bound during the subsequent wash step (as above). 100 µL of working conjugate 2 was then added and the wells incubated for a further 30 min. This allows the avidin-HRP (horseradish peroxidase) in conjugate 2 to specifically bind to the complexes that are bound to the well. Wells are subsequently washed to remove any unbound conjugate.

100 µL of working TMB solution, which contains hydrogen peroxide (a substrate of HRP), was added to each well and incubated in the dark for 30 min at room temperature. A blue-green colour change develops during oxidation of the TMB solution which is proportional to the amount of HIV-1 p24 antigen bound to the well. This was measured spectrophotometrically using an ELX808 ultra microplate reader at a wavelength of 450/630 nm after the addition of 100 µL of the stop solution.

Each assay was run with a p24 standard (AG6054 recombinant HIV-1-p24gag, Aalto Bio Reagents Ltd) of 2-fold dilutions starting at 100 pg/mL down to 6.25 pg/mL. A negative control (DMEM) was also added to one of the 8 wells in every strip used. An assay was deemed successful if the mean absorbance of the positive controls was greater than or equal to 0.5 absorbance units (AU) and the mean absorbance of the negative controls was greater than 0.0 AU but less than or equal to 0.1 AU. The AU values of the p24 standard were used to create a standard curve, from which the p24 values of the pseudoviruses were extrapolated.

2.4 Protein analysis

2.4.1 Antibodies

Primary antibodies, human HIV Ig, (catalogue #3957) and mouse monoclonal antibody to HIV-1 p24 (catalogue #4121) were obtained from the NIH AIDS Research and Reference Reagents Program. The goat anti-human Ig (catalogue #AH10305, Life Technologies) secondary antibody was donated by Sara Bisset, HPV, PHE, Colindale.

2.4.2 Western blotting

HIV-1 proteins were detected either from cell lysates of pseudovirus-infected cells or from virions in supernatants of transfected cells. Viral supernatants were harvested as described in sections 2.3.2.1 and 2.3.2.2. Cell lysates were harvested from 293T infected cells seeded in a 6-well plate 12 and 24 hours post-infection by washing the cell monolayer with 1 mL PBS followed by incubation with 0.5 mL Trypsin-EDTA for 1 min at 37°C. 1 mL DMEM growth media was then added to neutralize the reaction and the cells were subsequently harvested in a 2 mL tube and centrifuged at 1520 x g for 4 min. The supernatant was removed and the cell pellet was frozen down at -20°C until required.

Harvested cell pellets were lysed with 200 µL Mammalian Protein Extraction Reagent® (M-PER®; Thermo Scientific) and centrifuged at 18626 x g for 7 min. The supernatant was removed and used for western blotting. Proteins were separated using the Novex NuPAGE Gel Electrophoresis System (Life Technologies). Briefly, 8 µL of each sample was denatured using 2 µL NuPAGE® Sample Reducing Agent and 10 µL Novex® Tris-Glycine SDS sample buffer at 85°C for 2 min. 17.5 µL of each sample was then loaded onto separate wells of a precast Novex® 4-20% Tris-Glycine Mini Gel as well as 10 µL of SeeBlue® Plus2 Pre-Stained Standard as a molecular weight marker. The gel was run in 1× Novex Tris-Glycine SDS Running Buffer at a constant voltage of 150 V for 1.5 hours. The separated proteins were then transferred onto a 0.2 µm PVDF membrane (Life Technologies) by electroblotting at constant voltage of 30 V for 2 hours. Subsequent blocking, antibody washing and membrane development steps was carried out with reagents from the WesternBreeze® Chromogenic Western Blot Immunodetection Kit (anti-mouse; Life Technologies).

After electroblotting, the membrane was rinsed in distilled water and blocked for 30 min using the kit blocking solution (5 mL distilled water, 2 mL Blocker Part A and 3 mL Blocker Part B) and washed two times with distilled water. The membrane was then incubated for 1 hour whilst shaking with a primary antibody diluted in the kit antibody diluent (7 mL distilled water, 2 mL Diluent Part A and 1 mL Diluent Part B), 1:5000 and 1:2000 for human Ig and mouse monoclonal respectively. The membrane was washed with PBS-Tween (PBS-T; in-house, Media Section, PHE, Colindale) four times and incubated for 30 min on a shaker with a secondary antibody. The goat anti-human Ig secondary antibody was diluted 1:10000 in the primary antibody diluent (as previously) whilst 10 mL of the Secondary Antibody Solution (containing anti-mouse antibodies) was incubated with the mouse monoclonal-treated membrane.

After incubation washing was carried out as before (four washes with PBS-T) with the addition of three extra washes with distilled water. Development of the membrane was carried out by adding 5 mL of chromogenic substrate to the membrane and incubation with shaking for 1 to 60 minutes until the development of purple bands. The membrane was washed three times with distilled water and allowed to air dry.

2.4.2.1 Band density quantification

The densities of western blot bands were quantified using ImageJ software v1.47 (Schneider et al., 2012). Individual band densities within a lane were quantified relative to each other and expressed as a proportion of 100%.

2.5 Phylogenetics and bioinformatics

2.5.1 Sequence assembly

Sequences were assembled using Sequencher 4.9 software (Gene Codes Corporation). Prior to assembly, contigs were trimmed to remove poor quality data at both the 5' and 3' ends of the contigs and secondary peaks were called at 25%. Contigs were assembled onto a reference sequence and manually trimmed to ensure the correct open reading frame. In addition contigs were manually edited at ambiguous bases depending on the state of the chromatogram. Mixed bases were defined as mixtures that occurred in more than one sequencing contig and which consisted of more than 25% of the dominant peak. For single genome sequence

analysis, sequences containing one or more mixtures were discarded and not used for further analysis.

2.5.2 Sequence alignment

Sequence contigs were imported into MEGA4.1 alignment software (Tamura et al., 2007). These were used to create a nucleotide sequence alignment using the Clustal W algorithm (Thompson et al., 1994) with default settings. Sequences were manually checked to ensure they were in the correct open reading frame by translating into protein.

2.5.2.1 Single genome sequences

Sequences were trimmed to only include the *pol* coding region and were manually checked for sequence abnormalities. Any sequences containing premature stop codons or mixed bases, were discarded from the analysis.

In addition to full-length *pol* alignments of the single genome sequences, alignments were made for each of the individual *pol* genes. *PR* alignments consisted of base pairs 1 to 297; *RT*, base pairs 298 to 1977; and *IN*, base pairs 1978 to 2844.

2.5.2.2 Subtype F sequences

The subtype F sequence alignment contained *PR* and full-length *RT*. The analysis of C terminal mutations in these sequences involved a manual search through the alignment noting the amino acid differences between RTI-resistant and RTI-sensitive patients (as determined by the presence or absence of N terminal RTI resistance mutations defined by the Stanford HIV drug resistance database).

2.5.3 Database analysis for C terminal mutations

To determine the frequency of each novel C terminal mutation found in RTI-sensitive and -resistant patients infected with subtype F we used the “Detailed RT Mutation Query” tool on the Stanford University HIV drug resistance database (Rhee et al., 2006). The frequency of mutations was calculated as a percent of the total number of deposited sequences. The

percent difference of mutation frequency between the two patient populations was calculated by subtracting the mutation frequency of RT-sensitive patients from RTI-treated patients.

2.5.4 Phylogenetic reconstruction of inpatient viral evolution

The phylogenies of the inpatient viral populations were estimated using a maximum likelihood (ML) approach with the program PhyML version 3.0 (Guindon and Gascuel, 2003). The ML method (Felsenstein, 1973; Felsenstein, 1981) evaluates the probability of observing a tree given the sequence alignment. It involves the calculation of likelihood for all theoretical trees for a sequence alignment given a specific nucleotide substitution model. The tree(s) with the highest or maximum likelihood is the inferred phylogenetic tree (Vandamme et al., 2009). The single genome sequence alignments of full-length *pol* or individual gene segments derived from patient A or B were generated using MEGA4.1 software (Tamura et al., 2007) and imported into PhyML 3.0 to construct ML phylogenetic trees under the GTR model of nucleotide substitution. The statistical robustness of the trees was evaluated by bootstrap analysis with 1,000 rounds of replication. The phylogenetic trees were visualized and edited using FigTree software version 1.3 (<http://tree.bio.ed.ac.uk/software/figtree/>) and MEGA5 software (Tamura et al., 2011).

2.5.5 Inferring the genetic diversity of the HIV-1 *pol* gene of patients A and B

To investigate the inpatient viral diversity we calculated the pairwise genetic distance of all single genomes from each patient at each time point using TREEPUZZLE software under the GTR model of nucleotide substitution (Schmidt et al., 2002). The analysis was performed on the full-length *pol* nucleotide sequence alignment as well as each individual *pol* gene nucleotide alignment. The data was then used to determine the mean pairwise genetic distance and plotted using GraphPad Prism 6 software (Graph Pad Software Inc.).

2.5.6 Analysis of inpatient viral evolution and phylodynamics using the coalescent theory

The inpatient evolutionary history was estimated using BEAST (Bayesian Evolutionary Analysis Sampling Trees) software version 1.6.1. BEAST is a statistical framework based on Bayesian inference and prior knowledge of the sequence data (Drummond and Rambaut, 2007). It searches for the tree best fitted to the sequence alignment and allows the user to test a variety of evolutionary models that best fit the data including nucleotide substitution models (models that define the rates at which different substitutions occur along the branch of a tree), molecular clock models (rate of mutation/ evolution) and demographic models (a mathematical function that describes the change in the effective population size over time) (Drummond et al., 2005). This is implemented by the use of the Metropolis-Hastings Markov Chain Monte Carlo (MCMC) methods (Hastings, 1970). BEAST xml files were created in BEAUTi v1.6.1 (part of the BEAST v1.6.1 package) for each individual gene alignment. Each sequence was given a tip date (e.g. 2007.88) which identifies the year and month of sampling.

Initially, we considered two nucleotide substitution models (SRD06 or GTR), two demographic models (constant and Bayesian skyline plot) and two molecular clock models (strict and uncorrelated lognormal relaxed). The only prior to be changed was the clock.rate. This was set as lognormal with an initial value and log(Mean) of 2.5×10^{-3} and a log(Stdev) of 0.1. The mean was also calculated in real space. The Markov chain Monte Carlo (MCMC) was set to run for 50 million generations, logging parameters every 1000th generation and the option to create a substitutions trees file and operator analysis file were selected.

The xml files were run in BEAST v.1.6.1. After BEAST analysis the created log files were opened and analysed in Tracer v1.5 (part of the BEAST v1.6.1 package) to determine if the analysis was successful by ensuring the effective sample size (ESS) values of each parameter lay between 150-200, ideally more than 200. Each alignment was run twice to ensure that the results were consistent across runs. Model comparison was achieved by calculating the Bayes Factor (BF), the ratio of marginal likelihoods of the two models being compared as implemented in Tracer v1.5. A BF > 20 was considered as strong evidence against the null model.

2.5.6.1 Bayesian Skyline reconstruction

Bayesian skyline reconstruction was performed in Tracer v1.5. The BEAST time.tree output file was used and the maximum time as the root height's median was opted for.

2.5.7 Analysis of selection pressure on the HIV-1 *pol* gene of patients A and B

The web-based Datamonkey software package was used to analyse the inpatient selection pressures on the HIV-1 *pol* gene in the patient samples. The software calculates the ratio (ω) of the rate of nonsynonymous substitutions per nonsynonymous site (dN) over the rate of synonymous substitutions per synonymous site (dS). If dN/dS is equal to 1 then no selection is inferred. If dN/dS is <1, the number of dS is greater than dN and therefore negative selection is inferred. If however dN/dS is >1, the number of dN is greater than dS and so positive selection is in play (Vandamme et al., 2009). Three different algorithms were used to calculate dN/dS ratios these being: FEL (Fixed Effects Likelihood), SLAC (Single Likelihood Ancestor Counting) and FUBAR (Fast Unbiased Bayesian AppRoximation) (Kosakovsky Pond and Frost, 2005; Murrell et al., 2013). FEL is an algorithm based on calculating dN/dS ratios for each codon in a given sequence. SLAC on the other hand calculates both global (full alignment) and codon-specific dN/dS ratios and is therefore a more conservative method (Vandamme et al., 2009). FUBAR is relatively new and can run up to 50 times faster than the previously mentioned algorithms. It detects selection under a model which allows substitution rate variations from site to site. However, FUBAR calculates the mean posterior distribution of synonymous (α) and non-synonymous (β) substitution rates (Kosakovsky Pond and Frost, 2005; Murrell et al., 2013).

Positive and negative selection analysis was performed on the individual *pol* gene amino acid alignments of single genome sequences from patient A and patient B. For analysis using DataMonkey each gene coding sequence alignment was uploaded along with a corresponding ML-inferred phylogenetic tree. Substitution models were determined using the automatic substitution model selection tool and a genetic algorithm for recombination detection (GARD) was performed. Subsequent analyses used the GARD generated trees. For both the PR and IN sequence alignments the selected model was the HKY85 model whilst the O12212 model was selected for the RT sequence alignment. SLAC, FEL, and FUBAR algorithms were subsequently run with each predetermined substitution model and positive or negative selection was estimated with confidence intervals of 1.0 and a significance level of <0.05 (SLAC and FEL) and 0.95 (FUBAR).

2.6 Statistics

2.6.1 Subtype F C terminal of *reverse transcriptase* analysis

Significance between the frequency of C terminal mutations present in RTI-resistant and RTI-sensitive patients was calculated using the Fisher's exact test tool available on www.graphpad.com/quickcalcs/contingency1.cfm.

2.6.2 Pairwise genetic distance

The one-way ANOVA test was used to calculate the significance between pairwise genetic distances across each time point. Subsequently the student's t test was used to test for significance between two different time points. Both tests were calculated using GraphPad Prism 6 Software (Graph Pad Software Inc.).

2.6.3 Phenotypic and replication capacity assays

The mean, standard deviation and standard error of EC₅₀ values and percent replicative capacity were all calculated using the function tool in Excel.

EC₅₀ values and the replication capacity of the patient-derived expressions vectors were calculated as statistically significant from the p8.9NSX wild-type and or the patient-derived *IN* wild-type control using the student's t test tool available on www.graphpad.com/quickcalcs/ttest1.cfm.

CHAPTER 3

Investigation of the development and genetic linkage of drug resistance mutations in full-length *pol* gene of patients on RAL-containing salvage therapy using single genome sequencing

3.1 Introduction

Patients undergoing RAL salvage therapy are also likely to have experienced or to be taking additional ARV drugs which target other genes within HIV-1 *pol* namely, PIs and RTIs. The development and evolution of RAL resistance mutations in the *IN* gene in patients failing RAL treatment has been widely investigated (Malet et al., 2008; Sichtig et al., 2009; Charpentier et al., 2008; Canducci et al., 2010; Malet et al., 2009; Reigadas et al., 2010; Ferns et al., 2009; Fransen et al., 2009a; Baldanti et al., 2010). However, few data exist on the co-evolution and linkage of PI, RTI and RAL resistance mutations within the HIV-1 *pol* gene.

To address this gap in knowledge, we developed a novel SGS assay which encompasses full-length HIV-1 *pol* and thus the 3 main genes targeted for HIV-1 therapy. The assay results in the amplification of a 3.5kb fragment that includes the *PR*, *RT* and *IN* genes. The use of this assay allows a more in depth analysis of sequence diversity compared to population-based genotyping as well as being able to identify the genetic linkage of mutations across full-length *pol*. The SGS assay used was a modified version of an assay developed by Sarah Palmer *et al.* that is based on limiting dilution of cDNA from the RT reaction of viral RNA (Palmer et al., 2005).

Palmer *et al.* used the assay to assess the extent of resistance mutations in treatment-experienced patients compared to standard population-based genotyping. They found PI and RTI resistance mutations in 24 patients that were missed by population-based genotyping and that the mutations that were present in less than 10% of single genomes were almost never detected in the population genotype. Additionally, mutations present in 10-35% of single

genomes were detected only 25% of the time in the population genotype. Thus, they concluded that population-based genotyping was inadequate for detecting low frequency drug resistance mutations and that SGS has a greater sensitivity and allows for the detection of genetic linkage.

SGS involves diluting the RT reaction product (cDNA) to a concentration low enough to ensure that each subsequent PCR reaction only amplifies a single cDNA template. According to Poisson's distribution, a dilution yielding a PCR product in 30% of the total number of reactions is likely to contain a single genome 80% of the time. Sequencing data from each reaction is thus generated from a single viral genome. SGS assays are sensitive and provide mutational context and genetic linkage information. However, they are labour intensive and relatively expensive to run compared to population-based genotyping (Palmer et al., 2005).

We used the assay to investigate the development and genetic linkage of RAL, PI and RTI resistance mutations in full-length *pol* in sequential samples obtained from patients undergoing RAL-containing salvage therapy. These aims are not achievable through the use of standard population-based sequencing and indeed other, more specialised technologies, for example next generation sequencing (NGS) and point mutation assays (as discussed in the General Introduction).

3.2 Results

3.2.1 Development of a single genome sequencing assay encompassing full length *pol*

Palmer *et al.* previously described an SGS assay based on limiting dilution with forward primers (1849+ and 1870+) in the p6 region of *gag* and reverse primers in the middle of *RT* that resulted in the amplification of a 1.3 kb fragment encompassing *PR* and the N terminal region of *RT*. We modified this assay by designing new reverse primers (5317- and 5252-) in the *vif* gene which would result in amplification of full-length *pol* (~3.5kb). A preliminary test of the primers in population-based PCR reactions using molecular clones and patient samples from 5 different subtypes (A, B, C, D and F) resulted in positive PCR products in 81% (13/16) of the reactions. This indicated that this primer set was suitable for the amplification of full-length HIV-1 *pol* gene from different subtypes.

The primer set was then used to generate single genomes from six patient samples infected with different HIV-1 subtypes. The first two were sequential plasma samples from a patient infected with HIV-1 subtype C. The samples were taken before the initiation of RAL therapy (pre-RAL) and after RAL treatment failure (post-RAL). Two single genomes were generated from the pre-RAL sample whereas 44 single genomes were amplified from the post-RAL sample. It is possible that only two single genomes were amplified from the pre-RAL sample because of reduced primer sensitivity or low viral load, however this could not be verified as viral load data were not available. Of the remaining four samples, two were from a patient infected with HIV-1 subtype AG, and two from a patient infected with HIV-1 subtype B. All four samples had viral loads greater than 2,000 copies/mL. However, no single genomes or population-based sequences were generated from all four samples. This indicated that the primers were probably not as robust as previously determined and therefore not suitable for SGS amplification. Subsequently, eight alternative reverse primers (Table 3.1), including KVL069 and KVL084 which were used in a genotypic assay for the amplification of *IN* from a diverse range of HIV-1 group M subtypes (Van et al., 2008), were designed and tested on a range of molecular clones and patient samples from different HIV-1 subtypes (A, AG, B, C, D, G). This showed that the majority of samples could be amplified in a population-based assay using the outer reverse primer KVL069 but alternative inner reverse primers had to be used for different samples. The different combinations of first round and nested reverse primers tested and the success rate for each combination is shown in Table 3.1.

This strategy was then employed to generate single genomes from six sequential samples from two patients enrolled in RAL-containing salvage therapy at the Mortimer Market Clinic, namely patients A and B.

Table 3.1 First round and nested reverse primer combinations tested for full-length *pol* genotyping

First round PCR primers	Nested PCR primers	No. of positive amplicons/Total no. of samples tested (%)
5317-	5254-	13/16 (81)
	KVL084	1/5 (20)
5175-	5125-	0/5 (0)
5254-	5175-	1/5 (20)
5450-	5095-	2/3 (67)
KVL069	KVL084	4/5 (80)
	5317-	2/5 (40)
	5095-	*5/6 (83)
	5222-deg	*1/1 (100)

*Indicates samples from different time points taken from patients A and B that were used for SGS and further analysis.

3.2.2 Clinical details of patients A and B

Patients A and B were attending the HIV Clinic at the Mortimer Market Centre, UCLH. They had both failed previous antiretroviral therapy (ART) containing PIs and RTIs and were undergoing salvage ART with RAL (600 mg daily) in combination with other ARVs. The viral load, CD4+ cell counts and treatment regimen of the two patients are shown in Figure 3.1. We were able to obtain sequential samples before, during, and after treatment with RAL. Samples were also received after RAL treatment was re-initiated in patient A. Both patients were infected with HIV-1 subtype B and the samples had previously undergone population-based sequencing of *IN* gene and additionally clonal sequencing for patient A (Ferns et al., 2009).

Patient A had initially started RAL salvage therapy in September 2007 and continued on RAL in combination with darunavir/ritonavir (DRV/r) and etravirine (ETR) until February 2008 when the patient experienced virological failure. The patient was then switched onto therapy containing tenofovir (TDF) and lamivudine (3TC), but the patient experienced virological failure again 2 months later (April 2008). RAL treatment was then re-started in combination with TDF/emtricitabine (FTC), DRV/r and ETR in September 2008. Six samples were obtained from patient A and these were pre-RAL therapy (sampled on 29.06.07; designated preRAL), 2, 4 and 5 months on RAL (sampled on 14.11.07, 18.01.08 and 01.02.08; designated 2RAL, 4RAL and 5RAL, respectively), 4 months after RAL was stopped (sampled on 25.06.08; designated 4post) and 0.5 months after RAL was re-started (sampled on 07.10.08; designated reRAL).

Patient B was initially started on RAL salvage therapy in May 2007 and continued taking RAL in combination with FTC, TDF and tipranavir/ritonavir (TPV/r). Virological failure was experienced in January 2008 and the patient was subsequently switched to treatment with FTC. Six samples were also obtained from patient B and these were preRAL (sampled on 30.03.07), 2RAL (sampled on 23.07.07), 8 months on RAL (sampled on 08.01.08; designated 8RAL) and 1, 3 and 13 months after RAL was stopped (sampled on 13.02.08, 25.04.08 and 12.02.09; designated 1post, 3post and 13post, respectively).

3.2.3 Single genome sequencing of patient A and B samples

Population-based sequencing demonstrated that all six sequential samples from patient A could be amplified with the primer sets indicated in Table 3.2. An average of 23 single genome sequences per sample was generated at the following time points: preRAL (n=16), 4RAL (n=26), 5RAL (n=23), 4post (n=39) and reRAL (n=13). No single genomes were obtained for the 2RAL time point, possibly due to low viral load of 140 copies/ mL.

On the other hand, population-based sequencing performed on six sequential samples from patient B generated sequences from three time points only, these being: 8RAL, 1post and 3post. The remaining three samples preRAL, 2RAL and 13post could not be amplified using the three different reverse primer sets used on samples from patient A or the long cDNA SGS protocol as described in Materials and Methods. Several reasons could explain the failure to amplify viral RNA from these samples including low viral load, (in particular for the 13post sample at 590 copies/ mL), inadequate handling and storage of the sample or reduced primer specificity. Nonetheless, an average of 20 single genomes per sample was generated as follows: 8RAL (n=25), 1post (n=13) and 3post (n=24).

Table 3.3 shows the drug resistance mutations that were found in single genomes at each time point in each patient.

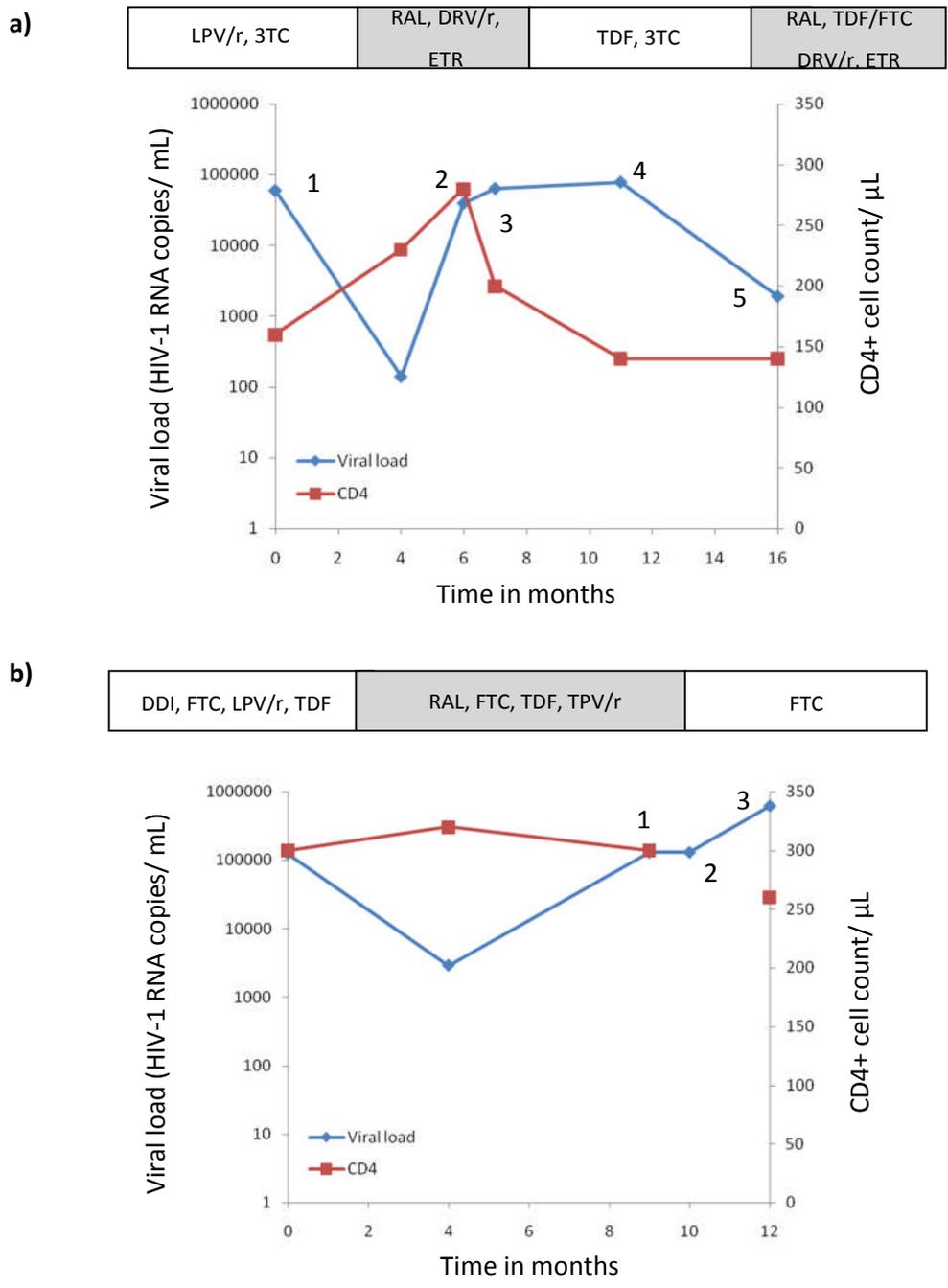


Figure 3.1 a) Patient A and b) Patient B viral load, CD4 count and antiretroviral therapy data. Each graph illustrates the clinical data of a) patient A and b) patient B and the samples obtained. Single genomes were amplified from five different time points from patient A as follows; 1) preRAL 2) 4RAL 3) 5RAL 4) 4post and 5) reRAL, whilst single genomes from patient B were amplified from three different time points as follows; 1) 8RAL 2) 1post and 3) 3post. LPV/r = ritonavir- boosted lopinavir, 3TC = lamivudine, RAL = raltegravir, DRV/r = ritonavir boosted darunavir, ETR = etravirine, TDF = tenofovir, FTC = emtricitabine, DDI = didanosine. NB. No CD4+ cell count data was available for the 3post time point of patient B.

Table 3.2 Primer combinations used in the amplification of each patient sample

Patients	Sample	First round primers	Nested primers
Patient A	Pre	KVL069	5095-
	2 months RAL ^a		KVL084
	4 months RAL		KVL084
	5 months RAL		5222-deg
	4 months post RAL		5095-
	*0.5 months RAL re-started		KVL084
Patient B	*8 months RAL	KVL069	5095-
	*1 month post RAL		
	*3 months post RAL		

^apopulation-based genotyping only

*single genome amplification using the long cDNA SGS protocol

Table 3.3 Drug resistance mutations found in each patient

Patient	Time point	PR	RT	IN
Patient A	Pre-RAL	V32I, M46I, I47V, I54L, I84V, L90M	M41L, D67N, K70R, L74V, M184V, T215Y/C, K219E, L100I, K103N, N348I	G163R
	4RAL	V32I, M46I, I47V, I54L, V82L, I84V, L90M	M41L, D67N, K70R, L74V, M184V, T215Y, K219E, L100I, K103N, N348I	E92Q, G140A, Y143R/C, Q148R, V151I, N155H, G163R/K
	5RAL	V32I, M46I, I47V, I54L, I84V, L90M	M41L, D67N, L74V, M184V, T215Y, K219E, L100I, K103N, N348I	T97A, G140A, Y143R/C, Q148R, G163R
	4post	V32I, M46I, I47V, I54L, I84V, L90M	M41L, D67N, K70R, L74V, M184V, T215Y, K219E, L100I, K103N, N348I	Y143G, G163R
	Re-RAL	V32I, M46I, I47V, I54L, I84V, L90M	M41L, D67N, K70R, L74V/I, M184V, T215Y, K219E, L100I, K103N, N348I	Y143G, G163R
Patient B	8RAL	V32I, M46I, I47V, I54M, V82T, L90M	M41L, L74V, M184V, L210W, T215Y/F, K101E, V108I, Y181C, G190A	E138D, V151I, N155H, G163R
	1post	V32I, M46I, I47V, I54M, V82T, L90M	M41L, L74V, M184V, L210W, T215Y/F, K101E, V108I, Y181C, G190A	E138D, V151I, N155H, G163R
	3post	V32I, M46I, I47V, I54M, V82T/A, L90M	M41L, K70R, L74V, M184V, L210W, T215Y/F, K101E, V108I, Y181C, G190A, N348I	E138D, V151I, N155H, G163R

3.2.4 Development and linkage of RAL resistance mutations in patient A

Analysis of the 16 single genomes generated from the preRAL time point revealed the presence of an amino acid substitution (G163E) in a single genome at a position associated with INSTI resistance (Figure 3.2a). This substitution is not classified as either a minor or major INSTI resistance substitution at position 163 according to the Stanford HIV Drug Resistance database (Rhee et al., 2003). Analysis of single genomes generated for 4RAL and 5RAL time points however, revealed that all the genomes contained major INSTI resistance mutations. Interestingly, the single genomes from the 4RAL time point contained resistance mutations at the three major positions associated with RAL resistance namely: Y143R/C (15 out of 26 genomes), Q148R (10/26) and N155H (1/26). The major INSTI resistance mutation pathways were reduced to two after 5 months on RAL therapy (5RAL) these being, Y143R/C (16 out of 23 genomes) and Q148R (7/23).

Although we observed all of the three major RAL resistance-associated mutations at positions 143, 148 and 155 in patient A, they were never genetically linked i.e. they never occurred on the same genome. However, all three major RAL resistance mutations were found to be linked to one of the following accessory mutations: E92Q, T97A, G140A, V151I and G163R/K. Analysis of the genetic linkage of primary and secondary RAL-associated resistance mutations revealed 7 different genotypes in the 49 single genomes during the first round of RAL treatment in patient A (4RAL and 5RAL) these being: Y143R+G163R (n=27), Q148R+G140A (17), Y143R+G163K (n=1), Y143C+E92Q (n=1), Y143C+G163R (n=1), Y143C+T97A (n=1) and N155H+V151I (n=1). Six of the genotypes were present at 4RAL and this decreased to 3 a month later at 5RAL. However, at both time points 2 genotypes, Y143R+G163R and Q148R+G140A, dominated the population constituting 46% and 39% of the genotypes at 4RAL and 65% and 30% at 5RAL, respectively.

Four months following the withdrawal of RAL treatment (4post), 97% (38/39) of single genomes contained no RAL resistance-associated mutations with only 1 single genome harbouring a novel substitution at position 143 (Y to G) linked to the G163R secondary mutation. This minor variant was not detected in any of the single genomes at 4RAL or 5RAL. Significantly, two weeks after RAL therapy was re-instated (ReRAL) this novel Y143G+G163R mutant dominated the viral population with all of the single genomes containing the double mutation (n= 13).

3.2.5 Development and linkage of RAL resistance mutations in patient B

As stated earlier, we were unable to amplify any viral sequences either by population-based or single-genome sequencing from the sample collected before initiation of RAL therapy in patient B. Similar to patient A, major RAL resistance mutations were observed during RAL treatment (Figure 3.2b). However, only one of the major RAL resistance mutations, N155H, in combination with different secondary mutations (E138D, V151I and/or G163R) was present in 100% (n=25) of the single genomes amplified from the 8RAL time point. One month after RAL was withdrawn (1post) all single genomes still harboured the N155H resistance mutation linked to the same secondary mutations E138D, V151I and/or G163R. However, three months after RAL was withdrawn (3post), the number of single genomes harbouring the N155H resistance mutation had declined to 63% (15/24) with the remainder containing no RAL resistance-associated mutations.

During RAL treatment (8RAL) 4 different RAL resistant genotypes were observed among the 25 single genomes generated, namely: N155H+V151I (n=8), N155H+E138D (n=7), N155H+E138D+V151I (n=8), N155H+V151I+G163R (n=2). During the off-treatment period (1post and 3post) one additional RAL resistant genotype was observed among the 37 single genomes generated with the distribution of the 5 different RAL resistant genotypes being: N155H+V151I (n=16), N155H+E138D (n=3), N155H+E138D+V151I (n=5), N155H+V151I+G163R (n=3) and N155H+E138D+V151I+G163R (n=1). The remaining 9 single genomes, all from the 3post sample, contained no RAL resistance mutations.

a)

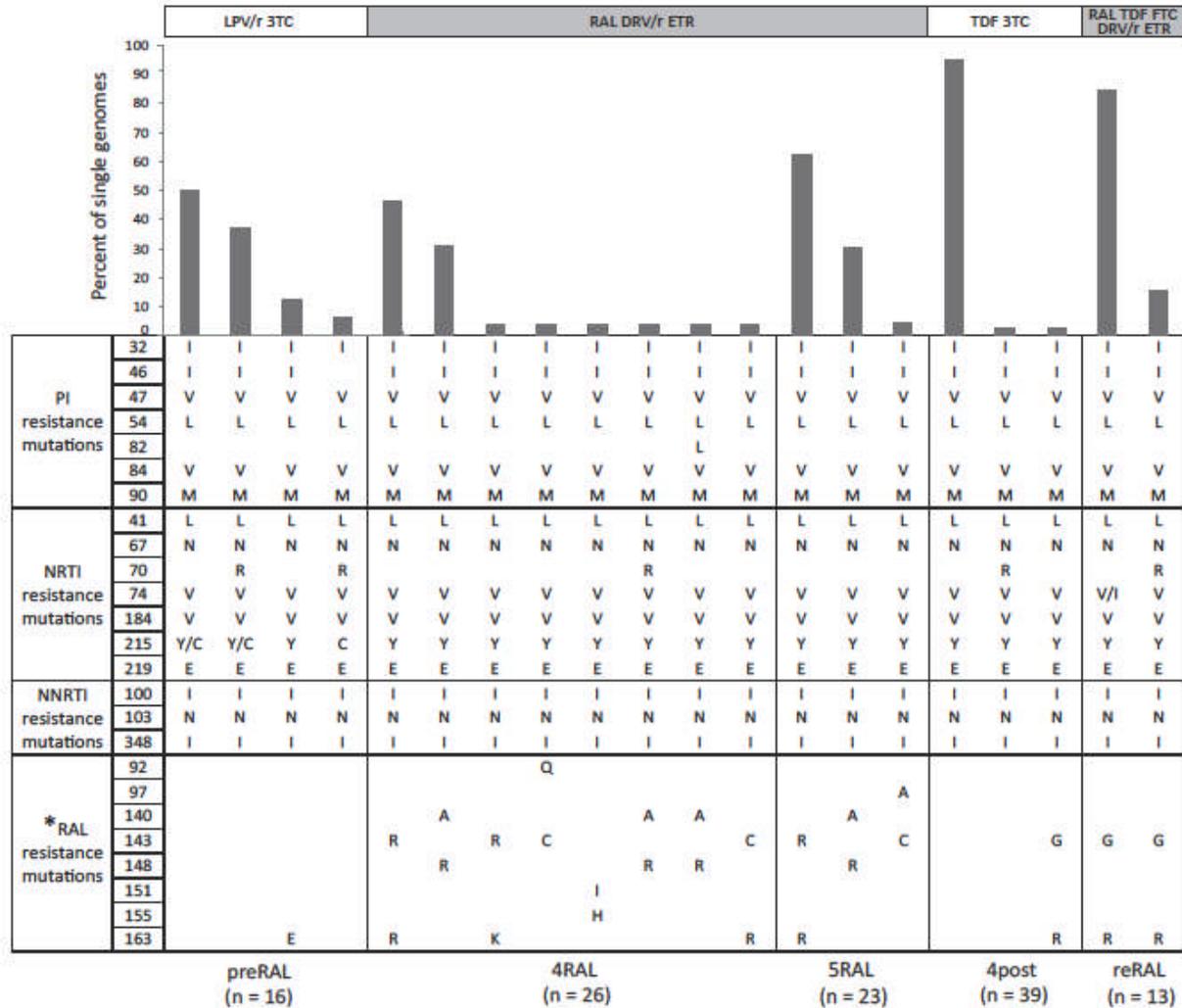


Figure 3.2 Viral variants present in a) patient A and b) patient B and the percent of single genomes they represent at each time point. The composition of major PI, RTI and RAL drug resistance mutations (as determined using the Stanford Drug Resistance Database) in a) patient A and b) patient B. Each bar in the graph above the table represents the percent of single genomes containing the linked resistance mutations at each time point shown in the table. Details of patient treatment are indicated above the bar graphs.

*Minor RAL resistance mutations are also included

3.2.6 Development of PI and RTI resistance mutations and genetic linkage to RAL resistance mutations in patient A and B

Before initiation of RAL-containing salvage therapy both patients were on a failing regimen containing RTIs and PIs; 3TC and LPV/r for patient A, and ddI, FTC, TDF and LPV/r for patient B. In comparison to the drug resistance mutations in the *IN* gene there was little variation in the composition of PI and RTI resistance mutations over the period of sampling. All of the single genomes amplified from patient A contained the following drug resistance mutations: L10F, V32I, I47V, I54L, A71T, I84V, L89V and L90M in *PR*, and M41L/I, D67N, L74V/I, M184V, T215Y/C, K219E, L100I, K103N and N348I in *RT* (Figure 3.2a and Table 3.4). Likewise, all single genomes amplified from patient B had the following drug resistance mutations L33F, I54M and L90M in *PR*, and M184V, T215Y/F, K101E, V108I, Y181C and G190A in *RT* (Figure 3.2b and Table 3.5). As this sequence data was obtained from single genomes it is expected that these mutations are genetically linked. Furthermore, in both patients, PI resistance mutations I54L/M and L90M and NRTI resistance mutations M184V and T215Y/C were found together in 100% (n=179) of the amplified single genomes enhancing the likelihood that these four resistance mutations are genetically linked.

Interestingly, these PI and RTI resistance mutations were maintained in the viral population even in the absence of their respective drugs. For example, in patient A the PI resistance mutations in the pre-RAL time point were still present when the patient was no longer on PIs at 4 months after RAL treatment was stopped. Additionally, NRTI (M41L/I, D67N, L74V/I, M184V, T215Y/C and K219E) and NNRTI (L100I, K103N and N348I) resistance mutations present before RAL-treatment was started were still present at 4 and 5 months on RAL, and 4 months post-RAL when the patient was no longer on NRTI or NNRTIs, respectively (Figure 3.2a). Similarly in patient B, PI resistance mutations L33F, I54M and L90M and NNRTI drug resistance mutations K101E, V108I, Y181C and G190A remained present in all single genomes, even when the patient was no longer on PI and NNRTI treatment at 1 and 3 months post-RAL treatment (Figure 3.2b).

In both patients A and B, various other major and secondary resistance mutations in *PR* and *RT* occurred and their numbers fluctuated throughout the sampling period (Figure 3.2 and Tables 3.4 and 3.5). In patient A there were a total of five dynamic resistance mutations in the *PR* and

RT genes. These were L33F, M46I and V82L in *PR* and K70R and V179T in *RT* and the presence of each varied between 0 and 100% of single genomes at each time point. Table 3.4 shows the number of single genomes containing major and minor drug resistance mutations at each time point from patient A.

In patient B there were nine dynamic resistance mutations in the *PR* gene (V32I, M46I, I47V and V82T/A) and 6 dynamic resistance mutations in the *RT* gene (M41L, K70R, L74V, L210W and N348I), which also ranged from 0 to 100% of single genomes at each time point. Table 3.5 shows the numbers of single genomes containing major and minor drug resistance mutations at each time point from patient B.

We also examined the genetic linkage of RAL resistance mutations with PI and RTI resistance mutations. The data revealed that 100% (n=125) of single genomes containing RAL resistance mutations occurred on the same genomes that contained PI and RTI resistance mutations. As multiple, different RAL resistance pathways developed on heavily PI and RTI resistant backbones (with many of the same resistance mutations present in each genome) it suggests that the presence of PI and or RTI resistance mutations does not constrain the type of RAL resistance pathways that can develop during RAL treatment.

3.2.7 Presence of other polymorphism in *IN*, *PR* and *RT*

Many other substitutions from the wild type B consensus (according to the Stanford Drug Resistance Database) also appear in viral sequences generated from both patients. No mutations appeared to be significantly enriched during the presence of RAL in both patients and there were no obvious trends in the appearance or disappearance of substitutions according to patient treatment. These mutations are likely to be stochastic mutations which fluctuate over time due to the error prone nature of the HIV replication process (data not shown).

Table 3.4 Variations in the numbers of major and minor PI and RTI resistance mutations occurring in patient A during the sampling period

Resistance mutation		Pre-RAL n=16	4 RAL n=26	5 RAL n=23	4 post RAL n=39	Re-RAL n=13
PR	L10F	16 (100) ^a	26 (100)	23 (100)	39 (100)	13 (100)
	L33F	4 (25)	26 (100)	23 (100)	29 (74)	13 (100)
	M46I	15 (94)	26 (100)	23 (100)	39 (100)	13 (100)
	A71T	16 (100)	36 (100)	23 (100)	39 (100)	13 (100)
	V82L	0 (0)	1 (4)	0 (0)	0 (0)	0 (0)
	L89V	16 (100)	26 (100)	23 (100)	39 (100)	13 (100)
RT	K70R	7 (44)	1 (4)	0 (0)	1 (3)	2 (15)
	V179T	0 (0)	0 (0)	0 (0)	22 (56)	12 (92)

^apercentage of single genomes containing resistance mutation at each time point. NB: Major PI and RTI resistance mutations that were present in 100% of single genomes are not included in this table.

Table 3.5 Variations in the numbers of major and minor PI and RTI resistance mutations occurring in patient B during the sampling period

Resistance mutation		8 RAL (n=25)	1 post RAL (n=13)	3 post RAL (n=24)
PR	L10I/V	25 (100) ^a	12 (92.3)	23 (95.8)
	V32I	25 (100)	13 (100)	21 (88)
	K43T	10 (40)	5 (38.5)	7 (29.2)
	M46I	1 (4)	2 (15)	0 (0)
	I47V	25 (100)	13 (100)	21 (88)
	A71V	0 (0)	0 (0)	3 (12.5)
	G73S	24 (96)	12 (92.3)	22 (91.7)
	V82T/A	25 (100)	13 (100)	21 (88)
RT	L89V	25 (100)	12 (92.3)	21 (87.5)
	M4IL	25 (100)	13 (100)	22 (92)
	K70R	0 (0)	0 (0)	2 (8)
	L74V/I	22 (88)	11 (85)	12 (50)
	A98G	25 (100)	13 (100)	21 (87.5)
	L210W	25 (100)	13 (100)	21 (88)
N348I	0 (0)	0 (0)	1 (4)	

^apercentage of single genomes containing resistance mutation at each time point. NB: Major PI and RTI resistance mutations that were present in 100% of single genomes are not included in this table.

3.3 Discussion

We established a SGS assay that amplifies the full-length *pol* gene of HIV-1 and used it to study the development and linkage of drug resistance mutations in viral *PR*, *RT* and *IN* genes of patients failing RAL-containing salvage therapy. To our knowledge, this is the first report to investigate the evolution of drug resistance using SGS of full-length *pol*. A SGS assay encompassing the full length *pol* has been used to study *ex vivo* recombination in HIV-1 vectors (Galli et al., 2010) whereas other studies investigating evolution of drug resistance have used either clonal analysis or SGS of parts of the *pol* gene such as *PR*, *RT*, *PR-RT* or *IN* (Palmer et al., 2005; Kearney et al., 2008; McKinnon et al., 2011; Mbisa et al., 2011b). The assay established here is a modification of previously described SGS assays with very low error rates (using high fidelity Taq polymerase) and negligible recombination rates due in part to reduced RNase H activity of RT enzyme used for cDNA synthesis, making it suitable for genetic linkage studies (Palmer et al., 2005; Galli et al., 2010). The assay amplifies a fragment of ~3.5 kb encompassing all three *pol* genes with forward primers in the p6 region of *gag* and reverse primers in *vif*. Nonetheless, the establishment of the assay proved to be a challenging task as no single pair of reverse primers could be used for cDNA synthesis and PCR amplifications of all samples collected from the same patient at different time points. This could be due to genetic diversity in *vif* regions that were used as reverse primer binding sites for RT and PCR steps or mutations elsewhere in the *pol* gene could have had an effect on viral RNA secondary structure which may have impacted on the cDNA synthesis step. As a result, different reverse primer pairs were used for different samples and this enabled us to determine the development and genetic linkage of PIs, RTIs and RAL resistance mutations in the context of the full-length *pol* gene. This is of primary importance as patients undergoing RAL salvage therapy will likely have been on or simultaneously undergoing treatment with PIs and RTIs.

This approach enabled us to identify minority viral variants, including variants containing rare mutations at positions in *IN* associated with RAL resistance namely, G163E and Y143G from the pre-RAL and 4post samples of patient A, respectively. The presence of a substitution at the known INSTI resistance associated site (G163E) before RAL was started may have influenced the rapid development of RAL resistance in patient A that ultimately could have resulted in viral variants exhibiting resistance to RAL as described previously (Ceccherini-Silberstein et al., 2008; Canducci et al., 2010; Myers and Pillay, 2008; Fransen et al., 2009a; Rhee et al., 2008).

Our data revealed the presence of all three major RAL resistance mutations at positions Q148, Y143 and N155 in a single patient during RAL treatment failure. We show that all three mutations can co-exist in a viral population from one patient albeit on different genomes, confirming the results of other studies (Ferns et al., 2009; Charpentier et al., 2008; Malet et al., 2009; Reigadas et al., 2010; Fransen et al., 2009a; Ferns et al., 2009; Charpentier et al., 2008; Malet et al., 2009; Reigadas et al., 2010; Fransen et al., 2009a). In this study these major RAL resistance mutations were always detected in the presence of other secondary mutations. This may be due to the time of sampling following initiation of RAL treatment which was 4 and 8 months in patients A and B, respectively. It is well established that major RAL resistance mutations can appear rapidly (sometimes within 1 month) after initiation of RAL treatment (Ferns et al., 2009; Reigadas et al., 2010; Malet et al., 2009; Canducci et al., 2010; Charpentier et al., 2008; Baldanti et al., 2010) with secondary mutations developing subsequently to compensate for a fitness loss and/or to increase drug resistance (Metifiot et al., 2010; Canducci et al., 2010; Reigadas et al., 2010; Fransen et al., 2009a; Malet et al., 2009; Delelis et al., 2009; Nakahara et al., 2009; Fun et al., 2010). Our data showing that the three major RAL resistance mutations are genetically linked to these accessory mutations is consistent with previous findings (Fransen et al., 2009a; Charpentier et al., 2008; Fun et al., 2010; Canducci et al., 2010; Malet et al., 2009) and all resistance patterns identified here have been published previously (Mbisa et al., 2011a) with the exception of the Y143G+G163R double mutation. This genetic linkage of primary and secondary mutations suggests that the accessory mutations compensate for a loss of fitness and/or can increase drug resistance of viral variants through cis-acting mechanisms. A study by Métifiot *et al.* confirmed this theory for the Q148H+G140S double mutant. They demonstrated that the catalytic properties of IN were greatly impaired by the single mutants at 140 and 148. However, the Q148H+G140S double mutant was able to fully restore the catalytic properties of IN and that this only occurred when the mutations were present on the same IN polypeptide (Metifiot et al., 2010).

INSTI resistance-associated mutations have been shown to quickly disappear following the withdrawal of RAL treatment (Ferns et al., 2009; Fun et al., 2010). Here, we show that RAL resistance mutations at positions 143 and 148 disappeared within 4 months of stopping RAL treatment in patient A, with the exception of only one minor variant harbouring the rare Y143G mutation. However, the disappearance of the N155H resistance mutation in patient B was relatively slow. Even after 3 months of no RAL treatment it was still present in the majority (63%) of sequences amplified. This could be due to the differences in sampling (3 months post RAL compared to 4 months post-RAL in patient A). Alternatively, it may be that this mutant has

less effect on viral fitness than the other resistant mutants. Indeed, work carried out by others has shown that substitutions at 143 and 148 are quick to disappear and that these mutations have a higher fitness cost than substitutions at position 155 (Ferns et al., 2009). Biochemical analysis has also shown that Y143 and Q148 mutants have more of an adverse impact on the catalytic activity of IN than N155 mutants (Malet et al., 2008; Delelis et al., 2009).

The development of the novel Y143G+G163R double mutant, its persistence in the absence of RAL treatment and re-emergence after only 2 weeks of patient A re-starting RAL suggests that this mutant may have a higher fitness and/or drug resistance effect compared to the other drug-resistant variants in patient A. This is further explored in Chapter 5.

Both patients used in this study were highly treatment experienced and previous treatment history was unavailable, therefore, extracting meaningful conclusions regarding specific PI or RTI resistance mutations from this data set proved to be difficult. As expected many PI and RTI resistance mutations were present and all 179 single genomes generated harboured between seventeen and twenty-two (patient A) and six and twenty-two (patient B) PI and RTI resistance mutations respectively. What can be gauged from this data set is that even in the absence of a specific drug class some PI and RTI resistance mutations remained in the viral population. For example, in patient A, PI resistance mutations in the pre-RAL time point are still present when the patient was no longer on PIs at 4 months after RAL treatment was stopped. Also in patient A, NRTI resistance mutations present before RAL treatment was started were still present at 4 to 5 months on RAL and 4 months post RAL when the patient was no longer on NRTI.

Similarly in patient B, PI and NNRTI drug resistance mutations remained present in all single genomes throughout the 3 time points, even when the patient was no longer on PI or NNRTI treatment at 1 and 3 months post-RAL treatment. From this one could conclude that these combinations of resistance mutations do not confer a disadvantage to the viral fitness in the absence of their respective drug class. Other studies have also found that in the absence of NNRTI treatment, NNRTI resistance mutations at positions K103 and Y181 were still maintained in the viral population (Gianotti et al., 2005; Joly et al., 2004). Conversely, another study found that M184V/I NRTI resistance mutations rapidly disappeared in the absence of NRTI treatment, probably due to the negative impact the mutation is known to have on viral fitness (Trignetti et al., 2009).

As with resistance mutations occurring within the *IN* gene it is possible to determine the genetic linkage of intra- and inter-gene mutations found in *PR*, *RT* and *IN*. In both patients, PI

resistance mutations I54L/M and L90M and NRTI resistance mutations M184V and T215Y/C were found together on all 179 single genomes. Additionally in each patient there were other resistance mutations in *PR* and *RT* that were linked. In patient A, PI resistance mutations L10F, V32I, I47V, I54L, A71T, I84V, L89V and L90M and RTI mutations M41L/I, D67N, L74V/I, M184V, T215Y/C, K219E, L100I, K103N and N348I were found together on all 117 single genomes. In patient B, PI resistance mutations L33F, I54M and L90M and RTI mutations M184V, T215Y/F, K101E, V108I, Y181C and G190A were found together on all 62 single genomes. The occurrence of linkage of these drug resistance mutations on all single genomes in the two patients indicates that this is unlikely to have occurred by recombination during the assay. Furthermore, the genetic linkage of some of these mutations was also seen in another study. Palmer *et al.* found that PI resistance mutations I84V and L90M were linked and also that RTI resistance mutations K101E, Y181C, G190A and T215Y were also linked (Palmer *et al.*, 2005). In addition, a study using a SGS assay to amplify *gag* and *PR* found that all the major PI resistance mutations present (including the M41I, I54V, V82A mutations seen here) and secondary mutations (including the A71V mutation seen here) were linked on the same viral genome (McKinnon *et al.*, 2011).

The linkage of some PI and RTI resistance mutations may also be an explanation as to why some resistance mutations remain in the viral population even in the absence of their respective drug class. These mutations may have been carried through the next round of viral replication purely because they were present on the same genome as mutations conferring resistance to drugs in the patient's current therapy but which themselves have no effect on resistance and may actually impact viral fitness. It is also feasible that other resistance mutations/ polymorphisms present elsewhere in the genome compensate for the loss of fitness conferred by disadvantageous mutations in the absence of drug. For example, it is known that there are mutations in *gag*, specifically in the *gag* cleavage sites, which are able to overcome the fitness loss conferred by some resistance mutations in *PR*. This is thought to be possibly due to the improved cleavage at these sites by the resistant PR (Clavel and Mammano, 2010). A study of PI-treated patients showed the development of mutations in the pNC/SP2 and SP2/p6 *gag* cleavage site in addition to PI resistance mutations M46L/I, I54V and V82A. They revealed that the replication of viruses with these PI resistance mutations was rescued when combined with the mutations found in the *gag* cleavage sites (Zhang *et al.*, 1997b). An *in vitro* passage study with PI structural analogues also found the emergence of *gag* cleavage site mutations along with PI resistance mutations. In the absence of the cleavage site

mutations PI resistant clones were reduced in their ability to replicate; however, they did not affect PI resistance confirming their role in viral fitness (Doyon et al., 1996).

Our data also clearly shows that linkage of drug resistance mutations occurs across the whole *pol* gene. 100% of RAL resistance mutations were found on genomes which also contained PI and RTI resistance mutations. As different RAL resistant genotypes developed on genomes with a consensus of resistance mutations in *PR* and *RT*, one could conclude that the presence of PI and RTI resistance mutations in the genome did not preclude the development of multiple RAL resistant genotypes. This was found also to be the case in another study involving the amplification of full-length *pol* from a population of viruses taken from patients failing EVG treatment using clonal analysis (Winters et al., 2012). Statistical analysis of the association of drug resistance mutations in *IN* and *RT* has shown that the M154L mutation in *IN*, which is associated with INSTI use, was significantly associated with resistance mutations, F227L and T215Y in the *RT* gene. In addition, *IN* mutations, V165I and T206S were positively associated with RTI resistance mutations F227L and L210W, respectively (Ceccherini-Silberstein et al., 2007).

To conclude, we have developed a SGS assay which amplifies the whole *pol* gene. Using this assay we have shown that the development of drug resistance in patients failing RAL-containing salvage treatment is a dynamic process. We show that there is linkage of resistance mutations across the three main genes targeted in ARV treatment. However, the effect of this on drug susceptibility and viral replication fitness remains to be determined.

CHAPTER 4

Investigation of inpatient viral evolution during the development of resistance to RAL salvage therapy

4.1 Introduction

Viral sequences can be used to infer transmission dynamics as well as reconstructing and defining epidemics at the population level. Furthermore, the high evolutionary rate of HIV allows for viral dynamics and evolution within a single patient to be studied in real time. In fact, investigating inpatient viral evolution is fundamental to the understanding of evolution at the population level as it is the definitive source of HIV diversity. Additionally, the investigation of inpatient viral evolution under antiviral selective pressure is critical to the understanding of possible mechanisms leading to the acquisition of drug resistance (Pybus and Rambaut, 2009).

Features of HIV-1 RT such as the lack of proofreading activity and its ability to facilitate recombination result in high HIV sequence diversity (Rambaut et al., 2004). The *in vitro* error rates of HIV-1 RT have been estimated to be 3 to 6 x 10⁻⁴ substitutions per site per round of replication (Roberts et al., 1988; Preston et al., 1988; Hubner et al., 1992; Boyer et al., 1992) whilst the recombination rate is thought to be around 3 recombination events per genome per round of replication (Jetzt et al., 2000; Zhuang et al., 2002). In addition, the overall evolutionary rate depends on the virus replication time and population size. For example, in a large population viral evolutionary dynamics are driven by competition between two or more beneficial variants whereas in a small population mutations that are considered to be deleterious could be given a chance to spread and may subsequently become the dominant species (Elena and Sanjuan, 2007). However, the major driving force of HIV evolution within a patient is most likely selective pressures such as ART or the host immune response (Pybus and Rambaut, 2009). Thus, the generation of multiple single genomes of full-length HIV-1 *pol*, which is approximately a third of the entire HIV-1 genome, from sequential samples obtained from two patients on RAL-containing salvage therapy gave us an opportunity to investigate

inpatient viral dynamics under selective drug pressure using a variety of phylogenetic methods.

4.2 Results

4.2.1 Analysis of inpatient viral population dynamics during development of resistance to RAL salvage therapy

Initially, we examined the inpatient viral evolution and dynamics during the development of resistance to RAL therapy in patients A and B by phylogenetic reconstruction using the single genome sequences generated in Chapter 4. The inferred maximum likelihood (ML) phylogeny obtained using the 117 single genomes from patient A showed that the sequences segregated into two distinct groups however, these were not highly supported by bootstrap (Figure 4.1). Of note, most of the nodes in the ML phylogenetic trees were not highly supported by bootstrap analysis which is expected from inpatient viral sequences. The first group, located at the base of the tree, contains single genome sequences from preRAL, 4post and reRAL time points. Firstly, this indicates that viruses emerging after withdrawal of RAL treatment (4post), which do not contain RAL resistance mutations, are closely related to sequences that were present prior to initiation of RAL therapy (preRAL). This suggests that 4post viruses are most likely recrudescence viruses from viral reservoirs established prior to initiation of RAL therapy rather than a result of reversion of RAL resistance mutations. Secondly, this shows that viruses emerging after RAL treatment was re-initiated (reRAL) are more closely related to those from the preRAL and 4post subpopulations than viruses from 4RAL and 5RAL time points. This is most likely due to the expansion of the minority Y143G+G163R double mutant that was observed at the 4post time point and became the predominant viral variant at the re-RAL time point.

The second group of sequences is located nearer the tip of the tree and contains viruses sampled from 4RAL and 5RAL time points. As anticipated, the inferred phylogeny illustrates that the founding viruses of those sampled at 4RAL and 5RAL originated from the pre-RAL pool as the branch subtending this second group emerges from the pre-RAL subpopulation. Furthermore, the group containing preRAL, 4post and reRAL single genome sequences displays a diversified branching topology in contrast to the 4RAL and 5RAL time points. This suggests a

decrease in diversity following RAL-containing therapy failure resulting in the selection of a limited number of RAL-resistant lineages. Overall, the phylogenetic reconstruction shows the fixation of beneficial mutations in the viral population and replacement of the less well adapted viruses.

On the other hand, the ML-inferred phylogeny obtained using the 62 single genome sequences from patient B (Figure 4.2) shows no definitive clustering or grouping of single genome sequences that is dependent on time of sampling. This could be due to the fact that the 3 time points sampled in patient B were within a 4 month period compared to 16 months for the 5 time points in patient A. Interestingly, some of the single genome sequences amplified from the 3post time point are located at the base of the tree despite 3post being the last time point of sampling. This is probably due to the fact that these single genome sequences, which contain no RAL resistance mutations, are most likely recrudescence viruses that were present prior to initiation of RAL treatment.

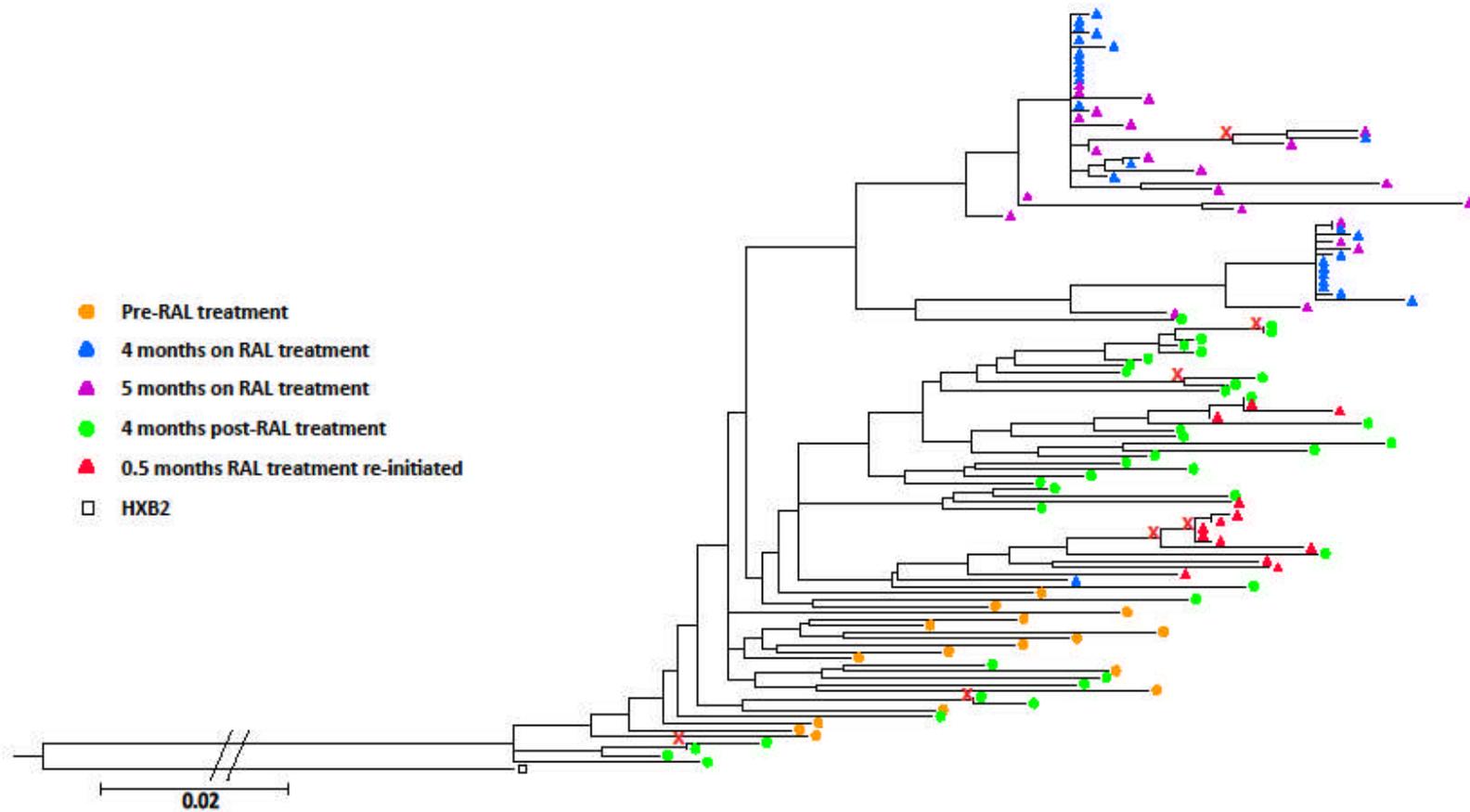


Figure 4.1 Maximum likelihood (ML) phylogeny of viral single genome data from patient A. The inferred ML phylogeny obtained using 117 single genome sequences of viruses amplified from five different time points during RAL-containing salvage therapy in patient A. The tree was constructed according to the GTR model of evolution and rooted against the HXB2 consensus sequence. Bootstrap values $\geq 90\%$ are indicated with X

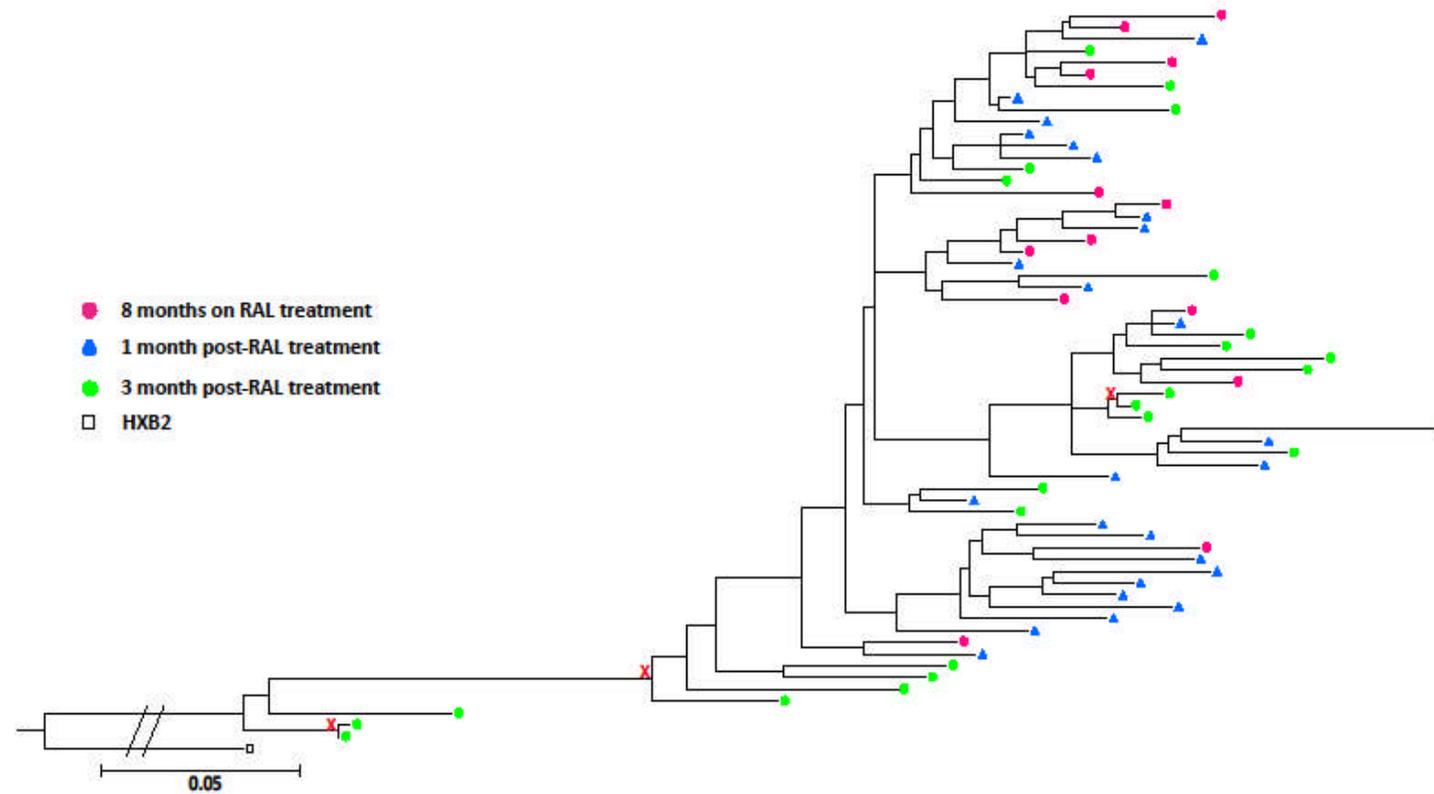


Figure 4.2 Maximum likelihood (ML) phylogeny of viral single genome data from patient B. The inferred ML phylogeny obtained using 62 single genome sequences of viruses amplified from three different time points during RAL-containing salvage therapy in patient B. The tree was constructed according to the GTR model of evolution and rooted against the HXB2 consensus sequence. Bootstrap values $\geq 90\%$ are indicated with X

4.2.2 Analysis of inpatient viral genetic diversity during RAL salvage therapy

To further explore the genetic variation of viral sequences from patients A and B we calculated the pairwise genetic distances of full-length *pol* from single genome sequences at each time point (Figure 4.3). The analysis of the data from patient A showed that the mean pairwise genetic distances were high for the sequences sampled from off RAL treatment time points (preRAL and 4post) at 0.0121 ± 0.0004 and 0.0109 ± 0.0001 nucleotide substitutions per site, respectively. In contrast, the mean pairwise genetic distances for single genome sequences sampled from time points during RAL treatment (4RAL, 5RAL and reRAL) were significantly lower at 0.0074 ± 0.0003 , 0.0073 ± 0.0003 and 0.0079 ± 0.0004 nucleotide substitutions per site, respectively ($p < 0.0001$; student's t test). This is in keeping with a bottleneck effect during RAL-containing treatment resulting in the selection of a limited number of drug resistant lineages.

On the other hand, analysis of patient B data showed that the mean pairwise genetic distance during RAL treatment (8RAL) was 0.0097 ± 0.0001 nucleotide substitutions per site which was similar to that of single genome sequences from the first off RAL treatment time point (1post) at 0.0096 ± 0.0003 ($p = 0.79$). This is consistent with a shorter interval in sampling time of 1 month between the two periods. However, this increased significantly to 0.013 ± 0.0003 ($p < 0.0001$) nucleotide substitutions per site by the second off RAL-treatment time point (3post) two months later. The emergence of recrudescing viruses at 3post from the period prior to initiation of RAL therapy probably accounts for the increase in mean pairwise genetic distance. Interestingly, the mean pairwise genetic distance values are similar during and off RAL-treatment in both patients.

Next, we determined the contribution of each individual gene segment in *pol* to the fluctuations in mean pairwise genetic distances observed using the full-length *pol* gene. We partitioned the single genome sequences into three fragments encompassing *PR*, *RT* and *IN*, and calculated the pairwise genetic distances for each individual gene segments at the different time points (Figure 4.4). The results from patient A revealed that the mean pairwise genetic distances for the *RT* gene closely resembled those observed for the full-length *pol* gene both in trend and absolute values (Figure 4.4a).

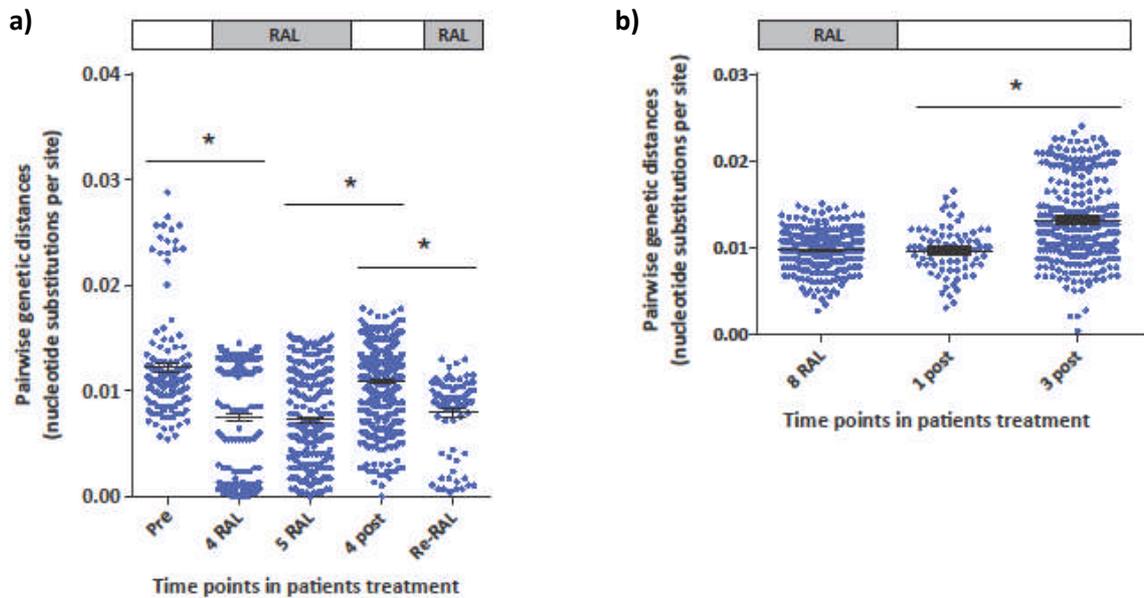


Figure 4.3 Pairwise genetic distances of the whole *pol* gene. Pairwise genetic distances for the whole *pol* gene were calculated for all single viral genomes amplified from a) patient A and b) patient B from each time point. Standard errors around the mean are shown for each time point. The bars above the graph indicated patient treatment; white boxes are indicative of no RAL-containing therapy; grey boxes indicate RAL-containing therapy. Pre = preRAL therapy, 4, 5, 8 RAL = 4, 5 and 8 months on RAL-containing therapy, 1, 3 and 4 post = 1, 3 and 4 months off RAL-containing therapy and re-RAL = 0.5 months after RAL-containing therapy was re-instated. All mean pairwise distances at each time point in patient A and patient B were found to be significantly different ($p \leq 0.05$) from those calculated at other time points (calculated using the one-way ANOVA test, not shown on graphs). * indicates significance ($p \leq 0.05$) between individual time points as tested by the student's t test.

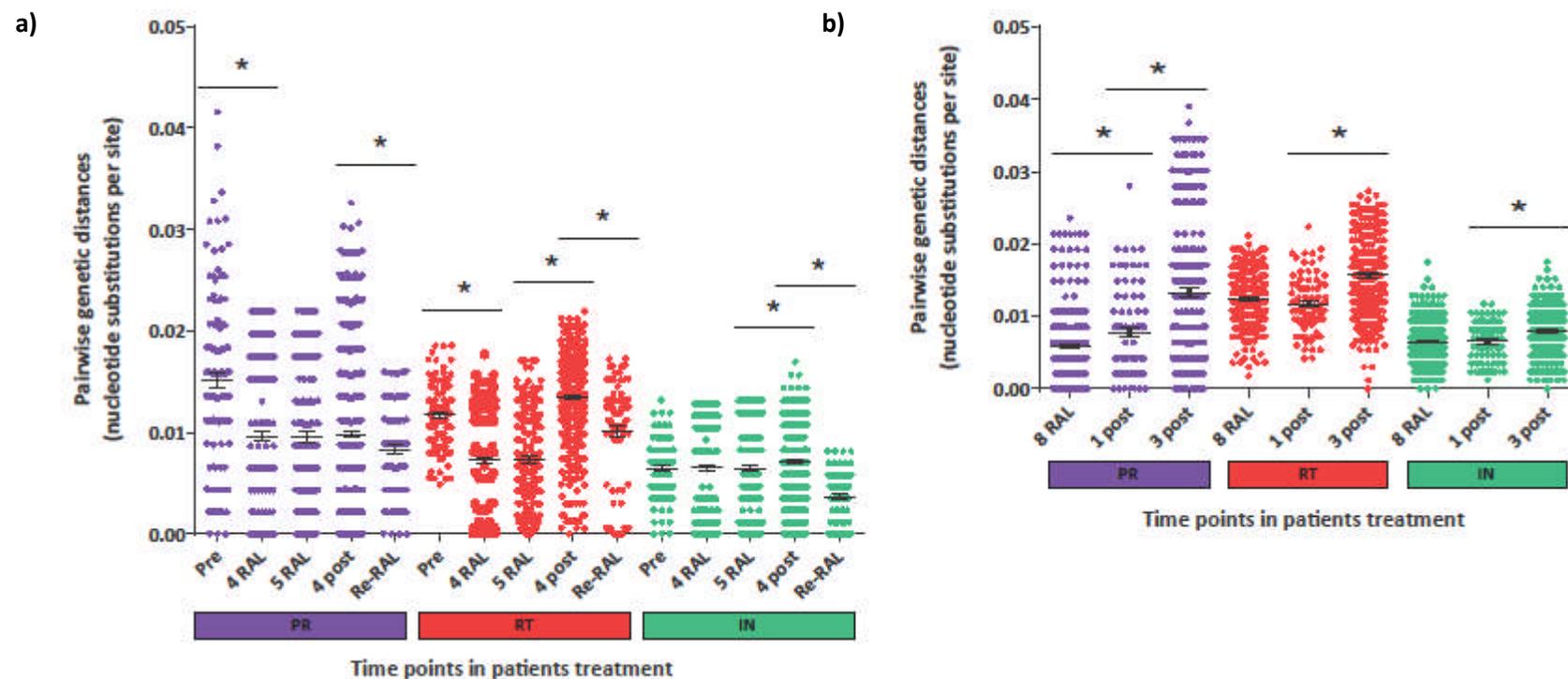


Figure 4.4 Pairwise genetic distances of individual *pol* genes. Pairwise genetic distances for each individual *pol* gene were calculated for all single viral genomes amplified from a) patient A and b) patient B. Standard errors around the mean are shown for each time point. PR in purple; RT in red and IN in green. Pre = preRAL therapy, 4, 5, 8 RAL = 4, 5 and 8 months on RAL-containing therapy, 1, 3 and 4 post = 1, 3 and 4 months off RAL-containing therapy and re-RAL = 0.5 months after RAL-containing therapy was re-instated. All mean pairwise distances at each time point in patient A and patient B were found to be significantly different ($p \leq 0.05$) from those calculated at other time points (calculated using the one-way ANOVA test, not shown on graphs). * indicates significance ($p \leq 0.05$) between individual time points as tested by the student's t test.

Thus, the mean pairwise genetic distances for the *RT* gene off RAL treatment were 0.0117 ± 0.0003 and 0.0134 ± 0.0001 nucleotide substitutions per site for preRAL and 4post, respectively. However, these values were significantly lower during RAL treatment at 0.0072 ± 0.0003 , 0.0073 ± 0.0003 and 0.0100 ± 0.0006 for 4RAL and 5RAL and reRAL, respectively ($p < 0.0001$).

Further investigation however indicated that for the first three time points the mean pairwise genetic distance for *PR* was significantly higher than for *RT* ($p < 0.0001$) ranging from 0.0096 ± 0.0005 to 0.0152 ± 0.0008 nucleotide substitutions per site for *PR* and 0.0073 ± 0.0003 to 0.0117 ± 0.0003 nucleotide substitutions per site for *RT*. This indicates that it is the *PR* gene that is contributing the most to the pairwise genetic distances of the full-length *pol* gene at preRAL and 4- and 5RAL time points. In contrast, during the 4post time point the mean pairwise genetic distance for *RT* was significantly higher compared to that of the *PR* gene (0.0134 ± 0.0001 and 0.0098 ± 0.0003 nucleotide substitutions per site respectively; $p < 0.0001$) indicating that at this time point *RT* is contributing the most. On the other hand during the reRAL time point there was no significant difference between the mean pairwise genetic distances calculated for *PR* or *RT* (0.0083 ± 0.0005 and 0.0100 ± 0.0006 nucleotide substitutions per site respectively; $p = 0.33$).

Of note, mean pairwise genetic distances for the *IN* gene were consistently low ranging from 0.0036 ± 0.0003 to 0.0071 ± 0.0001 nucleotide substitutions per site compared to 0.0072 ± 0.0003 to 0.0134 ± 0.0001 for *RT* and 0.0083 ± 0.0005 to 0.0151 ± 0.0008 for *PR*. In addition, it was interesting to find that the *PR* and *IN* genes were more clonal i.e. contained a higher number of two or more identical sequences, at 55% and 50% for the *PR* and *IN* genes, respectively, compared to 28% for *RT*.

When this analysis was carried out on data from patient B, the results were a bit different. All three genes seemed to follow the general trend observed for the full-length *pol* gene with a significant increase in the mean pairwise genetic distances from 8RAL to 3post (Figure 4.4b). However, the *IN* gene still showed consistently low mean pairwise genetic distances compared to the other *pol* gene regions. The mean pairwise genetic distances for *PR*, *RT* and *IN* in patient B ranged from 0.0059 ± 0.0003 to 0.0132 ± 0.0006 , 0.0117 ± 0.0004 to 0.0157 ± 0.0003 and 0.0065 ± 0.0002 to 0.0080 ± 0.0002 , respectively, with the *RT* gene having a significantly higher mean pairwise genetic distance across the three time points compared to *PR* and *IN* ($p \leq 0.001$).

To ensure that the calculation of mean pairwise genetic distances was not influenced by the number of sequences amplified at each time point we examined the relationship between the mean pairwise genetic distance and number of single genome sequences generated at each time point by linear regression analysis. This showed that there was no significant correlation between the two factors using the data from both patients (Figure 4.5a and b).

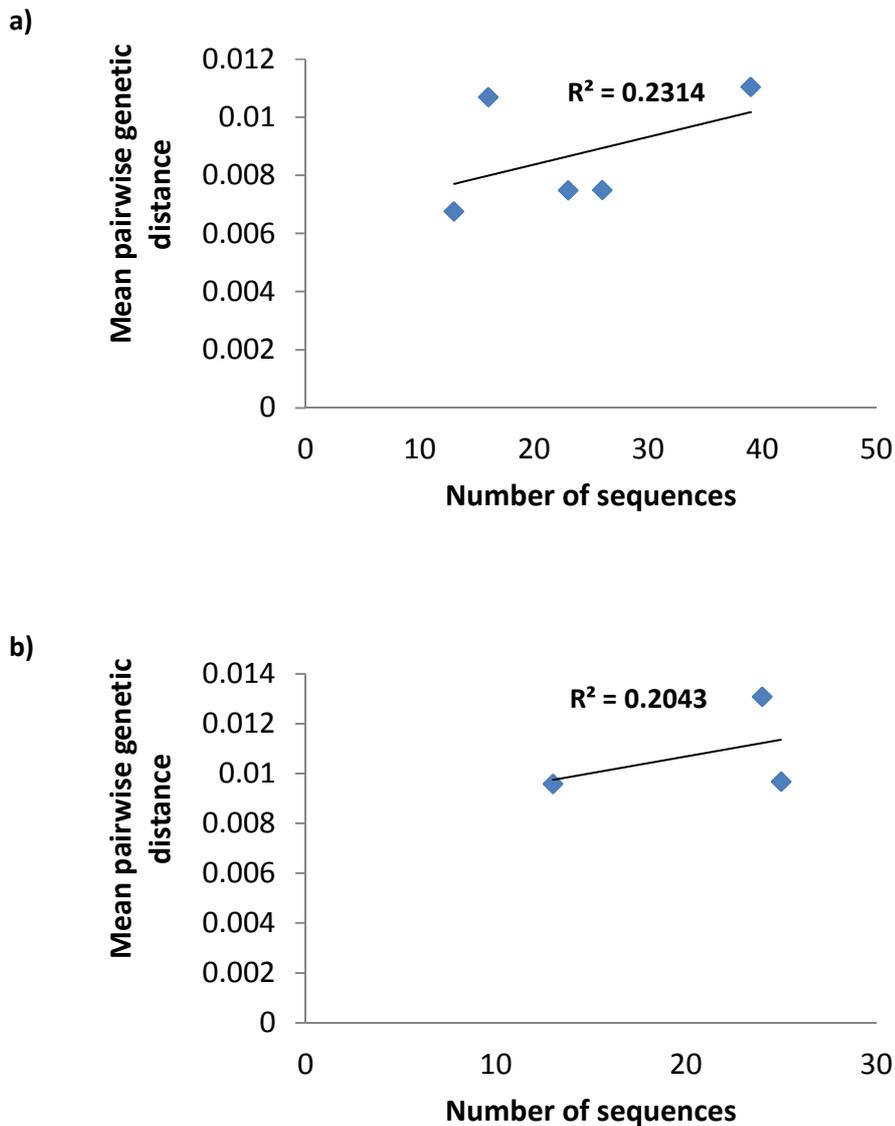


Figure 4.5 Linear regression analysis; correlation between pairwise genetic distance and number of sequences amplified at each time point. Linear regression analysis to determine the relationship between pairwise genetic distance calculated and the number of viral sequences that were amplified from a) patient A and b) patient B. R^2 is the measure of the slope of the line of best fit and indicates the correlation between mean pairwise genetic distance and number of viral sequences amplified.

We further investigated the inpatient viral genetic diversity over time in patient A using a statistical framework based on coalescent theory as implemented in the BEAST v1.6.1 software. Initially, we separated the alignments of the single genome sequences into the individual *pol* gene segments thereby creating three datasets containing *PR*, *RT* or *IN* sequences. We then performed model testing to determine the substitution, molecular clock and demographic models that best described the evolutionary history of the single genome sequencing data for each gene segment separately. This was done by calculating the approximate marginal likelihoods of two substitution models (GTR or SRD06), two demographic models (BSP or constant population size) and two molecular clock models (strict or relaxed lognormal). Model comparison was then achieved by calculating the Bayes Factor (BF) to determine which model combination to use in the analysis (Table 4.1). A BF above 20 was considered strong support for the favoured model. The BF favoured the use of the SRD06 substitution model, the BSP demographic model and relaxed molecular clock for the *RT* and *IN* genes (BF \geq 21.9). Although the SRD06 substitution model was also favoured for the *PR* gene (BF = 2511.7), the BF did not favour one model when comparing BSP and constant demographic models or strict and relaxed molecular clock models (BF \leq 18.8). This implies that either of the demographic or molecular clock models equally fit the *PR* single genome sequence data. However, as more information can be derived using the BSP demographic model this was also used for the analysis of the *PR* gene dataset together with the strict molecular clock model as it is the simpler of the two molecular clock models tested.

Table 4.1 Bayes Factors between different nucleotide substitution, demographic and molecular clock models to determine best fit models for patient A viral *PR*, *RT* and *IN* sequences

Models tested		Gene alignment		
		<i>PR</i>	<i>RT</i>	<i>IN</i>
Substitution models	GTR	0	0	0
	SRD06	2511.705	1.561E+28	2.0696E+14
Demographic models	BSP	18.755	1111022.216	251.427
	Constant	0.053	0	0.004
Clock models	Strict	0.517	0.046	0.021
	Relaxed lognormal	1.933	21.931	48.38

Highlighted values indicate BF \geq 20, indicating strong evidence for a better fit of one model over the other.

The analysis estimated the most recent common ancestor (MRCA) for the *PR* and *RT* genes to be 2003.8 and 2003.6 [2000.8-2005.9 and 2002-2004.9 ; 95% highest probability density (HPD) respectively] whereas the *IN* gene analysis dated the MRCA to 2006.2 [2005.5-2006.8; 95% HPD] (Figure 4.6). This discrepancy could be due to the fact that lineages containing INSTI resistance mutations evolved from variants present at the later stages of infection i.e. immediately before initiation of RAL therapy, compared to lineages containing *PR* and *RT* resistance mutations which would have developed earlier during infection after the failure of previous PI and RTI therapies. On the other hand, the skyline plot shows oscillatory population dynamics during the sampling period with two trough periods that roughly coincide with RAL salvage therapy time points. The trough periods for the *RT* dataset had median estimates of effective population size (N_e ; the fraction of the total population that is successful in proliferation) of 19 [6-54; 95% HPD] and 14 [2-61; 95% HPD] in November 2007 and October 2008, respectively compared to a peak median estimate of 85 [40-208; 95% HPD] in June 2007. The trend for the *IN* and *PR* datasets were similar to that of *RT*. However, the values for the median estimate of N_e were lower at 3 [1-12; 95% HPD] and 2 [0.5-6; 95% HPD] during the trough periods in November 2007 and October 2008, respectively, and a peak median estimate of 8 [4-20; 95% HPD] in April 2008 for *IN* whereas *PR* showed a median estimate of N_e of 2 [1-7; 95% HPD] and 2 [1-10; 95% HPD] during the trough periods in November 2007 and in March 2008, respectively, and a peak median estimate of 12 [2-107; 95% HPD] in October 2008. These trends are consistent with the mean pairwise genetic distance data and the finding of higher numbers of clonal sequences in *PR* and *IN* datasets described previously.

We also examined the rates of evolution for each gene from the coalescent theory analysis. The data shows that rates of evolution of *PR* and *RT* are relatively similar at 2.75×10^{-3} and 2.71×10^{-3} substitutions per site per year [$2.25-3.25 \times 10^{-3}$ and $2.16-3.29 \times 10^{-3}$; 95% HPD], respectively, whilst the rate of evolution of the *IN* gene was slightly lower at 2.9×10^{-3} substitutions per site per year [$2.37-3.59 \times 10^{-3}$; 95% HPD]. This is consistent with previous estimates using HIV-1 *pol*, and *env* genes, which were all in the range of 10^{-3} substitutions per site per year (Hue et al., 2005; Korber et al., 2000; Mbisa et al., 2012; Leitner et al., 1996; Robbins et al., 2003).

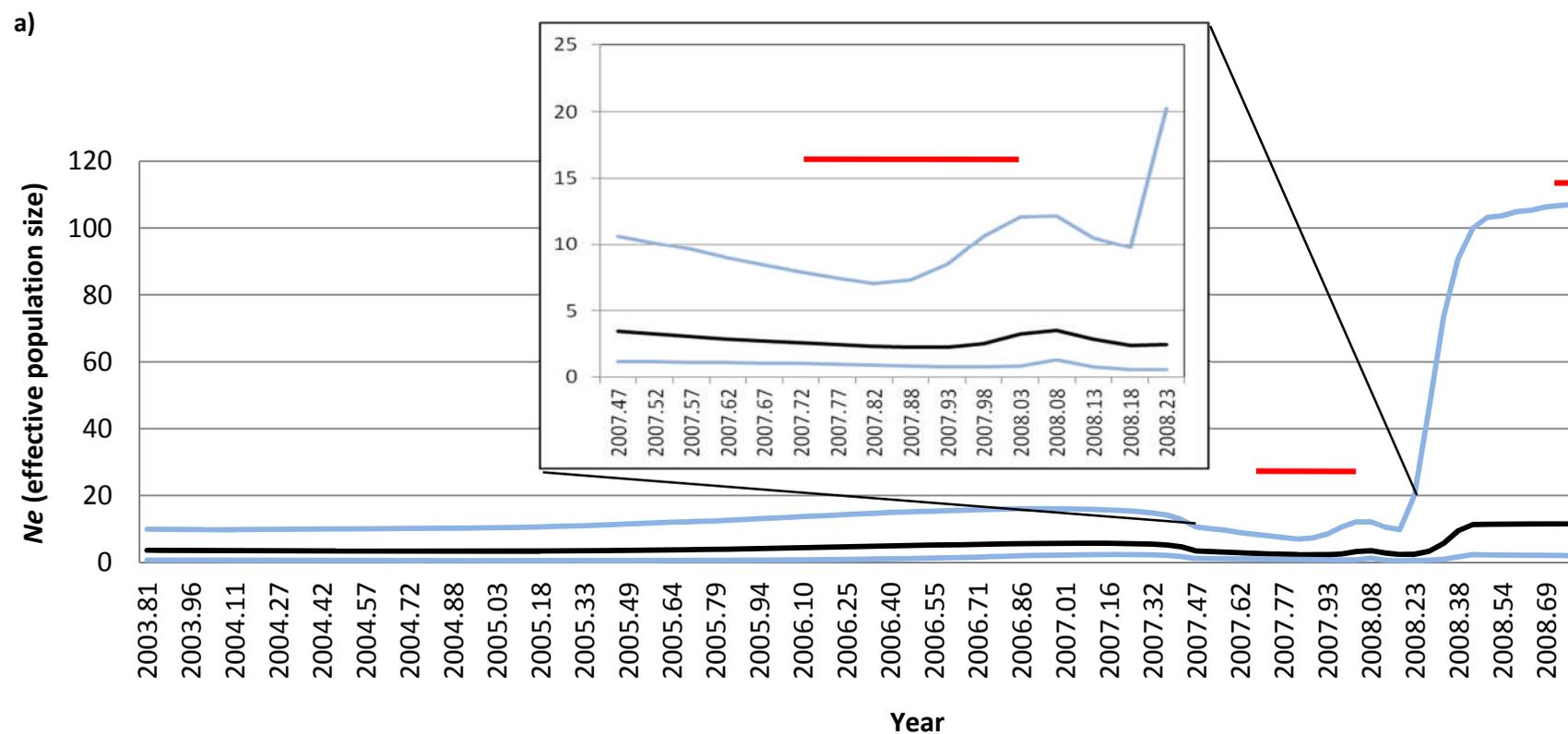


Figure 4.6 Bayesian skyline reconstruction of patient A viral *pol* genes. Bayesian skyline plots created using the time.tree file from the BEAST analysis using the maximum time as the root height's median. The plots illustrate the change in the effective population size (N_e), i.e. viral diversity, over time in each viral *pol* gene; a) *protease*, b) *reverse transcriptase* and c) *integrase* with respect to the patients treatment. The time at which the x axis intercepts the y axis is indicative of the most recent common ancestor. — Median; — 95% highest probability density (upper and lower limits); — RAL-containing therapy.

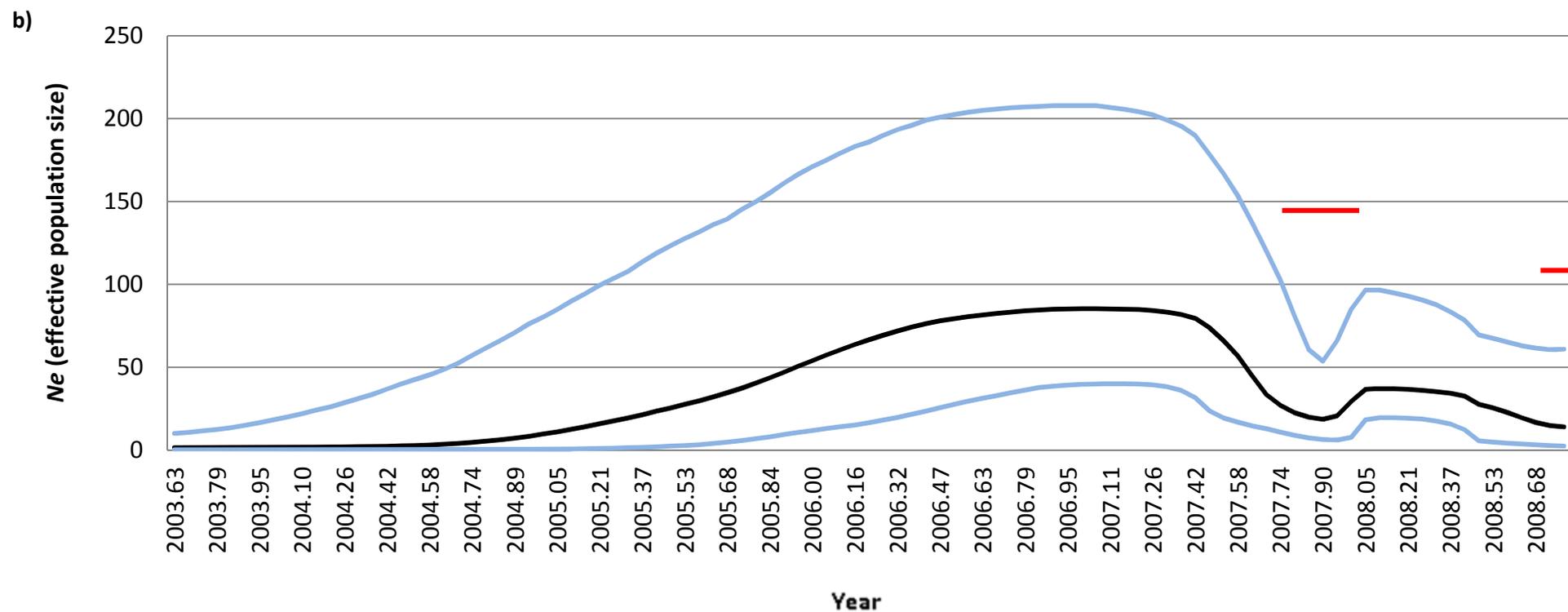


Figure 4.6 Bayesian skyline reconstruction of patient A viral *pol* genes (continued). Bayesian skyline plots created using the time.tree file from the BEAST analysis using the maximum time as the root height's median. The plots illustrates the change in the effective population size (N_e), i.e. viral diversity, over time in each viral *pol* gene; a) *protease*, b) *reverse transcriptase* and c) *integrase* with respect to the patients treatment. The time at which the x axis intercepts the y axis is indicative of the most recent common ancestor. — Median; — 95% highest probability density (upper and lower limits); — RAL-containing therapy.

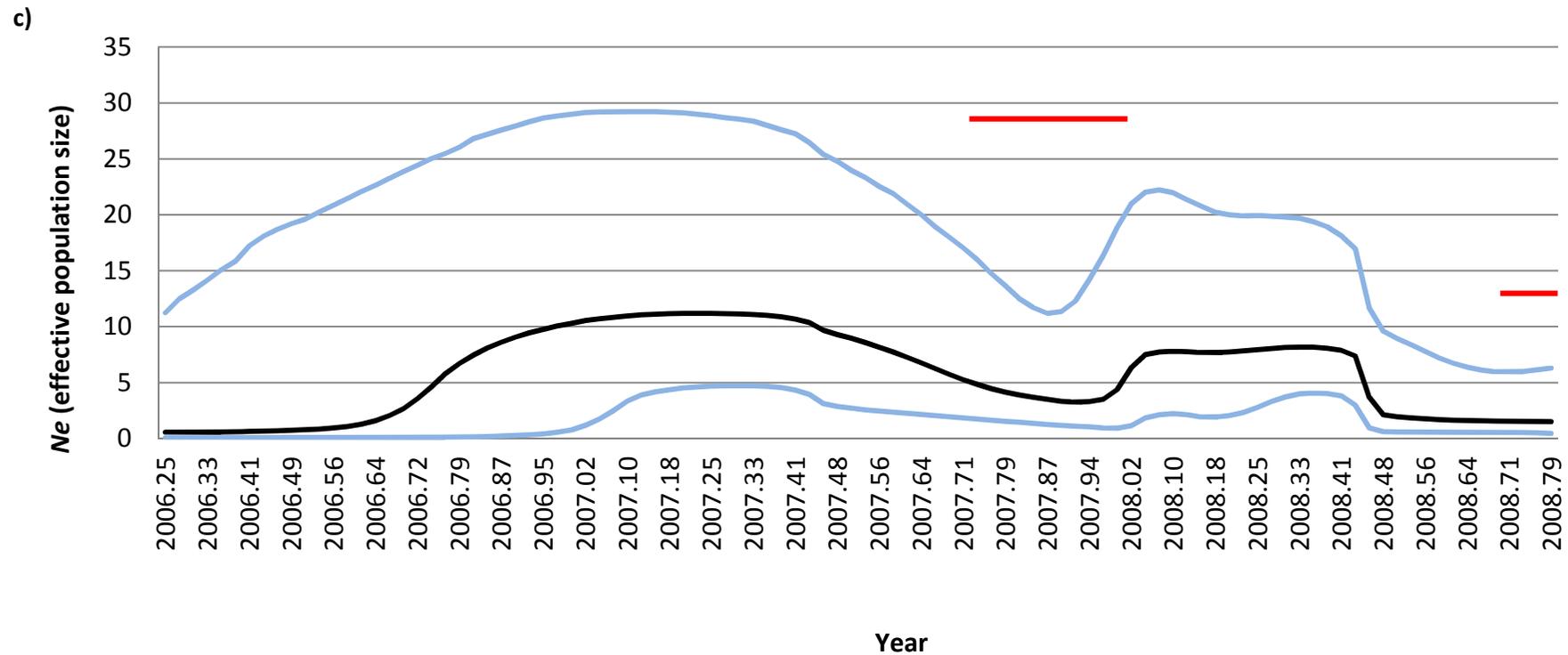


Figure 4.6 Bayesian skyline reconstruction of patient A viral *pol* genes (continued). Bayesian skyline plots created using the time.tree file from the BEAST analysis using the maximum time as the root height's median. The plots illustrate the change in the effective population size (N_e), i.e. viral diversity, over time in each viral *pol* gene; a) *protease*, b) *reverse transcriptase* and c) *integrase* with respect to the patient's treatment. The time at which the x axis intercepts the y axis is indicative of the most recent common ancestor. — Median; — 95% highest probability density (upper and lower limits); — RAL-containing therapy.

4.2.3 Analysis of drug selective pressure on the HIV-1 *pol* gene during development of resistance to RAL salvage therapy

To determine whether there were selective forces acting on specific codon positions in the HIV-1 *pol* gene during RAL-containing salvage therapy in patients A and B we examined the single genome datasets using three different algorithms implemented in DataMonkey, namely SLAC, FEL and FUBAR.

The analyses identified a total of 2 positively selected positions in patient A (2 by FEL, and 1 by FUBAR; Table 4.2) and 11 positively selected positions in patient B (1 by FEL and 11 by FUBAR; Table 4.3). Of these, 1 position in patient A and 1 position in patient B were identified by 2 or more algorithms, and neither of the positions were common to both patients.

For patient A, the positively selected position was located in *IN* (Figure 4.7) whereas for patient B it was located in *RT* (Figure 4.8). Of note, the positively selected position in the *IN* gene of patient A was codon 143 which is a major drug resistance site with a tyrosine residue in the wild-type virus. This was mutated to either arginine or cysteine earlier in the development of RAL resistance in patient A and then to glycine during the second round of RAL-containing treatment. On the other hand, the positively selected position in the *RT* gene of patient B was not at a known drug resistance site.

In contrast, the analyses identified a total of 60 negatively selected positions in patient A (25 by SLAC, 38 by FEL and 41 by FUBAR; Table 4.4) and 41 negatively selected positions in patient B (18 by SLAC, 33 by FEL and 30 by FUBAR; Table 4.5). Of these, 48 positions in patient A and 30 positions in patient B were identified by 2 or more algorithms with 1 position (codon 227 in *RT*) common to both patients. In patient A, 1 position was located in *PR*, 38 positions were found in *RT* and 9 of the positions were located in *IN*. For patient B, 1 position was located in *PR*, 24 located in *RT* and 5 located in *IN*. Of interest, 3 of negatively selected positions in *RT* of patient A were at known RTI resistance sites, namely 106, 190 and 215. However, no known RTI resistance sites were found to be negatively selected in patient B.

Table 4.2 Positively selected codons by algorithm in viral *PR*, *RT* and *IN* sequences from patient A

Codon position		Normalised dN-dS		β - α^a
		SLAC ^b	FEL ^b	FUBAR ^c
<i>PR</i>		No positively selected sites		
<i>RT</i>	195	NS	10.95	NS
<i>IN</i>	143 ^R	NS	41.79	2.39

NS indicates a position not selected by that particular algorithm; ^aposterior distribution of non-synonymous (β) and synonymous rates (α); ^bdN-dS values with a p value <0.05; ^c β - α values with a posterior probability >0.95. Shaded rows indicate positions selected by 2 or more algorithms; ^A indicates known accessory mutation positions; ^R indicates known resistance mutation positions.

Table 4.3 Positively selected codons by algorithm in viral *PR*, *RT* and *IN* sequences from patient B

Codon position		Normalised dN-dS		β - α^a
		SLAC ^b	FEL ^b	FUBAR ^c
<i>PR</i>	43 ^A	NS	NS	2.04
<i>RT</i>	74 ^R	NS	NS	3.14
	162	NS	NS	1.91
	203	NS	48.35	2.80
	215 ^R	NS	NS	1.60
	312	NS	NS	2.55
	359	NS	NS	2.01
	388	NS	NS	2.25
	466	NS	NS	2.91
	554	NS	NS	2.14
<i>IN</i>	138 ^A	NS	NS	3.62

NS indicates a position not selected by that particular algorithm; ^aposterior distribution of non-synonymous (β) and synonymous rates (α); ^bdN-dS values with a p value <0.05; ^c β - α values with a posterior probability >0.95. Shaded rows indicate positions selected by 2 or more algorithms; ^A indicates known accessory mutation positions; ^R indicates known resistance mutation positions.

	113			143			172
PreRAL1	IHTDNGTNFT	SNVKAACWW	AGIKQEFQIP	Y	NPQSQGVVE	SMNKELKKII	GQVRDQAEHL
PreRAL2
PreRAL3
PreRAL4
PreRAL5	E.....
PreRAL6
PreRAL7
PreRAL8
PreRAL9
PreRAL10
PreRAL11	R.....
PreRAL12	A.....
PreRAL13	A.....
PreRAL14	A.....
PreRAL15
PreRAL16
4RAL1	R	R.....
4RAL2	R	R.....
4RAL3	R	R.....
4RAL4	A.....	R.....
4RAL5	R	R.....
4RAL6	R	R.....
4RAL7	C
4RAL8	R	R.....
4RAL9	I.....	H.....
4RAL10	R	K.....
4RAL11	A.....	R.....
4RAL12	R	R.....
4RAL13	A.....	R.....
4RAL14	A.....	R.....
4RAL15	R	R.....

Figure 4.7 Alignment showing the positively selected position in patient A. *Integrase* gene alignment (amino acids 113 to 172) of all single genomes amplified from patient A. Position 143 (highlighted in red) was positively selected by two or more algorithms. PreRAL = preRAL-containing therapy, 4RAL and 5RAL = 4 and 5 months on RAL-containing therapy 4Pt = 4 months off RAL-containing therapy and ReRAL = 0.5 months after RAL-containing therapy was re-instated.

4RAL16	R	R
4RAL17	R	R
4RAL18	R	R
4RAL19	C	R
4RAL20	A.	.	R
4RAL21	A.	.	R
4RAL22	A.	.	R
4RAL23	A.	.	R
4RAL24	A.	.	R
4RAL25	R	R
4RAL26	A.	.	R
5RAL1	R	R
5RAL2	R	R
5RAL3	R	R
5RAL4	A.	.	R
5RAL5	R	R
5RAL6	R	R
5RAL7	C	Q
5RAL8	R	R
5RAL9	C
5RAL10	R	R
5RAL11	R	R
5RAL12	R	R
5RAL13	R	R
5RAL14	A.	.	R
5RAL15	A.	.	R
5RAL16	A.	.	R
5RAL17	R	R
5RAL18	R	R
5RAL19	A.	.	R
5RAL20	A.	.	R
5RAL21	R	R

5RAL22	R	R
4Pt1
4Pt2
4Pt3
4Pt4
4Pt5
4Pt6
4Pt7
4Pt8	A
4Pt9
4Pt10
4Pt11
4Pt12
4Pt13
4Pt14
4Pt15
4Pt16
4Pt17
4Pt18
4Pt19
4Pt20	A
4Pt21
4Pt22
4Pt23	G	R
4Pt24	A
4Pt25
4Pt26
4Pt27
4Pt28
4Pt29
4Pt30	I
4Pt31

	173			203			232
8RAL1	KQNPDIVICQ	YVDDLYIASD	LEIGQHRTKI	KELREHLWKW	GFYTPDKKHQ	KEPPFLWMGY	
8RAL2	V E
8RAL3	E
8RAL4V	E	..F	..Y
8RAL5	E
8RAL6V	A	D
8RAL7	E	..R
8RAL8	E
8RAL9V	E
8RAL10	E
8RAL11	E
8RAL12V	E
8RAL13V	E	..F	..Y
8RAL14V	E	..F	..Y
8RAL15	E
8RAL16V	E	..F	..Y
8RAL17	E
8RAL18	E
8RAL19	E
8RAL20	E
8RAL21	V	G
8RAL22V	E	..F	..Y
8RAL23	V	E
8RAL24	E
8RAL25	E
1Pt1	E
1Pt2	E
1Pt3	E	..R
1Pt4	E
1Pt5	E
1Pt6	E

Figure 4.8 Alignment showing the positively selected position in patient B. Reverse transcriptase gene alignment (amino acids 173 to 232) of all single genomes amplified from patient B. Position 203 (highlighted in red) was positively selected by two or more algorithms. 8RAL = 8 months on RAL-containing therapy, 1Pt and 3Pt = 1 and 3 months off RAL-containing therapy

1Pt7V.....	E
1Pt8	E
1Pt9V.....	E	F	Y
1Pt10	E
1Pt11V.....	E	F	Y
1Pt12V.....	E	F
1Pt13	E
3Pt1V.....	E	F	Y
3Pt2V.....	E	F
3Pt3	V	E
3Pt4V.....	E	F	Y
3Pt5	E
3Pt6V.....	E	F	Y
3Pt7	E
3Pt8V.....	E	F	Y
3Pt9	E	L	F
3Pt10	E
3Pt12V.....	E	F	Y
3Pt12V.....	E
3Pt13V.....	E	F	Y
3Pt14V.....	E	F	Y
3Pt15V.....	E	F	Y
3Pt16	E	L	F
3Pt17V.....	E
3Pt18	E
3Pt19V.....	A	D	F	Y
3Pt20	E
3Pt21	E	F
3Pt22V.....	E	F
3Pt23V.....	E	F	Y
3Pt24	E	L	F

Table 4.4 Negatively selected codons by algorithm in viral *PR*, *RT* and *IN* sequences from patient A

Codon position		Normalised dN-dS		β - α^a
		SLAC ^b	FEL ^b	FUBAR ^c
<i>PR</i>	30 ^R	NS	-64.56	NS
	53 ^A	NS	-64.57	NS
	83 ^A	NS	-39.69	NS
	94	NS	-77.55	-3.78
<i>RT</i>	7	NS	-20.60	-1.78
	22	NS	-14.43	-1.03
	60	-8.66	-32.99	-3.73
	78	-4.60	-11.67	-1.04
	81	NS	-25.71	-2.19
	93	-4.95	-12.77	-1.13
	106 ^R	-3.71	-18.59	-1.71
	108 ^R	-3.71	-10.10	-0.74
	125	NS	-12.09	-0.87
	139	-5.58	-16.97	-2.02
	141	NS	-5.93	-0.43
	162	NS	-25.75	-2.18
	168	NS	-13.72	-1.36
	190 ^R	-3.71	-9.71	-0.77
	198	-4.25	-42.27	-4.87
	215 ^R	-12.77	-62.88	-9.85
	218	-5.44	-19.45	-1.93
	227 ^A	-4.38	-37.57	-4.27
	232	-7.23	-65.95	-9.07
	237	-12.80	-82.02	-11.71
	243	-3.07	-34.35	-3.79
	249	-3.74	-28.79	-2.77
	264	NS	-18.51	-1.97
	273	-4.09	-16.40	-1.51
	289	-4.33	-26.04	-2.68
	302	NS	-17.56	-1.27
	323	NS	-19.802	-1.48
	331	-4.36	-16.93	-1.57
	367	NS	NS	-0.97
	369	-4.09	-28.26	-2.95
	403	-2.82	-11.03	-0.88
	407	NS	NS	-1.45
	418	NS	-39.99	-3.87
	423	-23.70	-15.38	-1.27
453	NS	-8.93	-0.60	
454	-56.08	-63.34	-7.85	
496	-39.49	-43.68	-5.29	
505	NS	-12.51	-0.93	
533	NS	NS	-1.18	

	547	-28.26	-27.93	-2.63
	557	-60.53	-42.10	-4.91
<i>IN</i>	2	NS	NS	-1.55
	18	-26.44	-95.33	-3.56
	38	-15.54	-60.74	-2.32
	43	-21.15	-82.75	-5.00
	45	NS	-45.58	-1.93
	62	NS	-65.39	-3.97
	68 ^A	NS	-64.56	-3.18
	81	NS	-33.52	NS
	89	-17.54	-104.57	-4.32
	116	NS	-36.56	NS
	128	NS	-28.78	NS
	162	NS	-76.13	NS
	163 ^A	-17.85	-87.38	-6.76
	166	NS	-57.47	-3.32
267	NS	-29.43	NS	

NS indicates a position not selected by that particular algorithm; ^aposterior distribution of non-synonymous (β) and synonymous rates (α); ^bdN-dS values with a p value <0.05; ^c β - α values with a posterior probability >0.95. Shaded rows indicate positions selected by 2 or more algorithms; ^A indicates known accessory mutation positions; ^R indicates known resistance mutation positions.

Table 4.5 Negatively selected codons by algorithm in viral *PR*, *RT* and *IN* sequences from patient B

Codon position		Normalised dN-dS		β - α ^a
		SLAC ^b	FEL ^b	FUBAR ^c
<i>PR</i>	2	NS	-61.35	NS
	16	NS	NS	NS
	24 ^R	-31.28	-73.81	-2.29
	50 ^R	NS	-101.91	NS
	66	NS	-69.26	NS
	69 ^R	NS	-126.50	NS
	93 ^R	NS	-70.06	NS
<i>RT</i>	2	NS	-99.61	-3.42
	26	NS	-56.72	-2.76
	40 ^A	-15.98	-96.78	-4.82
	72	-14.46	-64.78	-2.99
	108 ^R	NS	-73.56	-2.68
	116	NS	-71.19	-2.66
	170	-24.95	-64.56	-2.76
	206	NS	-43.06	-1.67
222	-38.81	-89.44	-4.51	

	227 ^A	-43.24	-106.01	-4.82
	237	NS	-77.86	-2.94
	250	-43.24	-99.26	-4.52
	261	-24.95	-26.99	-1.08
	282	-24.95	-54.59	-2.39
	292	NS	-32.04	NS
	305	-48.62	-107.32	-5.54
	309	-31.75	-100.12	-3.98
	332	-11.57	-73.08	-3.61
	341	NS	-74.60	-2.52
	378	-11.42	-63.09	-3.11
	430	-20.30	-122.00	-6.31
	460	-64.44	-243.22	-12.03
	474	NS	-102.32	-3.94
	513	NS	-86.47	-3.23
	549	NS	-106.94	-5.03
<i>IN</i>	57	NS	-86.51	NS
	67	-58.30	-201.83	-4.80
	94	NS	-50.07	NS
	123	NS	-86.51	NS
	128	-63.12	-207.47	-4.98
	153 ^R	NS	-156.01	-3.96
	156	-21.66	-211.77	-7.50
	164	NS	NS	-3.49
	170	-21.43	-154.09	-5.68

NS indicates a position not selected by that particular algorithm; ^aposterior distribution of non-synonymous (β) and synonymous rates (α); ^bdN-dS values with a p value <0.05; ^c β - α values with a posterior probability >0.95. Shaded rows indicate positions selected by 2 or more algorithms; ^A indicates known accessory mutation positions; ^R indicates known resistance mutation positions.

In summary, all three genes from both patients resulted in more negatively selected than positively selected codon positions, that is, more nucleotide substitutions resulted in synonymous rather than non-synonymous changes resulting in an amino acid change.

4.3 Discussion

We investigated the inpatient viral dynamics during RAL-containing salvage therapy using a variety of phylogenetic methods. A ML-inferred phylogeny of the single genome data from patient A revealed three phenomena. Firstly, the grouping of variants from the 4post time point together with variants sampled before RAL treatment was initiated indicates recrudescence of viruses from the pre-RAL pool after withdrawal of RAL selective pressure rather than reversion through back mutation of INSTI resistance mutations to wild-type residues. This re-emergence of predecessor strains following treatment change has also been described elsewhere (Kitchen et al., 2004; Deeks et al., 2001; Boucher et al., 2005; Imamichi et al., 2001).

Secondly, the evolution of reRAL variants from 4post variants rather than another recrudescence event from previous time points on RAL (4RAL and 5RAL) is probably due to the expansion of the novel Y143G+G163R minority variant which was first amplified from the 4post time point. This variant dominated the viral population during the re-RAL time point (described in Chapter 3). This suggests that the variant confers an advantage during RAL therapy, either in terms of viral fitness and/ or resistance to RAL. This is consistent with previous studies that have described the expansion of minority drug resistant variants that confer a replicative advantage during antiviral treatment (Deeks et al., 2001; Boucher et al., 2005; Charpentier et al., 2004). One such study evaluated the role of minority populations during PI therapy failure. Charpentier *et al.* found that in some cases the dominance of a viral variant during PI treatment failure was the result of the expansion of a previously minority variant population. This expansion was accompanied by an increase in PI resistance and in some instances, a replicative advantage (Charpentier et al., 2004).

Lastly, the branch lengths of 4 and 5RAL sequences were shorter than those of preRAL, 4post and reRAL sequences suggests a decrease in viral diversity caused by the patient's treatment regimen. HAART is well known to reduce genetic diversity, not only causing genetic constrictions in the genes targeted by therapy but also in other areas of the HIV genome (Harada et al., 2013; Kitrinis et al., 2005; Ibanez et al., 2000).

In contrast, little could be inferred from the ML-reconstructed phylogeny of patient B viral sequences. This could be due to the lack of a sample before the initiation of RAL-containing therapy containing wild-type in *IN* viruses. In addition, single genome sequences could only be generated from 3 samples in patient B over a 4 month sampling period compared to 5 samples

over a 16 month sampling period in patient A. Together, this might not have provided enough genetic information to reconstruct the phylogeny and dynamics of the infection during RAL salvage therapy in patient B.

To further investigate the viral diversity in patient A we calculated the pairwise genetic distances of full-length *pol* and individual genes separately at the different sampling times. In addition, we used coalescent theory methods to estimate past viral population dynamics using the sequence data. We found that the mean pairwise genetic distances for the whole *pol* gene were reduced during RAL treatment indicating a reduction in viral diversity. This is in agreement with the ML-inferred tree of patient A data which showed a bottleneck effect for sequences sampled during RAL treatment (4RAL and 5RAL). This reduction could be occurring during RAL-containing therapy as RAL was the only fully active drug in the regimen during the sampling period causing a restriction in the viral load followed by the selection of a few RAL resistant lineages.

Analysis of the mean pairwise genetic distances for the individual *pol* gene segments indicated that both the *PR* and *RT* genes were mainly responsible for the diversity observed for the whole *pol* gene. This was also confirmed by the coalescent theory analysis, in particular for the *RT* gene, which showed oscillating population dynamics exhibiting trough periods that coincided with RAL treatment periods as well as a higher median estimates of effective population size peaking at 85 [40-208; 95% HPD] compared to 12 [2-107; 95% HPD] and 8 [4-20; 95% HPD] for *PR* and *IN*. The association of *RT* and *PR* diversity to the overall full-length *pol* gene diversity could be due to PI and RTI treatment which continuously changed during the 16-month sampling period; from 3TC and LPV/r during preRAL to ETR and DRV/r during 4RAL and 5RAL to TDF, 3TC and no PI during 4post and finally TDF, FTC, ETR and DRV/r at reRAL. However, this phenomenon does not appear to be a result of differences in evolutionary rates of the individual *pol* genes as all three genes had similar rates of viral evolution.

On the other hand, we observed that the mean pairwise genetic distances for the *IN* gene were consistently lower compared to *PR* and *RT*. This could be attributed to *IN* being under more selective pressure than the other *pol* genes on RAL-containing therapy. Resistance had not yet developed in the *IN* gene and so is likely to undergo selection to increase viral fitness therefore resulting in a significant bottleneck effect. Drug resistance mutations were already present in *PR* and *RT* and therefore these two genes are unlikely to be under the same amount of selective pressure which is reflected in the higher pairwise genetic distances in both genes.

Our analyses revealed the positive selection of one amino acid site in each of the patients. Intriguingly there was no overlap in the selected sites between the patients. Several factors could account for this observation including differences in the treatment regimens, the length of sampling period (4 months compared to 16 months for patient A) or lack of a pre-RAL sample for patient B. The amino acid selected in patient A was at the known site of drug resistance (position 143 in IN), however this was not the case in patient B (position 203 in RT). The positive selection of 143 in IN was expected due to the presence of the selective pressure of the treatment regimen (Napravnik et al., 2005; Shi et al., 2010). In patient A, position 143 in IN is continuously evolving (from Y to R/C to G) and mutations at this position eventually emerge as the dominant resistant genotype due to RAL selective pressure.

The presence of positively selected amino acids other than those at known drug resistance positions does suggest that there are pressures being exerted on the virus other than those conferred by treatment. For example, evolution towards immune escape mutants exhibited by reduced binding of HIV antigens to the human leukocyte antigen (HLA) class I molecules involved in antigen presentation to the immune system has been described (Liu et al., 2011; Allen et al., 2005; Liu et al., 2006). In addition, their selection could indicate a possible role in drug resistance and/or viral fitness (Chen et al., 2004; Svicher et al., 2005). For example, whilst residue 203 in RT is not known to cause resistance on its own, it sits in the palm domain of RT which plays a part in drug binding (Kohlstaedt et al., 1992). One study has also shown this position to be significantly associated with the accumulation of TAMs and primary NNRTI resistance mutations (Koning et al., 2013b).

There were 48 and 30 amino acid positions that were negatively selected in patient A and B, respectively. Only one position was the same in both patients, this being amino acid position 227 in RT. Again the reason for the discrepancy may be due to the different patient regimens or the shorter sampling time period and lack of pre-RAL sample for patient B. None of the positions identified in patient B were at sites of known drug resistance, however all were recognised as sites of negative selection in the HIV positive Selection Database where the data existed (there was little or no information on sites after amino acid position 339 in RT and no data regarding the *IN* gene). These sites may represent essential sites for viral function (Chen et al., 2004). For example, codon 43 in the *IN* gene was negatively selected in viral sequences amplified from patient A. This residue is part of the zinc binding motif in the NTD of IN involved in the dimerization of IN monomers and the binding of cellular factors required for IN function (McColl and Chen, 2010). Substitution of this residue could potentially disrupt IN function.

In contrast, three negatively selected sites from patient A in our analyses were at positions of known drug resistance. Substitutions at amino acid positions 106 and 190 in RT were selected to remain as the wild-type residues, valine and glycine respectively. Mutations at these residues are known to cause high levels of resistance to efavirenz and nevirapine (neither of which was present in the patient's regimen) and only low levels of resistance to ETR (Johnson et al., 2013). These resistance mutations therefore would have had little impact in the resistance of the virus in this case. The other site was at the NRTI resistance position 215 and was selected to remain as the resistance substitution, tyrosine, and can also be explained by the patient's regimen. The patient was on NRTIs before and after RAL therapy and as this mutation causes high levels of resistance to all NRTIs when combined with other TAMs (M41L, D67N and K219E) (Johnson et al., 2013) it would have provided the virus with a resistance advantage during the sampling period.

Our data suggests that for the patients studied here negative selection is a major player in the evolution of the *pol* gene, and synonymous substitutions account for most of the observed viral diversity. This is in keeping with other studies (Gojobori et al., 1990; Kils-Hutten et al., 2001). The evolution of the virus is likely constrained by the small genome of HIV and is probably inhibited by protein functionality. The negative selection of known sites of essential enzyme function is indicative of strong natural selection (imposed by enzymatic functional and structural constraints) as is the positive selection of known drug resistance mutations (imposed by antiviral therapy). It may be possible that due to limitations of the data set (lack of pre-HAART samples) we were unable to ascertain the true selective pressure dynamics that contributed to viral diversity. Furthermore, it is likely that recombination also played a part in viral evolution (Brown et al., 2011; Temin, 1993; Vartanian et al., 1991), but this was not extensively studied here. Our data emphasise the importance of understanding inpatient viral evolution with regards to drug selective pressure.

CHAPTER 5

Inpatient relationship between viral drug susceptibility and replication fitness in the development of resistance to RAL therapy

5.1 Introduction

The development and evolution of drug resistance mutations in HIV is a result of the interplay between viral resistance and fitness. It has repeatedly been shown that primary drug resistance mutations develop in order to confer reduced drug susceptibility; however, these mutations usually come at a fitness cost to the virus. Consequently, additional mutations develop to compensate for the fitness loss and/or to increase further the resistance to a drug (Canducci et al., 2010; Fransen et al., 2009b; Fun et al., 2010; Malet et al., 2008; Goethals et al., 2010; Baldanti et al., 2010; Reigadas et al., 2010; Cooper et al., 2008). As described in previous chapters, the drug susceptibility of a viral isolate can be determined using genotypic assays and this is an essential part of patient care before initiating therapy and following treatment failure (NIH, 2012a; Hirsch et al., 2008). However, genotypic assays only give an estimate of the susceptibility of the major viral variants present in the patient. This is predicted using an algorithm which interprets the combination of resistance mutations present in the sequenced gene or gene fragments (Rhee et al., 2003). In contrast, cell-based phenotypic drug susceptibility assays give a more accurate picture of the drug resistance of a patient's virus as it directly measures the complex interactions of the different mutations and/or heterogeneity of viruses present in the patient (Hertogs et al., 1998; Kellam and Larder, 1994; Petropoulos et al., 2000; Hertogs et al., 1998; Kellam and Larder, 1994; Petropoulos et al., 2000).

With the establishment of multi-class drug use as standard in clinical practice it is important to investigate the combined effects on drug susceptibility and viral fitness of resistance mutations in all viral genes being targeted for treatment of HIV-1 infection. Here, we utilised a single-cycle replication assay to investigate the RAL and EVG susceptibilities of recombinant viruses expressing either patient-derived *IN* only or full-length *pol* containing drug resistance

mutations in all three *pol* genes that are targeted for ART. This enabled us to determine whether drug resistance mutations in the *PR* and *RT* genes have any effect on RAL or EVG susceptibility and replication fitness. In addition, this assay was used to determine the replication fitness of the recombinant viruses in the absence of INSTIs. This has allowed us to investigate the relationship between susceptibility to INSTIs and viral fitness in the development and evolution of INSTI, PI and RTI resistance mutations.

5.2 Results

5.2.1 Development of an integrase inhibitor phenotypic drug susceptibility assay

A recently described single-cycle replication assay was adapted to investigate the susceptibility of different patient-derived *pol* gene fragments to INSTIs. The assay has previously been used to determine the PI and RTI susceptibility and replication fitness of patient-derived *gag*, *PR* or *RT* genes (Parry et al., 2009; Mbisa et al., 2011b; Gupta et al., 2010a). The p8.9NSX *gag-pol* expression vector which contains a unique *Apal* restriction site in *gag* (upstream of the *PR* start codon) and an *EcoRI* restriction site in *vif* (downstream of the *IN* stop codon) was modified by introducing a unique *Clal* restriction site at the beginning of *IN* (flanking amino acids 4/5) creating the p8.9NSXClal+ vector (Figure 5.1). The unique restriction sites were then used for the cloning of patient-derived *PR+RT* (*Apal* and *Clal*) or *IN* (*Clal* and *EcoRI*). The generation of patient-derived full-length *pol* vectors was achieved by cloning of patient-derived *IN* fragments into a vector expressing patient-derived *PR+RT* using the *Clal* and *EcoRI* restriction sites.

To produce recombinant viruses the p8.9NSXClal+ *gag-pol* expression vector containing patient-derived *pol* gene fragments was co-transfected with two other vectors into 293T cells as described in Materials and Methods; these being, the pMDG vector that expresses the VSV-G envelope glycoprotein and the HIV-based pCSFLW vector expressing the *luciferase* reporter gene. After 48 hours, VSV-G pseudotyped virions capable of a single-round of replication were harvested from the culture supernatants and used to infect fresh 293T cells in the absence or presence of serial dilutions of the INSTIs; RAL and EVG. We used previously reported EC₅₀ and CC₅₀ (50% cytotoxicity concentration) values of RAL and EVG (Shimura et al., 2008; Kobayashi et al., 2008) as a guide to determine the range of serial dilutions to perform for each drug that would result in optimal dose-response curves. Initially, the maximum concentrations (highest

concentration of drug) were set at 3000 nM and 600 nM for RAL and EVG, respectively. Three-fold serial dilutions of the drugs were performed and used to determine the percent inhibition of a wild-type B virus (p8.9NSXClal+) or recombinant viruses expressing RAL resistance mutations Y143R or Q148R.

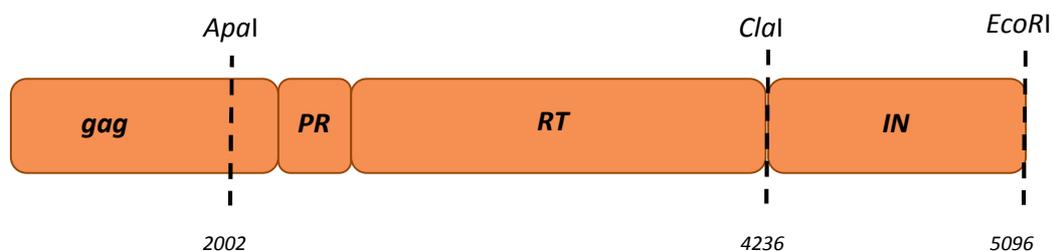


Figure 5.1 Restriction sites present in the p8.9NSXClal+ *gag-pol* region that were used for cloning. Schematic representation of the restriction sites in p8.9NSXClal+ that were used for the cloning of patient-derived *PR+RT* (*ApaI* and *Clal*), *IN* only (*Clal* and *EcoRI*) and the subcloning of *IN* into *PR+RT* containing vectors to generate full-length *pol* constructs (*Clal* and *EcoRI*). Numbers in italics indicate the nucleotide position in HXB2.

As illustrated in Figure 5.2a, this dilution series was sufficient for the determination of the RAL EC_{50} values of both the wild-type and resistant viruses. In addition, the RAL EC_{50} value of the wild-type virus (4.5 ± 0.37 nM; average of twelve independent experiments) is relatively consistent with previously published data that ranges from 7-10 nM (Kobayashi et al., 2008; Delelis et al., 2009; Malet et al., 2009). On the other hand, the EVG EC_{50} for the resistant variants and the wild-type *IN* control were at the lower end or outside the limit of the dilution series (Figure 5.2b). As a result further serial dilutions were carried out with EVG maximum concentrations of 100 nM and 1 nM. Optimal dose-response curves were obtained for the resistant viruses using the 100 nM maximum concentration dilution series whereas the 1 nM maximum concentration dilution series was required for the wild-type virus (Figure 5.2c and d respectively). Thus, the two dilution series were run in parallel when performing EVG susceptibility assays. The EVG EC_{50} value of the wild-type virus (0.10 ± 0.02 nM) was slightly lower than previously reported data that ranges from 0.7-1.4 nM (Shimura et al., 2008; Kobayashi et al., 2008). Reasons for this discrepancy could include differences in the source of EVG and the cell lines and wild-type viruses used. The assays showed good reproducibility

resulting in statistically equivalent EC₅₀ values for wild-type virus produced from two independent transfections for both RAL (3.5 ± 0.87 and 4.7 ± 0.43 nM; p = 0.28) and EVG (0.14 ± 0.049 and 0.12 ± 0.038 nM; student's t test, p = 0.76) (Table 5.1).

Table 5.1 Reproducibility of the single-cycle assay in determining the RAL and EVG EC₅₀ of the p8.9NSX wild-type B control

	RAL EC₅₀ (nM)	EVG EC₅₀ (nM)
Transfection 1	2.8	0.08
	5.2	0.10
	2.4	0.24
Average	3.5	0.14
Standard error	0.87	0.049
Transfection 2	4.8	0.19
	5.3	0.075
	3.8	0.088
Average	4.7	0.12
Standard Error	0.43	0.038

Six different RAL and EVG EC₅₀ values obtained during the development of the single-cycle susceptibility assay. Each EC₅₀ value is an average from one experiment in which each pseudovirus was tested in duplicate. The average EC₅₀ were determined for two independently transfected wild-type p8.9NSX control viruses and were found to be statistically insignificant (p = 0.28 for RAL and p = 0.76 for EVG).

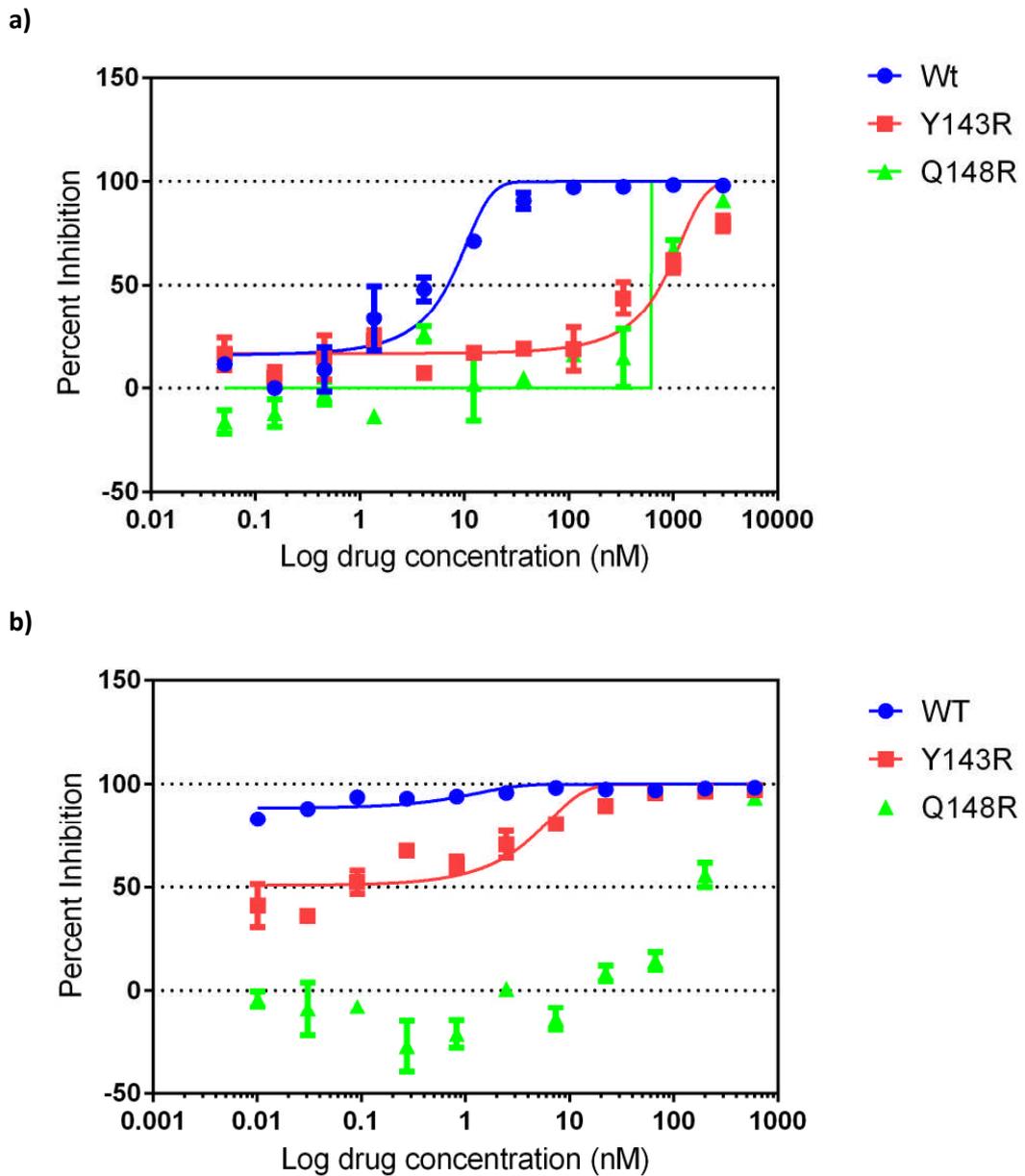


Figure 5.2 Dose-response curves obtained using different maximum concentrations of INSTI. Dose-response curves obtained for wild-type *IN* (Wt; p8.9NSX) and two RAL resistant recombinant viruses containing either Y143R or Q148R using a maximum concentration of a) 3000 nM RAL b) 600 nM EVG c) 100 nM EVG and d) 1 nM EVG during the development of the single-cycle replication INSTI susceptibility assay. Values are expressed as percent inhibition compared to the no drug control. A maximum concentration of 3000 nM for RAL was sufficient for the determination of RAL EC_{50} for both wild-type and RAL-resistant viruses whereas maximum concentrations of 100 nM and 1 nM EVG were required for the determination of EVG EC_{50} for RAL-resistant and wild-type viruses, respectively.

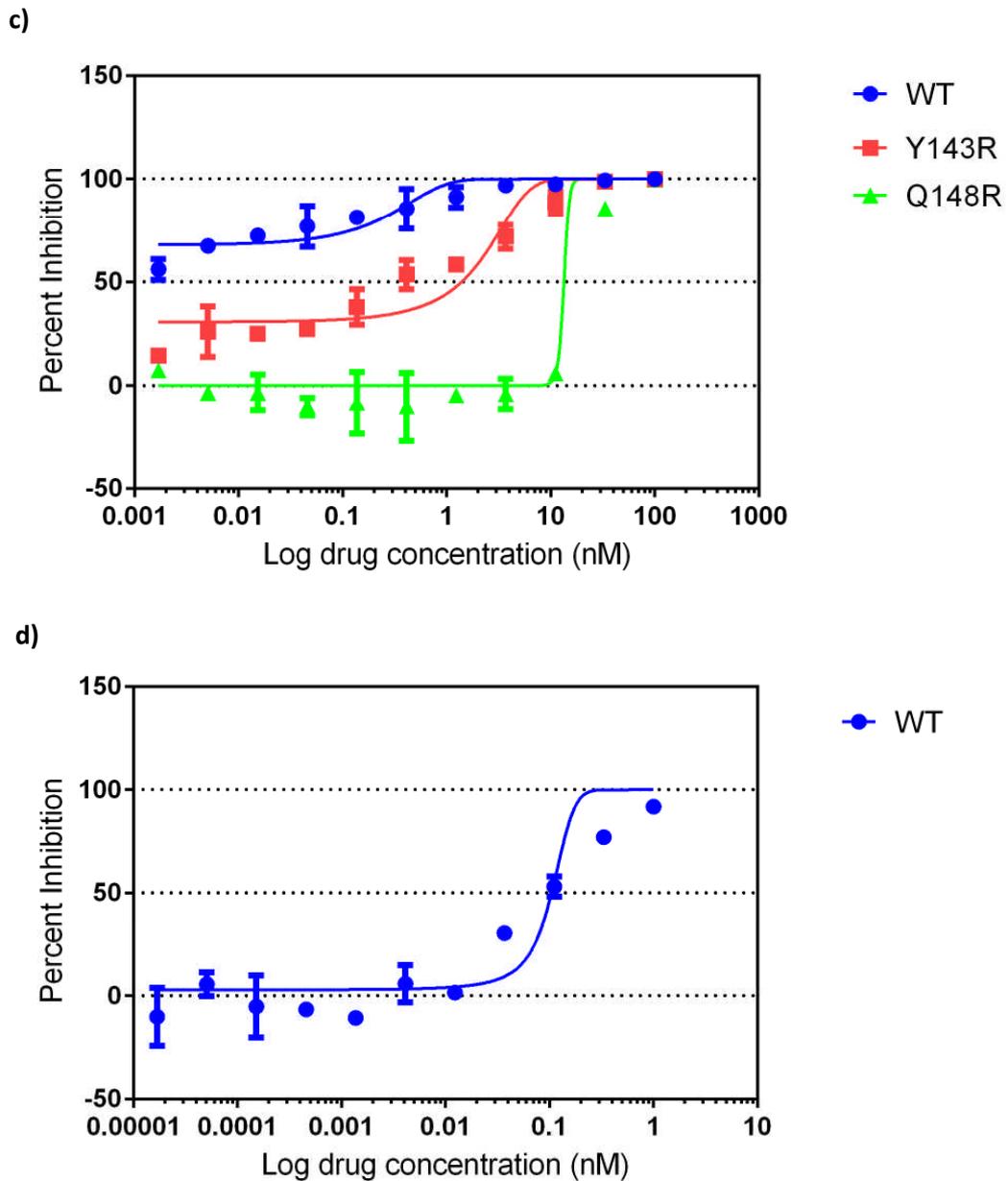


Figure 5.2 Dose-response curves obtained using different maximum concentrations of INSTI (continued). Dose-response curves obtained for wild-type *IN* (Wt; p8.9NSX) and two RAL resistant recombinant viruses containing either Y143R or Q148R using a maximum concentration of a) 3000 nM RAL b) 600 nM EVG c) 100 nM EVG and d) 1 nM EVG during the development of the single-cycle replication INSTI susceptibility assay. Values are expressed as percent inhibition compared to the no drug control. A maximum concentration of 3000 nM for RAL was sufficient for the determination of RAL EC_{50} for both wild-type and RAL-resistant viruses whereas maximum concentrations of 100 nM and 1 nM EVG were required for the determination of EVG EC_{50} for RAL-resistant and wild-type viruses, respectively.

5.2.2 Construction of patient-derived viral vectors

Nineteen viral vectors containing *PR+RT*, *IN* or full-length *pol* gene fragments from patient A were generated to investigate their effects on INSTI susceptibility and replicative fitness. Nine single genome variants were selected from different on- and off-RAL treatment time points that included all the different RAL resistance mutation combinations observed in patient A (Figure 5.3), these being: N155H+V151I, Q148R+G140A, Y143R+G163R, Y143R+G163K, Y143C+E92Q, Y143C+T97A, Y143C+G163R, Y143G+G163R as well as a wild-type *IN* from the pre-RAL time point (designated ptA_WT*IN*). All single genomes selected also contained mutations A23V, V31I, T112I, S119T, T124N, T125V and S230N in their *IN* gene. Of note, one single genome (containing the Y143C+T97A resistance mutations) contained an additional mutation in *IN* (E170Q) which was not present in any other single genomes.

The *IN* genes from the single genomes were cloned into the p8.9NSXClal+ vector using *Clal* and *EcoRI* restriction sites. However, several attempts at cloning of full-length *pol* or *PR+RT* fragments into the p8.9NSXClal+ vector using *ApaI* and *EcoRI* or *Clal* restriction sites proved unsuccessful or resulted in undesired mutations in the cloned fragments. Only a single *PR+RT* variant cloned into the p8.9NSXClal+ vector was free of undesired mutations (designated ptA_*PR+RT*), as a result we used this as a backbone to clone all of the other nine *IN* gene variants creating full-length *pol* patient-derived vectors. The *PR+RT* variant used was from the single genome expressing the N155H+V151I RAL resistance mutations and contained the resistance mutations L10F, V32I, L33F, M46I, I47V, I54L, A71T, I84V, L89V and L90M in *PR*, and M41L, D67N, L74V, M184V, T215Y, K219E, L100I, K103N and N348I in *RT*. These nineteen resistance mutations were present in all single genomes chosen for the analysis; however, two single genomes contained additional resistance mutations in *RT*; these were the genomes containing Y143C+T97A and Y143G+G163R which contained the minor NRTI resistance mutation T69A and the minor NNRTI resistance mutation V179T, respectively. In addition there were two additional mutations in both *PR* (N37D and R57K) and *RT* (V292I and E396D) that were not present in the 8 other single genomes. There were also 3 and 12 other mutations in *PR* (K14T, 15V and I72V) and *RT* (E44A, S48A, T139M, E194D, I195L, H221Y, D177N, I195L, E224D, R461K V466I and E529K) respectively that were not present in ptA_*PR+RT* but were found in one or more of the other 8 single genomes.

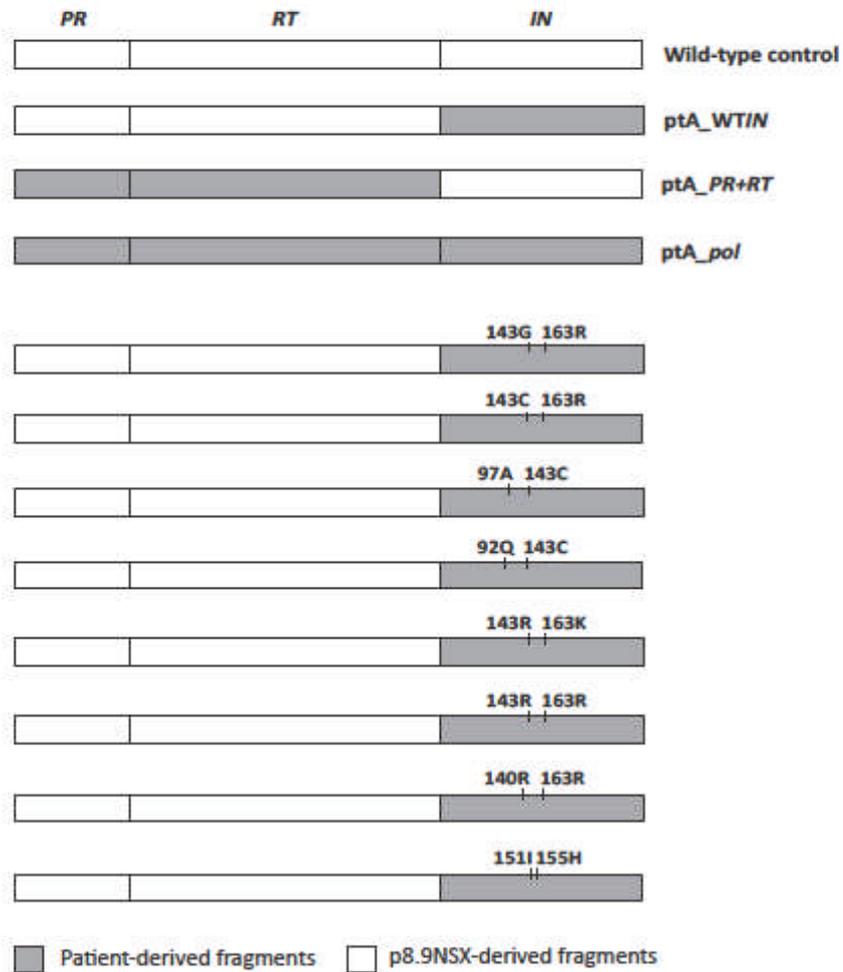


Figure 5.3 Patient-derived *IN* only and full-length *pol* control vectors studied in the single-cycle replication assay for phenotypic and replicative fitness determination. Wild-type control = p8.9NSX wild-type subtype B; ptA_WT/IN = patient-derived wild-type in *IN*; ptA_PR+RT = patient-derived *PR* and *RT*; ptA_pol = patient-derived full-length *pol*, wild-type in *IN*. Each patient-derived *IN* only variant was sub-cloned into the ptA_PR+RT vector to generate patient-derived full-length *pol* vectors. The patient-derived *PR*+*RT* fragment contained the PI resistance mutations L10F, V32I, I47V, I54L, A71T, I84V and L90M and RTI resistance mutations M41L, D67N, I74V, M184V, T215Y, K219E, L100I, K103N ad N348I.

5.2.3 Effects of patient-derived *IN* on RAL susceptibility

Initially, we determined the RAL susceptibility of the viruses expressing patient-derived *IN* genes only. As expected, the RAL EC₅₀ of the virus containing the patient-derived *IN* from the pre-RAL time-point which contained no RAL resistance mutations (ptA_WT/*IN*) was equivalent to that of the wild-type control virus (p8.9NSX) at 4.2 ± 0.11 vs 4.5 ± 0.37 nM ($p = 0.76$; Figure 5.4). In contrast, all viruses containing patient-derived *IN* from single genomes sampled during RAL treatment exhibited significant decreases in RAL susceptibility compared to that of the wild-type *IN* controls with EC₅₀ values ranging from 84.6 ± 10.5 to 900 ± 145.2 nM ($p \leq 0.0001$).

Of interest, the different amino acid substitutions at Y143 had significantly different susceptibilities to RAL with the virus expressing the Y143R+G163R mutations exhibiting the lowest RAL susceptibility (EC₅₀ of 900 ± 145.2 nM) whereas the virus expressing the Y143C+G163R mutation showed the highest RAL susceptibility (EC₅₀ of 84.6 ± 10.5 nM).

RAL susceptibility was not only influenced by the type of amino acid substitution at the primary RAL resistance sites but also by the type of accessory mutation linked to the primary mutation. For example, the Y143C primary resistance mutation was found linked to three different accessory mutations (E92Q, T97A and G163R) on separate genomes (see Chapter 3). When viruses generated using these three different genomes were analysed for RAL susceptibility, we observed that the Y143C mutation linked to E92Q was the least susceptible to RAL with an EC₅₀ of 822.9 ± 30 nM compared to Y143C+T97A and Y143C+G163R mutants at 495.7 ± 82 and 84.6 ± 10.5 nM, respectively ($p < 0.0038$). However, in the instance where different amino acid substitutions were detected at an accessory position, e.g. G163R and G163K, linked to the same primary RAL resistance mutation (Y143R) we did not observe any significant differences in RAL susceptibilities, namely 900 ± 145.2 and 714.3 ± 81.7 nM for Y143R+G163R and Y143R+G163K, respectively ($p = 0.29$).

Interestingly, the virus expressing the patient-derived *IN* containing the rare Y143G+G163R mutation had a RAL susceptibility comparable to that of the virus expressing N155H+V151I (EC₅₀ of 227.2 ± 39.6 nM vs 160.6 ± 32.7 nM, respectively; $p = 0.22$) which was significantly higher than that of the virus expressing the Y143C+G163R mutation (84.6 ± 10.5 nM; $p = 0.006$). Table 5.2 lists the RAL susceptibilities (EC₅₀ values and fold-changes) for all the viruses expressing patient-derived *IN* only in the order of lowest to highest RAL susceptibility.

Table 5.2 RAL EC₅₀ of patient-derived *IN* only viruses

<i>IN</i> genotype	RAL EC₅₀ (nM)	FC EC₅₀
Y143R+G163R	900	200
Y143C+E92Q	822.9	183
Y143R+G163K	714.3	159
Q148R+G140A	664.5	147
Y143C+T97A	495.7	110
Y143G+G163R	227.2	50
N155H+V151I	160.6	36
Y143C+G163R	84.6	19
ptA_WT/ <i>IN</i> *	4.2	1

RAL EC₅₀ and fold-changes in RAL EC₅₀ for all viruses expressing patient-derived *IN* only, listed in order from the lowest to the highest susceptibility to RAL. FC = fold-change; *ptA_WT/*IN* indicates the vector containing patient-derived *IN* only from the pre-RAL time point.

5.2.4 Effects of patient-derived full-length *pol* on RAL susceptibility

Next, we investigated whether the coevolved *PR+RT* gene segment from patient A affected the RAL susceptibility of the different patient-derived *IN* genes. As expected, the virus expressing the patient-derived full-length *pol* (designated ptA_*pol*; with resistance mutations in *PR* and *RT* but wild-type in *IN*) as well as virus expressing the patient-derived *PR+RT* only (designated ptA_*PR+RT*) had RAL susceptibilities comparable to that of the wild-type control virus (p8.9NSX) with EC₅₀ values of 5.5 ± 1.1 nM, 4.1 ± 0.48 and 4.5 ± 0.37 nM, respectively (p ≥ 0.30; Figure 5.5). Overall, the viruses expressing patient-derived full-length *pol* from time points during RAL treatment exhibited decreased RAL susceptibilities that were similar to viruses expressing respective patient-derived *IN* only except for the Y143R+G163R, Y143R+G163K and Y143C+T97A mutation combinations. For these three viruses, the patient-derived *IN* only showed significantly greater decreases in RAL susceptibilities than full-length *pol* (p ≤ 0.032). This suggests that the coevolved *PR+RT* can confer a negative effect on resistance to RAL which is dependent on the combination of resistance mutations in *IN*.

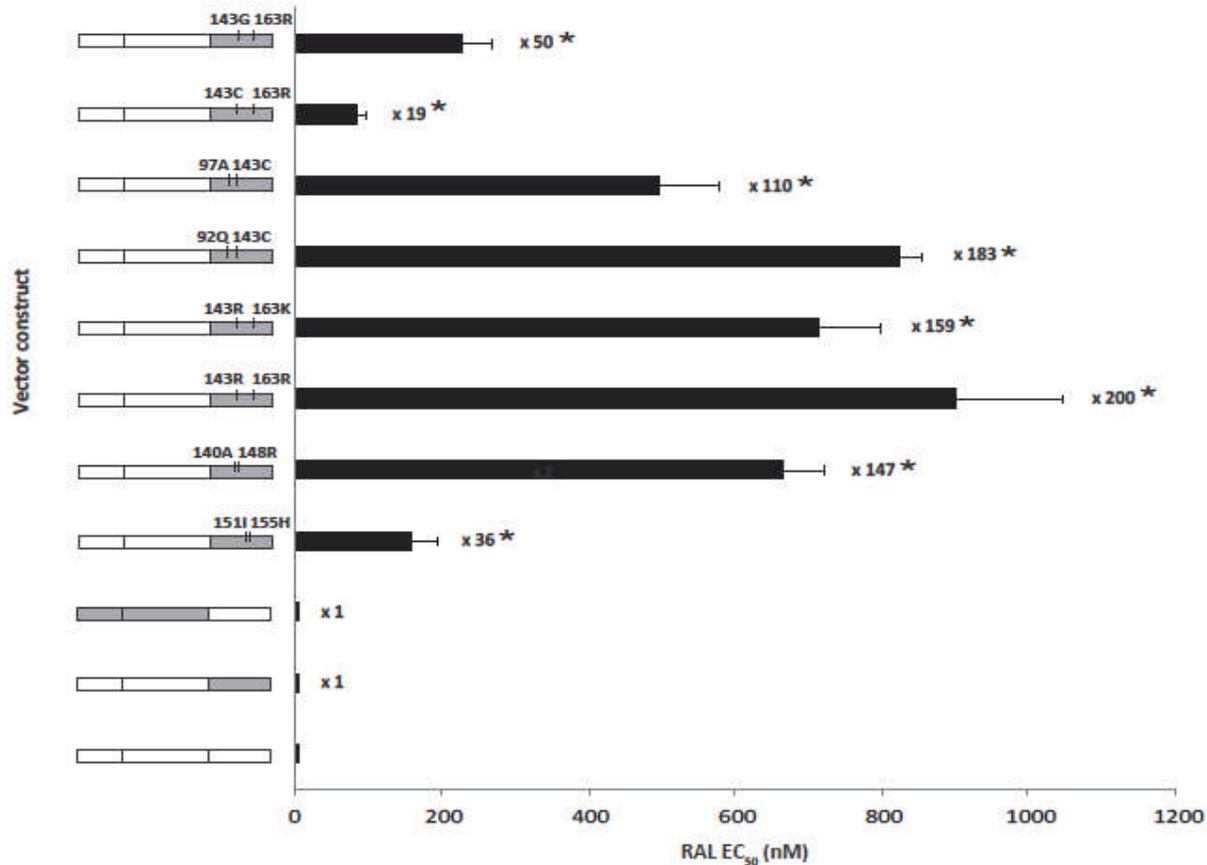


Figure 5.4 RAL susceptibilities of patient-derived *IN* only viruses. Each result is an average of 6 to 12 independent experiments. Fold changes compared to the p8.9NSX wild-type control are indicated next to each bar. Viruses that have a significantly higher ($p < 0.0001$) RAL EC₅₀ compared to the wild-type control are indicated with *

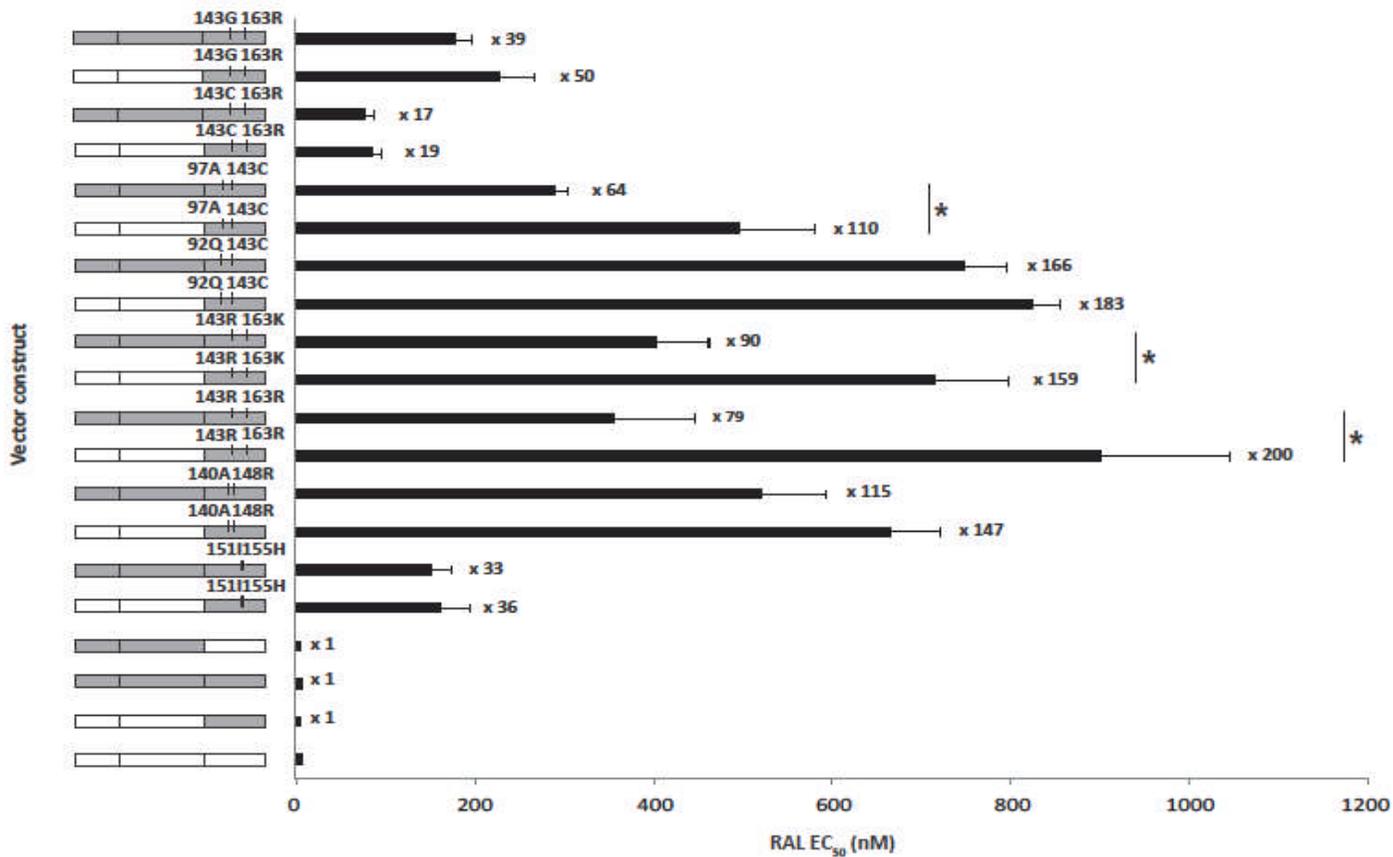


Figure 5.5 Comparison of RAL susceptibilities of patient-derived *IN* only and full-length *pol* viruses. Each result is an average of 6 to 12 independent experiments. Fold changes compared to the p8.9NSX wild-type control are indicated next to each bar. Viruses that have a significantly higher ($p \leq 0.032$) RAL EC₅₀ compared to their full-length *pol* counterpart are indicated with *

5.2.5 Effects of patient-derived *IN* on EVG susceptibility

Broad cross-resistance between RAL and EVG has previously been reported (Metifiot et al., 2011; Abram et al., 2012). Therefore, we next investigated the susceptibilities to EVG of the viruses expressing patient-derived *IN* only. The EVG EC₅₀ of each of the patient-derived *IN* only vectors are shown in Figure 5.6. Interestingly, the EVG EC₅₀ of the virus expressing the patient-derived *IN* from pre-RAL time point (ptA_WT/*IN*) was significantly higher compared to the wild-type control (p8.9NSX) virus at 0.24 ± 0.026 vs 0.1 ± 0.016 nM ($p = 0.0002$). This is consistent with other studies that have shown a 2 to 3-fold increase in EVG EC₅₀ for viruses expressing *IN* genes from INSTI-naïve patients compared to wild-type controls (Low et al., 2009; Garrido et al., 2012).

Two viruses expressing patient-derived *IN* containing the RAL resistance mutations Y143C+G163R and Y143G+G163R also exhibited EVG EC₅₀ values similar to that of the wild-type control virus at 0.06 ± 0.022 and 0.22 ± 0.088 nM, respectively ($p \geq 0.12$). The remaining six viruses expressing patient-derived *IN* only were found to have significantly higher EVG EC₅₀ values compared to the wild-type control virus ranging from 0.48 ± 0.064 to 23.31 ± 0.69 nM ($p < 0.05$). This includes the virus expressing the Y143R+G163R mutation combination, which exhibited the highest fold-change in RAL EC₅₀ (200-fold), but only resulted in a 5-fold change in EVG susceptibility. The highest fold change in EVG susceptibility (227-fold) was exhibited by the virus expressing the Q148R+G140A mutation combination which also resulted in a very high fold change to RAL susceptibility (147-fold). Table 5.3 lists the EVG susceptibilities (EC₅₀ values and fold changes) for all the viruses expressing patient-derived *IN* only in the order of lowest to highest EVG susceptibility.

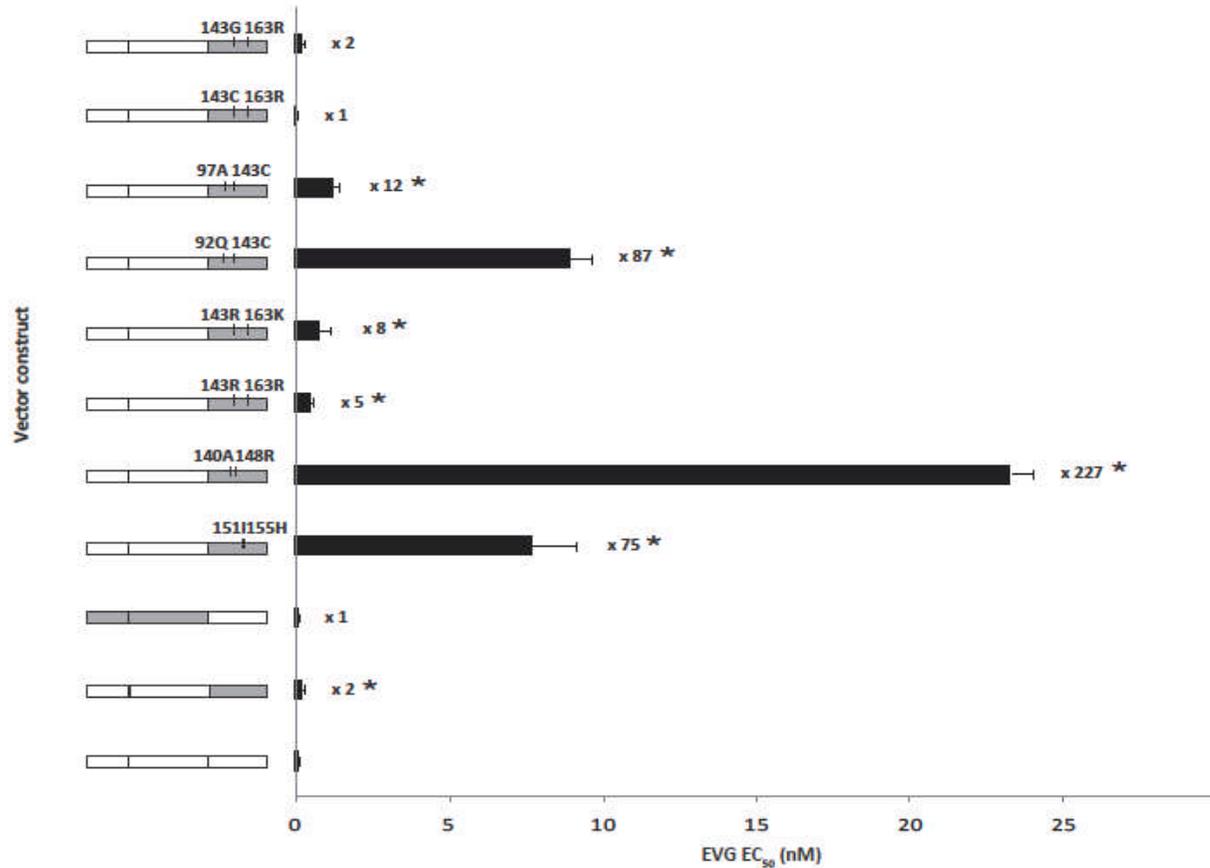


Figure 5.6 EVG susceptibilities of patient-derived *IN* only viruses. Each result is an average of 6 to 12 independent experiments. Fold changes compared to the p8.9NSX wild-type control are indicated next to each bar. Viruses that have a significantly higher ($p < 0.021$) EVG EC₅₀ compared to the wild-type control are indicated with *

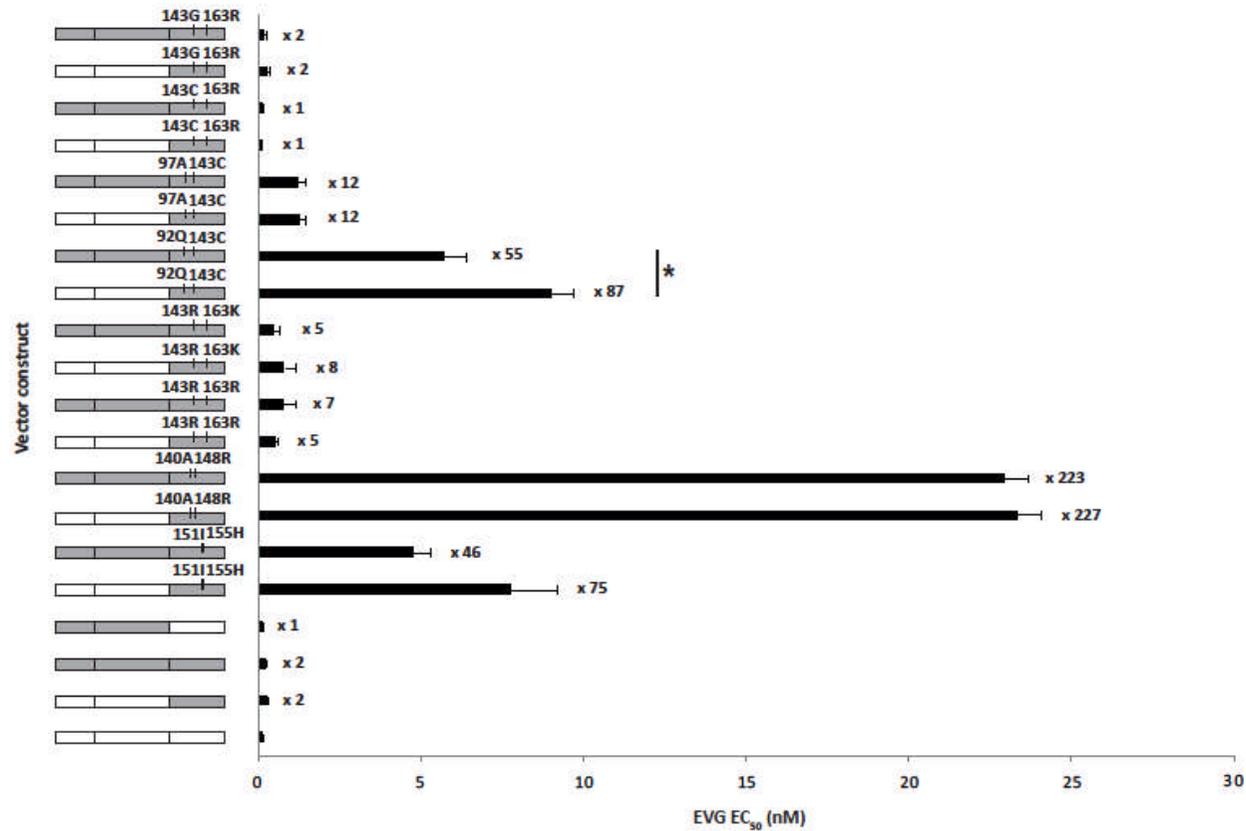


Figure 5.7 Comparison of EVG susceptibilities of patient-derived *IN* only and full-length *pol* viruses. Each result is an average of 6 to 12 independent experiments. Fold changes compared to the p8.9NSX wild-type control are indicated next to each bar. Viruses that have a significantly higher ($p = 0.0064$) EVG EC₅₀ compared to their full-length *pol* counterpart are indicated with *

Table 5.3 EVG susceptibilities of patient-derived *IN* only viruses

<i>IN</i> genotype	EVG EC ₅₀ (nM)	FC EC ₅₀
Q148R+G140A	23.31	227
Y143C+E92Q	8.96	87
N155H+V151I	7.70	75
Y143C+T97A	1.21	12
Y143R+G163K	0.78	8
Y143R+G163R	0.48	5
Y143G+G163R	0.22	2
pt8_WT/ <i>IN</i> *	0.24	2
Y143C+G163R	0.059	1

EVG EC₅₀ and fold changes in EVG EC₅₀ for all viruses expressing patient-derived *IN* only, listed in order from the lowest to the highest susceptibility to EVG. FC = fold change; *ptA_WT/*IN* indicates the vector containing patient-derived *IN* only from the pre-RAL time point

5.2.6 Effects of patient-derived full-length *pol* on EVG susceptibility

To determine whether the coevolved *PR+RT pol* gene segment affected the EVG susceptibilities of patient-derived *IN*, we tested the EVG susceptibilities of patient-derived full-length *pol*. Similar to the virus expressing patient-derived wild-type *IN* only (ptA_WT/*IN*), the virus expressing patient-derived full-length *pol* with wild-type *IN* (ptA_*pol*) also exhibited a significantly higher EVG EC₅₀ compared to the wild-type control (p9.8NSX) at 0.18 ± 0.033 vs 0.1 ± 0.016 nM ($p = 0.03$; Figure 5.7). Overall, the EVG EC₅₀ values of viruses expressing patient-derived full-length *pol* were similar to that of viruses expressing patient-derived *IN* only with the exception of the virus expressing Y143C+E92Q mutations which showed a significant decrease in EVG susceptibility for patient-derived *IN* only compared to full-length *pol* ($p = 0.0064$). Again, this may indicate that the patient coevolved *PR+RT* influences the susceptibility to EVG which is dependent on the combination of INSTI resistance mutations; however, this effect is different between EVG and RAL.

5.2.7 Effects of patient-derived *pol* gene fragments on viral replication fitness

To determine whether there is a relationship between RAL susceptibilities associated with the patient-derived *pol* fragments and viral replication fitness, we tested the viral infectivity of the

viruses expressing full-length *pol* or *IN* gene only using a single-cycle replication assay. For this analysis the infectivity of the viruses was compared to that of the virus expressing the patient derived *IN* only from the pre-RAL time point (ptA_WT/*IN*) which was set to 100%. As expected, the wild-type control (p8.9NSX) and the virus expressing patient-derived *PR+RT* (ptA_*PR+RT*) showed similar infectivity levels to ptA_WT/*IN* at $110.5 \pm 7.7\%$ and $109.7 \pm 12.3\%$, respectively ($p \geq 0.39$; Figure 5.8).

Interestingly, the only viruses expressing INSTI resistance mutations that showed infectivity comparable to that of ptA_WT/*IN* was the patient-derived full-length *pol* or *IN* only virus containing the rare Y143G+G163R mutation combination in *IN* at 86.9 ± 7.6 and $102 \pm 6.6\%$ respectively ($p = 0.28$ and 0.85). All other viruses expressing patient-derived full-length *pol* or *IN* only had significantly lower infectivity levels than ptA_WT/*IN* ranging from 12.9 ± 0.6 to $72.9 \pm 3.4\%$ ($p \leq 0.01$).

Of interest, different substitutions at the major RAL resistance site 143 resulted in different effects on infectivity, as did the combination of different accessory mutations with a major resistance mutation such as Y143C. For example, the virus expressing the patient-derived full-length *pol* with the Y143C+G163R mutations had a significantly lower infectivity ($12.9 \pm 0.61\%$) compared to other substitutions at Y143 which also had the G163R accessory mutation ($61.4 \pm 2.69\%$ and $86.9 \pm 7.6\%$ for the Y143R+G163R and Y143G+G163R viruses, respectively; $p < 0.0001$). On the other hand, the nature of the amino acid at the accessory site 163 also differentially affected the infectivity of viruses expressing the Y143R mutation. Thus, the patient-derived full-length *pol* virus containing Y143R+G163R in *IN* had a significantly lower infectivity than the Y143R+G163K virus ($61.4 \pm 2.7\%$ vs. $72.9 \pm 3.4\%$; $p = 0.025$). In addition, for full-length *pol* viruses expressing the Y143C mutation, linkage to E92Q resulted in a significantly higher infectivity than linkage to T97A and G163R ($39.2 \pm 2.3\%$ vs. $28.3 \pm 3.8\%$ and $12.9 \pm 0.61\%$, respectively; $p \leq 0.034$).

In general, the infectivity of viruses expressing full-length *pol* was greater than or comparable to that of viruses expressing the respective *IN* gene only. The viruses showing increased infectivity upon expression of patient-derived full-length *pol* relative to *IN* gene only were those with following RAL resistance mutation combinations: Q148R+G140A ($p = 0.0023$), Y143R+G163R ($p = 0.024$), Y143R+G163K ($p \leq 0.0001$) and Y143C+G163R ($p = 0.013$). This indicates that the coevolved *PR+RT* not only affects susceptibility to INSTIs but also confers a replicative advantage to the virus containing INSTI resistance mutations.

Table 5.4 lists the infectivity levels of viruses expressing patient-derived full-length *pol* in order of highest to lowest infectivity.

Table 5.4 Infectivity levels of patient-derived full-length *pol* viruses

<i>IN</i> genotype	Infectivity (% of ptA_WT/<i>IN</i>)
ptA_ <i>pol</i> *	102.9
Y143G+G163R	86.9
Y143R+G163K	72.9
N155H+V151I	64.8
Y143R+G163R	61.4
Q148R+G140A	54.6
Y143C+E92Q	39.2
Y143C+T97A	28.3
Y143C+G163R	12.9

Levels of infectivity of vectors expressing patient-derived full-length *pol* compared to the ptA_WT/*IN* control (vector containing patient-derived *IN* only from the pre-RAL time point), listed in order from the highest to the lowest infectivity; *ptA_*pol* indicates the patient-derived full length *pol*, with wild-type *IN* from pre-RAL time point.

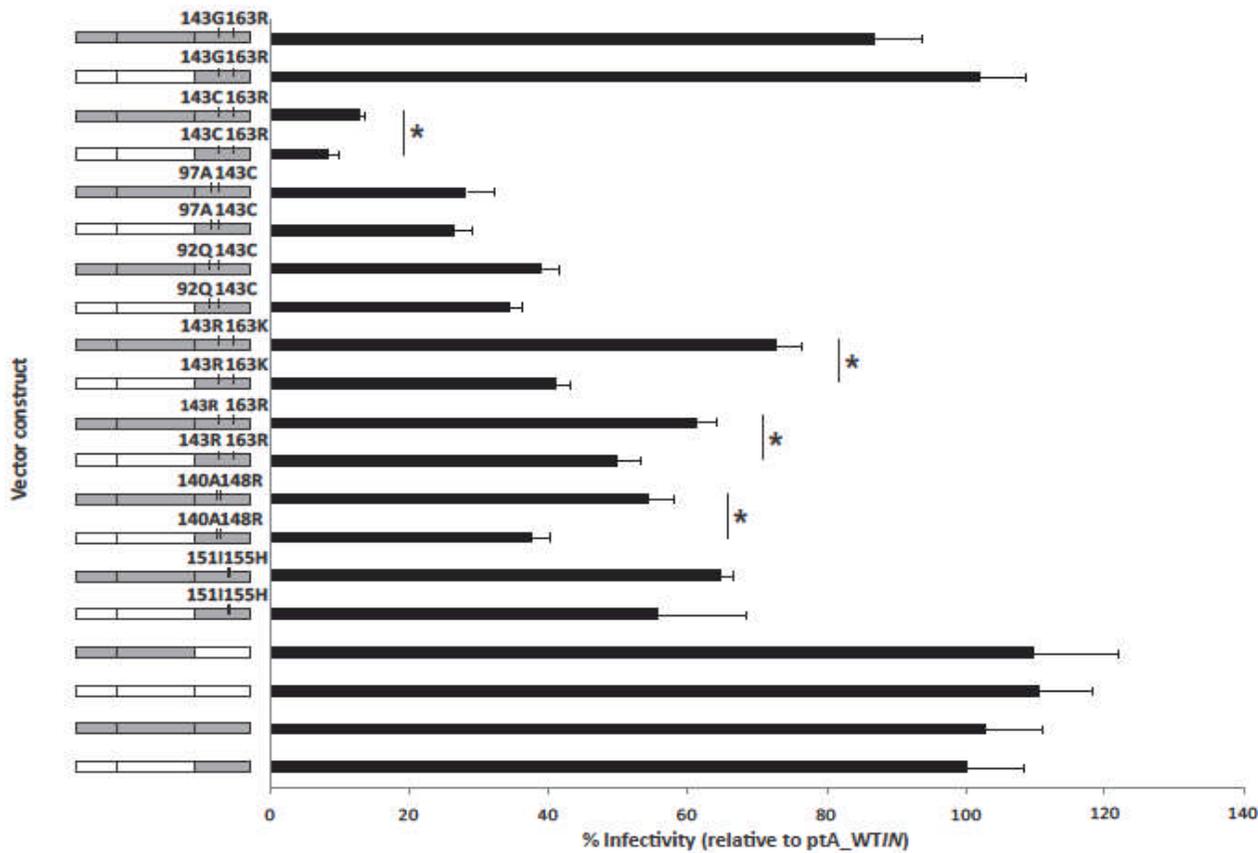


Figure 5.8 Effects of patient-derived *IN* only and full-length *pol* viruses on replicative fitness. Viral infectivity was used to determine the replicative fitness of viruses and was expressed as a percentage of the ptA_WT/IN control (vector containing patient-derived *IN* only from the pre-RAL time point). Significant differences ($p \leq 0.024$) between patient-derived *IN* only viruses and their full-length *pol* counterparts are indicated with *

5.3 Discussion

The majority of phenotypic drug susceptibility assays have been developed to study the drug susceptibility of patient-derived *gag+PR* (Parry et al., 2009), *PR+RT* (Hertogs et al., 1998; Petropoulos et al., 2000), *RT* (Kellam and Larder, 1994), *RT+IN* (Fun et al., 2010; Van et al., 2009a) or *IN* only regions (Hu and Kuritzkes, 2010; Delelis et al., 2009; Goethals et al., 2010; Goethals et al., 2008; Buzon et al., 2008). However, few phenotypic drug susceptibility assays have been developed to investigate patient-derived full-length *pol* gene. Therefore, little data exist on the effects of different combinations of PI, RTI and INSTI resistance mutations on drug resistance. We modified a previously described single-cycle replication assay developed by Parry *et al.*, which has been used to investigate PI and RTI drug susceptibility of patient-derived *PR*, *PR+RT* and *gag+PR* (Parry et al., 2009; Mbisa et al., 2011b; McCormick et al., 2011; Gupta et al., 2010a) by introducing unique restriction sites to facilitate the cloning of patient-derived full-length *pol*. To our knowledge, only two other groups have reported the development of an assay enabling the simultaneous quantification of multi-class drug resistance of patient-derived recombinant vectors containing the whole *pol* gene (Weber et al., 2011; Fransen et al., 2005). Assays encompassing full-length *pol* are rare due to technical difficulties. These include problems amplifying (especially from low viral load samples) (Weber et al., 2011) and cloning long fragments as well as induced toxicities of cloned genes in bacteria resulting in the cloned genes being mutated or the introduction of deletions (Dudley et al., 2009).

The first assay employed the homologous recombination of one or two patient-derived fragments in yeast cells (Weber et al., 2011) whereas the second was based on the PhenoSense assay with the exception of cloning of full-length patient-derived *pol* rather than *PR* and the N terminal region of *RT* (Fransen et al., 2005). A major benefit of our assay is the capacity for it to be performed in a Category 2 laboratory rather than Category 3. In addition, as drug susceptibility is being studied in a single round of replication, it eliminates the problem of selection during *in vitro* viral passage which is unlikely to reflect the phenotype of the virus isolated from the patient. The assay developed by Weber *et al.* however used a multiple replication cycle approach (Weber et al., 2011).

We used the assay to investigate the RAL and EVG susceptibilities of viral variants that were isolated from a patient failing RAL-containing salvage therapy with multi-drug resistance. Patient-derived *IN* only or full-length *pol* vectors were generated to study the effect of a highly resistant patient-derived *PR+RT* fragment on the INSTI susceptibility and infectivity of vectors

harbouring INSTI resistance mutations. Limited data exists on the effects of drug resistance mutations in the *PR* and *RT* genes on susceptibility to INSTIs and on the fitness of viruses harbouring drug resistance mutations in all three *pol* genes. Most studies have only looked at susceptibility or fitness of patient-derived *IN* only and rarely *RT+IN* (Delelis et al., 2009; Goethals et al., 2008; Goethals et al., 2010; Buzon et al., 2008; Hu and Kuritzkes, 2010; Fun et al., 2010; Fransen et al., 2012). However, in the era of salvage therapy that includes ARVs targeting not only *PR* and *RT* but also *IN* it is important to understand the implications of multi-class drug resistance that involves the newest class of these ARV (i.e. INSTIs) and whether mutations in the whole *pol* gene interact with each other to influence drug resistance or viral fitness.

The three residues where major RAL resistance mutations occur, N155, Q148 and Y143, are found in the catalytic core domain of *IN* (Mbisa et al., 2011a; Mbisa et al., 2011a), as such these mutations reduce *IN* enzymatic activity and thereby viral fitness (Metifiot et al., 2010; Delelis et al., 2009; Reigadas et al., 2010; Fun et al., 2010; Hu and Kuritzkes, 2010; Canducci et al., 2010; Fransen et al., 2009a). In addition, all three are also known to result in high levels of resistance to RAL, even in the absence of other accessory mutations (Cooper et al., 2008). It is well established that there is variation in the level of RAL susceptibility of each primary RAL resistance mutation (Mbisa et al., 2011a). The patient-derived *IN* only vectors in this study resulted in decreases in RAL susceptibilities of between 19- and 200-fold relative to wild-type virus (p8.9NSX control). Generally, it has been shown that N155H mutants exhibit a higher RAL susceptibility than Q148R and Y143R/C mutants, and that mutations at position 143 commonly have a lower RAL susceptibility than the Q148R mutation (Fransen et al., 2009a; Fransen et al., 2012; Delelis et al., 2010; Reigadas et al., 2010). This has been shown to be a possible cause of primary drug resistance pathway switches in patients on RAL (Malet et al., 2009; Reigadas et al., 2010; Sichtig et al., 2009; Canducci et al., 2010). A study by Quercia *et al.* however concluded that after prolonged failure, N155H is replaced by Q148H+G140S due to its better selective advantage compared N155H in the presence of continued RAL pressure (Quercia et al., 2009).

In our study, the variants harbouring the Y143R and Q148R major resistance mutations were the majority viral variants during the first round of RAL therapy even though variants containing mutations at all three major resistance sites in *IN* were observed (Chapter 3). Our phenotypic data reveal that this is possibly due to their reduced susceptibility to RAL; the N155H mutation resulted in a 36-fold decrease in RAL susceptibility compared to 147-fold for

Q148R and 159- to 200-fold for Y143R (depending on the nature of accessory mutations present). Others have shown that these mutants often dominate over N155H mutants later during RAL therapy because of the increased viral fitness of Y143 and Q148R mutants (when combined with accessory mutations) in the presence of RAL (Malet et al., 2009; Reigadas et al., 2010; Sichtig et al., 2009; Canducci et al., 2010).

The resistance patterns with regards to substitutions at Y143 have not been well described and little is known about the effects of secondary mutations on this resistance pathway and on the replication fitness of the virus. However, it is widely considered that these mutations play a positive role in IN activity and/or RAL resistance (Canducci et al., 2010; Reigadas et al., 2010; Delelis et al., 2010). Our data revealed that the type of accessory mutation (E92Q, G163R or T97A) combined with the primary resistance mutation Y143C significantly influenced the levels of susceptibility to RAL. We found that the Y143C+E92Q mutant had the lowest susceptibility to RAL (183-fold decrease compared to wild-type virus), whereas the Y143C+G163R mutant had the highest (19-fold decrease). In addition, the replication fitness of these mutants also varied notably with the Y143C+E92Q mutant that exhibited the lowest RAL susceptibility having the highest replication fitness. All three of these accessory mutations are found in the vicinity of the catalytic active site of IN (Mbisa et al., 2011a) therefore it is envisioned that they could be affecting susceptibility and replication fitness by directly influencing the structure of the active site. In contrast, the R or K substitution of G163 in combination with Y143R resulted in similar effects on RAL susceptibility yet showed a significant difference in replication fitness. Both arginine and lysine have positively charged side chains and therefore it would be expected that their effects on IN catalytic activity would be similar. It may be that the bulkier side chain of arginine could have more of an effect on the fitness of the virus than it does on the binding of RAL.

Furthermore, we found that the nature of the amino acid at the primary resistance position 143 (R, C or G) when combined with the accessory mutation G163R exhibited varying RAL susceptibilities and replication fitness. This finding is contrary to another study which showed that the RAL susceptibilities of both the Y143R and Y143C mutants were similar (Delelis et al., 2010). However, in that study the Y143 mutants were not linked to the G163R accessory mutation, which may partially explain the differences between the two studies. This suggests that accessory mutations not only affect replication fitness but also drug susceptibility of the mutant viruses and that a balance between the two effects could play a role in determining the development and evolution of primary and accessory mutations. To emphasize that point,

it was interesting to observe that the Y143G+G163R mutant, which was present as a minority variant in patient A after RAL treatment was stopped, emerged as the dominant viral variant upon re-initiating RAL therapy (see Chapter 3). The Y143G+G163R variant exhibited a significantly higher RAL EC₅₀ than the Y143C+G163R mutant and had significantly higher replication fitness compared to both the Y143R+G163R and Y143C+G163R mutants. The replicative fitness of Y143G+G163R was in fact the highest of all RAL resistant viruses sampled from patient A and was comparable to that of control viruses (ptA_WT/N and p8.9NSX). It is therefore likely that the persistence and subsequent outgrowth of this viral variant after the re-initiation of RAL therapy was due to its higher replicative fitness compared to the other RAL resistant viral variants present in the quasispecies which rapidly disappeared following cessation of RAL therapy.

The late and rare emergence of the Y143G substitution could be due to an unfavourable genetic barrier. For patient A, the wild-type codon for position 143 in IN was TAC. One nucleotide change was required for the Y to C substitution (TAC to TGC) whilst two nucleotide changes were needed to generate the G substitution (TAC to GGC) or R substitution (TAC to CGC). The second change required for the Y to R substitution (T to C) is a transition which occurs at a higher frequency compared to the T to G transversion required for the Y to G substitution. Thus, although more advantageous to the virus, the Y143G substitution is likely to occur less frequently compared to Y143R/C substitutions. It is likely that this mutation developed in patient A because the patient was kept on a failing RAL-containing therapy for a long time (5 months). This illustrates that continuous selective drug pressure during a failing regimen will force the virus to continuously evolve towards a fitter resistant virus that is then more likely to persist in the absence of drug pressure. To our knowledge, the Y143G mutant has only been previously described by one other study (Canducci et al., 2010). This study showed that the Y143G mutant also appeared in a RAL-treated patient after the development of Y143C, albeit in the presence of different accessory mutations (L74M and T97A). However, the Y143G mutation then developed into Y143R before reverting back to Y143C during a 64 week period.

Structural studies have shown that RAL and EVG have similar binding models at the IN active site and therefore cross-resistance is anticipated (Krishnan et al., 2010; Hare et al., 2010). With regards to EVG susceptibility, the most common major resistance mutations that were selected *in vivo*, at virological failure in a Phase 2 study were Q148H/R/H, N155H, E92Q and E138K (McColl et al., 2007). Similar to Q148 and N155, residues E92 and E138 are located in

the catalytic core domain of IN (Goethals et al., 2008; Mbisa et al., 2011a) and the development of mutations at both positions are also associated with significant reductions in viral replication capacity as well as EVG susceptibility (McColl et al., 2007; Goethals et al., 2008). A study carried out by Goethals *et al.* (Goethals et al., 2008) using site directed mutants of Q148R, E92Q, N155H and E138K revealed that the Q148R mutant had the highest fold change in susceptibility to EVG, followed by E92Q, N155H and finally E138K. As patient A was treated with RAL and not EVG, it was anticipated that only mutations conferring dual resistance will have any impact on EVG susceptibility, these being Q148R, N155H and E92Q. This proved to be the case, with the patient-derived *IN* only vectors harbouring N155H+V151I, Q148R+G140A and Y143C+E92Q showing significant fold changes in EVG susceptibility; ranging from 75 to 227 compared to 1 to 12 fold changes for the other patient-derived *IN* only vectors. In addition, the EVG susceptibilities determined here follow the same trend as described by Goethals *et al.* with the Q148R+G140A mutant displaying the lowest EVG susceptibility, followed by Y143C+E92Q and N155H+V151I (Goethals et al., 2008).

The effect of PI and RTI resistance mutations on INSTI resistance has rarely been studied. In addition, the effect of these mutations on viral fitness when combined with resistance mutations in the *IN* gene is yet to be fully elucidated. The few studies undertaken have used site-directed mutagenesis in a wild-type background to create vectors with resistance mutations in their *RT* and *IN* genes (Hu and Kuritzkes, 2012; Gupta et al., 2009b). Interestingly, this showed that certain combinations of NNRTI and INSTI resistance mutations affect viral fitness and susceptibility to some ARV drugs (Hu and Kuritzkes, 2012). Hu and Kuritzkes showed that site-directed mutants harbouring NNRTI and INSTI resistance mutations were less fit than mutants harbouring a single class of resistance mutation in the absence of drug. However, viruses with the K103N resistance mutation in *RT* when combined with N155H or Q148H+G140S resistance mutations in *IN* were fitter than the K103N virus alone in the presence of EFV whereas the NNRTI resistance mutations E138K and Y181C improved the fitness of Q148H+G140S viruses in the presence of RAL. Furthermore, viruses expressing K103N+Q148H+G140S or E138K+Q148H+G140S displayed a significantly higher fold-increase in EFV EC₅₀ than the NNRTI mutants alone. In addition, the E138K+Q148H+G140S virus also had a significantly greater fold-increase in RAL EC₅₀ than the *IN* mutant alone. Conversely, another study showed that different combinations of NNRTI and INSTI resistance mutations did not alter the RAL and NNRTI susceptibility of the site-directed mutants but that they did reduce their fitness. The fitness of viruses harbouring the Q148R/H/K INSTI resistance mutation and K103N, Y181C or G190A/S NNRTI resistance mutations was reduced when compared to the

Q148R/H/K mutation alone whereas viruses harbouring the INSTI resistance associated mutation E92Q, together with K103N was non-infectious (Gupta et al., 2009b).

Our assay allowed us to directly measure the effect of the coevolved multi-drug resistant *PR+RT* on the INSTI susceptibility and viral fitness of different patient-derived *IN* genes harbouring different INSTI resistance genotypes. Interestingly, our data showed that upon the addition of the patient-derived *PR+RT* fragment, the INSTI susceptibilities were significantly increased for three INSTI resistant genotypes (Y143R+G163R, Y143R+G163K and Y143C+T97A) for RAL and one genotype (Y143C+E92Q) for EVG. Furthermore, the viral infectivity was significantly increased upon the addition of the patient-derived *PR+RT* fragment of four INSTI resistant genotypes (Q148R+G140A, Y143R+G163R, Y143R+G163K and Y143C+G163R). These data indicate that there may be mutation(s) in the *PR* and/or *RT* gene(s) which are increasing the INSTI drug susceptibility and/or viral fitness depending on the INSTI resistant genotype. In our study the patient-derived multi-drug resistant *PR+RT* fragment alone had no effect on RAL or EVG susceptibility suggesting that these effects are possibly a result of interactions between mutations in the different genes making up HIV-1 *pol*. These interactions could either be protein-protein in reverse transcription complexes (RTCs) and/or PICs during DNA synthesis and/or integration as both complexes contain RT and IN (Miller et al., 1997; Bukrinsky et al., 1993a; Farnet and Haseltine, 1991; Chakraborty et al., 2013; Fassati and Goff, 2001; Karageorgos et al., 1993), or through the secondary or tertiary structure of the HIV-1 RNA genome which could affect viral replication or interactions in the PR processing of the gag-pol polyproteins (Sundquist and Krausslich, 2012).

Taken together, our data indicate that phenotypic drug susceptibility testing of patient-derived *IN* only might not give the full picture considering that the majority of patients receiving RAL have experienced and failed on PI- and RTI-containing therapy. Our results suggest that mutation(s) in other parts of the *pol* gene may contribute to the fitness and INSTI resistance of the virus. In addition, we show that the dominance of certain variants in a quasispecies is likely the result of a complex interplay between drug susceptibility and viral fitness. Further investigation and analysis of different combinations of PI, RTI and INSTI resistance mutations may result in the design of better therapeutic strategies and may improve our ability to predict the emergence of drug resistance in the era of HAART.

CHAPTER 6

The role of the C terminal region of reverse transcriptase in drug resistance of HIV-1 Subtype F

6.1 Introduction

Typically genotypic and phenotypic assays for RT inhibitors only include the first 320 amino acids of RT in the N terminal region. This is because the majority of the major RTI resistance mutations are located proximal to the polymerase active site (the target for NRTIs) or within the NNRTI binding pocket, both of which are located in the N terminal region of RT (Johnson et al., 2011). However, studies have shown that there are mutations in the C terminal region of *RT* that can affect RTI susceptibility alone or in combination with mutations in the N terminal region (Paredes et al., 2011; Yap et al., 2007; Brehm et al., 2007; Sluis-Cremer et al., 2010; Harrigan et al., 2002; Nikolenko et al., 2007; Kemp et al., 1998). This has evoked discussion about the inclusion of the C terminal region in standard resistance genotyping and phenotyping (Gotte, 2007).

Studies exploring the presence of mutations in the C terminal region of *RT* involved in resistance have predominantly involved subtype B sequences (Brehm et al., 2007; Kemp et al., 1998; Brehm et al., 2008; Nikolenko et al., 2010; Delviks-Frankenberry et al., 2008; Lengruher et al., 2011; Zelina et al., 2008) with only a handful of studies investigating their presence and respective resistance mechanisms in non-B subtypes including subtypes C and A and the circulating recombinant form, CRF01_AE (Mbisa et al., 2011b; Delviks-Frankenberry et al., 2009; Paredes et al., 2011; McCormick et al., 2011; Delviks-Frankenberry et al., 2013). However, 89% of HIV-1 infections worldwide between 2004 and 2007 were caused by non-B subtypes (Hemelaar et al., 2011). Therefore, it is important to determine the role of these mutations in non-B subtypes especially as it has been shown that the nature and evolution of drug resistance mutations can be subtype-specific (see General Introduction chapter).

As part of a EuroCoord CHAIN study (www.eurocoord.net/collaborative_projects/chain.aspx) we obtained samples from children infected with HIV-1 subtype F1 from Romania who were

undergoing ART. HIV-1 subtype F accounted for approximately 150,000 HIV-1 infections globally between 2004 and 2007. The majority of subtype F infections occur in Latin America, Europe, Western and Central Africa and Central Asia and were responsible for between 0.47 and 3.53% of the total infections in these regions between 2004 and 2007 (Hemelaar et al., 2011). The HIV-1 subtype F1 epidemic in Romanian children occurred in the 1980s through the parenteral route by the use of unsterilized needles in institutionalized children (Hersh et al., 1991). This gave us an opportunity not only to investigate the development of resistance mutations in the C terminal region of *RT* in a non-B HIV-1 subtype but also to investigate the resistance profile to RTIs in children who had acquired the infection parenterally. Studies have shown that resistance in children is more extensive than in adults probably due to higher viral loads, problems with drug dosage and adherence resulting in higher rates of virological failure (Gupta et al., 2009a).

Therefore, our goal in this study was to amplify full-length *RT* from the HIV-1 subtype F1-infected children to investigate the appearance of novel mutations in the C terminal region of *RT* associated with known resistance mutations in the N terminal region. In addition, we aimed to investigate the effect of any C terminal region mutations identified on drug susceptibility and/or RT activities *in vitro* using a subtype-F specific *gag-pol* containing plasmid vector in a single-cycle replication assay.

6.2 Results

6.2.1 Identification of mutations in the C terminal region of *RT* of HIV-1 subtype F associated with primary resistance mutations

Previously, thirty-nine samples from children infected with HIV-1 subtype F1 were sequenced from codon 5 of *PR* to codon 320 of *RT*. Eleven of the samples were found to contain known primary RTI resistance mutations. We therefore set out to determine if the primary RTI resistance mutations were associated with other mutations in the C terminal region. To that end, we performed population-based sequencing of full-length *PR* and *RT* of twenty-seven of the samples. This confirmed that eleven of these samples had known primary RTI resistance mutations in the N terminal region which included NRTI resistance mutations at residues D67, T69, K70, V75, F77, M184, L210, T215 and K219 and NNRTI resistance mutations at residues

V90, K101, K103, E138, Y188 and P225 (Table 6.1). Analysis of the C terminal region of these sequences compared to the remaining sixteen sequences that did not contain any primary resistance mutations revealed eleven mutations that exclusively occurred in the presence of N terminal region resistance mutations, namely S332T, N348I/L, M357R/T, T386I, N447S, S468A, T470S/N, D471E, K530R, A554S and I556V (Table 6.1).

Next, we examined the association of the eleven C terminal region mutations with the development of RTI resistance in subtype F by analyzing sequences from the Stanford University HIV drug resistance database. Using these sequences we compared the mutation frequencies between RTI-naive and RTI-experienced patients (Table 6.2). The number of subtype F sequences available for analysis that were sequenced beyond the first 320 amino acids of RT was limited ($n \leq 149$). As mentioned before this is the result of most genotypic assays only encompassing the first 300 amino acids of RT. That said, eight out of the eleven mutations identified above were found at different frequencies in RTI-naive and -experienced patients. Three out of the eight mutations, namely N348I, N447S and I556V, appeared to occur more frequently in the treatment-experienced than treatment-naive patients suggesting that they could potentially be associated with drug resistance. In contrast, the remaining five mutations (S332T, M357R/T, T386I, K530R and A554S) seemed to be present more frequently in the RTI-naive group suggesting that they could be polymorphisms. However, only two of the mutations were calculated to be significantly associated with one group than the other, these being N348I and M357R with p values of 0.01 and 0.0001 (Fisher's exact test), respectively.

Table 6.1 Mutations present in the C terminal of *RT* of HIV-1 subtype F viruses isolated from patients in the PENTA-EPPIC network

Reference No.		C terminal <i>RT</i> Mutations	
^a RTI-sensitive patients	RS0900 0181	T345P K366R M378Q A435T L517I N519S	
	RS0900 0183	Y342F T345P Q452I L517I	
	RS0900 0184	K366R	
	RS0900 0189	M378I C379S E404D A435T Q452T L517I	
	RS0900 0192	I329L F346Y M378Q E413D A435T K454R	
	RS0900 0196	T345Q K366R E370D M378L E404D A435I Q452L K512Q L517I Q524H A534T	
	RS0900 0197	K512Q	
	RS0900 0198	V467I S468P N519S	
	RS0900 0203	F346Y M378L	
	RS0900 0204	F346Y K390R Q524L	
	RS0900 0206	A371V I375V Q480E Q520K	
	RS0900 0207	A371V I375V Q520K Q524K Q527E	
	RS0900 0209	F346Y	
	RS0900 0212	I329V Q334L A371V I375V M378L S446A T470N L517I	
	RS0900 0216	T345P T369A A376S K390R T403L E404D A435I Q452E V467I Q509K	
RS0900 0218	T345P M378R A400T E404D A435I S446A Q452K V467I T470A I482M A502T Q509K K512Q		
		C terminal <i>RT</i> Mutations	N terminal Resistance Mutations
^b RTI-resistant patients	RS0900 0180	S322T T345A I375V N447S	T69N K219Q
	RS0900 0195	T345P K366R M378Q C379S E413D Q452L	D67N T69N K70KR L74V M184V K219Q Y188L
	RS0900 0199	S322T F346Y	K103KR
	RS0900 0200	T345A F346Y M378Q C379S A435T S446A A554S	K103KN
	RS0900 0201	S322T I329V T345A F346Y N348I T386I N447S	D67N K70R K219Q N348I V90I K103N P225H
	RS0900 0202	Q334L F346Y M357R A400T E471D K530R I556V	K101Q
	RS0900 0205	T345M N348L A371V I375V S446A N519S	E44DE V75M F77FL M184V L210LRW T215FY N348L V90I K103N P225H
	RS0900 0208	N348I M378R I435T S446A	M184V N348I V106A F227L
	RS0900 0210	M357T E370D M378R A435I T470N K512Q	E138G
	RS0900 0217	I329V A376S M378Q S468A T470S D471E	T69AT
	RS0900 0219	T369A E404D S468A D471E Q509K	E138G

^aPatients with no known RTI resistance mutations in the N terminal of *RT*; ^bPatients with known RTI resistance mutations in the N terminal of *RT*. Amino acid mutations in **bold** are present only in patients with RTI resistance mutations in the N terminal of *RT* (which are also indicated in the table).

Table 6.2 Analysis of the frequency of C terminal mutations in RTI-naive and -experienced patients infected with Subtype F from the Stanford University Database

RT domain	Consensus B residue	RTI-naive patients		RTI-experienced patients		% difference in mutations ^b	P value
		No. of sequences available	% mutation frequency ^a	No. of sequences available	% mutation frequency ^a		
Cn	S322	149	9.4 (T)	147	8.2 (T)	-1.2	0.84
	N348	55	0 (I/L)	40	12.5 (I) 0 (L)	+12.5 (I) 0 (L)	0.01 (I) 1 (L)
	M357	53	56.6 (R) 3.8 (T)	40	5 (R) 0(T)	-51.6 (R) -3.8 (T)	0.0001 (R) 0.15 (T)
	T386	52	23.1 (I)	39	20.5 (I)	-3.1	0.30
Rh	N447	15	6.7 (S)	19	10.5 (S)	+3.8	1
	S468	15	0 (A)	19	0 (A)	0	1
	T470	15	0 (S/N)	19	0 (S/N)	0	1
	D471	15	100 (E)	19	100 (E)	0	1
	K530	15	33.3 (R)	19	0 (R)	-33.3	1
	A554	15	13.3 (S)	19	0 (S)	-13.3	0.15
	I556	15	0 (V)	19	5.3(V)	+5.3	1

^aThe nature of the amino acid mutation from wild-type is indicated in (); ^b% difference is calculated by subtracting the mutation frequency of RTI-naive patients from RTI-experienced patients, plus sign indicates an increase and minus sign a decrease in mutation frequency in RTI-experienced patients compared to RTI-naive. Mutations calculated to be statistically more frequent in one patient group over another are indicated in **bold** with p values ≤ 0.01 (Fisher's exact test).

6.2.2 Phenotypic characterization of C terminal *RT* mutations identified in subtype F to be associated with drug resistance

As discussed previously, studies have shown that subtype-specific polymorphisms elsewhere in the genome may affect drug resistance. Thus, we sought to determine the effects of the mutations we identified at RT positions N348 and M357 on RTI susceptibility and RT activity using a subtype-specific vector. A wild-type subtype F specific expression vector (p8.9NSX-F) was created to ensure that any subtype F specific polymorphisms in *gag-pol* were taken into consideration during the analysis of subtype F patient-derived *RT* expression vectors. This vector was constructed by subcloning the whole *gag-pol* region (~4.3kb) of p93BR020.1 (a wild-type subtype F molecular clone) into the p8.9NSX expression vector using *NotI* and *EcoRI* restriction sites upstream of *gag* and downstream of *IN*, respectively. p8.9NSX-F was modified to accommodate *RT* domain-swapping by introducing three restriction sites; *HpaI* (between the *pol* domain and *cn* subdomain in *RT*), *SpeI* (between the *cn* subdomain and RNase H domain in *RT*) and *Clal* (at the beginning of the *IN* gene). Full-length *PR+RT* patient-derived PCR products were also modified to facilitate *RT* domain-swapping into the p8.9NSX-F expression vector using the same three restriction sites. In addition, internal *HpaI* and *SpeI* restriction sites were knocked out to ensure the introduced sites were unique. Patient-derived *PR+RT* fragments were subsequently subcloned into p8.9NSX-F to generate patient-derived *PR+RT* vectors with a wild-type subtype F backbone.

Unfortunately, the co-transfection of p8.9NSX-F or p8.9NSX-F expressing patient-derived *PR+RT* with pCSFLW and pMDG into 293T cells resulted in minimal or undetectable virus production in cell culture supernatants determined by infectivity assay (data not shown). As previous attempts to clone smaller *gag-pol* fragments from different non-B HIV-1 subtypes into p8.9NSX have resulted in infectious virus production, we therefore attempted the cloning of wild-type subtype F or patient-derived fragments encompassing the *PR+RT* region or *RT* only directly into the p8.9NSX vector. The transfection of these constructs into 293T cells also resulted in minimal or undetectable virus production in cell culture supernatants compared to the subtype B p8.9NSX vector (Figure 6.1). Sequencing of the whole *gag-pol* region of all constructed vectors revealed no sequence abnormalities.

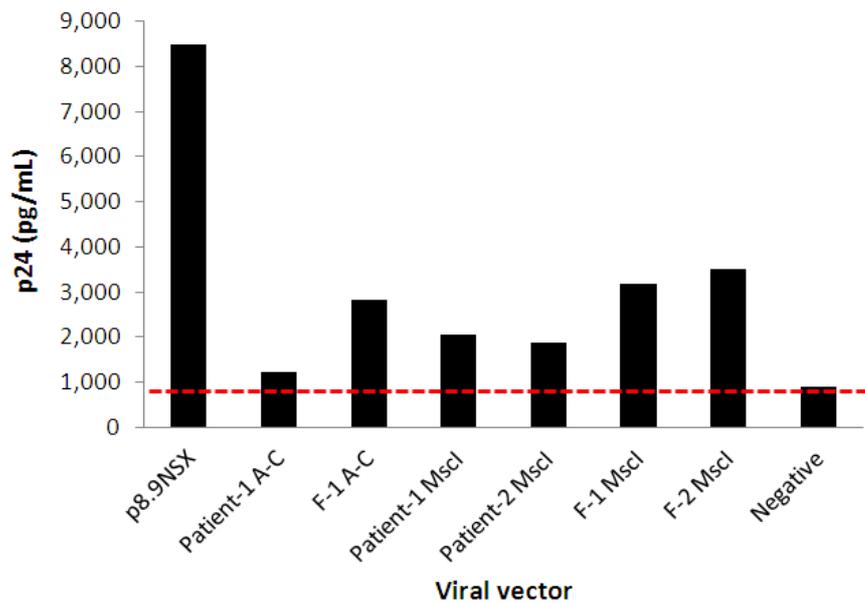


Figure 6.1 Amount of p24 present 48 hours post transfection of retroviral vectors. All viral vectors depicted here contain the p8.9NSX vector backbone. All pseudoviruses were harvested 48 hours post transfection from cell-free supernatant. A-C indicates the *ApaI-ClaI* fragment (*PR+RT* fragment), *MscI* indicates the *MscI-MscI* (*RT* only fragment). “Patient” indicates a patient-derived fragment and “F” indicates a wild-type subtype F-derived fragment. DMEM was used as a negative control; red dotted line indicates the limit of detection of the p24 ELISA. Numbers indicate different patient / wild-type subtype F clones used to generate the expression vector.

It is possible that the subtype F sequences were incompatible with the p8.9NSX vector system. Therefore, we next attempted the cloning of the wild-type and patient-derived subtype F sequences into a different expression vector, pHL(WT) (Nikolenko et al., 2007). As well as containing the *gag-pol* region, the pHL(WT) vector contains the complete HIV-1 genome with the exception of a deleted *env* gene and the expression of a *luciferase* reporter gene in the *nef* open reading frame (Figure 6.2). This eliminates the need for the separate retroviral-based vector expressing luciferase (pCSFLW). In addition, this vector is under the influence of an HIV promoter rather than a CMV and MLV promoter as in p8.9NSX and pCSFLW, respectively. Wild-type subtype F (p93BR020.1) and patient-derived *RT* fragments (~1.6kb) were subcloned into pHL(WT) using the restriction sites *MscI* (flanking *RT* amino acids 25/26) and *ClaI* (flanking *IN*

amino acids 4/5). The vectors were co-transfected with pMDG into 293T cells to produce pseudotyped recombinant viruses. However, this also resulted in minimal or undetectable virus production when assessed using an infectivity assay (data not shown).

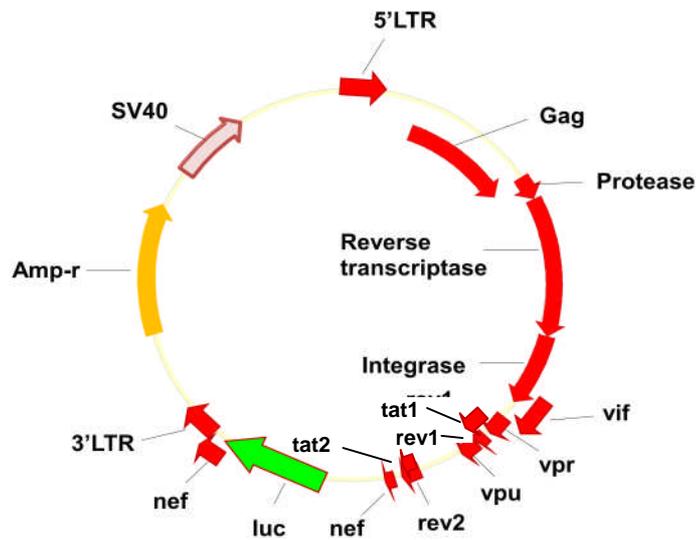


Figure 6.2 The pHL(WT) HIV expression vector. The pHL(WT) vector expresses all HIV genes, (with the exception of *env*) and the *firefly luciferase* reporter gene under an HIV LTR promoter. Similar to p8.9NSX, it also contains the SV40 Ori and an *ampicillin resistance* gene (*Amp-r*).

6.2.3 Analysis of viral protein expression and proteolytic processing of subtype F expressing recombinant viruses

Next, we investigated whether biological defects were responsible for the defective virus production by the subtype F vectors. We performed western blot analysis on cell lysates from transfected cells to detect if HIV proteins were expressed by the vectors and to determine the extent of polyprotein processing. Cell lysates were collected from cells transfected with wild-type subtype B, wild-type subtype F and patient-derived vectors at 12 and 24 hours post-transfection. Unsurprisingly, western blotting analysis of the cell lysates using human polyclonal sera or mouse p24 antibodies showed minimal or undetectable viral protein expression at 12 hours post-transfection. However, at 24 hours post-transfection viral proteins were detected in all cell lysates transfected with either wild-type or patient-derived vectors (Figure 6.3a). The analysis shows that only the final gag cleavage product, CA (p24 viral antigen), could be detected in the pseudovirus control samples (p8.9NSX and pHL[WT]) harvested from the supernatant at 48 hours post-transfection. In contrast, the cell lysates

show different degrees of cleavage of the viral gag polyprotein, Pr55^{gag}, into MA-CA, CA-SP1/CA in all six samples (lanes 1-6; Figure 6.3b). The wild-type subtype B (p8.9NSX and pHL[WT]; lanes 3 and 6) and the patient-derived subtype F (MC-pHL; lane 4) vectors had a higher efficiency of cleavage into the final cleavage product CA-SP1/CA with $\geq 50\%$ of the overall lane density being attributed to the CA-SP1/CA band. In contrast, all other vectors (lanes 1, 2 and 5) had $\leq 50\%$ of the overall lane density being attributed to the CA-SP1/CA bands indicating a lower efficiency of cleavage.

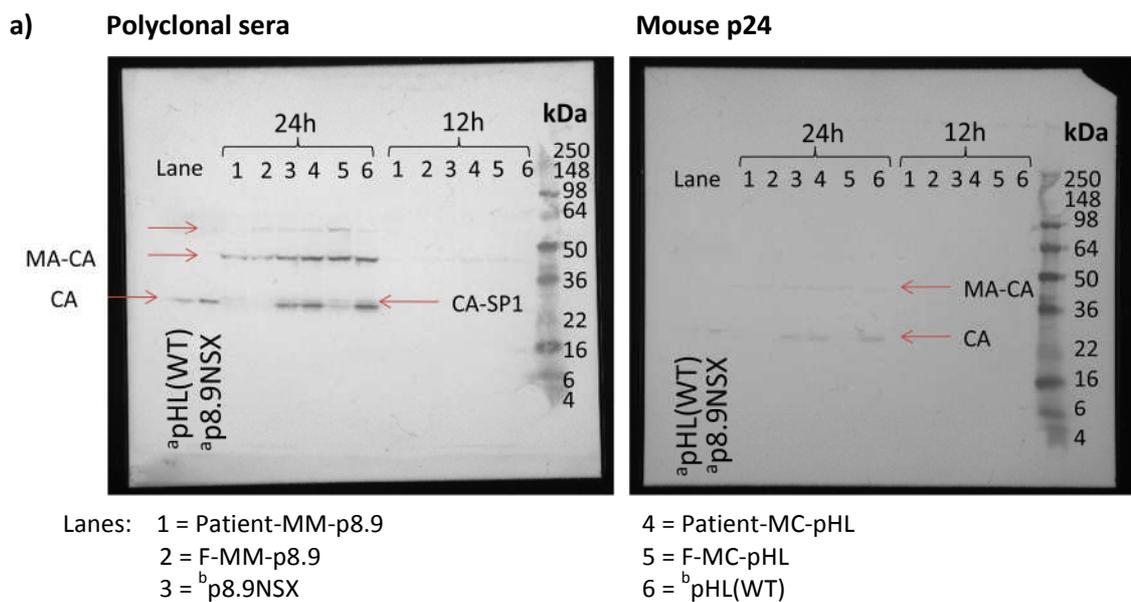
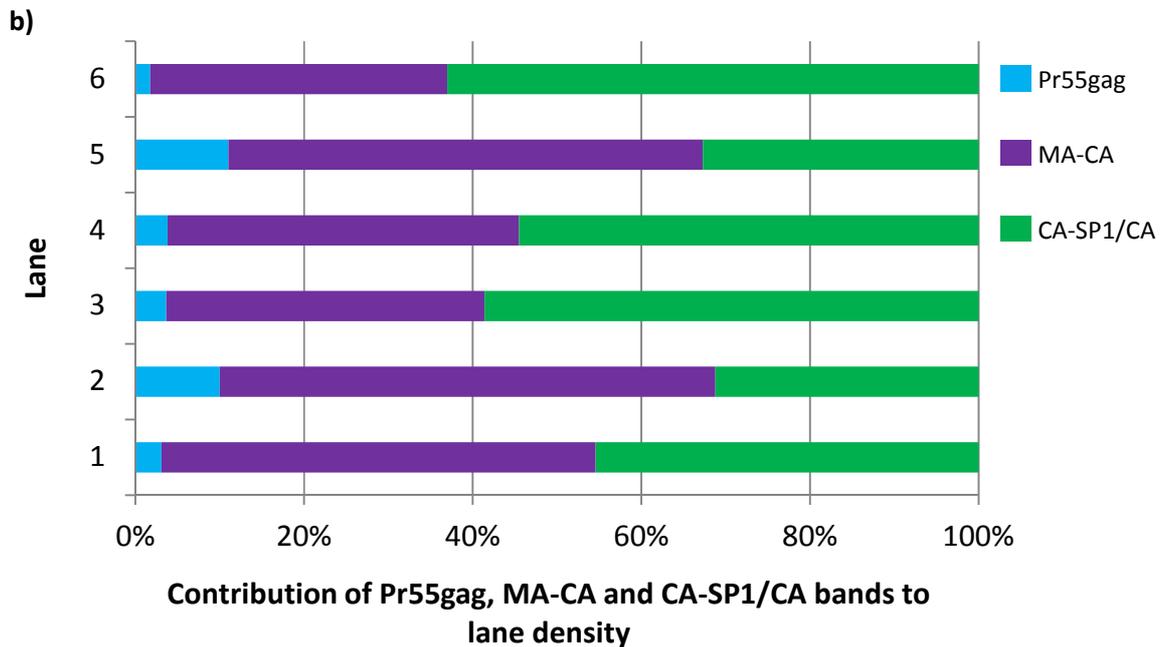


Figure 6.3 Western blot analysis to a) confirm the presence of viral proteins in cell lysates 12 and 24 hours post-transfection and b) determine the cleavage efficiency of pseudoviruses 24 hours post-transfection. ^apseudovirus controls harvested 48 hours post-transfection from cell-free supernatant. ^bpseudovirus controls harvested from cell lysates. p8.9 indicates a p8.9NSX viral vector backbone and pHL indicates a pHL(WT) viral vector backbone. MM indicates the *MscI-MscI* fragment (*RT* only fragment), MC indicates the *MscI-ClaI* (*RT* only fragment). “Patient” indicates a patient-derived subtype F fragment and “F” indicates a wild-type molecular clone subtype F-derived fragment. Pr55^{gag}; viral gag polyprotein, MA; Matrix, CA; Capsid/p24, SP1; Spacer Peptide 1. NB: CA-SP1 and CA bands were not sufficiently separated on the gel used, therefore the two bands were quantified as one band.



Lanes: 1 = Patient-MM-p8.9
 2 = F-MM-p8.9
 3 = ^bp8.9NSX

4 = Patient-MC-pHL
 5 = F-MC-pHL
 6 = ^bpHL(WT)

Figure 6.3 Western blot analysis to a) confirm the presence of viral proteins in cell lysates 12 and 24 hours post-transfection and b) determine the cleavage efficiency of pseudoviruses 24 hours post-transfection (continued). ^apseudovirus controls harvested 48 hours post-transfection from cell-free supernatant. ^bpseudovirus controls harvested from cell lysates. p8.9 indicates a p8.9NSX viral vector backbone and pHL indicates a pHL(WT) viral vector backbone. MM indicates the *MscI-MscI* fragment (*RT* only fragment), MC indicates the *MscI-ClaI* (*RT* only fragment). “Patient” indicates a patient-derived subtype F fragment and “F” indicates a wild-type molecular clone subtype F-derived fragment. Pr55gag; viral gag polyprotein, MA; Matrix, CA; Capsid/p24, SP1; Spacer Peptide 1. NB: CA-SP1 and CA bands were not sufficiently separated on the gel used, therefore the two bands were quantified as one band.

6.3 Discussion

Using an in-house RT-PCR assay we were able to amplify twenty-seven full-length *PR* and *RT* sequences from subtype F1 infected children. Eleven of these 27 sequences contained known resistance mutations in the N terminal region of *RT* including NRTI resistance mutations at residues D67, T69, K70, V75, F77, M184, L210, T215 and K219 and NNRTI resistance mutations at residues V90, K101, K103, E138, Y188 and P225. Further analysis of these sequences revealed that there were eleven novel mutations in the C terminal region of *RT* which only occurred in sequences that contained at least one known RTI resistance mutation in the N terminal region of *RT* namely, S332T, N348I/L, M357R/T, T386I, N447S, S468R, T470S, D471E, K530R, A554S and I556V. To verify our findings we analysed the frequency of these mutations in subtype F sequences on the Stanford HIV drug resistance database. The strength of our analysis was limited by the number of subtype F sequences in the database in general and the number sequenced beyond amino acid 300 in particular. For example, from residue 447 onwards there were only thirty-four sequences from RTI-naive and –experienced patients combined. Nonetheless, the analysis showed that two of the eleven mutations occurred at significantly higher frequencies, these being N348I which occurred more frequently in RTI-experienced patients and M357R which occurred more frequently in RTI-naive patients.

The N348I C terminal mutation is known to cause dual resistance to NRTIs and NNRTIs (Yap et al., 2007; Ehteshami and Gotte, 2008; Sluis-Cremer et al., 2010; Gupta et al., 2010b) and was present in two out of three sequences containing a mutation at this position. The remaining sequence harboured a novel leucine substitution at this position. To our knowledge this is the first description of this novel N348L in subtype F sequences. It has however been described in three subtype B viruses isolated from one RTI-treated and two RTI-naive patients (Rhee et al., 2003). Both mutations are associated with decreased susceptibility to nevirapine and AZT due to a decrease in RNase H activity and/or increase in AZT-MP excision (Radzio and Sluis-Cremer, 2011). The rare occurrence of the leucine substitution could be due to the requirement of two or more nucleotide changes from asparagine (AAU or AAC) to leucine (CUU, CUC, CUA, CUG, UUA or UUG) whereas the change to isoleucine (AUU, AUC or AUA) can be achieved through one nucleotide change. Thus, this mutation most likely takes longer to develop during treatment failure or it could develop as a revertant when treatment is stopped.

Little is known about the effects of substitutions occurring at position 357 in RT on RTI resistance or viral fitness. M357 is found in the connection domain of the RT enzyme and can

be highly polymorphic with substitutions to lysine, threonine and arginine occurring at high frequencies in subtypes A, B, C, F, G, AE and AG naive to RTI treatment (Rhee et al., 2003). In subtype F specifically, deviations from the subtype B consensus methionine residue to arginine are found in 56.6% of RT-naive sequenced viruses found on the Stanford Database, whilst viruses harbouring threonine at this position, the second substitution detected in our patients samples, are found in 3.8% of these viruses.

To take into account the effect of subtype-specific polymorphisms occurring elsewhere in the viral *pol* gene we generated a subtype F-specific full length *gag-pol* vector backbone to investigate the effects of these mutations (including the novel N348L mutation) on RTI susceptibility and RT activity. To this end a wild-type F fragment encompassing *gag-pol* was subcloned into the p8.9NSX viral vector to create p8.9NSX-F. Mechanistic assays were learnt in Dr Pathak's lab at the NCI (Frederick Maryland, USA) during an HPA funded travel fellowship in 2010 and involved the *in vitro* monitoring of three RT activities; polymerization, RNase H cleavage and AZT excision-extension (Delviks-Frankenberry et al., 2008) using viral-associated RT and radioactively labelled RNA and/or DNA oligonucleotides.

RT polymerisation of DNA is measured over a time course using a 5' radioactively labelled 18 mer DNA primer and a 42 mer RNA or DNA template and the addition of dNTPs. Products are then visualised using polyacrylamide gel electrophoresis (PAGE); a defect in polymerisation is indicated by a reduced amount of final product (42 mer) compared to a wild-type control (Delviks-Frankenberry et al., 2008). On the other hand, RNase H activity of RT was measured using an 18 mer RNA/DNA hybrid in which the 5' end of the RNA was radioactively labelled. After a 30 minute incubation period products are visualised by PAGE. A reduced RNase H activity is indicated by a reduction in the amount of 18 mer cleaved into 14/15 mer products compared to a wild-type control. Measuring RNAase H secondary cleavage can also be monitored. This uses a 41 mer RNA/DNA hybrid in which the 5' end of the RNA has also been radioactively labelled. Aliquots are taken at intervals during a 30 minute incubation period and, the cleavage products are similarly detected by PAGE. Normal RNase H activity results in an increase in the amount of secondary cleavage products (8, 15 and 18 mer products) with increasing time and complete cleavage of the 41 mer starting product after a 30 minutes incubation. However, a reduced RNase H activity may result in very little cleavage of the 41 mer starting product and a reduction in the 8 mer cleavage product (Delviks-Frankenberry et al., 2008).

AZT-excision was monitored using a 5' radioactively labelled 18 mer DNA primer hybridised to a 42 mer RNA or DNA template. AZT was used to block the template and halt RT polymerisation. With the addition of ATP, dNTPs and more AZT, aliquots of the reaction were taken at multiple time points over 60 minutes and products were visualised by PAGE. AZT-excision can be seen by the gradual increase in full length 42 mer product at the end of the 60 minutes incubation period (Delviks-Frankenberry et al., 2008).

The intention was to use these assays to determine the mechanisms of action for RTI resistance of the novel C terminal mutations in subtype F, such as the N348L mutation, (potential drug resistance mutations), which were identified in this study. Unfortunately however, the p8.9NSX-F vector did not produce detectable virus by infectivity assay so I was unable to perform the drug resistance phenotyping or biochemical studies.

Sequencing of the whole *gag-pol* region of p8.9NSX-F revealed no sequence abnormalities. Smaller subtype F and patient-derived fragments encompassing *PR+RT* and *RT* only were subsequently subcloned into the p8.9NSX backbone; however, this also resulted in no detectable virus production. We therefore hypothesise that there may be some incompatibility of subtype F sequences with the p8.9NSX single-cycle replication assay system. For this reason the subcloning and the single-cycle replication of wild-type subtype F- and patient-derived *RT* only was performed using an alternative vector; pHL(WT). This vector contains full-length HIV genome under HIV LTR promoter with a deleted *env* gene and the *luciferase* reporter gene encoded in the *nef* open reading frame. However, this strategy also produced no detectable virus by infectivity assay.

We speculate that there may be some incompatibility with both the p8.9NSX and pHL(WT) single-cycle replication assays and our subtype F sequences. All *gag-pol* sequences from each viral vector were found to be in the correct reading frame with no obvious sequence abnormalities. It is possible that recombinant pseudoviruses could not be produced by both vector systems due to events in the producer cells. To analyze this further we determined whether transfection of the viral vectors into 293T producer cells resulted in viral protein expression and proteolytic processing using western blotting. This revealed that both subtype F wild-type and patient-derived viral vectors resulted in viral protein expression. However, the analysis also showed that the viral proteins expressed in cells transfected with subtype F vectors had partial proteolytic processing of the viral *gag* polyprotein into its smaller functional subunits with the exception of the patient-derived *MscI* to *ClaI* fragment in the pHL(WT) backbone (Figure 6, lane 4). This suggests that a possible explanation for the reduced or

undetectable virus production with the subtype F vectors could be due to defects in proteolytic processing of viral polyproteins. This process is required for budding and maturation of progeny viral particles that are then released into the culture supernatant. Sequence comparisons of the *pol* region of patient-derived and wild-type subtype F vectors with the wild-type B vector sequences revealed no differences in cleavage sites at the *pol*/*rh* site in RT or at the RT/*IN* site and so this is unlikely to contribute to the defect in polyprotein processing.

Interestingly, almost all published full length sequences of circulating BF recombinant HIV-1 viruses on the Los Alamos Database seem to contain subtype B sequences at the N terminal region of *gag*, a region at the end of *PR* and beginning of *RT* in addition to large portions of *RT* (Thomson et al., 2002; Carr et al., 2001). This suggests that BF recombinants may require subtype B sequences in these regions to facilitate replication and maintain stable recombinants. This may also be a factor contributing to the non-infectious pseudovirions generated with our systems as both p8.9NSX and pHL(WT) vectors are derived from subtype B sequences.

In conclusion we have amplified full length *RT* to investigate mutations in the C terminal end of *RT*. We found eleven mutations (S332T, N348I/L, M357R/T, T386I, N447S, S468R, T470S, D471E, K530R, A554S and I556V) that were found only in the presence of N terminal RTI resistance mutations in our data set. Analysis of additional subtype F sequences from RT-naive and –experienced patients revealed that only two substitutions at position 348 and 357 had statistically different frequencies amongst the two groups of patients. Substitutions at 348 and 357 were found to be more frequent in RTI-experienced and –naive patients respectively. Further investigation of these C terminal mutations was pursued however all wild-type subtype F- and patient-derived viral vectors generated non-infectious pseudovirions. We hypothesise that there may be some incompatibility between our subtype F sequences and the subtype B sequences of both viral vectors used.

CHAPTER 7

General discussion and future work

7.1 Drug resistance in the era of HAART

The primary goal of this study was to investigate the inpatient development and evolution of drug resistance in the full-length HIV-1 *pol* gene. Drug resistance arises in the genes that are targeted for therapy, and in some cases other areas of the viral genomes, resulting in the inhibition of drug action. Mutations conferring resistance to all clinically approved drugs have been identified, and this information is useful for the treatment and clinical management of HIV-1 infection (Menendez-Arias, 2010). However, little is yet known about the interactions between drug resistance mutations located on the three main genes in HIV-1 *pol* targeted by HAART and how this could affect patient treatment outcomes. The establishment of a SGS assay amplifying the full-length *pol* gene in combination with phylogenetic analyses allowed the determination of the dynamics and evolution of drug resistance development in two patients failing RAL salvage therapy. A novel single-replication cycle phenotypic drug susceptibility assay was also developed that allowed the cloning of either patient-derived full-length *pol* or *IN* gene only fragments. This was used to investigate the INSTI susceptibility and viral replicative fitness of the single genomes isolated from one of the patients.

Using these methods, I have shown that the development of drug resistance in the *pol* gene is an extremely dynamic and complex process. It is a fine balance between maintaining adequate viral replicative fitness and attaining high levels of drug resistance within a highly volatile environment of treatment. In our study the evolution of a virus that had both high levels of RAL resistance and high levels of replicative fitness proved to be difficult (see Figure 7.1; upper right quadrant). The exception was the full-length *pol* vector expressing Q148R+G140A in the *IN* gene that is barely situated in the upper right quadrant with relatively moderate levels of resistance and fitness (Figure 7.1). In addition, the analysis suggests that replicative fitness seems to exert more of an influence on the evolution of the virus because all of the primary RAL resistance mutations that dominated at the different RAL treatment periods (Q148R, Y143R and Y143G) had a replicative fitness >50% (Figure 7.1; upper two quadrants) whereas

none of the variants in the lower two quadrants comprised the majority of the population at any time point. This is supported further by the final expansion and emergence of the virus containing the rare Y143G+G163R *IN* mutation, which had a replicative fitness comparable to wild-type virus but lower RAL resistance levels.

Our data also show that analysis of only one part of the genome is not sufficient to gauge the true dynamics in the evolution of drug resistance in the era of HAART. For example, the replicative fitness and susceptibility profiles of the two variants that dominated the first round of RAL-containing salvage therapy (Q148R+G140A and Y134R+G163R) are different when patient-derived *IN* or full-length *pol* fragments are assessed (Figure 7.1). Our data underscore the importance of taking a wider, full-length *pol* approach when investigating resistance to INSTI. We hope that this may provide some insight into devising better treatment regimens and improving the prediction of drug resistance emergence and subsequent treatment failure, specifically during INSTI salvage therapy.

Although our results are from a single patient, they concur with a recent study that showed an effect on viral fitness and susceptibility to EFV and RAL for certain combinations of NNRTI and INSTI resistance mutations (Hu and Kuritzkes, 2012). On the other hand, other studies have shown that mutations in *PR* and *RT* have little effect on the susceptibility to INSTIs but reduce viral replicative fitness of a resistant *IN* gene (Gupta et al., 2009b). Different experimental approaches, such as the use of site-directed mutants compared to patient-derived fragments or differences in the combination of resistance mutations and/or accessory mutations, could explain the contradictory outcomes. Therefore further studies are required to elucidate the interactions between mutations in full-length HIV-1 *pol* gene and their effects on susceptibility and viral fitness.

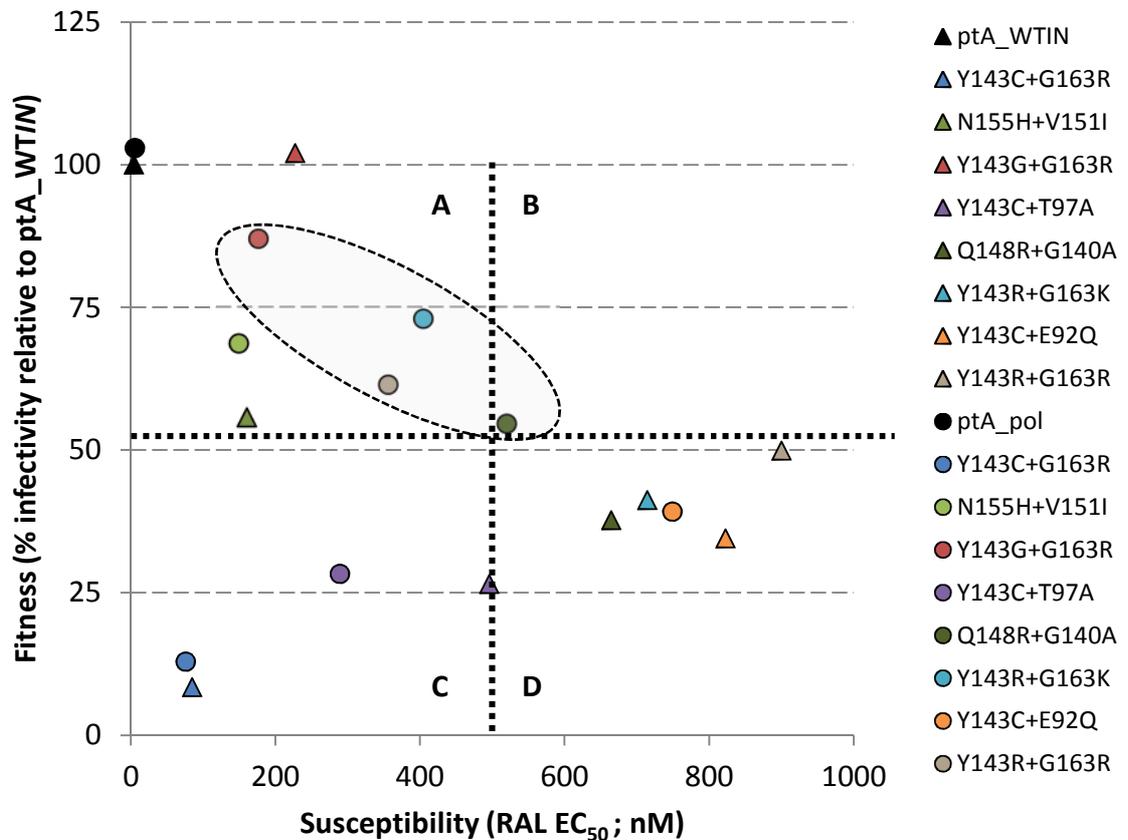


Figure 7.1 Graphical representation of the relationship between the replicative fitness and susceptibility of recombinant viruses from patient A. The graph is equally divided into four hypothetical quadrants: **(A)** high replicative fitness (>50%) and low RAL resistance (<450 nM), **(B)** high replicative fitness (>50%) and RAL resistance (>450 nM), **(C)** low replicative fitness (<50%) and low RAL resistance (<450 nM), **(D)** low replicative fitness (<50%) and high RAL resistance (>450 nM). Triangles represent patient-derived *IN* only vectors; circles represent the respective full length *pol* vectors. The oval represents viruses that dominated the viral population during RAL-containing salvage therapy.

7.1.1 The single genome sequencing assay

The development of the SGS assay was technically challenging; however, we opted to use it rather than cloning because it has less assay-related errors and avoids re-sampling. Developing the SGS assay was challenging in part because the amplification of full-length *pol*, and long templates in general, has previously proved to be difficult (Weber et al., 2011). As a result it is rarely used in the study or routine monitoring of drug resistance. Amplification of long HIV-1 templates can be affected by several factors such as the viral load and RNA degradation (due to sample handling). In addition, the assay is subject to the normal RT-PCR limitations such as sampling bias during viral RNA extraction which can result in approximately 10% of viral genomes being recovered using current HIV RNA extraction techniques. This means to get a good representation of viral diversity in a population a viral load of greater than 10,000 copies/mL is required (Le et al., 2009; Wang et al., 2007; Shafer, 2009). This can be compounded further by primer bias due to the inherent diversity of HIV. These factors may have resulted in the selective amplification of particular variants from each time point and may explain why some resistance mutations amplified in this study were not previously detected (Ferns et al., 2009).

That aside, an advantage of using the SGS assay is that it permits the investigation of the linkage of drug resistance mutations throughout the HIV-1 *pol* gene. In addition, the procedure facilitated the cloning of the isolated genomes into a *gag-pol* vector for use in a phenotypic drug susceptibility assay. These studies showed that resistance mutations in the *PR*, *RT* and *IN* genes are linked. Furthermore, the *PR* and *RT* mutational background did not seem to constrain the development of different INSTI resistant genotypes. These PI and RTI resistance mutations showed no impact on either INSTI susceptibility or replication capacity in the absence of INSTI resistance mutations. However, they were found to have a significant effect on INSTI susceptibility and replication capacity when combined with some coevolved INSTI resistant genotypes which resulted in an increase in the susceptibility to INSTIs and an increase in replication capacity.

Furthermore, the SGS assay enabled the detection of minority species present at <20% of the viral population, the limit of detection of bulk Sanger-based sequencing methods. However, recent technology such as NGS enables the detection of variants down to the 1% level (Gega and Kozal, 2011). This technology was not used here as it compromised the detection of genetic linkage. To date the maximum read length for most NGS technologies is less than 500

bp (Shendure and Ji, 2008). It is likely that as the technology becomes more advanced the amplification of the full-length *pol* in a single fragment will be possible enabling its use to determine mutation linkage and diversity. However, the error rates for NGS assays can be high, up to 1% In contrast, error rates for the SGS assay have been determined to be minimal at 0.011% (Palmer et al., 2005). Of these, recombination errors are of great concern for the analysis of genetic linkage; however, it has been shown that SGS assay-associated recombination is less than one crossover in 66,000 bp analyzed. Besides, recombination errors during SGS can only occur during the cDNA synthesis step as the subsequent dilution of cDNA to a single template precludes recombination between different viral genomes during PCR, unlike cloning techniques. In addition, the use of an RT enzyme with reduced RNase H activity reduces the frequency of template switching during RT and the use of a high fidelity *Taq* DNA polymerase minimises the number of substitutions errors during each amplification step.

7.1.2 Analysis of inpatient viral evolution and the role of recombination

In terms of viral evolution it is important to discuss the role of recombination. Recombination is known to affect the evolution of the virus (Kellam and Larder, 1995; Brown et al., 2011; Mild et al., 2007) therefore may bias the results of this study. Unfortunately due to the similarity of inpatient sequences, detecting recombination has proven to be difficult (Vandamme et al., 2009). Some studies have found obvious traces of recombination between samples obtained from different bodily compartments from the same patient where different evolutionary pressures may be present (Brown et al., 2011; Kemal et al., 2012; Philpott et al., 2005). In addition, recombination between two different viruses in a patient during superinfection has also been relatively easily detected (Kraft et al., 2012; Pacold et al., 2012; Templeton et al., 2009; Koning et al., 2013a). All viral RNA analysed in this study were extracted from plasma samples and there was no evidence of superinfection in these patients. We attempted to detect recombination between the single genome sequences isolated from patient A using split-decomposition networks inferred with Neighbor Net algorithm in SplitsTree software. This suggested the presence of recombinants; however, the detection of putative breakpoints by bootscanning using the Simplot software proved difficult due to the similarity of the sequences (data not shown). Therefore, the role recombination may play on viral evolution in these patients was not fully determined.

7.1.3 Phenotypic drug susceptibility cut offs

The clinical relevance of the RAL and EVG EC₅₀s for each pseudovirus determined in this study, in particular for the rare Y143G+G163R variant, is not known as BCOs were not determined and little data is available for CCOs. However, a BCO for RAL determined by Weber *et al.* using their ViralARTS HIV assay was reported as a fold change of 1.3 (Weber *et al.*, 2013), whilst for the PhenoSense Integrase assay (Monogram Biosciences, Inc.) is a fold change of 1.5 (Vavro *et al.*, 2013). The EVG BCO as determined by the PhenoSense Integrase assay is a fold change of 2.5 (Abram *et al.*, 2012).

What's more, we found that the RAL susceptibility of the rare Y143G+G163R variant was comparable to that of the N155H+V151I variant ($p = 0.22$), which is a well established RAL resistance pattern (Mbisa *et al.*, 2011a). In addition we can ascertain that because the patients were failing on RAL-salvage therapy (determined by their respective viral loads and CD4+ T cell counts) the values seen here are clinically relevant.

7.1.4 Single-cycle replication assay limitations

Whilst *in vitro* assays are useful tools to investigate viral replication and drug susceptibility in a predetermined environment they cannot fully mimic the conditions in a human host as infected cell types and microenvironments vary and can have different selective constraints (Quinones-Mateu and Arts, 2002). For instance, the antiviral activity profile in macrophages can be significantly different compared to CD4+ T lymphocytes (Aquaro *et al.*, 1997). On the contrary, a recent study compared EVG and DTG resistant phenotypes of viral variants in primary macrophages and lymphocytes in which the majority of phenotypes were comparable in both cell types (Canducci *et al.*, 2013).

In addition, significant differences in viral fitness and susceptibility between single- and multi-cycle assays have been described. Dykes *et al.* proposed that this disparity could be due to variations in the late stages of the virus life cycle that are not encompassed in a single cycle assay for example cell to cell transmission (Dykes *et al.*, 2010). Whilst viral multi-cycle growth competition assays are considered to be the gold standard method to measure viral fitness (Weber *et al.*, 2013) we were unable to perform these assays due to laboratory restrictions.

7.2 Drug resistance in the C terminal of *RT*

Recent studies have shown that mutations occurring in the C terminal region of *RT* can affect susceptibility to RTIs alone or in combination with mutations in the N terminal region (Paredes et al., 2011; Gupta et al., 2010b; Yap et al., 2007; Brehm et al., 2007; Sluis-Cremer et al., 2010; Harrigan et al., 2002; Nikolenko et al., 2007; Kemp et al., 1998). As standard genotypic and phenotypic assays typically only include the first 320 residues of *RT* these mutations tend to be missed. What's more, although more than 80% of HIV-1 infections worldwide are attributed to non-B subtype viruses (Hemelaar et al., 2011), studies investigating the presence of resistance associated mutations in the C terminal region have predominantly involved subtype B sequences (Delviks-Frankenberry et al., 2008; Brehm et al., 2007; Kemp et al., 1998; Brehm et al., 2008; Nikolenko et al., 2010; Lengruher et al., 2011).

To further our understanding of resistance associated mutations in the C terminal region of *RT* in a less defined HIV-1 subtype we amplified 27 full-length *RT* viral sequences from children infected with subtype F HIV-1. We identified several novel mutations occurring in the C terminal region of *RT* which were associated with RTI resistance mutations in the N terminal region in our dataset. Unfortunately we were unable to phenotypically characterize these mutations further as the cloning of patient-derived fragments resulted in undetectable virus production.

Nonetheless, as part of my PhD I undertook a visiting fellowship in Dr Pathak's lab at the NCI in Frederick, Maryland, USA. During the visit I learnt how to perform three different assays (described in Chapter 6) that are used to investigate the biochemical effects on *RT* of C terminal region mutations and therefore the mechanisms involved in drug resistance. The intention was to employ these techniques on my return to PHE to elucidate the mechanisms of action for RTI resistant mutations that may have been identified during the investigation of novel C terminal mutations in subtype F, such as the N348L mutation. However, as I was unable to generate any vectors which produced detectable virus expressing patient-derived fragments I didn't get a chance to perform these studies.

7.3 Future work

Continuing the work on INSTI resistance mutations to include the study of replication competent viruses in growth competition assays may prove valuable in understanding the

clinical relevance of minority variants and the linkage of drug resistance mutations whilst also increasing our knowledge of INSTI resistance mutations in general. These methods may provide a more accurate reflection of viral dynamics *in vivo* than single-cycle assays. Other groups have already used growth competition assays to reveal the development of RAL resistance patterns, specifically observing resistance (both primary and secondary) pathway switches (Canducci et al., 2010; Hu and Kuritzkes, 2010; Buzon et al., 2008). To the best of our knowledge growth competition assays determining the effects of secondary mutations on the Y143R/C/ primary resistance mutation have not been carried out, this analysis would therefore help fill the gap in this knowledge. This can be augmented by biochemical assays to determine the effect of the different combinations of primary and accessory mutations on IN activities such as 3' processing and strand transfer.

Our results suggest a discrepancy between amino acid substitutions at IN position 143 with C, G and R in relation to RAL and EVG sensitivity and to some extent the replication capacity of the virus. In addition the results also suggest different effects of the secondary mutations E92Q, T97A and G163R when combined with the Y143C primary INSTI resistance mutation. These differences in susceptibility and replication capacity have already been described for the primary INSTI resistance mutations Q148R/H/K and N155H with their respective secondary mutations (Fransen et al., 2009a; Fun et al., 2010). Investigating the influence of the nature of the substitution at 143 and the effects of different secondary mutations by using different starting concentrations of each mutant and allowing each to compete in the presence and absence of RAL or EVG will allow us to determine which mutant is better evolved to a particular environment.

Further work could also be done using a single-cycle replication assay (Mbisa et al., 2011b; Parry et al., 2009; Gupta et al., 2010a) to determine the effects of patient-derived *IN* only and full-length *pol* on resistance to commonly used PIs and RTIs. One study reported an increase in EFV resistance when NNRTI mutations K103N and E138K were combined with the INSTI resistance mutations Q148H and G140S relative to the NNRTI mutants alone (Hu and Kuritzkes, 2012). This could help to extract further information about the clinical relevance of the genetic linkage of PI, RTI and INSTI resistance mutations.

Work to overcome the problems encountered during the investigation of the role of the C terminal of RT in HIV-1 subtype F could include using an alternative viral vector pseudo-typing system. Using 293T cells that constitutively express the murine leukaemia virus (MLV) envelop, for example Phoenix cells (Pear et al., 1993; Zeilfelder et al., 2007) to determine whether it is

the VSV-G envelop which is incompatible with our wild-type and patient-derived subtype F viral vectors. An alternative strategy to investigate the RTI susceptibility and RT activity of the novel C terminal mutations we identified in our patients, specifically the N348L mutation, would be to use site-directed mutagenesis to introduce N348L into a subtype B background. This however does lose the value of investigating the mutation in the viral mutational context as well as not being able to determine the effect of subtype F polymorphisms and their contribution to the phenotype.

To conclude, as genotypic assays develop, it is likely that whole HIV genome sequencing will become available and this is already being made easier by the advances in NGS technologies. The bigger challenge is to develop phenotypic methods which will complement the generated genotypic data. As evidenced in this study, this would further the knowledge of the development of HIV drug resistance in the era of HAART and help to inform the best way to use the different classes of ARVs.

Reference List

Abbondanzieri,E.A., Bokinsky,G., Rausch,J.W., Zhang,J.X., Le Grice,S.F., and Zhuang,X. (2008). Dynamic binding orientations direct activity of HIV reverse transcriptase. *Nature*. 453, 184-189.

Abram, M. E., Hluhanich, R. M., Goodman, D. D., Andreatta, K. N., Margot, N. A., Ye, L., Cavanaugh, J., Majka, A., Novikov, N., Chehn, X., Svarovskaia, E. S., McColl, D. J., White, K. L., and Miller, M. D. Effect of primary elvitegravir resistance mutations in HIV-1 integrase on drug susceptibility and viral replicative fitness. *Antiviral Therapy* . 2012.

Ref Type: Abstract

Adamson,C.S., Waki,K., Ablan,S.D., Salzwedel,K., and Freed,E.O. (2009). Impact of human immunodeficiency virus type 1 resistance to protease inhibitors on evolution of resistance to the maturation inhibitor bevirimat (PA-457). *J. Virol*. 83, 4884-4894.

Aiken,C., Konner,J., Landau,N.R., Lenburg,M.E., and Trono,D. (1994). Nef induces CD4 endocytosis: requirement for a critical dileucine motif in the membrane-proximal CD4 cytoplasmic domain. *Cell*. 76, 853-864.

Alaeus,A., Lidman,K., Bjorkman,A., Giesecke,J., and Albert,J. (1999). Similar rate of disease progression among individuals infected with HIV-1 genetic subtypes A-D. *AIDS*. 13, 901-907.

Ali,A., Bandaranayake,R.M., Cai,Y., King,N.M., Kolli,M., Mittal,S., Murzycki,J.F., Nalam,M.N., Nalivaika,E.A., Ozen,A., Prabu-Jeyabalan,M.M., Thayer,K., and Schiffer,C.A. (2010). Molecular Basis for Drug Resistance in HIV-1 Protease. *Viruses*. 2, 2509-2535.

Allen,T.M., Altfeld,M., Geer,S.C., Kalife,E.T., Moore,C., O'sullivan,K.M., Desouza,I., Feeney,M.E., Eldridge,R.L., Maier,E.L., Kaufmann,D.E., Lahaie,M.P., Reyor,L., Tanzi,G., Johnston,M.N., Brander,C., Draenert,R., Rockstroh,J.K., Jessen,H., Rosenberg,E.S., Mallal,S.A., and Walker,B.D. (2005). Selective escape from CD8+ T-cell responses represents a major driving force of human immunodeficiency virus type 1 (HIV-1) sequence diversity and reveals constraints on HIV-1 evolution. *J. Virol*. 79, 13239-13249.

Allers,K., Hutter,G., Hofmann,J., Loddenkemper,C., Rieger,K., Thiel,E., and Schneider,T. (2011). Evidence for the cure of HIV infection by CCR5Delta32/Delta32 stem cell transplantation. *Blood*. *117*, 2791-2799.

Aquaro,S., Perno,C.F., Balestra,E., Balzarini,J., Cenci,A., Francesconi,M., Panti,S., Serra,F., Villani,N., and Calio,R. (1997). Inhibition of replication of HIV in primary monocyte/macrophages by different antiviral drugs and comparative efficacy in lymphocytes. *J. Leukoc. Biol.* *62*, 138-143.

Ayala,F.J. (1997). Vagaries of the molecular clock. *Proc. Natl. Acad. Sci. U. S. A.* *94*, 7776-7783.

Baeten,J.M., Chohan,B., Lavreys,L., Chohan,V., McClelland,R.S., Certain,L., Mandaliya,K., Jaoko,W., and Overbaugh,J. (2007). HIV-1 subtype D infection is associated with faster disease progression than subtype A in spite of similar plasma HIV-1 loads. *J. Infect. Dis.* *195*, 1177-1180.

Bainbridge,J.W., Stephens,C., Parsley,K., Demaison,C., Halfyard,A., Thrasher,A.J., and Ali,R.R. (2001). In vivo gene transfer to the mouse eye using an HIV-based lentiviral vector; efficient long-term transduction of corneal endothelium and retinal pigment epithelium. *Gene Ther.* *8*, 1665-1668.

Baldanti,F., Paolucci,S., Gulminetti,R., Brandolini,M., Barbarini,G., and Maserati,R. (2010). Early emergence of raltegravir resistance mutations in patients receiving HAART salvage regimens. *J. Med. Virol.* *82*, 116-122.

Bennett,D.E., Camacho,R.J., Otelea,D., Kuritzkes,D.R., Fleury,H., Kiuchi,M., Heneine,W., Kantor,R., Jordan,M.R., Schapiro,J.M., Vandamme,A.M., Sandstrom,P., Boucher,C.A., van,d., V, Rhee,S.Y., Liu,T.F., Pillay,D., and Shafer,R.W. (2009). Drug resistance mutations for surveillance of transmitted HIV-1 drug-resistance: 2009 update. *PLoS. One.* *4*, e4724.

Benyamini,H., Loyter,A., and Friedler,A. (2011). A structural model of the HIV-1 Rev-integrase complex: the molecular basis of integrase regulation by Rev. *Biochem. Biophys. Res. Commun.* *416*, 252-257.

Berkhout,B. (1999). Proposed alternatives for the use of anti-HIV drugs. *Drug Resist. Updat.* *2*, 69-70.

BHIVA writing group. (2012). BHIVA guidelines for the treatment of HIV-1 positive adults with antiretroviral therapy. *HIV Medicine*.

Blackard,J.T., Renjifo,B., Fawzi,W., Hertzmark,E., Msamanga,G., Mwakagile,D., Hunter,D., Spiegelman,D., Sharghi,N., Kagoma,C., and Essex,M. (2001). HIV-1 LTR subtype and perinatal transmission. *Virology*. 287, 261-265.

Boccaro,F., Lang,S., Meuleman,C., Ederhy,S., Mary-Krause,M., Costagliola,D., Capeau,J., and Cohen,A. (2013). HIV and coronary heart disease: time for a better understanding. *J. Am. Coll. Cardiol*. 61, 511-523.

Bolduan,S., Hubel,P., Reif,T., Lodermeier,V., Hohne,K., Fritz,J.V., Sauter,D., Kirchhoff,F., Fackler,O.T., Schindler,M., and Schubert,U. (2013). HIV-1 Vpu affects the anterograde transport and the glycosylation pattern of NTB-A. *Virology*. 440, 190-203.

Boucher,S., Recordon-Pinson,P., Neau,D., Ragnaud,J.M., Titier,K., Faure,M., Fleury,H., and Masquelier,B. (2005). Clonal analysis of HIV-1 variants in proviral DNA during treatment interruption in patients with multiple therapy failures. *J. Clin. Virol*. 34, 288-294.

Boyd,S.D., Maldarelli,F., Sereti,I., Ouedraogo,G.L., Rehm,C.A., Boltz,V., Shoemaker,D., and Pau,A.K. (2011). Transmitted raltegravir resistance in an HIV-1 CRF_AG-infected patient. *Antivir. Ther*. 16, 257-261.

Boyer,J.C., Bebenek,K., and Kunkel,T.A. (1992). Unequal human immunodeficiency virus type 1 reverse transcriptase error rates with RNA and DNA templates. *Proc. Natl. Acad. Sci. U. S. A*. 89, 6919-6923.

Brehm,J.H., Koontz,D., Meteer,J.D., Pathak,V., Sluis-Cremer,N., and Mellors,J.W. (2007). Selection of mutations in the connection and RNase H domains of human immunodeficiency virus type 1 reverse transcriptase that increase resistance to 3'-azido-3'-dideoxythymidine. *J. Virol*. 81, 7852-7859.

Brehm,J.H., Mellors,J.W., and Sluis-Cremer,N. (2008). Mechanism by which a glutamine to leucine substitution at residue 509 in the ribonuclease H domain of HIV-1 reverse transcriptase confers zidovudine resistance. *Biochemistry*. 47, 14020-14027.

Britten,R.J. (1986). Rates of DNA sequence evolution differ between taxonomic groups. *Science*. 231, 1393-1398.

Brown,P.O., Bowerman,B., Varmus,H.E., and Bishop,J.M. (1989). Retroviral integration: structure of the initial covalent product and its precursor, and a role for the viral IN protein. *Proc. Natl. Acad. Sci. U. S. A.* *86*, 2525-2529.

Brown,R.J., Peters,P.J., Caron,C., Gonzalez-Perez,M.P., Stones,L., Ankghuambom,C., Pondei,K., McClure,C.P., Alemnji,G., Taylor,S., Sharp,P.M., Clapham,P.R., and Ball,J.K. (2011). Intercompartmental recombination of HIV-1 contributes to env intrahost diversity and modulates viral tropism and sensitivity to entry inhibitors. *J. Virol.* *85*, 6024-6037.

Bryant,M. and Ratner,L. (1990). Myristoylation-dependent replication and assembly of human immunodeficiency virus 1. *Proc. Natl. Acad. Sci. U. S. A.* *87*, 523-527.

Bukrinsky,M.I., Sharova,N., McDonald,T.L., Pushkarskaya,T., Tarpley,W.G., and Stevenson,M. (1993a). Association of integrase, matrix, and reverse transcriptase antigens of human immunodeficiency virus type 1 with viral nucleic acids following acute infection. *Proc. Natl. Acad. Sci. U. S. A.* *90*, 6125-6129.

Bukrinsky,M.I., Sharova,N., McDonald,T.L., Pushkarskaya,T., Tarpley,W.G., and Stevenson,M. (1993b). Association of integrase, matrix, and reverse transcriptase antigens of human immunodeficiency virus type 1 with viral nucleic acids following acute infection. *Proc. Natl. Acad. Sci. U. S. A.* *90*, 6125-6129.

Bushman,F.D. (1999). Host proteins in retroviral cDNA integration. *Adv. Virus Res.* *52:301-17.*, 301-317.

Bushman,F.D. and Craigie,R. (1991). Activities of human immunodeficiency virus (HIV) integration protein in vitro: specific cleavage and integration of HIV DNA. *Proc. Natl. Acad. Sci. U. S. A.* *88*, 1339-1343.

Buzon,M.J., Marfil,S., Puertas,M.C., Garcia,E., Clotet,B., Ruiz,L., Blanco,J., Martinez-Picado,J., and Cabrera,C. (2008). Raltegravir susceptibility and fitness progression of HIV type-1 integrase in patients on long-term antiretroviral therapy. *Antivir. Ther.* *13*, 881-893.

Buzon,M.J., Massanella,M., Llibre,J.M., Esteve,A., Dahl,V., Puertas,M.C., Gatell,J.M., Domingo,P., Paredes,R., Sharkey,M., Palmer,S., Stevenson,M., Clotet,B., Blanco,J., and Martinez-Picado,J. (2010). HIV-1 replication and immune dynamics are affected by raltegravir intensification of HAART-suppressed subjects. *Nat. Med.* *16*, 460-465.

Canducci,F., Ceresola,E.R., Saita,D., Castagna,A., Gianotti,N., Underwood,M., Burioni,R., Lazzarin,A., and Clementi,M. (2013). In vitro phenotypes to elvitegravir and dolutegravir in primary macrophages and lymphocytes of clonal recombinant viral variants selected in patients failing raltegravir. *J. Antimicrob. Chemother.* 68, 2525-2532.

Canducci,F., Marinozzi,M.C., Sampaolo,M., Boeri,E., Spagnuolo,V., Gianotti,N., Castagna,A., Paolucci,S., Baldanti,F., Lazzarin,A., and Clementi,M. (2010). Genotypic/phenotypic patterns of HIV-1 integrase resistance to raltegravir. *J. Antimicrob. Chemother.* 65, 425-433.

Carr,J.K., Avila,M., Gomez,C.M., Salomon,H., Hierholzer,J., Watanaveeradej,V., Pando,M.A., Negrete,M., Russell,K.L., Sanchez,J., Birx,D.L., Andrade,R., Vinales,J., and McCutchan,F.E. (2001). Diverse BF recombinants have spread widely since the introduction of HIV-1 into South America. *AIDS.* 15, F41-F47.

Carrillo,A., Stewart,K.D., Sham,H.L., Norbeck,D.W., Kohlbrenner,W.E., Leonard,J.M., Kempf,D.J., and Molla,A. (1998). In vitro selection and characterization of human immunodeficiency virus type 1 variants with increased resistance to ABT-378, a novel protease inhibitor. *J. Virol.* 72, 7532-7541.

Ceccherini-Silberstein, F., Armenia, D., D'Arrigo, R., Micheli, V., Fabeni, L., Meraviglia, P., Capetti, A., Zaccarelli, M., Trotta, M. P., Narciso, P., Antinori, A., and Perno, C. F. Virological response and resistance in multi-experienced patients treated with raltegravir. *Antivir.Ther.* 2008.

Ref Type: Abstract

Ceccherini-Silberstein, F., Malet, I., Fabeni, L., Svicher, V., Gori, C., Dimonte, S., Bono, S., Artese, A., D'Arrigo, R., Katlama, C., Antinori, A., d'Arminio Monforte, A., Calvez, V., Marcelin, A. G., and Perno, C. F. Specific mutations related to resistance to HIV-1 integrase inhibitors are associated with reverse transcriptase mutations in HAART-treated patients. *Antivir.Ther.* 2007.

Ref Type: Abstract

Chakraborty,A., Sun,G.Q., Mustavich,L., Huang,S.H., and Li,B.L. (2013). Biochemical interactions between HIV-1 integrase and reverse transcriptase. *FEBS Lett.* 587, 425-429.

Charpentier,C., Dwyer,D.E., Mammano,F., Lecossier,D., Clavel,F., and Hance,A.J. (2004). Role of minority populations of human immunodeficiency virus type 1 in the evolution of viral resistance to protease inhibitors. *J. Virol.* 78, 4234-4247.

Charpentier,C., Karmochkine,M., Laureillard,D., Tisserand,P., Belec,L., Weiss,L., Si-Mohamed,A., and Piketty,C. (2008). Drug resistance profiles for the HIV integrase gene in patients failing raltegravir salvage therapy. *HIV. Med.* 9, 765-770.

Checroune,F., Yao,X.J., Gottlinger,H.G., Bergeron,D., and Cohen,E.A. (1995). Incorporation of Vpr into human immunodeficiency virus type 1: role of conserved regions within the P6 domain of Pr55gag. *J. Acquir. Immune. Defic. Syndr. Hum. Retrovirol.* 10, 1-7.

Chen,J.C., Krucinski,J., Miercke,L.J., Finer-Moore,J.S., Tang,A.H., Leavitt,A.D., and Stroud,R.M. (2000). Crystal structure of the HIV-1 integrase catalytic core and C-terminal domains: a model for viral DNA binding. *Proc. Natl. Acad. Sci. U. S. A.* 97, 8233-8238.

Chen,L., Perlina,A., and Lee,C.J. (2004). Positive selection detection in 40,000 human immunodeficiency virus (HIV) type 1 sequences automatically identifies drug resistance and positive fitness mutations in HIV protease and reverse transcriptase. *J. Virol.* 78, 3722-3732.

Cherepanov,P., Maertens,G., Proost,P., Devreese,B., Van,B.J., Engelborghs,Y., De,C.E., and Debysers,Z. (2003). HIV-1 integrase forms stable tetramers and associates with LEDGF/p75 protein in human cells. *J. Biol Chem.* 278, 372-381.

Christ,F., Voet,A., Marchand,A., Nicolet,S., Desimmie,B.A., Marchand,D., Bardiot,D., Van der Veken,N.J., Van,R.B., Strelkov,S.V., De,M.M., Chaltin,P., and Debysers,Z. (2010). Rational design of small-molecule inhibitors of the LEDGF/p75-integrase interaction and HIV replication. *Nat. Chem. Biol.* 6, 442-448.

Chun,T.W., Carruth,L., Finzi,D., Shen,X., DiGiuseppe,J.A., Taylor,H., Hermankova,M., Chadwick,K., Margolick,J., Quinn,T.C., Kuo,Y.H., Brookmeyer,R., Zeiger,M.A., Barditch-Crovo,P., and Siliciano,R.F. (1997a). Quantification of latent tissue reservoirs and total body viral load in HIV-1 infection. *Nature.* 387, 183-188.

Chun,T.W., Davey,R.T., Jr., Ostrowski,M., Shawn,J.J., Engel,D., Mullins,J.I., and Fauci,A.S. (2000). Relationship between pre-existing viral reservoirs and the re-emergence of plasma viremia after discontinuation of highly active anti-retroviral therapy. *Nat. Med.* 6, 757-761.

Chun,T.W., Stuyver,L., Mizell,S.B., Ehler,L.A., Mican,J.A., Baseler,M., Lloyd,A.L., Nowak,M.A., and Fauci,A.S. (1997b). Presence of an inducible HIV-1 latent reservoir during highly active antiretroviral therapy. *Proc. Natl. Acad. Sci. U. S. A.* 94, 13193-13197.

- Cichutek,K., Merget,H., Norley,S., Linde,R., Kreuz,W., Gahr,M., and Kurth,R. (1992). Development of a quasispecies of human immunodeficiency virus type 1 in vivo. *Proc. Natl. Acad. Sci. U. S. A.* *89*, 7365-7369.
- Ciuffi,A., Llano,M., Poeschla,E., Hoffmann,C., Leipzig,J., Shinn,P., Ecker,J.R., and Bushman,F. (2005). A role for LEDGF/p75 in targeting HIV DNA integration. *Nat. Med.* *11*, 1287-1289.
- Clavel,F., Hoggan,M.D., Willey,R.L., Strebel,K., Martin,M.A., and Repaske,R. (1989). Genetic recombination of human immunodeficiency virus. *J. Virol.* *63*, 1455-1459.
- Clavel,F. and Mammano,F. (2010). Role of Gag in HIV Resistance to Protease Inhibitors. *Viruses.* *2*, 1411-1426.
- Coffin,J.M., Hughes,S.H., and Varmus,H.E. (1997). *Retroviruses.*
- Cohen,C., Elion,R., Ruane,P., Shambraw,D., DeJesus,E., Rashbaum,B., Chuck,S.L., Yale,K., Liu,H.C., Warren,D.R., Ramanathan,S., and Kearney,B.P. (2011). Randomized, phase 2 evaluation of two single-tablet regimens elvitegravir/cobicistat/emtricitabine/tenofovir disoproxil fumarate versus efavirenz/emtricitabine/tenofovir disoproxil fumarate for the initial treatment of HIV infection. *AIDS.* *25*, F7-12.
- Coiras,M., Lopez-Huertas,M.R., Perez-Olmeda,M., and Alcami,J. (2009). Understanding HIV-1 latency provides clues for the eradication of long-term reservoirs. *Nat. Rev. Microbiol.* *7*, 798-812.
- Colgrove,R.C., Pitt,J., Chung,P.H., Welles,S.L., and Japour,A.J. (1998). Selective vertical transmission of HIV-1 antiretroviral resistance mutations. *AIDS.* *12*, 2281-2288.
- Collins,K.L., Chen,B.K., Kalams,S.A., Walker,B.D., and Baltimore,D. (1998). HIV-1 Nef protein protects infected primary cells against killing by cytotoxic T lymphocytes. *Nature.* *391*, 397-401.
- Conlon,C.P., Klenerman,P., Edwards,A., Larder,B.A., and Phillips,R.E. (1994). Heterosexual transmission of human immunodeficiency virus type 1 variants associated with zidovudine resistance. *J. Infect. Dis.* *169*, 411-415.

Conway,B., Montessori,V., Rouleau,D., Montaner,J.S., O'Shaughnessy,M.V., Fransen,S., Shillington,A., Weislow,O., and Mayers,D.L. (1999). Primary lamivudine resistance in acute/early human immunodeficiency virus infection. *Clin. Infect. Dis.* 28, 910-911.

Cooper,D.A., Steigbigel,R.T., Gatell,J.M., Rockstroh,J.K., Katlama,C., Yeni,P., Lazzarin,A., Clotet,B., Kumar,P.N., Eron,J.E., Schechter,M., Markowitz,M., Loutfy,M.R., Lennox,J.L., Zhao,J., Chen,J., Ryan,D.M., Rhodes,R.R., Killar,J.A., Gilde,L.R., Strohmaier,K.M., Meibohm,A.R., Miller,M.D., Hazuda,D.J., Nessly,M.L., DiNubile,M.J., Isaacs,R.D., Teppler,H., and Nguyen,B.Y. (2008). Subgroup and resistance analyses of raltegravir for resistant HIV-1 infection. *N. Engl. J. Med.* 359, 355-365.

Croteau,D., Letendre,S., Best,B.M., Ellis,R.J., Breidinger,S., Clifford,D., Collier,A., Gelman,B., Marra,C., Mbeo,G., McCutchan,A., Morgello,S., Simpson,D., Way,L., Vaida,F., Ueland,S., Capparelli,E., and Grant,I. (2010). Total Raltegravir Concentrations in Cerebrospinal Fluid Exceed the 50-Percent Inhibitory Concentration for Wild-Type HIV-1. *Antimicrob. Agents Chemother.* 54, 5156-5160.

Dalglish,A.G., Beverley,P.C., Clapham,P.R., Crawford,D.H., Greaves,M.F., and Weiss,R.A. (1984). The CD4 (T4) antigen is an essential component of the receptor for the AIDS retrovirus. *Nature.* 312, 763-767.

Das,K., Sarafianos,S.G., Clark,A.D., Jr., Boyer,P.L., Hughes,S.H., and Arnold,E. (2007). Crystal structures of clinically relevant Lys103Asn/Tyr181Cys double mutant HIV-1 reverse transcriptase in complexes with ATP and non-nucleoside inhibitor HBY 097. *J. Mol. Biol.* 365, 77-89.

daSilva,L.L., Sougrat,R., Burgos,P.V., Janvier,K., Mattera,R., and Bonifacino,J.S. (2009). Human immunodeficiency virus type 1 Nef protein targets CD4 to the multivesicular body pathway. *J. Virol.* 83, 6578-6590.

Dau,B., Ayers,D., Singer,J., Harrigan,P.R., Brown,S., Kyriakides,T., Cameron,D.W., Angus,B., and Holodniy,M. (2010). Connection domain mutations in treatment-experienced patients in the OPTIMA trial. *J. Acquir. Immune. Defic. Syndr.* 54, 160-166.

Dawson,L. and Yu,X.F. (1998). The role of nucleocapsid of HIV-1 in virus assembly. *Virology.* 251, 141-157.

Dayton,A.I., Sodroski,J.G., Rosen,C.A., Goh,W.C., and Haseltine,W.A. (1986). The trans-activator gene of the human T cell lymphotropic virus type III is required for replication. *Cell*. *44*, 941-947.

de,R.A., Schuurman,R., Goudsmit,J., van den Hoek,A., and Boucher,C. (1996). First case of new infection with zidovudine-resistant HIV-1 among prospectively studied intravenous drug users and homosexual men in Amsterdam, The Netherlands. *AIDS*. *10*, 231-232.

Deeks,S.G., Wrin,T., Liegler,T., Hoh,R., Hayden,M., Barbour,J.D., Hellmann,N.S., Petropoulos,C.J., McCune,J.M., Hellerstein,M.K., and Grant,R.M. (2001). Virologic and immunologic consequences of discontinuing combination antiretroviral-drug therapy in HIV-infected patients with detectable viremia. *N. Engl. J. Med.* *344*, 472-480.

Delelis,O., Malet,I., Na,L., Tchertanov,L., Calvez,V., Marcelin,A.G., Subra,F., Deprez,E., and Mouscadet,J.F. (2009). The G140S mutation in HIV integrases from raltegravir-resistant patients rescues catalytic defect due to the resistance Q148H mutation. *Nucleic Acids Res.* *37*, 1193-1201.

Delelis,O., Thierry,S., Subra,F., Simon,F., Malet,I., Alloui,C., Sayon,S., Calvez,V., Deprez,E., Marcelin,A.G., Tchertanov,L., and Mouscadet,J.F. (2010). Impact of Y143 HIV-1 integrase mutations on resistance to raltegravir in vitro and in vivo. *Antimicrob. Agents Chemother.* *54*, 491-501.

Delviks-Frankenberry,K.A., Lengruher,R.B., Santos,A.F., Silveira,J.M., Soares,M.A., Kearney,M.F., Maldarelli,F., and Pathak,V.K. (2013). Connection subdomain mutations in HIV-1 subtype-C treatment-experienced patients enhance NRTI and NNRTI drug resistance. *Virology*. *435*, 433-441.

Delviks-Frankenberry,K.A., Nikolenko,G.N., Boyer,P.L., Hughes,S.H., Coffin,J.M., Jere,A., and Pathak,V.K. (2008). HIV-1 reverse transcriptase connection subdomain mutations reduce template RNA degradation and enhance AZT excision. *Proc. Natl. Acad. Sci. U. S. A.* *105*, 10943-10948.

Delviks-Frankenberry,K.A., Nikolenko,G.N., Maldarelli,F., Hase,S., Takebe,Y., and Pathak,V.K. (2009). Subtype-specific differences in the human immunodeficiency virus type 1 reverse transcriptase connection subdomain of CRF01_AE are associated with higher levels of resistance to 3'-azido-3'-deoxythymidine. *J. Virol.* *83*, 8502-8513.

Delviks-Frankenberry,K.A., Nikolenko,G.N., and Pathak,V.K. (2010). The "Connection" Between HIV Drug Resistance and RNase H. *Viruses* 2, 1476-1503.

Deng,H., Liu,R., Ellmeier,W., Choe,S., Unutmaz,D., Burkhart,M., Di,M.P., Marmon,S., Sutton,R.E., Hill,C.M., Davis,C.B., Peiper,S.C., Schall,T.J., Littman,D.R., and Landau,N.R. (1996). Identification of a major co-receptor for primary isolates of HIV-1. *Nature*. %20;381, 661-666.

Descamps,D., Collin,G., Letourneur,F., Apetrei,C., Damond,F., Loussert-Ajaka,I., Simon,F., Saragosti,S., and Brun-Vezinet,F. (1997). Susceptibility of human immunodeficiency virus type 1 group O isolates to antiretroviral agents: in vitro phenotypic and genotypic analyses. *J. Virol.* 71, 8893-8898.

Deval,J., Selmi,B., Boretto,J., Egloff,M.P., Guerreiro,C., Sarfati,S., and Canard,B. (2002). The molecular mechanism of multidrug resistance by the Q151M human immunodeficiency virus type 1 reverse transcriptase and its suppression using alpha-boranophosphate nucleotide analogues. *J. Biol. Chem.* 277, 42097-42104.

DiGiusto,D.L., Krishnan,A., Li,L., Li,H., Li,S., Rao,A., Mi,S., Yam,P., Stinson,S., Kalos,M., Alvarnas,J., Lacey,S.F., Yee,J.K., Li,M., Couture,L., Hsu,D., Forman,S.J., Rossi,J.J., and Zaia,J.A. (2010). RNA-based gene therapy for HIV with lentiviral vector-modified CD34(+) cells in patients undergoing transplantation for AIDS-related lymphoma. *Sci. Transl. Med.* 2, 36ra43.

Divita,G., Rittinger,K., Geourjon,C., Deleage,G., and Goody,R.S. (1995). Dimerization kinetics of HIV-1 and HIV-2 reverse transcriptase: a two step process. *J. Mol. Biol.* 245, 508-521.

Doehle,B.P., Chang,K., Rustagi,A., McNevin,J., McElrath,M.J., and Gale,M., Jr. (2012). Vpu mediates depletion of interferon regulatory factor 3 during HIV infection by a lysosome-dependent mechanism. *J. Virol.* 86, 8367-8374.

Dorr,P., Westby,M., Dobbs,S., Griffin,P., Irvine,B., Macartney,M., Mori,J., Rickett,G., Smith-Burchnell,C., Napier,C., Webster,R., Armour,D., Price,D., Stammen,B., Wood,A., and Perros,M. (2005). Maraviroc (UK-427,857), a potent, orally bioavailable, and selective small-molecule inhibitor of chemokine receptor CCR5 with broad-spectrum anti-human immunodeficiency virus type 1 activity. *Antimicrob. Agents Chemother.* 49, 4721-4732.

- Doyon,L., Croteau,G., Thibeault,D., Poulin,F., Pilote,L., and Lamarre,D. (1996). Second locus involved in human immunodeficiency virus type 1 resistance to protease inhibitors. *J. Virol.* *70*, 3763-3769.
- Drummond,A.J., Pybus,O.G., Rambaut,A., Forsberg,R., and Rodrigo,A.G. (2003). Measurably evolving populations. *Trends Ecol. Evol.* 481-488.
- Drummond,A.J. and Rambaut,A. (2007). BEAST: Bayesian evolutionary analysis by sampling trees. *BMC. Evol. Biol.* *7*:214., 214.
- Drummond,A.J., Rambaut,A., Shapiro,B., and Pybus,O.G. (2005). Bayesian coalescent inference of past population dynamics from molecular sequences. *Mol. Biol Evol.* *22*, 1185-1192.
- Dudley,D.M., Gao,Y., Nelson,K.N., Henry,K.R., Nankya,I., Gibson,R.M., and Arts,E.J. (2009). A novel yeast-based recombination method to clone and propagate diverse HIV-1 isolates. *Biotechniques.* *46*, 458-467.
- Dussupt,V., Javid,M.P., Abou-Jaoude,G., Jadwin,J.A., de La,C.J., Nagashima,K., and Bouamr,F. (2009). The nucleocapsid region of HIV-1 Gag cooperates with the PTAP and LYPXnL late domains to recruit the cellular machinery necessary for viral budding. *PLoS. Pathog.* *5*, e1000339.
- Dyda,F., Hickman,A.B., Jenkins,T.M., Engelman,A., Craigie,R., and Davies,D.R. (1994). Crystal structure of the catalytic domain of HIV-1 integrase: similarity to other polynucleotidyl transferases. *Science.* *266*, 1981-1986.
- Dykes,C., Wu,H., Sims,M., Holden-Wiltse,J., and Demeter,L.M. (2010). Human immunodeficiency virus type 1 protease inhibitor drug-resistant mutants give discordant results when compared in single-cycle and multiple-cycle fitness assays. *J. Clin. Microbiol.* *48*, 4035-4043.
- Edwards,C.T., Holmes,E.C., Wilson,D.J., Viscidi,R.P., Abrams,E.J., Phillips,R.E., and Drummond,A.J. (2006). Population genetic estimation of the loss of genetic diversity during horizontal transmission of HIV-1. *BMC. Evol. Biol.* *6*:28., 28.
- Ehteshami,M., Beilhartz,G.L., Scarth,B.J., Tchesnokov,E.P., McCormick,S., Wynhoven,B., Harrigan,P.R., and Gotte,M. (2008). Connection domain mutations N348I and A360V in HIV-1

reverse transcriptase enhance resistance to 3'-azido-3'-deoxythymidine through both RNase H-dependent and -independent mechanisms. *J. Biol. Chem.* **283**, 22222-22232.

Ehteshami,M. and Gotte,M. (2008). Effects of mutations in the connection and RNase H domains of HIV-1 reverse transcriptase on drug susceptibility. *AIDS Rev.* **10**, 224-235.

Elena,S.F. and Sanjuan,R. (2007). Virus Evolution: Insights from an Experimental Approach. *The Annual Review of Ecology, Evolution, and Systematics* **38**, 27-52.

Engelman,A., Hickman,A.B., and Craigie,R. (1994). The core and carboxyl-terminal domains of the integrase protein of human immunodeficiency virus type 1 each contribute to nonspecific DNA binding. *J. Virol.* **68**, 5911-5917.

Erice,A., Mayers,D.L., Strike,D.G., Sannerud,K.J., McCutchan,F.E., Henry,K., and Balfour,H.H., Jr. (1993). Brief report: primary infection with zidovudine-resistant human immunodeficiency virus type 1. *N. Engl. J. Med.* **328**, 1163-1165.

Esposito,F., Corona,A., and Tramontano,E. (2012). HIV-1 Reverse Transcriptase Still Remains a New Drug Target: Structure, Function, Classical Inhibitors, and New Inhibitors with Innovative Mechanisms of Actions. *Mol. Biol Int.* **2012:586401**. doi: 10.1155/2012/586401. Epub;2012 Jun;20., 586401.

Farnet,C.M. and Haseltine,W.A. (1991). Determination of viral proteins present in the human immunodeficiency virus type 1 preintegration complex. *J. Virol.* **65**, 1910-1915.

Fassati,A. (2012). Multiple roles of the capsid protein in the early steps of HIV-1 infection. *Virus Res.* **170**, 15-24.

Fassati,A. and Goff,S.P. (2001). Characterization of intracellular reverse transcription complexes of human immunodeficiency virus type 1. *J. Virol.* **75**, 3626-3635.

FDA. FDA approves new drug to treat HIV infection. 12-8-2013.

Ref Type: Online Source

Felsenstein,J. (1973). Maximum-likelihood estimation of evolutionary trees from continuous characters. *Am. J. Hum. Genet.* **25**, 471-492.

Felsenstein,J. (1981). Evolutionary trees from DNA sequences: a maximum likelihood approach. *J. Mol. Evol.* **17**, 368-376.

Feng,J.Y., Myrick,F.T., Margot,N.A., Mulamba,G.B., Rimsky,L., Borroto-Esoda,K., Selmi,B., and Canard,B. (2006). Virologic and enzymatic studies revealing the mechanism of K65R- and Q151M-associated HIV-1 drug resistance towards emtricitabine and lamivudine. *Nucleosides Nucleotides Nucleic Acids.* 25, 89-107.

Ferns,R.B., Kirk,S., Bennett,J., Williams,I., Edwards,S., and Pillay,D. (2009). The dynamics of appearance and disappearance of HIV-1 integrase mutations during and after withdrawal of raltegravir therapy. *AIDS.* 23, 2159-2164.

Finzi,D., Hermankova,M., Pierson,T., Carruth,L.M., Buck,C., Chaisson,R.E., Quinn,T.C., Chadwick,K., Margolick,J., Brookmeyer,R., Gallant,J., Markowitz,M., Ho,D.D., Richman,D.D., and Siliciano,R.F. (1997). Identification of a reservoir for HIV-1 in patients on highly active antiretroviral therapy. *Science.* 278, 1295-1300.

Fisher,A.G., Feinberg,M.B., Josephs,S.F., Harper,M.E., Marselle,L.M., Reyes,G., Gonda,M.A., Aldovini,A., Debouk,C., Gallo,R.C., and . (1986). The trans-activator gene of HTLV-III is essential for virus replication. *Nature.* 320, 367-371.

Flexner,C. and Saag,M. (2013). The antiretroviral drug pipeline: prospects and implications for future treatment research. *Curr. Opin. HIV. AIDS.* 8, 572-578.

Flynn,N.M., Forthal,D.N., Harro,C.D., Judson,F.N., Mayer,K.H., and Para,M.F. (2005). Placebo-controlled phase 3 trial of a recombinant glycoprotein 120 vaccine to prevent HIV-1 infection. *J. Infect. Dis.* 191, 654-665.

Franke,E.K., Yuan,H.E., Bossolt,K.L., Goff,S.P., and Luban,J. (1994). Specificity and sequence requirements for interactions between various retroviral Gag proteins. *J. Virol.* 68, 5300-5305.

Fransen, S, Gupta, S., Paxinos, E. E., Huang, W., Parkin, N., Miller, M., Hazuda, D., and Petropoulos, C. J. Integrase (IN) Inhibitor Susceptibility Can Be Measured Using Recombinant Viruses That Express Patient Virus IN Alone, or in Combination with Protease (PR) and Reverse Transcriptase (RT). 12th Conference on Retroviruses and Opportunistic Infections. 2005.

Ref Type: Conference Proceeding

Fransen, S, Young, B, Frantzell, a., and et al. Control of viral replication following transmission of HIV-1 exhibiting resistance to reverse transcriptase, protease and integrase inhibitors. *Antivir.Ther.* 2010.

Ref Type: Abstract

Fransen,S., Gupta,S., Danovich,R., Hazuda,D., Miller,M., Witmer,M., Petropoulos,C.J., and Huang,W. (2009a). Loss of raltegravir susceptibility by human immunodeficiency virus type 1 is conferred via multiple nonoverlapping genetic pathways. *J. Virol.* *83*, 11440-11446.

Fransen,S., Gupta,S., Frantzell,a., Petropoulos,C.J., and Huang,W. (2012). Substitutions at amino acid positions 143, 148, and 155 of HIV-1 integrase define distinct genetic barriers to raltegravir resistance in vivo. *J. Virol.* *86*, 7249-7255.

Fransen,S., Karmochkine,M., Huang,W., Weiss,L., Petropoulos,C.J., and Charpentier,C. (2009b). Longitudinal analysis of raltegravir susceptibility and integrase replication capacity of human immunodeficiency virus type 1 during virologic failure. *Antimicrob. Agents Chemother.* *53*, 4522-4524.

Frater,A.J., Beardall,A., Ariyoshi,K., Churchill,D., Galpin,S., Clarke,J.R., Weber,J.N., and McClure,M.O. (2001). Impact of baseline polymorphisms in RT and protease on outcome of highly active antiretroviral therapy in HIV-1-infected African patients. *AIDS.* *15*, 1493-1502.

Freed,E.O. (2001). HIV-1 replication. *Somat. Cell Mol. Genet.* *26*, 13-33.

Frost,S.D.W., Nijhuis,M., Schuurman,R., Boucher,C.A.B., and Brown,A.J.L. (2000). Evolution of Lamivudine Resistance in Human Immunodeficiency Virus Type 1-Infected Individuals: the Relative Roles of Drift and Selection. *J. Virol.* *74*, 6262-6268.

Fun,A., Van,B.K., van Lelyveld,S.F., Schipper,P.J., Stuyver,L.J., Wensing,A.M., and Nijhuis,M. (2010). Mutation Q95K enhances N155H-mediated integrase inhibitor resistance and improves viral replication capacity. *J. Antimicrob. Chemother.* *65*, 2300-2304.

Fun,A., Wensing,A.M., Verheyen,J., and Nijhuis,M. (2012). Human Immunodeficiency Virus Gag and protease: partners in resistance. *Retrovirology.* *9:63. doi: 10.1186/1742-4690-9-63.*, 63-69.

Galli,A., Kearney,M., Nikolaitchik,O.A., Yu,S., Chin,M.P.S., Maldarelli,F., Coffin,J.M., Pathak,V.K., and Hu,W.S. (2010). Patterns of Human Immunodeficiency Virus Type 1 Recombination Ex Vivo Provide Evidence for Coadaptation of Distant Sites, Resulting in Purifying Selection for Intersubtype Recombinants during Replication. *J. Virol.* *84*, 7651-7661.

Gamble,T.R., Vajdos,F.F., Yoo,S., Worthylake,D.K., Houseweart,M., Sundquist,W.I., and Hill,C.P. (1996). Crystal structure of human cyclophilin A bound to the amino-terminal domain of HIV-1 capsid. *Cell*. 87, 1285-1294.

Gandhi,R.T., Coombs,R.W., Chan,E.S., Bosch,R.J., Zheng,L., Margolis,D.M., Read,S., Kallungal,B., Chang,M., Goecker,E.A., Wiegand,A., Kearney,M., Jacobson,J.M., D'Aquila,R., Lederman,M.M., Mellors,J.W., and Eron,J.J. (2012). No effect of raltegravir intensification on viral replication markers in the blood of HIV-1-infected patients receiving antiretroviral therapy. *J. Acquir. Immune. Defic. Syndr.* 59, 229-235.

Gao,F., Chen,Y., Levy,D.N., Conway,J.A., Kepler,T.B., and Hui,H. (2004). Unselected mutations in the human immunodeficiency virus type 1 genome are mostly nonsynonymous and often deleterious. *J. Virol.* 78, 2426-2433.

Gao,H.Q., Boyer,P.L., Sarafianos,S.G., Arnold,E., and Hughes,S.H. (2000). The role of steric hindrance in 3TC resistance of human immunodeficiency virus type-1 reverse transcriptase. *J. Mol. Biol.* 300, 403-418.

Garrido,C., Villacian,J., Zahonero,N., Pattery,T., Garcia,F., Gutierrez,F., Caballero,E., Van,H.M., Soriano,V., and De,M.C. (2012). Broad phenotypic cross-resistance to elvitegravir in HIV-1 infected patients failing on raltegravir-containing regimens. *Antimicrob. Agents Chemother.* 56, 2873-2878.

Gega, A. and Kozal, M. J. New Technology to Detect Low-level Drug-resistant HIV Variants. *Future Virology* 6[1], 17-26. 2011.

Ref Type: Abstract

Geretti,A.M. (2006). HIV-1 subtypes: epidemiology and significance for HIV management. *Curr. Opin. Infect. Dis.* 19, 1-7.

Gianotti,N., Galli,L., Boeri,E., Maillard,M., Serra,G., Ratti,D., Gallotta,G., Vacchini,D., Tremolada,Y., Lazzarin,A., Clementi,M., and Castagna,A. (2005). In vivo dynamics of the K103N mutation following the withdrawal of non-nucleoside reverse transcriptase inhibitors in Human Immunodeficiency Virus-infected patients. *New Microbiol.* 28, 319-326.

Gifford,R.J., de,O.T., Rambaut,A., Pybus,O.G., Dunn,D., Vandamme,A.M., Kellam,P., and Pillay,D. (2007). Phylogenetic surveillance of viral genetic diversity and the evolving molecular epidemiology of human immunodeficiency virus type 1. *J. Virol.* *81*, 13050-13056.

Gilead. FDA Advisory Committee Supports Approval of Gilead's Once-Daily Quad Single Tablet Regimen for HIV. 11-5-2012.

Ref Type: Online Source

Goethals,O., Clayton,R., Van,G.M., Vereycken,I., Wagemans,E., Geluykens,P., Dockx,K., Strijbos,R., Smits,V., Vos,A., Meersseman,G., Jochmans,D., Vermeire,K., Schols,D., Hallenberger,S., and Hertogs,K. (2008). Resistance mutations in human immunodeficiency virus type 1 integrase selected with elvitegravir confer reduced susceptibility to a wide range of integrase inhibitors. *J. Virol.* *82*, 10366-10374.

Goethals,O., Vos,A., Van,G.M., Geluykens,P., Smits,V., Schols,D., Hertogs,K., and Clayton,R. (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.* *402*, 338-346.

Gojobori,T., Moriyama,E.N., and Kimura,M. (1990). Molecular clock of viral evolution, and the neutral theory. *Proc. Natl. Acad. Sci. U. S. A.* *87*, 10015-10018.

Goody,R.S., Muller,B., and Restle,T. (1991). Factors contributing to the inhibition of HIV reverse transcriptase by chain-terminating nucleotides in vitro and in vivo. *FEBS Lett.* *291*, 1-5.

Gotte,M. (2007). Should we include connection domain mutations of HIV-1 reverse transcriptase in HIV resistance testing. *PLoS. Med.* *4*, e346.

Gottlinger,H.G., Dorfman,T., Cohen,E.A., and Haseltine,W.A. (1993). Vpu protein of human immunodeficiency virus type 1 enhances the release of capsids produced by gag gene constructs of widely divergent retroviruses. *Proc. Natl. Acad. Sci. U. S. A.* *90*, 7381-7385.

Gottlinger,H.G., Dorfman,T., Sodroski,J.G., and Haseltine,W.A. (1991). Effect of mutations affecting the p6 gag protein on human immunodeficiency virus particle release. *Proc. Natl. Acad. Sci. U. S. A.* *88*, 3195-3199.

Grenfell,B.T., Pybus,O.G., Gog,J.R., Wood,J.L., Daly,J.M., Mumford,J.A., and Holmes,E.C. (2004). Unifying the epidemiological and evolutionary dynamics of pathogens. *Science.* *303*, 327-332.

Grobler,J.A., Stillmock,K., Hu,B., Witmer,M., Felock,P., Espeseth,A.S., Wolfe,A., Egbertson,M., Bourgeois,M., Melamed,J., Wai,J.S., Young,S., Vacca,J., and Hazuda,D.J. (2002). Diketo acid inhibitor mechanism and HIV-1 integrase: implications for metal binding in the active site of phosphotransferase enzymes. *Proc. Natl. Acad. Sci. U. S. A.* 99, 6661-6666.

Grobler, J. A., Stillmock, K., Miller, M. D., and Hazuda, D. Mechanism by which the HIV Integrase Active Site Mutation N155H Confers Resistance to Raltegravir. *Antivir. Ther.* 2008.

Ref Type: Abstract

Grossman,Z., Vardinon,N., Chemtob,D., Alkan,M.L., Bentwich,Z., Burke,M., Gottesman,G., Istomin,V., Levi,I., Maayan,S., Shahar,E., and Schapiro,J.M. (2001). Genotypic variation of HIV-1 reverse transcriptase and protease: comparative analysis of clade C and clade B. *AIDS.* 15, 1453-1460.

Guindon,S. and Gascuel,O. (2003). A simple, fast, and accurate algorithm to estimate large phylogenies by maximum likelihood. *Syst. Biol.* 52, 696-704.

Gupta,R.K., Gibb,D.M., and Pillay,D. (2009a). Management of paediatric HIV-1 resistance. *Curr. Opin. Infect. Dis.* 22, 256-263.

Gupta,R.K., Kohli,A., McCormick,A.L., Towers,G.J., Pillay,D., and Parry,C.M. (2010a). Full-length HIV-1 Gag determines protease inhibitor susceptibility within in vitro assays. *AIDS.* 24, 1651-1655.

Gupta, S., Fransen, S., Frantzell, A., Chappey, C., Petropoulos, C. J., and Huang, W. Combinations of Primary NNRTI and INI Resistance Mutations do not Alter HIV-1 Drug Susceptibility but Impair Replication Capacity. 16th Conference on Retroviruses and Opportunistic Infections. 2009b.

Ref Type: Conference Proceeding

Gupta,S., Fransen,S., Paxinos,E.E., Stawiski,E., Huang,W., and Petropoulos,C.J. (2010b). Combinations of mutations in the connection domain of human immunodeficiency virus type 1 reverse transcriptase: assessing the impact on nucleoside and nonnucleoside reverse transcriptase inhibitor resistance. *Antimicrob. Agents Chemother.* 54, 1973-1980.

Hachiya,A., Kodama,E.N., Sarafianos,S.G., Schuckmann,M.M., Sakagami,Y., Matsuoka,M., Takiguchi,M., Gatanaga,H., and Oka,S. (2008). Amino acid mutation N348I in the connection

subdomain of human immunodeficiency virus type 1 reverse transcriptase confers multiclass resistance to nucleoside and nonnucleoside reverse transcriptase inhibitors. *J. Virol.* **82**, 3261-3270.

Haffar,O.K., Popov,S., Dubrovsky,L., Agostini,I., Tang,H., Pushkarsky,T., Nadler,S.G., and Bukrinsky,M. (2000). Two nuclear localization signals in the HIV-1 matrix protein regulate nuclear import of the HIV-1 pre-integration complex. *J. Mol. Biol.* **299**, 359-368.

Hamel,D.J., Sankale,J.L., Eisen,G., Meloni,S.T., Mullins,C., Gueye-Ndiaye,A., MBoup,S., and Kanki,P.J. (2007). Twenty years of prospective molecular epidemiology in Senegal: changes in HIV diversity. *AIDS Res. Hum. Retroviruses.* **23**, 1189-1196.

Hammer,S.M., Eron,J.J., Jr., Reiss,P., Schooley,R.T., Thompson,M.A., Walmsley,S., Cahn,P., Fischl,M.A., Gatell,J.M., Hirsch,M.S., Jacobsen,D.M., Montaner,J.S.G., Richman,D.D., Yeni,P.G., and Volberding,P.A. (2008). Antiretroviral Treatment of Adult HIV Infection: 2008 Recommendations of the International AIDS Society-USA Panel. *JAMA* **300**, 555-570.

Hansen,M., Jelinek,L., Whiting,S., and Barklis,E. (1990). Transport and assembly of gag proteins into Moloney murine leukemia virus. *J. Virol.* **64**, 5306-5316.

Harada,S., Yoshimura,K., Yamaguchi,A., Boonchawalit,S., Yusa,K., and Matsushita,S. (2013). Impact of antiretroviral pressure on selection of primary HIV-1 envelope sequences in vitro. *J. Gen. Virol.*

Hare,S., Gupta,S.S., Valkov,E., Engelman,A., and Cherepanov,P. (2010). Retroviral intasome assembly and inhibition of DNA strand transfer. *Nature.* **464**, 232-236.

Harrigan,P.R., Salim,M., Stammers,D.K., Wynhoven,B., Brumme,Z.L., McKenna,P., Larder,B., and Kemp,S.D. (2002). A mutation in the 3' region of the human immunodeficiency virus type 1 reverse transcriptase (Y318F) associated with nonnucleoside reverse transcriptase inhibitor resistance. *J. Virol.* **76**, 6836-6840.

Hasegawa,M. and Kishino,H. (1989). Heterogeneity of tempo and mode of mitochondrial DNA evolution among mammalian orders. *Jpn. J. Genet.* **64**, 243-258.

Hastings,W.K. (1970). Monte Carlo sampling methods using Markov chains and their applications. *Biometrika* **57**, 97-109.

Haubrich,R.H. (2004). Resistance and replication capacity assays: clinical utility and interpretation. *Top. HIV. Med.* 12, 52-56.

Hazuda,D.J., Felock,P., Witmer,M., Wolfe,A., Stillmock,K., Grobler,J.A., Espeseth,A., Gabryelski,L., Schleif,W., Blau,C., and Miller,M.D. (2000). Inhibitors of strand transfer that prevent integration and inhibit HIV-1 replication in cells. *Science.* 287, 646-650.

Hecht,F.M., Grant,R.M., Petropoulos,C.J., Dillon,B., Chesney,M.A., Tian,H., Hellmann,N.S., Bandrapalli,N.I., Digilio,L., Branson,B., and Kahn,J.O. (1998). Sexual transmission of an HIV-1 variant resistant to multiple reverse-transcriptase and protease inhibitors. *N. Engl. J. Med.* 339, 307-311.

Hecht,R.e.a. (2009). Critical choices in financing the response to the global HIV/AIDS pandemic. *Health Affairs* 28, 1591-1605.

Heinzinger,N.K., Bukinsky,M.I., Haggerty,S.A., Ragland,A.M., Kewalramani,V., Lee,M.A., Gendelman,H.E., Ratner,L., Stevenson,M., and Emerman,M. (1994). The Vpr protein of human immunodeficiency virus type 1 influences nuclear localization of viral nucleic acids in nondividing host cells. *Proc. Natl. Acad. Sci. U. S. A.* 91, 7311-7315.

Hemelaar,J. (2012). The origin and diversity of the HIV-1 pandemic. *Trends Mol. Med.* 18, 182-192.

Hemelaar,J., Gouws,E., Ghys,P.D., and Osmanov,S. (2011). Global trends in molecular epidemiology of HIV-1 during 2000-2007. *AIDS.* 25, 679-689.

Henderson,L.E., Krutzsch,H.C., and Oroszlan,S. (1983). Myristyl amino-terminal acylation of murine retrovirus proteins: an unusual post-translational proteins modification. *Proc. Natl. Acad. Sci. U. S. A.* 80, 339-343.

Hersh,B.S., Popovici,F., Apetrei,R.C., Zolotusca,L., Beldescu,N., Calomfirescu,A., Jezek,Z., Oxtoby,M.J., Gromyko,A., and Heymann,D.L. (1991). Acquired immunodeficiency syndrome in Romania. *Lancet.* 338, 645-649.

Hertogs,K., de Bethune,M.P., Miller,V., Ivens,T., Schel,P., Van,C.A., Van Den Eynde,C., Van,G., V, Azijn,H., Van,H.M., Peeters,F., Staszewski,S., Conant,M., Bloor,S., Kemp,S., Larder,B., and Pauwels,R. (1998). A rapid method for simultaneous detection of phenotypic resistance to inhibitors of protease and reverse transcriptase in recombinant human immunodeficiency

virus type 1 isolates from patients treated with antiretroviral drugs. *Antimicrob. Agents Chemother.* **42**, 269-276.

Hightower, K.E., Wang, R., Deanda, F., Johns, B.A., Weaver, K., Shen, Y., Tomberlin, G.H., Carter, H.L., III, Broderick, T., Sigethy, S., Seki, T., Kobayashi, M., and Underwood, M.R. (2011). Dolutegravir (S/GSK1349572) exhibits significantly slower dissociation than raltegravir and elvitegravir from wild-type and integrase inhibitor-resistant HIV-1 integrase-DNA complexes. *Antimicrob. Agents Chemother.* **55**, 4552-4559.

Hirsch, M.S., Gunthard, H.F., Schapiro, J.M., Brun-Vezinet, F., Clotet, B., Hammer, S.M., Johnson, V.A., Kuritzkes, D.R., Mellors, J.W., Pillay, D., Yeni, P.G., Jacobsen, D.M., and Richman, D.D. (2008). Antiretroviral drug resistance testing in adult HIV-1 infection: 2008 recommendations of an International AIDS Society-USA panel. *Clin. Infect. Dis.* **47**, 266-285.

Hirsch, V.M., Olmsted, R.A., Murphey-Corb, M., Purcell, R.H., and Johnson, P.R. (1989). An African primate lentivirus (SIVsm) closely related to HIV-2. *Nature.* **339**, 389-392.

Holmes, E.C. and de, A.Z.P. (1998). Genetic drift of human immunodeficiency virus type 1? *J. Virol.* **72**, 886-887.

Holt, N., Wang, J., Kim, K., Friedman, G., Wang, X., Taupin, V., Crooks, G.M., Kohn, D.B., Gregory, P.D., Holmes, M.C., and Cannon, P.M. (2010). Human hematopoietic stem/progenitor cells modified by zinc-finger nucleases targeted to CCR5 control HIV-1 in vivo. *Nat. Biotechnol.* **28**, 839-847.

Hu, D.J., Buve, A., Baggs, J., van der Groen, G., and Dondero, T.J. (1999). What role does HIV-1 subtype play in transmission and pathogenesis? An epidemiological perspective. *AIDS.* **13**, 873-881.

Hu, W.S. and Temin, H.M. (1990). Genetic consequences of packaging two RNA genomes in one retroviral particle: pseudodiploidy and high rate of genetic recombination. *Proc. Natl. Acad. Sci. U. S. A.* **87**, 1556-1560.

Hu, Z. and Kuritzkes, D.R. (2010). Effect of raltegravir resistance mutations in HIV-1 integrase on viral fitness. *J. Acquir. Immune. Defic. Syndr.* **55**, 148-155.

Hu, Z.X. and Kuritzkes, D.R. (2012). The combinations of NNRTI mutations and IN inhibitor (INI) mutations in HIV-1 affect viral fitness and drug susceptibility. *Antivir. Ther.*

Huang,M., Orenstein,J.M., Martin,M.A., and Freed,E.O. (1995). p6Gag is required for particle production from full-length human immunodeficiency virus type 1 molecular clones expressing protease. *J. Virol.* *69*, 6810-6818.

Hubner,A., Kruhoffer,M., Grosse,F., and Krauss,G. (1992). Fidelity of human immunodeficiency virus type I reverse transcriptase in copying natural RNA. *J. Mol. Biol.* *223*, 595-600.

Hudson,R.R. (1983). Properties of a neutral allele model with intragenic recombination. *Theor. Popul. Biol.* *23*, 183-201.

Hue,S., Pillay,D., Clewley,J.P., and Pybus,O.G. (2005). Genetic analysis reveals the complex structure of HIV-1 transmission within defined risk groups. *Proc. Natl. Acad. Sci. U. S. A.* *102*, 4425-4429.

Huelsenbeck,J.P., Larget,B., and Swofford,D. (2000). A compound poisson process for relaxing the molecular clock. *Genetics.* *154*, 1879-1892.

Hughes,A., Barber,T., and Nelson,M. (2008). New treatment options for HIV salvage patients: an overview of second generation PIs, NNRTIs, integrase inhibitors and CCR5 antagonists. *J. Infect.* *57*, 1-10.

Hutter,G., Nowak,D., Mossner,M., Ganepola,S., Mussig,A., Allers,K., Schneider,T., Hofmann,J., Kucherer,C., Blau,O., Blau,I.W., Hofmann,W.K., and Thiel,E. (2009). Long-term control of HIV by CCR5 Delta32/Delta32 stem-cell transplantation. *N. Engl. J. Med.* *360*, 692-698.

Ibanez,A., Clotet,B., and Martinez,M.A. (2000). Human immunodeficiency virus type 1 population bottleneck during indinavir therapy causes a genetic drift in the env quasispecies. *J. Gen. Virol.* *81*, 85-95.

Imamichi,H., Crandall,K.A., Natarajan,V., Jiang,M.K., Dewar,R.L., Berg,S., Gaddam,A., Bosche,M., Metcalf,J.A., Davey,R.T., Jr., and Lane,H.C. (2001). Human immunodeficiency virus type 1 quasi species that rebound after discontinuation of highly active antiretroviral therapy are similar to the viral quasi species present before initiation of therapy. *J. Infect. Dis.* *183*, 36-50.

Isaka,Y., Miki,S., Kawauchi,S., Suyama,A., Sugimoto,H., Adachi,A., Miura,T., Hayami,M., Yoshie,O., Fujiwara,T., and Sato,A. (2001). A single amino acid change at Leu-188 in the reverse

transcriptase of HIV-2 and SIV renders them sensitive to non-nucleoside reverse transcriptase inhibitors. *Arch. Virol.* 146, 743-755.

Izumi,K., Kodama,E., Shimura,K., Sakagami,Y., Watanabe,K., Ito,S., Watabe,T., Terakawa,Y., Nishikawa,H., Sarafianos,S.G., Kitauro,K., Oishi,S., Fujii,N., and Matsuoka,M. (2009). Design of peptide-based inhibitors for human immunodeficiency virus type 1 strains resistant to T-20. *J. Biol Chem.* %20;284, 4914-4920.

Jacks,T., Power,M.D., Masiarz,F.R., Luciw,P.A., Barr,P.J., and Varmus,H.E. (1988). Characterization of ribosomal frameshifting in HIV-1 gag-pol expression. *Nature.* 331, 280-283.

Jacobo-Molina,A., Ding,J., Nanni,R.G., Clark,A.D., Jr., Lu,X., Tantillo,C., Williams,R.L., Kamer,G., Ferris,A.L., Clark,P., and . (1993). Crystal structure of human immunodeficiency virus type 1 reverse transcriptase complexed with double-stranded DNA at 3.0 A resolution shows bent DNA. *Proc. Natl. Acad. Sci. U. S. A.* 90, 6320-6324.

Japour,A.J., Mayers,D.L., Johnson,V.A., Kuritzkes,D.R., Beckett,L.A., Arduino,J.M., Lane,J., Black,R.J., Reichelderfer,P.S., D'Aquila,R.T., and . (1993). Standardized peripheral blood mononuclear cell culture assay for determination of drug susceptibilities of clinical human immunodeficiency virus type 1 isolates. The RV-43 Study Group, the AIDS Clinical Trials Group Virology Committee Resistance Working Group. *Antimicrob. Agents Chemother.* 37, 1095-1101.

Jetzt,A.E., Yu,H., Klarmann,G.J., Ron,Y., Preston,B.D., and Dougherty,J.P. (2000). High rate of recombination throughout the human immunodeficiency virus type 1 genome. *J. Virol.* 74, 1234-1240.

Johns, B et al. The discovery of S/GSK1349572: a once-daily next generation integrase inhibitor with a superior resistance profile. 17th Conference on Retroviruses and Opportunistic Infections . 2010.

Ref Type: Abstract

Johnson,J.A., Li,J.F., Wei,X., Lipscomb,J., Bennett,D., Brant,A., Cong,M.E., Spira,T., Shafer,R.W., and Heneine,W. (2007). Simple PCR assays improve the sensitivity of HIV-1 subtype B drug resistance testing and allow linking of resistance mutations. *PLoS. One.* 2, e638.

Johnson,J.A., Li,J.F., Wei,X., Lipscomb,J., Irlbeck,D., Craig,C., Smith,A., Bennett,D.E., Monsour,M., Sandstrom,P., Lanier,E.R., and Heneine,W. (2008). Minority HIV-1 drug resistance mutations are present in antiretroviral treatment-naive populations and associate with reduced treatment efficacy. *PLoS. Med.* 5, e158.

Johnson,V.A., Calvez,V., Gunthard,H.F., Paredes,R., Pillay,D., Shafer,R., Wensing,A.M., and Richman,D.D. (2011). 2011 update of the drug resistance mutations in HIV-1. *Top. Antivir. Med.* 19, 156-164.

Johnson,V.A., Calvez,V., Gunthard,H.F., Paredes,R., Pillay,D., Shafer,R.W., Wensing,A.M., and Richman,D.D. (2013). Update of the drug resistance mutations in HIV-1: March 2013. *Top. Antivir. Med.* 21, 6-14.

Johnson,V.A., Petropoulos,C.J., Woods,C.R., Hazelwood,J.D., Parkin,N.T., Hamilton,C.D., and Fiscus,S.A. (2001). Vertical transmission of multidrug-resistant human immunodeficiency virus type 1 (HIV-1) and continued evolution of drug resistance in an HIV-1-infected infant. *J. Infect. Dis.* 183, 1688-1693.

Joly,V., Descamps,D., Peytavin,G., Touati,F., Mentre,F., Duval,X., Delarue,S., Yeni,P., and Brun-Vezinet,F. (2004). Evolution of human immunodeficiency virus type 1 (HIV-1) resistance mutations in nonnucleoside reverse transcriptase inhibitors (NNRTIs) in HIV-1-infected patients switched to antiretroviral therapy without NNRTIs. *Antimicrob. Agents Chemother.* 48, 172-175.

Ju,S.M., Song,H.Y., Lee,J.A., Lee,S.J., Choi,S.Y., and Park,J. (2009). Extracellular HIV-1 Tat up-regulates expression of matrix metalloproteinase-9 via a MAPK-NF-kappaB dependent pathway in human astrocytes. *Exp. Mol. Med.* 41, 86-93.

Jung,M., Leye,N., Vidal,N., Fargette,D., Diop,H., Toure,K.C., Gascuel,O., and Peeters,M. (2012). The Origin and Evolutionary History of HIV-1 Subtype C in Senegal. *PLoS. One.* 7, .

Kanki,P.J., Hamel,D.J., Sankale,J.L., Hsieh,C., Thior,I., Barin,F., Woodcock,S.A., Gueye-Ndiaye,A., Zhang,E., Montano,M., Siby,T., Marlink,R., NDoye,I., Essex,M.E., and MBoup,S. (1999). Human immunodeficiency virus type 1 subtypes differ in disease progression. *J. Infect. Dis.* 179, 68-73.

Kao,S., Khan,M.A., Miyagi,E., Plishka,R., Buckler-White,A., and Strebel,K. (2003). The human immunodeficiency virus type 1 Vif protein reduces intracellular expression and inhibits

packaging of APOBEC3G (CEM15), a cellular inhibitor of virus infectivity. *J. Virol.* 77, 11398-11407.

Karageorgos,L., Li,P., and Burrell,C. (1993). Characterization of HIV replication complexes early after cell-to-cell infection. *AIDS Res. Hum. Retroviruses.* 9, 817-823.

Kearney,M., Palmer,S., Maldarelli,F., Shao,W., Polis,M.A., Mican,J., Rock-Kress,D., Margolick,J.B., Coffin,J.M., and Mellors,J.W. (2008). Frequent polymorphism at drug resistance sites in HIV-1 protease and reverse transcriptase. *AIDS.* 22, 497-501.

Kellam,P. and Larder,B.A. (1994). Recombinant virus assay: a rapid, phenotypic assay for assessment of drug susceptibility of human immunodeficiency virus type 1 isolates. *Antimicrob. Agents Chemother.* 38, 23-30.

Kellam,P. and Larder,B.A. (1995). Retroviral recombination can lead to linkage of reverse transcriptase mutations that confer increased zidovudine resistance. *J. Virol.* 69, 669-674.

Kemal,K.S., Ramirez,C.M., Burger,H., Foley,B., Mayers,D., Klimkait,T., Hamy,F., Anastos,K., Petrovic,K., Minin,V.N., Suchard,M.A., and Weiser,B. (2012). Recombination between variants from genital tract and plasma: evolution of multidrug-resistant HIV type 1. *AIDS Res. Hum. Retroviruses.* 28, 1766-1774.

Kemp,S.D., Shi,C., Bloor,S., Harrigan,P.R., Mellors,J.W., and Larder,B.A. (1998). A novel polymorphism at codon 333 of human immunodeficiency virus type 1 reverse transcriptase can facilitate dual resistance to zidovudine and L-2',3'-dideoxy-3'-thiacytidine. *J. Virol.* 72, 5093-5098.

Kempf,D.J., Marsh,K.C., Kumar,G., Rodrigues,A.D., Denissen,J.F., McDonald,E., Kukulka,M.J., Hsu,A., Granneman,G.R., Baroldi,P.A., Sun,E., Pizzuti,D., Plattner,J.J., Norbeck,D.W., and Leonard,J.M. (1997). Pharmacokinetic enhancement of inhibitors of the human immunodeficiency virus protease by coadministration with ritonavir. *Antimicrob. Agents Chemother.* 41, 654-660.

Kiernan,R.E., Ono,A., Englund,G., and Freed,E.O. (1998). Role of matrix in an early postentry step in the human immunodeficiency virus type 1 life cycle. *J. Virol.* 72, 4116-4126.

Kilby,J.M., Hopkins,S., Venetta,T.M., DiMassimo,B., Cloud,G.A., Lee,J.Y., Alldredge,L., Hunter,E., Lambert,D., Bolognesi,D., Matthews,T., Johnson,M.R., Nowak,M.A., Shaw,G.M., and Saag,M.S.

(1998). Potent suppression of HIV-1 replication in humans by T-20, a peptide inhibitor of gp41-mediated virus entry. *Nat. Med.* *4*, 1302-1307.

Kils-Hutten,L., Cheynier,R., Wain-Hobson,S., and Meyerhans,A. (2001). Phylogenetic reconstruction of intrapatient evolution of human immunodeficiency virus type 1: predominance of drift and purifying selection. *J. Gen. Virol.* *82*, 1621-1627.

Kim,J., Roberts,A., Yuan,H., Xiong,Y., and Anderson,K.S. (2012). Nucleocapsid protein annealing of a primer-template enhances (+)-strand DNA synthesis and fidelity by HIV-1 reverse transcriptase. *J. Mol. Biol.* *415*, 866-880.

Kimura,M. (1968). Evolutionary rate at the molecular level. *Nature.* *217*, 624-626.

Kingman,J.F.C. (1982a). On the genealogy of large populations. *Journal of Applied Probability* *27*-43.

Kingman,J.F.C. (1982b). The coalescent. *Stochastic Processes and Their Applications* 235-248.

Kitchen,C.M., Philpott,S., Burger,H., Weiser,B., Anastos,K., and Suchard,M.A. (2004). Evolution of human immunodeficiency virus type 1 coreceptor usage during antiretroviral Therapy: a Bayesian approach. *J. Virol.* *78*, 11296-11302.

Kitrinos,K.M., Nelson,J.A.E., Resch,W., and Swanstrom,R. (2005). Effect of a Protease Inhibitor-Induced Genetic Bottleneck on Human Immunodeficiency Virus Type 1 env Gene Populations. *J. Virol.* *79*, 10627-10637.

Kiwanuka,N., Laeyendecker,O., Robb,M., Kigozi,G., Arroyo,M., McCutchan,F., Eller,L.A., Eller,M., Makumbi,F., Birx,D., Wabwire-Mangen,F., Serwadda,D., Sewankambo,N.K., Quinn,T.C., Wawer,M., and Gray,R. (2008). Effect of human immunodeficiency virus Type 1 (HIV-1) subtype on disease progression in persons from Rakai, Uganda, with incident HIV-1 infection. *J. Infect. Dis.* *197*, 707-713.

Kliger,Y., Peisajovich,S.G., Blumenthal,R., and Shai,Y. (2000). Membrane-induced conformational change during the activation of HIV-1 gp41. *J. Mol. Biol.* *301*, 905-914.

Kobayashi,M., Nakahara,K., Seki,T., Miki,S., Kawauchi,S., Suyama,A., Wakasa-Morimoto,C., Kodama,M., Endoh,T., Oosugi,E., Matsushita,Y., Murai,H., Fujishita,T., Yoshinaga,T., Garvey,E., Foster,S., Underwood,M., Johns,B., Sato,A., and Fujiwara,T. (2008). Selection of diverse and

clinically relevant integrase inhibitor-resistant human immunodeficiency virus type 1 mutants. *Antiviral Res.* *80*, 213-222.

Kobayashi,M., Yoshinaga,T., Seki,T., Wakasa-Morimoto,C., Brown,K.W., Ferris,R., Foster,S.A., Hazen,R.J., Miki,S., Suyama-Kagitani,A., Kawauchi-Miki,S., Taishi,T., Kawasuji,T., Johns,B.A., Underwood,M.R., Garvey,E.P., Sato,A., and Fujiwara,T. (2011). In Vitro antiretroviral properties of S/GSK1349572, a next-generation HIV integrase inhibitor. *Antimicrob. Agents Chemother.* *55*, 813-821.

Kohlstaedt,L.A., Wang,J., Friedman,J.M., Rice,P.A., and Steitz,T.A. (1992). Crystal structure at 3.5 Å resolution of HIV-1 reverse transcriptase complexed with an inhibitor. *Science.* *256*, 1783-1790.

Koning,F.A., Badhan,A., Shaw,S., Fisher,M., Mbisa,J.L., and Cane,P.A. (2013a). Dynamics of HIV Type 1 Recombination Following Superinfection. *AIDS Res. Hum. Retroviruses.* *29*, 963-970.

Koning,F.A., Castro,H., Dunn,D., Tilston,P., Cane,P.A., and Mbisa,J.L. (2013b). Subtype-specific differences in the development of accessory mutations associated with high-level resistance to HIV-1 nucleoside reverse transcriptase inhibitors. *J. Antimicrob. Chemother.* *68*, 1220-1236.

Korber,B., Gaschen,B., Yusim,K., Thakallapally,R., Kesmir,C., and Detours,V. (2001). Evolutionary and immunological implications of contemporary HIV-1 variation. *Br. Med. Bull.* *58:19-42.*, 19-42.

Korber,B., Muldoon,M., Theiler,J., Gao,F., Gupta,R., Lapedes,A., Hahn,B.H., Wolinsky,S., and Bhattacharya,T. (2000). Timing the ancestor of the HIV-1 pandemic strains. *Science.* *288*, 1789-1796.

Kosakovsky Pong,S.L. and Frost,S.D. (2005). Not so different after all: a comparison of methods for detecting amino acid sites under selection. *Mol. Biol Evol.* *22*, 1208-1222.

Kovaleski,B.J., Kennedy,R., Khorchid,A., Kleiman,L., Matsuo,H., and Musier-Forsyth,K. (2007). Critical role of helix 4 of HIV-1 capsid C-terminal domain in interactions with human lysyl-tRNA synthetase. *J. Biol Chem.* *282*, 32274-32279.

Kozak,S.L., Platt,E.J., Madani,N., Ferro,F.E., Jr., Peden,K., and Kabat,D. (1997). CD4, CXCR-4, and CCR-5 dependencies for infections by primary patient and laboratory-adapted isolates of human immunodeficiency virus type 1. *J. Virol.* *71*, 873-882.

Kraft,C.S., Basu,D., Hawkins,P.A., Hraber,P.T., Chomba,E., Mulenga,J., Kilembe,W., Khu,N.H., Derdeyn,C.A., Allen,S.A., Manigart,O., and Hunter,E. (2012). Timing and source of subtype-C HIV-1 superinfection in the newly infected partner of Zambian couples with disparate viruses. *Retrovirology*. %20;9:22. doi: 10.1186/1742-4690-9-22., 22-29.

Krishnan,L., Li,X., Naraharisetty,H.L., Hare,S., Cherepanov,P., and Engelman,A. (2010). Structure-based modeling of the functional HIV-1 intasome and its inhibition. *Proc. Natl. Acad. Sci. U. S. A.* 107, 15910-15915.

Kunanusont,C., Foy,H.M., Kreiss,J.K., Rerks-Ngarm,S., Phanuphak,P., Raktham,S., Pau,C.P., and Young,N.L. (1995). HIV-1 subtypes and male-to-female transmission in Thailand. *Lancet*. 345, 1078-1083.

Kusumi,K., Conway,B., Cunningham,S., Berson,A., Evans,C., Iversen,A.K., Colvin,D., Gallo,M.V., Coutre,S., Shpaer,E.G., and . (1992). Human immunodeficiency virus type 1 envelope gene structure and diversity in vivo and after cocultivation in vitro. *J. Virol.* 66, 875-885.

Lapadat-Tapolsky,M., De,R.H., Van,G.D., Roques,B., Plasterk,R., and Darlix,J.L. (1993). Interactions between HIV-1 nucleocapsid protein and viral DNA may have important functions in the viral life cycle. *Nucleic Acids Res.* 21, 831-839.

Larder,B.A., Purifoy,D.J., Powell,K.L., and Darby,G. (1987). Site-specific mutagenesis of AIDS virus reverse transcriptase. *Nature*. 327, 716-717.

Lataillade,M., Chiarella,J., Yang,R., Schnittman,S., Wirtz,V., Uy,J., Seekins,D., Krystal,M., Mancini,M., McGrath,D., Simen,B., Egholm,M., and Kozal,M. (2010). Prevalence and clinical significance of HIV drug resistance mutations by ultra-deep sequencing in antiretroviral-naive subjects in the CASTLE study. *PLoS. One.* 5, e10952.

Le Grice,S.F., Naas,T., Wohlgensinger,B., and Schatz,O. (1991). Subunit-selective mutagenesis indicates minimal polymerase activity in heterodimer-associated p51 HIV-1 reverse transcriptase. *EMBO J.* 10, 3905-3911.

Le,T., Chiarella,J., Simen,B.B., Hanczaruk,B., Egholm,M., Landry,M.L., Dieckhaus,K., Rosen,M.I., and Kozal,M.J. (2009). Low-abundance HIV drug-resistant viral variants in treatment-experienced persons correlate with historical antiretroviral use. *PLoS. One.* 4, e6079.

Leavitt,A.D., Rose,R.B., and Varmus,H.E. (1992). Both substrate and target oligonucleotide sequences affect in vitro integration mediated by human immunodeficiency virus type 1 integrase protein produced in *Saccharomyces cerevisiae*. *J. Virol.* *66*, 2359-2368.

Lee,M.S. and Craigie,R. (1994). Protection of retroviral DNA from autointegration: involvement of a cellular factor. *Proc. Natl. Acad. Sci. U. S. A.* *91*, 9823-9827.

Leigh Brown,A.J. and Cleland,A. (1996). Independent evolution of the env and pol genes of HIV-1 during zidovudine therapy. *AIDS.* *10*, 1067-1073.

Leitner,T., Escanilla,D., Franzen,C., Uhlen,M., and Albert,J. (1996). Accurate reconstruction of a known HIV-1 transmission history by phylogenetic tree analysis. *Proc. Natl. Acad. Sci. U. S. A.* *93*, 10864-10869.

Lemey,P., Pybus,O.G., Wang,B., Saksena,N.K., Salemi,M., and Vandamme,A.M. (2003). Tracing the origin and history of the HIV-2 epidemic. *Proc. Natl. Acad. Sci. U. S. A.* *100*, 6588-6592.

Lengruber,R.B., Delviks-Frankenberry,K.A., Nikolenko,G.N., Baumann,J., Santos,A.F., Pathak,V.K., and Soares,M.A. (2011). Phenotypic characterization of drug resistance-associated mutations in HIV-1 RT connection and RNase H domains and their correlation with thymidine analogue mutations. *J. Antimicrob. Chemother.* *66*, 702-708.

Lewin,S.R. and Rouzioux,C. (2011). HIV cure and eradication: how will we get from the laboratory to effective clinical trials? *AIDS.* *25*, 885-897.

Li,F., Goila-Gaur,R., Salzwedel,K., Kilgore,N.R., Reddick,M., Matallana,C., Castillo,A., Zoumplis,D., Martin,D.E., Orenstein,J.M., Allaway,G.P., Freed,E.O., and Wild,C.T. (2003). PA-457: a potent HIV inhibitor that disrupts core condensation by targeting a late step in Gag processing. *Proc. Natl. Acad. Sci. U. S. A.* *100*, 13555-13560.

Li,J.C., Lee,D.C., Cheung,B.K., and Lau,A.S. (2005a). Mechanisms for HIV Tat upregulation of IL-10 and other cytokine expression: kinase signaling and PKR-mediated immune response. *FEBS Lett.* *579*, 3055-3062.

Li,M.J., Kim,J., Li,S., Zaia,J., Yee,J.K., Anderson,J., Akkina,R., and Rossi,J.J. (2005b). Long-term inhibition of HIV-1 infection in primary hematopoietic cells by lentiviral vector delivery of a triple combination of anti-HIV shRNA, anti-CCR5 ribozyme, and a nucleolar-localizing TAR decoy. *Mol. Ther.* *12*, 900-909.

Lin,C.W. and Engelman,A. (2003). The barrier-to-autointegration factor is a component of functional human immunodeficiency virus type 1 preintegration complexes. *J. Virol.* *77*, 5030-5036.

Liu,R., Paxton,W.A., Choe,S., Ceradini,D., Martin,S.R., Horuk,R., MacDonald,M.E., Stuhlmann,H., Koup,R.A., and Landau,N.R. (1996). Homozygous defect in HIV-1 coreceptor accounts for resistance of some multiply-exposed individuals to HIV-1 infection. *Cell.* *86*, 367-377.

Liu,Y., McNevin,J., Cao,J., Zhao,H., Genowati,I., Wong,K., McLaughlin,S., McSweyn,M.D., Diem,K., Stevens,C.E., Maenza,J., He,H., Nickle,D.C., Shriner,D., Holte,S.E., Collier,A.C., Corey,L., McElrath,M.J., and Mullins,J.I. (2006). Selection on the human immunodeficiency virus type 1 proteome following primary infection. *J. Virol.* *80*, 9519-9529.

Liu,Y., McNevin,J.P., Holte,S., McElrath,M.J., and Mullins,J.I. (2011). Dynamics of viral evolution and CTL responses in HIV-1 infection. *PLoS. One.* *6*, e15639.

Lobritz,M.A., Lassen,K.G., and Arts,E.J. (2011). HIV-1 replicative fitness in elite controllers. *Curr. Opin. HIV. AIDS.* *6*, 214-220.

Low,A., Prada,N., Topper,M., Vaida,F., Castor,D., Mohri,H., Hazuda,D., Muesing,M., and Markowitz,M. (2009). Natural polymorphisms of human immunodeficiency virus type 1 integrase and inherent susceptibilities to a panel of integrase inhibitors. *Antimicrob. Agents Chemother.* *53*, 4275-4282.

Malet,I., Delelis,O., Soulie,C., Wirden,M., Tchertanov,L., Mottaz,P., Peytavin,G., Katlama,C., Mouscadet,J.F., Calvez,V., and Marcelin,A.G. (2009). Quasispecies variant dynamics during emergence of resistance to raltegravir in HIV-1-infected patients. *J. Antimicrob. Chemother.* *63*, 795-804.

Malet,I., Delelis,O., Valantin,M.A., Montes,B., Soulie,C., Wirden,M., Tchertanov,L., Peytavin,G., Reynes,J., Mouscadet,J.F., Katlama,C., Calvez,V., and Marcelin,A.G. (2008). Mutations associated with failure of raltegravir treatment affect integrase sensitivity to the inhibitor in vitro. *Antimicrob. Agents Chemother.* *52*, 1351-1358.

Marcelin,A.G., Reynes,J., Yerly,S., Ktorza,N., Segondy,M., Piot,J.C., Delfraissy,J.F., Kaiser,L., Perrin,L., Katlama,C., and Calvez,V. (2004). Characterization of genotypic determinants in HR-1

and HR-2 gp41 domains in individuals with persistent HIV viraemia under T-20. *AIDS*. *18*, 1340-1342.

Marchand,C., Johnson,A.A., Karki,R.G., Pais,G.C., Zhang,X., Cowansage,K., Patel,T.A., Nicklaus,M.C., Burke,T.R., Jr., and Pommier,Y. (2003). Metal-dependent inhibition of HIV-1 integrase by beta-diketo acids and resistance of the soluble double-mutant (F185K/C280S). *Mol. Pharmacol.* *64*, 600-609.

Marchand,C., Maddali,K., Metifiot,M., and Pommier,Y. (2009). HIV-1 IN inhibitors: 2010 update and perspectives. *Curr. Top. Med. Chem.* *9*, 1016-1037.

Margottin,F., Bour,S.P., Durand,H., Selig,L., Benichou,S., Richard,V., Thomas,D., Strebel,K., and Benarous,R. (1998). A novel human WD protein, h-beta TrCp, that interacts with HIV-1 Vpu connects CD4 to the ER degradation pathway through an F-box motif. *Mol. Cell.* *1*, 565-574.

Marinello,J., Marchand,C., Mott,B.T., Bain,A., Thomas,C.J., and Pommier,Y. (2008). Comparison of raltegravir and elvitegravir on HIV-1 integrase catalytic reactions and on a series of drug-resistant integrase mutants. *Biochemistry.* *47*, 9345-9354.

Martinez-Cajas,J.L., Pant-Pai,N., Klein,M.B., and Wainberg,M.A. (2008). Role of genetic diversity amongst HIV-1 non-B subtypes in drug resistance: a systematic review of virologic and biochemical evidence. *AIDS Rev.* *10*, 212-223.

Martinez-Cajas,J.L. and Wainberg,M.A. (2007). Protease inhibitor resistance in HIV-infected patients: molecular and clinical perspectives. *Antiviral Res.* *76*, 203-221.

Martinez-Picado,J., Sutton,L., De Pasquale,M.P., Savara,A.V., and D'Aquila,R.T. (1999). Human immunodeficiency virus type 1 cloning vectors for antiretroviral resistance testing. *J. Clin. Microbiol.* *37*, 2943-2951.

Mascolini, M. No viral rebound after ART stops in two stem-cell transplant patients. Seventh International AIDS Society (IAS) Conference on HIV Pathogenesis, Treatment and Prevention. 5-7-2013.

Ref Type: Conference Proceeding

Massanella,M., Negredo,E., Puig,J., Puertas,M.C., Buzon,M.J., Perez-Alvarez,N., Carrillo,J., Clotet,B., Martinez-Picado,J., and Blanco,J. (2012). Raltegravir intensification shows differing

effects on CD8 and CD4 T cells in HIV-infected HAART-suppressed individuals with poor CD4 T-cell recovery. *AIDS*. 26, 2285-2293.

Maxfield,L.F., Fraize,C.D., and Coffin,J.M. (2005). Relationship between retroviral DNA-integration-site selection and host cell transcription. *Proc. Natl. Acad. Sci. U. S. A.* 102, 1436-1441.

Mbisa,J.L., Martin,S.A., and Cane,P.A. (2011a). Patterns of resistance development with integrase inhibitors in HIV. *Infection and Drug Resistance* 2011, 65-76.

Mbisa,J.L., Gupta,R.K., Kabamba,D., Mulenga,V., Kalumbi,M., Chintu,C., Parry,C.M., Gibb,D.M., Walker,S.A., Cane,P.A., and Pillay,D. (2011b). The evolution of HIV-1 reverse transcriptase in route to acquisition of Q151M multi-drug resistance is complex and involves mutations in multiple domains. *Retrovirology*. 8:31. doi: 10.1186/1742-4690-8-31., 31-38.

Mbisa,J.L., Hue,S., Buckton,A.J., Myers,R.E., Duiculescu,D., Ene,L., Oprea,C., Tardei,G., Rugina,S., Mardarescu,M., Floch,C., Notheis,G., Zohrer,B., Cane,P.A., and Pillay,D. (2012). Phylodynamic and phylogeographic patterns of the HIV type 1 subtype F1 parenteral epidemic in Romania. *AIDS Res. Hum. Retroviruses*. 28, 1161-1166.

McColl,D.J. and Chen,X. (2010). Strand transfer inhibitors of HIV-1 integrase: bringing IN a new era of antiretroviral therapy. *Antiviral Res.* 85, 101-118.

McColl, D. J., Miller, M. D., Nguyen, B. Y., and Zhao, J. Resistance and cross-resistance to first generation integrase inhibitors: insights from a phase II study of elvitegravir (GS-9137). *Antivir.Ther.* 2007.

Ref Type: Abstract

McCormick,A.L., Parry,C.M., Crombe,A., Goodall,R.L., Gupta,R.K., Kaleebu,P., Kityo,C., Chirara,M., Towers,G.J., and Pillay,D. (2011). Impact of the N348I mutation in HIV-1 reverse transcriptase on nonnucleoside reverse transcriptase inhibitor resistance in non-subtype B HIV-1. *Antimicrob. Agents Chemother.* 55, 1806-1809.

McKinnon,J.E., Delgado,R., Pulido,F., Shao,W., Arribas,J.R., and Mellors,J.W. (2011). Single genome sequencing of HIV-1 gag and protease resistance mutations at virologic failure during the OK04 trial of simplified versus standard maintenance therapy. *Antivir. Ther.* 16, 725-732.

McNearney,T., Hornickova,Z., Markham,R., Birdwell,A., Arens,M., Saah,A., and Ratner,L. (1992). Relationship of human immunodeficiency virus type 1 sequence heterogeneity to stage of disease. *Proc. Natl. Acad. Sci. U. S. A.* *89*, 10247-10251.

Menendez-Arias,L. (2010). Molecular basis of human immunodeficiency virus drug resistance: an update. *Antiviral Res.* *85*, 210-231.

Menzo,S., Castagna,A., Monachetti,A., Hasson,H., Danise,A., Carini,E., Bagnarelli,P., Lazzarin,A., and Clementi,M. (2004). Genotype and phenotype patterns of human immunodeficiency virus type 1 resistance to enfuvirtide during long-term treatment. *Antimicrob. Agents Chemother.* *48*, 3253-3259.

Mesplede,T., Quashie,P.K., Osman,N., Han,Y., Singhroy,D.N., Lie,Y., Petropoulos,C.J., Huang,W., and Wainberg,M.A. (2013). Viral fitness cost prevents HIV-1 from evading dolutegravir drug pressure. *Retrovirology.* *10:22. doi: 10.1186/1742-4690-10-22.*, 22-10.

Metifiot,M., Maddali,K., Naumova,A., Zhang,X., Marchand,C., and Pommier,Y. (2010). Biochemical and pharmacological analyses of HIV-1 integrase flexible loop mutants resistant to raltegravir. *Biochemistry.* *49*, 3715-3722.

Metifiot,M., Vandegraaff,N., Maddali,K., Naumova,A., Zhang,X., Rhodes,D., Marchand,C., and Pommier,Y. (2011). Elvitegravir overcomes resistance to raltegravir induced by integrase mutation Y143. *AIDS.* *25*, 1175-1178.

Metzner,K.J., Rauch,P., Walter,H., Boesecke,C., Zollner,B., Jessen,H., Schewe,K., Fenske,S., Gellermann,H., and Stellbrink,H.J. (2005). Detection of minor populations of drug-resistant HIV-1 in acute seroconverters. *AIDS.* *19*, 1819-1825.

Meyer,P.R., Matsuura,S.E., So,A.G., and Scott,W.A. (1998). Unblocking of chain-terminated primer by HIV-1 reverse transcriptase through a nucleotide-dependent mechanism. *Proc. Natl. Acad. Sci. U. S. A.* *95*, 13471-13476.

Mild,M., Esbjornsson,J., Fenyo,E.M., and Medstrand,P. (2007). Frequent inpatient recombination between human immunodeficiency virus type 1 R5 and X4 envelopes: implications for coreceptor switch. *J. Virol.* *81*, 3369-3376.

Miller,M.D., Farnet,C.M., and Bushman,F.D. (1997). Human immunodeficiency virus type 1 preintegration complexes: studies of organization and composition. *J. Virol.* *71*, 5382-5390.

Mitchell,R.S., Beitzel,B.F., Schroder,A.R., Shinn,P., Chen,H., Berry,C.C., Ecker,J.R., and Bushman,F.D. (2004). Retroviral DNA integration: ASLV, HIV, and MLV show distinct target site preferences. *PLoS. Biol.* *2*, E234.

Molina,J.M., Lamarca,A., Andrade-Villanueva,J., Clotet,B., Clumeck,N., Liu,Y.P., Zhong,L., Margot,N., Cheng,A.K., and Chuck,S.L. (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. *Lancet Infect. Dis.* *12*, 27-35.

Munier,C.M., Andersen,C.R., and Kelleher,A.D. (2011). HIV vaccines: progress to date. *Drugs.* *71*, 387-414.

Murrell,B., Moola,S., Mabona,A., Weighill,T., Sheward,D., Kosakovsky Pond,S.L., and Scheffler,K. (2013). FUBAR: a fast, unconstrained bayesian approximation for inferring selection. *Mol. Biol Evol.* *30*, 1196-1205.

Myers,R.E. and Pillay,D. (2008). Analysis of natural sequence variation and covariation in human immunodeficiency virus type 1 integrase. *J. Virol.* *82*, 9228-9235.

Nakahara,K., Wakasa-Morimoto,C., Kobayashi,M., Miki,S., Noshi,T., Seki,T., Kanamori-Koyama,M., Kawauchi,S., Suyama,A., Fujishita,T., Yoshinaga,T., Garvey,E.P., Johns,B.A., Foster,S.A., Underwood,M.R., Sato,A., and Fujiwara,T. (2009). Secondary mutations in viruses resistant to HIV-1 integrase inhibitors that restore viral infectivity and replication kinetics. *Antiviral Res.* *81*, 141-146.

Naldini,L., Blomer,U., Gallay,P., Ory,D., Mulligan,R., Gage,F.H., Verma,I.M., and Trono,D. (1996). In vivo gene delivery and stable transduction of nondividing cells by a lentiviral vector. *Science.* *272*, 263-267.

Napravnik,S., Edwards,D., Stewart,P., Stalzer,B., Matteson,E., and Eron,J.J., Jr. (2005). HIV-1 drug resistance evolution among patients on potent combination antiretroviral therapy with detectable viremia. *J. Acquir. Immune. Defic. Syndr.* *40*, 34-40.

Navia,M.A., Fitzgerald,P.M., McKeever,B.M., Leu,C.T., Heimbach,J.C., Herber,W.K., Sigal,I.S., Darke,P.L., and Springer,J.P. (1989). Three-dimensional structure of aspartyl protease from human immunodeficiency virus HIV-1. *Nature.* *337*, 615-620.

Neil,S.J., Zang,T., and Bieniasz,P.D. (2008). Tetherin inhibits retrovirus release and is antagonized by HIV-1 Vpu. *Nature*. 451, 425-430.

Nelson,K.E., Costello,C., Suriyanon,V., Sennun,S., and Duerr,A. (2007). Survival of blood donors and their spouses with HIV-1 subtype E (CRF01 A_E) infection in northern Thailand, 1992-2007. *AIDS*. 21 *Suppl 6*:S47-54., S47-S54.

NIH. Organization of the HIV-1 virion. 2004.

Ref Type: Online Source

NIH. AIDS info: Guidelines for the use of ART agents in HIV-1-infected adults and adolescents. 2012a.

Ref Type: Report

NIH. HIV Replication Cycle. 2012b.

Ref Type: Online Source

Nikolenko,G.N., Delviks-Frankenberry,K.A., Palmer,S., Maldarelli,F., Fivash,M.J., Jr., Coffin,J.M., and Pathak,V.K. (2007). Mutations in the connection domain of HIV-1 reverse transcriptase increase 3'-azido-3'-deoxythymidine resistance. *Proc. Natl. Acad. Sci. U. S. A.* 104, 317-322.

Nikolenko,G.N., Delviks-Frankenberry,K.A., and Pathak,V.K. (2010). A novel molecular mechanism of dual resistance to nucleoside and nonnucleoside reverse transcriptase inhibitors. *J. Virol.* 84, 5238-5249.

Nikolenko,G.N., Palmer,S., Maldarelli,F., Mellors,J.W., Coffin,J.M., and Pathak,V.K. (2005). Mechanism for nucleoside analog-mediated abrogation of HIV-1 replication: balance between RNase H activity and nucleotide excision. *Proc. Natl. Acad. Sci. U. S. A.* 102, 2093-2098.

Pacold,M.E., Pond,S.L., Wagner,G.A., Delport,W., Bourque,D.L., Richman,D.D., Little,S.J., and Smith,D.M. (2012). Clinical, virologic, and immunologic correlates of HIV-1 intraclade B dual infection among men who have sex with men. *AIDS*. 26, 157-165.

Page,R.D.M. and Holmes,E.C. (1998). *Molecular Evolution: A Phylogenetic Approach*.

Palmer,S., Kearney,M., Maldarelli,F., Halvas,E.K., Bixby,C.J., Bazmi,H., Rock,D., Falloon,J., Davey,R.T., Jr., Dewar,R.L., Metcalf,J.A., Hammer,S., Mellors,J.W., and Coffin,J.M. (2005).

Multiple, linked human immunodeficiency virus type 1 drug resistance mutations in treatment-experienced patients are missed by standard genotype analysis. *J. Clin. Microbiol.* *43*, 406-413.

Pantaleo,G., Graziosi,C., and Fauci,A.S. (1993). New concepts in the immunopathogenesis of human immunodeficiency virus infection. *N. Engl. J. Med.* *328*, 327-335.

Paredes,R., Puertas,M.C., Bannister,W., Kisic,M., Cozzi-Lepri,A., Pou,C., Bellido,R., Betancor,G., Bogner,J., Gargalianos,P., Banhegyi,D., Clotet,B., Lundgren,J., Menendez-Arias,L., and Martinez-Picado,J. (2011). A376S in the Connection Subdomain of HIV-1 Reverse Transcriptase Confers Increased Risk of Virological Failure to Nevirapine Therapy. *J. Infect. Dis.* *204*, 741-752.

Parry,C.M., Kohli,A., Boinett,C.J., Towers,G.J., McCormick,A.L., and Pillay,D. (2009). Gag determinants of fitness and drug susceptibility in protease inhibitor-resistant human immunodeficiency virus type 1. *J. Virol.* *83*, 9094-9101.

Pastore,C., Ramos,A., and Mosier,D.E. (2004). Intrinsic obstacles to human immunodeficiency virus type 1 coreceptor switching. *J. Virol.* *78*, 7565-7574.

Patel,P.H., Jacobo-Molina,A., Ding,J., Tantillo,C., Clark,A.D., Jr., Raag,R., Nanni,R.G., Hughes,S.H., and Arnold,E. (1995). Insights into DNA polymerization mechanisms from structure and function analysis of HIV-1 reverse transcriptase. *Biochemistry.* *34*, 5351-5363.

Pear,W.S., Nolan,G.P., Scott,M.L., and Baltimore,D. (1993). Production of high-titer helper-free retroviruses by transient transfection. *Proc. Natl. Acad. Sci. U. S. A.* *90*, 8392-8396.

Peeters,M., Aghokeng,A.F., and Delaporte,E. (2010). Genetic diversity among human immunodeficiency virus-1 non-B subtypes in viral load and drug resistance assays. *Clin. Microbiol. Infect.* *16*, 1525-1531.

Peeters,M. and Delaporte,E. (2012). Simian retroviruses in African apes. *Clin. Microbiol. Infect.* *18*, 514-520.

Perelson,A.S., Neumann,A.U., Markowitz,M., Leonard,J.M., and Ho,D.D. (1996). HIV-1 dynamics in vivo: virion clearance rate, infected cell life-span, and viral generation time. *Science.* *271*, 1582-1586.

Perno,C.F. and Bertoli,A. (2006). Clinical cut-offs in the interpretation of phenotypic resistance.

Petropoulos,C.J., Parkin,N.T., Limoli,K.L., Lie,Y.S., Wrin,T., Huang,W., Tian,H., Smith,D., Winslow,G.A., Capon,D.J., and Whitcomb,J.M. (2000). A novel phenotypic drug susceptibility assay for human immunodeficiency virus type 1. *Antimicrob. Agents Chemother.* *44*, 920-928.

Philpott,S., Burger,H., Tsoukas,C., Foley,B., Anastos,K., Kitchen,C., and Weiser,B. (2005). Human immunodeficiency virus type 1 genomic RNA sequences in the female genital tract and blood: compartmentalization and inpatient recombination. *J. Virol.* *79*, 353-363.

Pieniazek,D., Rayfield,M., Hu,D.J., Nkengasong,J., Wiktor,S.Z., Downing,R., Biryahwaho,B., Mastro,T., Tanuri,A., Soriano,V., Lal,R., and Dondero,T. (2000). Protease sequences from HIV-1 group M subtypes A-H reveal distinct amino acid mutation patterns associated with protease resistance in protease inhibitor-naive individuals worldwide. HIV Variant Working Group. *AIDS.* *14*, 1489-1495.

Pillay,D., Walker,A.S., Gibb,D.M., de,R.A., Kaye,S., Ait-Khaled,M., Munoz-Fernandez,M., and Babiker,A. (2002). Impact of human immunodeficiency virus type 1 subtypes on virologic response and emergence of drug resistance among children in the Paediatric European Network for Treatment of AIDS (PENTA) 5 trial. *J. Infect. Dis.* *186*, 617-625.

Pitisuttithum,P., Gilbert,P., Gurwith,M., Heyward,W., Martin,M., van,G.F., Hu,D., Tappero,J.W., and Choopanya,K. (2006). Randomized, double-blind, placebo-controlled efficacy trial of a bivalent recombinant glycoprotein 120 HIV-1 vaccine among injection drug users in Bangkok, Thailand. *J. Infect. Dis.* *194*, 1661-1671.

Plikat,U., Nieselt-Struwe,K., and Meyerhans,A. (1997). Genetic drift can dominate short-term human immunodeficiency virus type 1 nef quasispecies evolution in vivo. *J. Virol.* *71*, 4233-4240.

Pollard,V.W. and Malim,M.H. (1998). The HIV-1 Rev protein. *Annu. Rev. Microbiol.* *52*:491-532., 491-532.

Pomerantz,R.J. (2002). Reservoirs of human immunodeficiency virus type 1: the main obstacles to viral eradication. *Clin. Infect. Dis.* *34*, 91-97.

Prabu-Jeyabalan,M., Nalivaika,E., and Schiffer,C.A. (2002). Substrate shape determines specificity of recognition for HIV-1 protease: analysis of crystal structures of six substrate complexes. *Structure.* *10*, 369-381.

- Prabu-Jeyabalan,M., Nalivaika,E.A., King,N.M., and Schiffer,C.A. (2003). Viability of a drug-resistant human immunodeficiency virus type 1 protease variant: structural insights for better antiviral therapy. *J. Virol.* 77, 1306-1315.
- Preston,B.D. and Dougherty,J.P. (1996). Mechanisms of retroviral mutation. *Trends Microbiol.* 4, 16-21.
- Preston,B.D., Poiesz,B.J., and Loeb,L.A. (1988). Fidelity of HIV-1 reverse transcriptase. *Science.* 242, 1168-1171.
- Pybus,O.G. and Rambaut,A. (2005). GENIE: estimating demographic history from molecular phylogenies. *Bioinformatics.* 1404-1405.
- Pybus,O.G. and Rambaut,A. (2009). Evolutionary analysis of the dynamics of viral infectious disease. *Nat. Rev. Genet.* 10, 540-550.
- Quashie,P.K., Mesplede,T., Han,Y.S., Oliveira,M., Singhroy,D.N., Fujiwara,T., Underwood,M.R., and Wainberg,M.A. (2012a). Characterization of the R263K mutation in HIV-1 integrase that confers low-level resistance to the second-generation integrase strand transfer inhibitor dolutegravir. *J. Virol.* 86, 2696-2705.
- Quashie,P.K., Sloan,R.D., and Wainberg,M.A. (2012b). Novel therapeutic strategies targeting HIV integrase. *BMC. Med.* 10:34. doi: 10.1186/1741-7015-10-34., 34-10.
- Quercia,R., Dam,E., Perez-Bercoff,D., and Clavel,F. (2009). Selective-advantage profile of human immunodeficiency virus type 1 integrase mutants explains in vivo evolution of raltegravir resistance genotypes. *J. Virol.* 83, 10245-10249.
- Quinones-Mateu,M.E. and Arts,E.J. (2002). Fitness of drug resistant HIV-1: methodology and clinical implications. *Drug Resist. Updat.* 5, 224-233.
- Quinones-Mateu,M.E., Ball,S.C., Marozsan,A.J., Torre,V.S., Albright,J.L., Vanham,G., van Der,G.G., Colebunders,R.L., and Arts,E.J. (2000). A dual infection/competition assay shows a correlation between ex vivo human immunodeficiency virus type 1 fitness and disease progression. *J. Virol.* 74, 9222-9233.
- Radzio,J. and Sluis-Cremer,N. (2011). Subunit-specific mutational analysis of residue N348 in HIV-1 reverse transcriptase. *Retrovirology.* 8:69. doi: 10.1186/1742-4690-8-69., 69-8.

Raffi,F., Jaeger,H., Quiros-Roldan,E., Albrecht,H., Belonosova,E., Gatell,J.M., Baril,J.G., Domingo,P., Brennan,C., Almond,S., and Min,S. (2013). Once-daily dolutegravir versus twice-daily raltegravir in antiretroviral-naive adults with HIV-1 infection (SPRING-2 study): 96 week results from a randomised, double-blind, non-inferiority trial. *Lancet Infect. Dis.* *13*, 927-935.

Raghavendra,N.K., Shkriabai,N., Graham,R.L., Hess,S., Kvaratskhelia,M., and Wu,L. (2010). Identification of host proteins associated with HIV-1 preintegration complexes isolated from infected CD4+ cells. *Retrovirology.* *7:66. doi: 10.1186/1742-4690-7-66.*, 66-67.

Rambaut,A., Posada,D., Crandall,K.A., and Holmes,E.C. (2004). The causes and consequences of HIV evolution. *Nat. Rev. Genet.* *5*, 52-61.

Reid,S.D., Fidler,S.J., and Cooke,G.S. (2013). Tracking the progress of HIV: the impact of point-of-care tests on antiretroviral therapy. *Clin. Epidemiol.* *5:387-396.*, 387-396.

Reigadas,S., Anies,G., Masquelier,B., Calmels,C., Stuyver,L.J., Parissi,V., Fleury,H., and Andreola,M.L. (2010). The HIV-1 integrase mutations Y143C/R are an alternative pathway for resistance to Raltegravir and impact the enzyme functions. *PLoS. One.* *5*, e10311.

Reigadas,S., Masquelier,B., Calmels,C., Laguerre,M., Lazaro,E., Vandenhende,M., Neau,D., Fleury,H., and Andreola,M.L. (2011). Structure-analysis of the HIV-1 integrase Y143C/R raltegravir resistance mutation in association with the secondary mutation T97A. *Antimicrob. Agents Chemother.* *55*, 3187-3194.

Ren,J., Esnouf,R., Garman,E., Somers,D., Ross,C., Kirby,I., Keeling,J., Darby,G., Jones,Y., Stuart,D., and . (1995). High resolution structures of HIV-1 RT from four RT-inhibitor complexes. *Nat. Struct. Biol.* *2*, 293-302.

Ren,J., Nichols,C., Bird,L., Chamberlain,P., Weaver,K., Short,S., Stuart,D.I., and Stammers,D.K. (2001). Structural mechanisms of drug resistance for mutations at codons 181 and 188 in HIV-1 reverse transcriptase and the improved resilience of second generation non-nucleoside inhibitors. *J. Mol. Biol.* *312*, 795-805.

Ren,J., Nichols,C.E., Chamberlain,P.P., Weaver,K.L., Short,S.A., and Stammers,D.K. (2004). Crystal structures of HIV-1 reverse transcriptases mutated at codons 100, 106 and 108 and mechanisms of resistance to non-nucleoside inhibitors. *J. Mol. Biol.* *336*, 569-578.

Ren,J., Nichols,C.E., Stamp,A., Chamberlain,P.P., Ferris,R., Weaver,K.L., Short,S.A., and Stammers,D.K. (2006). Structural insights into mechanisms of non-nucleoside drug resistance for HIV-1 reverse transcriptases mutated at codons 101 or 138. *FEBS J.* 273, 3850-3860.

Renjifo,B., Gilbert,P., Chaplin,B., Msamanga,G., Mwakagile,D., Fawzi,W., and Essex,M. (2004). Preferential in-utero transmission of HIV-1 subtype C as compared to HIV-1 subtype A or D. *AIDS.* 18, 1629-1636.

Rerks-Ngarm,S., Pitisuttithum,P., Nitayaphan,S., Kaewkungwal,J., Chiu,J., Paris,R., Prensri,N., Namwat,C., de,S.M., Adams,E., Benenson,M., Gurunathan,S., Tartaglia,J., McNeil,J.G., Francis,D.P., Stablein,D., Birx,D.L., Chunsuttiwat,S., Khamboonruang,C., Thongcharoen,P., Robb,M.L., Michael,N.L., Kunasol,P., and Kim,J.H. (2009). Vaccination with ALVAC and AIDSVAX to prevent HIV-1 infection in Thailand. *N. Engl. J. Med.* 361, 2209-2220.

Rhee,S.S. and Hunter,E. (1987). Myristylation is required for intracellular transport but not for assembly of D-type retrovirus capsids. *J. Virol.* 61, 1045-1053.

Rhee,S.Y., Gonzales,M.J., Kantor,R., Betts,B.J., Ravela,J., and Shafer,R.W. (2003). Human immunodeficiency virus reverse transcriptase and protease sequence database. *Nucleic Acids Res.* 31, 298-303.

Rhee,S.Y., Kantor,R., Katzenstein,D.A., Camacho,R., Morris,L., Sirivichayakul,S., Jorgensen,L., Brigido,L.F., Schapiro,J.M., and Shafer,R.W. (2006). HIV-1 pol mutation frequency by subtype and treatment experience: extension of the HIVseq program to seven non-B subtypes. *AIDS.* 20, 643-651.

Rhee,S.Y., Liu,T.F., Kiuchi,M., Zioni,R., Gifford,R.J., Holmes,S.P., and Shafer,R.W. (2008). Natural variation of HIV-1 group M integrase: implications for a new class of antiretroviral inhibitors. *Retrovirology.* 5:74., 74.

Robbins,K.E., Lemey,P., Pybus,O.G., Jaffe,H.W., Youngpairoj,A.S., Brown,T.M., Salemi,M., Vandamme,A.M., and Kalish,M.L. (2003). U.S. Human immunodeficiency virus type 1 epidemic: date of origin, population history, and characterization of early strains. *J. Virol.* 77, 6359-6366.

Roberts,J.D., Bebenek,K., and Kunkel,T.A. (1988). The accuracy of reverse transcriptase from HIV-1. *Science.* 242, 1171-1173.

Roehr, B. Doctors claim to have cured girl of HIV. *BMJ.* 5-3-2013.

Ref Type: Journal (Full)

Rogel,M.E., Wu,L.I., and Emerman,M. (1995). The human immunodeficiency virus type 1 vpr gene prevents cell proliferation during chronic infection. *J. Virol.* *69*, 882-888.

Ross,H.A. and Rodrigo,A.G. (2002). Immune-mediated positive selection drives human immunodeficiency virus type 1 molecular variation and predicts disease duration. *J. Virol.* *76*, 11715-11720.

Rossenkhan,R., Novitsky,V., Sebuya,T.K., Musonda,R., Gashe,B.A., and Essex,M. (2012). Viral Diversity and Diversification of Major Non-Structural Genes vif, vpr, vpu, tat exon 1 and rev exon 1 during Primary HIV-1 Subtype C Infection. *PLoS. One.* *7*, .

Sadiq,S.K., Noe,F., and De,F.G. (2012). Kinetic characterization of the critical step in HIV-1 protease maturation. *Proc. Natl. Acad. Sci. U. S. A.* *109*, 20449-20454.

Saez-Cirion,A., Bacchus,C., Hocqueloux,L., Avettand-Fenoel,V., Girault,I., Lecuroux,C., Potard,V., Versmisse,P., Melard,A., Prazuck,T., Descours,B., Guernon,J., Viard,J.P., Boufassa,F., Lambotte,O., Goujard,C., Meyer,L., Costagliola,D., Venet,A., Pancino,G., Autran,B., and Rouzioux,C. (2013). Post-treatment HIV-1 controllers with a long-term virological remission after the interruption of early initiated antiretroviral therapy ANRS VISCONTI Study. *PLoS. Pathog.* *9*, e1003211.

Santos,A.F., Lengruher,R.B., Soares,E.A., Jere,A., Sprinz,E., Martinez,A.M., Silveira,J., Sion,F.S., Pathak,V.K., and Soares,M.A. (2008). Conservation patterns of HIV-1 RT connection and RNase H domains: identification of new mutations in NRTI-treated patients. *PLoS. One.* *3*, e1781.

Sarafianos,S.G., Das,K., Clark,A.D., Jr., Ding,J., Boyer,P.L., Hughes,S.H., and Arnold,E. (1999). Lamivudine (3TC) resistance in HIV-1 reverse transcriptase involves steric hindrance with beta-branched amino acids. *Proc. Natl. Acad. Sci. U. S. A.* *96*, 10027-10032.

Sarafianos,S.G., Marchand,B., Das,K., Himmel,D.M., Parniak,M.A., Hughes,S.H., and Arnold,E. (2009). Structure and function of HIV-1 reverse transcriptase: molecular mechanisms of polymerization and inhibition. *J. Mol. Biol.* *385*, 693-713.

Schmalzbauer,E., Strack,B., Dannull,J., Guehmann,S., and Moelling,K. (1996). Mutations of basic amino acids of NCp7 of human immunodeficiency virus type 1 affect RNA binding in vitro. *J. Virol.* *70*, 771-777.

Schmidt,H.A., Strimmer,K., Vingron,M., and von,H.A. (2002). TREE-PUZZLE: maximum likelihood phylogenetic analysis using quartets and parallel computing. *Bioinformatics*. *18*, 502-504.

Schneider,C.A., Rasband,W.S., and Eliceiri,K.W. (2012). NIH Image to ImageJ: 25 years of image analysis. *Nat. Methods*. *9*, 671-675.

Schroder,A.R., Shinn,P., Chen,H., Berry,C., Ecker,J.R., and Bushman,F. (2002). HIV-1 integration in the human genome favors active genes and local hotspots. *Cell*. *110*, 521-529.

Shafer, R. W. Rationale and Uses of a Public HIV Drug-Resistance Database. *Journal of Infectious Diseases* . 2006.

Ref Type: Abstract

Shafer,R.W. (2009). Low-abundance drug-resistant HIV-1 variants: finding significance in an era of abundant diagnostic and therapeutic options. *J. Infect. Dis*. *199*, 610-612.

Shafer,R.W., Kantor,R., and Gonzales,M.J. (2000). The Genetic Basis of HIV-1 Resistance to Reverse Transcriptase and Protease Inhibitors. *AIDS Rev*. *2*, 211-228.

Shendure,J. and Ji,H. (2008). Next-generation DNA sequencing. *Nat. Biotechnol*. *26*, 1135-1145.

Shi,B., Kitchen,C., Weiser,B., Mayers,D., Foley,B., Kemal,K., Anastos,K., Suchard,M., Parker,M., Brunner,C., and Burger,H. (2010). Evolution and recombination of genes encoding HIV-1 drug resistance and tropism during antiretroviral therapy. *Virology*. *404*, 5-20.

Shimura,K., Kodama,E., Sakagami,Y., Matsuzaki,Y., Watanabe,W., Yamataka,K., Watanabe,Y., Ohata,Y., Doi,S., Sato,M., Kano,M., Ikeda,S., and Matsuoka,M. (2008). Broad antiretroviral activity and resistance profile of the novel human immunodeficiency virus integrase inhibitor elvitegravir (JTK-303/GS-9137). *J. Virol*. *82*, 764-774.

Shriner,D., Shankarappa,R., Jensen,M.A., Nickle,D.C., Mittler,J.E., Margolick,J.B., and Mullins,J.I. (2004). Influence of random genetic drift on human immunodeficiency virus type 1 env evolution during chronic infection. *Genetics*. *166*, 1155-1164.

Sichtig,N., Sierra,S., Kaiser,R., Daumer,M., Reuter,S., Schuler,E., Altmann,A., Fatkenheuer,G., Dittmer,U., Pfister,H., and Esser,S. (2009). Evolution of raltegravir resistance during therapy. *J. Antimicrob. Chemother*. *64*, 25-32.

Siliciano,J.D., Kajdas,J., Finzi,D., Quinn,T.C., Chadwick,K., Margolick,J.B., Kovacs,C., Gange,S.J., and Siliciano,R.F. (2003). Long-term follow-up studies confirm the stability of the latent reservoir for HIV-1 in resting CD4+ T cells. *Nat. Med.* 9, 727-728.

Simen,B.B., Simons,J.F., Hullsiek,K.H., Novak,R.M., Macarthur,R.D., Baxter,J.D., Huang,C., Lubeski,C., Turenchalk,G.S., Braverman,M.S., Desany,B., Rothberg,J.M., Egholm,M., and Kozal,M.J. (2009). Low-abundance drug-resistant viral variants in chronically HIV-infected, antiretroviral treatment-naive patients significantly impact treatment outcomes. *J. Infect. Dis.* 199, 693-701.

Simons, J. F., Egholm, M., Lanza, J., Turenchalk, G., and et al. Ultradeep sequencing of HIV clinical samples. *Antivir.Ther.* 2005.

Ref Type: Abstract

Sluis-Cremer,N., Moore,K., Radzio,J., Sonza,S., and Tachedjian,G. (2010). N348I in HIV-1 reverse transcriptase decreases susceptibility to tenofovir and etravirine in combination with other resistance mutations. *AIDS.* 24, 317-319.

Smith,P.F., Ogundele,A., Forrest,A., Wilton,J., Salzwedel,K., Doto,J., Allaway,G.P., and Martin,D.E. (2007). Phase I and II study of the safety, virologic effect, and pharmacokinetics/pharmacodynamics of single-dose 3-o-(3',3'-dimethylsuccinyl)betulinic acid (bevirimat) against human immunodeficiency virus infection. *Antimicrob. Agents Chemother.* 51, 3574-3581.

Sundquist,W.I. and Krausslich,H.G. (2012). HIV-1 Assembly, Budding, and Maturation. *Cold Spring Harb. Perspect. Med.* 2, a006924.

Svicher,V., Ceccherini-Silberstein,F., Erba,F., Santoro,M., Gori,C., Bellocchi,M.C., Giannella,S., Trotta,M.P., Monforte,A., Antinori,A., and Perno,C.F. (2005). Novel human immunodeficiency virus type 1 protease mutations potentially involved in resistance to protease inhibitors. *Antimicrob. Agents Chemother.* 49, 2015-2025.

Tajima,F. (1989). Statistical Method for Testing the Neutral Mutation Hypothesis by DNA Polymorphism. *Genetics.* 123, 585-595.

Tamura,K., Dudley,J., Nei,M., and Kumar,S. (2007). MEGA4: Molecular Evolutionary Genetics Analysis (MEGA) software version 4.0. *Mol. Biol Evol.* 24, 1596-1599.

Tamura,K., Peterson,D., Peterson,N., Stecher,G., Nei,M., and Kumar,S. (2011). MEGA5: molecular evolutionary genetics analysis using maximum likelihood, evolutionary distance, and maximum parsimony methods. *Mol. Biol Evol.* 28, 2731-2739.

Tazi,L., Imamichi,H., Hirschfeld,S., Metcalf,J.A., Orsega,S., Perez-Losada,M., Posada,D., Lane,H.C., and Crandall,K.A. (2011). HIV-1 infected monozygotic twins: a tale of two outcomes. *BMC. Evol. Biol.* 11:62. doi: 10.1186/1471-2148-11-62., 62-11.

Temin,H.M. (1993). Retrovirus variation and reverse transcription: abnormal strand transfers result in retrovirus genetic variation. *Proc. Natl. Acad. Sci. U. S. A.* 90, 6900-6903.

Templeton,A.R., Kramer,M.G., Jarvis,J., Kowalski,J., Gange,S., Schneider,M.F., Shao,Q., Zhang,G.W., Yeh,M.F., Tsai,H.L., Zhang,H., and Markham,R.B. (2009). Multiple-infection and recombination in HIV-1 within a longitudinal cohort of women. *Retrovirology.* 6:54. doi: 10.1186/1742-4690-6-54., 54-56.

Thompson,J.D., Higgins,D.G., and Gibson,T.J. (1994). CLUSTAL W: improving the sensitivity of progressive multiple sequence alignment through sequence weighting, position-specific gap penalties and weight matrix choice. *Nucleic Acids Res.* 22, 4673-4680.

Thomson,M.M., Delgado,E., Herrero,I., Villahermosa,M.L., Vazquez-de,P.E., Cuevas,M.T., Carmona,R., Medrano,L., Perez-Alvarez,L., Cuevas,L., and Najera,R. (2002). Diversity of mosaic structures and common ancestry of human immunodeficiency virus type 1 BF intersubtype recombinant viruses from Argentina revealed by analysis of near full-length genome sequences. *J. Gen. Virol.* 83, 107-119.

Thorne,J.L., Kishino,H., and Painter,I.S. (1998). Estimating the rate of evolution of the rate of molecular evolution. *Mol. Biol Evol.* 15, 1647-1657.

Tie,Y., Boross,P.I., Wang,Y.F., Gaddis,L., Liu,F., Chen,X., Tozser,J., Harrison,R.W., and Weber,I.T. (2005). Molecular basis for substrate recognition and drug resistance from 1.1 to 1.6 angstroms resolution crystal structures of HIV-1 protease mutants with substrate analogs. *FEBS J.* 272, 5265-5277.

Trignetti,M., Sing,T., Svicher,V., Santoro,M.M., Forbici,F., D'Arrigo,R., Bellocchi,M.C., Santoro,M., Marconi,P., Zaccarelli,M., Trotta,M.P., Bellagamba,R., Narciso,P., Antinori,A.,

Lengauer,T., Perno,C.F., and Ceccherini-Silberstein,F. (2009). Dynamics of NRTI resistance mutations during therapy interruption. *AIDS Res. Hum. Retroviruses*. 25, 57-64.

Tuailon,E., Gueudin,M., Lemee,V., Gueit,I., Roques,P., Corrigan,G.E., Plantier,J.C., Simon,F., and Braun,J. (2004). Phenotypic susceptibility to nonnucleoside inhibitors of virion-associated reverse transcriptase from different HIV types and groups. *J. Acquir. Immune. Defic. Syndr*. 37, 1543-1549.

UNAIDS. AIDS Epidemic Update, 2009. 2009.

Ref Type: Report

UNAIDS. Global Report (report on the global AIDS epidemic). 2010.

Ref Type: Report

Vallejo,A., Gutierrez,C., Hernandez-Novoa,B., Diaz,L., Madrid,N., Abad-Fernandez,M., Drona,F., Perez-Elias,M.J., Zamora,J., Munoz,E., Munoz-Fernandez,M.A., and Moreno,S. (2012). The effect of intensification with raltegravir on the HIV-1 reservoir of latently infected memory CD4 T cells in suppressed patients. *AIDS*. 26, 1885-1894.

van der Loeff,M.F., Awasana,A.A., Sarge-Njie,R., van der Sande,M., Jaye,A., Sabally,S., Corrah,T., McConkey,S.J., and Whittle,H.C. (2006). Sixteen years of HIV surveillance in a West African research clinic reveals divergent epidemic trends of HIV-1 and HIV-2. *Int. J. Epidemiol*. 35, 1322-1328.

Van,B.K., Rondelez,E., Van,E., V, Arien,K., Clynhens,M., Van den Zegel,P., Winters,B., and Stuyver,L.J. (2009a). A combined genotypic and phenotypic human immunodeficiency virus type 1 recombinant virus assay for the reverse transcriptase and integrase genes. *J. Virol. Methods*. 161, 231-239.

Van,B.K., Salzwedel,K., Rondelez,E., Van,E., V, De,V.S., Verheyen,A., Steegen,K., Verlinden,Y., Allaway,G.P., and Stuyver,L.J. (2009b). Susceptibility of human immunodeficiency virus type 1 to the maturation inhibitor bevirimat is modulated by baseline polymorphisms in Gag spacer peptide 1. *Antimicrob. Agents Chemother*. 53, 2185-2188.

Van,L.K., Schrooten,Y., Covens,K., Dekeersmaecker,N., De,M.P., Van,W.E., Van,R.M., and Vandamme,A.M. (2008). A genotypic assay for the amplification and sequencing of integrase from diverse HIV-1 group M subtypes. *J. Virol. Methods*. 153, 176-181.

Vandamme,A.M., Bottu,G., Van Ranst,M., Lemey,P., Strimmer,K., von,H.A., Salemi,M., Van de Peer,Y., Schmidt,H.A., Ronquist,F., van der Mark,P., Huelsenbeck,J.P., Swofford,D., Sullivan,J.S., Opperdoes,F.R., Posada,D., Pybus,O.G., Shapiro,B., Kosakovsky Pond,S.L., Poon,A.F.Y., Frost,S.D.W., Salminen,M., Martin,D., Rodrigo,A., Drummond,A.J., Rambaut,A., Kuhner,M.K., Xia,X., Moulton,V., and Huber,K.T. (2009). *The Phylogenetic Handbook: A Practical Approach to Phylogenetic Analysis and Hypothesis Testing*.

Vanitharani,R., Mahalingam,S., Rafaeli,Y., Singh,S.P., Srinivasan,A., Weiner,D.B., and Ayyavoo,V. (2001). HIV-1 Vpr transactivates LTR-directed expression through sequences present within -278 to -176 and increases virus replication in vitro. *Virology*. 289, 334-342.

Vartanian,J.P., Meyerhans,A., Asjo,B., and Wain-Hobson,S. (1991). Selection, recombination, and G----A hypermutation of human immunodeficiency virus type 1 genomes. *J. Virol.* 65, 1779-1788.

Vavro,C., Hasan,S., Madsen,H., Horton,J., Deanda,F., Martin-Carpenter,L., Sato,A., Cuffe,R., Chen,S., Underwood,M., and Nichols,G. (2013). Prevalent polymorphisms in wild-type HIV-1 integrase are unlikely to engender drug resistance to dolutegravir (S/GSK1349572). *Antimicrob. Agents Chemother.* 57, 1379-1384.

Veenstra,J., Schuurman,R., Cornelissen,M., van't Wout,A.B., Boucher,C.A., Schuitemaker,H., Goudsmit,J., and Coutinho,R.A. (1995). Transmission of zidovudine-resistant human immunodeficiency virus type 1 variants following deliberate injection of blood from a patient with AIDS: characteristics and natural history of the virus. *Clin. Infect. Dis.* 21, 556-560.

Vergne,L., Peeters,M., Mpoudi-Ngole,E., Bourgeois,A., Liegeois,F., Toure-Kane,C., Mboup,S., Mulanga-Kabeya,C., Saman,E., Jourdan,J., Reynes,J., and Delaporte,E. (2000). Genetic diversity of protease and reverse transcriptase sequences in non-subtype-B human immunodeficiency virus type 1 strains: evidence of many minor drug resistance mutations in treatment-naive patients. *J. Clin. Microbiol.* 38, 3919-3925.

Vink,C., van Gent,D.C., Elgersma,Y., and Plasterk,R.H. (1991). Human immunodeficiency virus integrase protein requires a subterminal position of its viral DNA recognition sequence for efficient cleavage. *J. Virol.* 65, 4636-4644.

Vogt,P.K. (1971). Genetically stable reassortment of markers during mixed infection with avian tumor viruses. *Virology*. 46, 947-952.

Wang,C., Mitsuya,Y., Gharizadeh,B., Ronaghi,M., and Shafer,R.W. (2007). Characterization of mutation spectra with ultra-deep pyrosequencing: application to HIV-1 drug resistance. *Genome Res.* 17, 1195-1201.

Wang,J.Y., Ling,H., Yang,W., and Craigie,R. (2001). Structure of a two-domain fragment of HIV-1 integrase: implications for domain organization in the intact protein. *EMBO J.* 20, 7333-7343.

Waters, J., Margot, N., Hluhanich, R., Svarovskaia, J., Harris, J., Borroto-Esoda, K., Miller, M. D., and McColl, D. J. Evolution of resistance to the HIV integrase inhibitor (INI) elvitegravir can involve genotypic switching among primary INI resistance patterns. *Antivir. Ther.* 2009.

Ref Type: Abstract

Weaver,T.A. and Panganiban,A.T. (1990). N myristoylation of the spleen necrosis virus matrix protein is required for correct association of the Gag polyprotein with intracellular membranes and for particle formation. *J. Virol.* 64, 3995-4001.

Weber,J., Rose,J.D., Vazquez,A.C., Winner,D., Margot,N., McColl,D.J., Miller,M.D., and Quinones-Mateu,M.E. (2013). Resistance mutations outside the integrase coding region have an effect on human immunodeficiency virus replicative fitness but do not affect its susceptibility to integrase strand transfer inhibitors. *PLoS. One.* 8, e65631.

Weber,J., Vazquez,A.C., Winner,D., Rose,J.D., Wylie,D., Rhea,A.M., Henry,K., Pappas,J., Wright,A., Mohamed,N., Gibson,R., Rodriguez,B., Soriano,V., King,K., Arts,E.J., Olivo,P.D., and Quinones-Mateu,M.E. (2011). Novel method for simultaneous quantification of phenotypic resistance to maturation, protease, reverse transcriptase, and integrase HIV inhibitors based on 3'Gag(p2/p7/p1/p6)/PR/RT/INT-recombinant viruses: a useful tool in the multitarget era of antiretroviral therapy. *Antimicrob. Agents Chemother.* 55, 3729-3742.

Wei,X., Decker,J.M., Liu,H., Zhang,Z., Arani,R.B., Kilby,J.M., Saag,M.S., Wu,X., Shaw,G.M., and Kappes,J.C. (2002). Emergence of resistant human immunodeficiency virus type 1 in patients receiving fusion inhibitor (T-20) monotherapy. *Antimicrob. Agents Chemother.* 46, 1896-1905.

Weiss,R.A., Mason,W.S., and Vogt,P.K. (1973). Genetic recombinants and heterozygotes derived from endogenous and exogenous avian RNA tumor viruses. *Virology.* 52, 535-552.

Westby,M., Smith-Burchnell,C., Mori,J., Lewis,M., Mosley,M., Stockdale,M., Dorr,P., Ciaramella,G., and Perros,M. (2007). Reduced Maximal Inhibition in Phenotypic Susceptibility

Assays Indicate that Viral Strains Resistant to the CCR5 Antagonist Maraviroc Utilize Inhibitor-Bound Receptor for Entry. *J. Virol.* *81*, 2359-2371.

Wild, C.T., Shugars, D.C., Greenwell, T.K., McDanal, C.B., and Matthews, T.J. (1994). Peptides corresponding to a predictive alpha-helical domain of human immunodeficiency virus type 1 gp41 are potent inhibitors of virus infection. *Proc. Natl. Acad. Sci. U. S. A.* *91*, 9770-9774.

Winters, M.A., Lloyd, R.M., Jr., Shafer, R.W., Kozal, M.J., Miller, M.D., and Holodniy, M. (2012). Development of elvitegravir resistance and linkage of integrase inhibitor mutations with protease and reverse transcriptase resistance mutations. *PLoS. One.* *7*, e40514.

Witvrouw, M., Pannecouque, C., Switzer, W.M., Folks, T.M., De, C.E., and Heneine, W. (2004). Susceptibility of HIV-2, SIV and SHIV to various anti-HIV-1 compounds: implications for treatment and postexposure prophylaxis. *Antivir. Ther.* *9*, 57-65.

Wlodawer, A. and Erickson, J.W. (1993). Structure-based inhibitors of HIV-1 protease. *Annu. Rev. Biochem.* *62*:543-85., 543-585.

Woodward, C.L., Prakobwanakit, S., Mosessian, S., and Chow, S.A. (2009). Integrase interacts with nucleoporin NUP153 to mediate the nuclear import of human immunodeficiency virus type 1. *J. Virol.* *83*, 6522-6533.

Wright, E., Temperton, N.J., Marston, D.A., McElhinney, L.M., Fooks, A.R., and Weiss, R.A. (2008). Investigating antibody neutralization of lyssaviruses using lentiviral pseudotypes: a cross-species comparison. *J. Gen. Virol.* *89*, 2204-2213.

Wu, X. and Burgess, S.M. (2004). Integration target site selection for retroviruses and transposable elements. *Cell Mol. Life Sci.* *61*, 2588-2596.

Xu, L., Pozniak, A., Wildfire, A., Stanfield-Oakley, S.A., Mosier, S.M., Ratcliffe, D., Workman, J., Joall, A., Myers, R., Smit, E., Cane, P.A., Greenberg, M.L., and Pillay, D. (2005). Emergence and evolution of enfuvirtide resistance following long-term therapy involves heptad repeat 2 mutations within gp41. *Antimicrob. Agents Chemother.* *49*, 1113-1119.

Yamashita, M. and Emerman, M. (2004). Capsid is a dominant determinant of retrovirus infectivity in nondividing cells. *J. Virol.* *78*, 5670-5678.

Yang,C., Li,M., Newman,R.D., Shi,Y.P., Ayisi,J., van Eijk,A.M., Otieno,J., Misore,A.O., Steketee,R.W., Nahlen,B.L., and Lal,R.B. (2003). Genetic diversity of HIV-1 in western Kenya: subtype-specific differences in mother-to-child transmission. *AIDS*. *17*, 1667-1674.

Yap,S.H., Sheen,C.W., Fahey,J., Zanin,M., Tyssen,D., Lima,V.D., Wynhoven,B., Kuiper,M., Sluis-Cremer,N., Harrigan,P.R., and Tachedjian,G. (2007). N348I in the connection domain of HIV-1 reverse transcriptase confers zidovudine and nevirapine resistance. *PLoS. Med.* *4*, e335.

Yerly,S., Kaiser,L., Race,E., Bru,J.P., Clavel,F., and Perrin,L. (1999). Transmission of antiretroviral-drug-resistant HIV-1 variants. *Lancet*. *354*, 729-733.

Yoder,K.E. and Bushman,F.D. (2000). Repair of gaps in retroviral DNA integration intermediates. *J. Virol.* *74*, 11191-11200.

Yu,X., Yu,Q.C., Lee,T.H., and Essex,M. (1992). The C terminus of human immunodeficiency virus type 1 matrix protein is involved in early steps of the virus life cycle. *J. Virol.* *66*, 5667-5670.

Yukl, S. A., Chun, T., Strain, M. C., Siliciano, J., Eisele, E., and et al. Challenges inherent in detecting HIV persistence during potentially curative interventions. *Antivir.Ther.* 2012.

Ref Type: Abstract

Zanotto,P.M., Kallas,E.G., de Souza,R.F., and Holmes,E.C. (1999). Genealogical evidence for positive selection in the nef gene of HIV-1. *Genetics*. *153*, 1077-1089.

Zeifelder,U., Frank,O., Sparacio,S., Schon,U., Bosch,V., Seifarth,W., and Leib-Mosch,C. (2007). The potential of retroviral vectors to cotransfer human endogenous retroviruses (HERVs) from human packaging cell lines. *Gene*. *390*, 175-179.

Zelina,S., Sheen,C.W., Radzio,J., Mellors,J.W., and Sluis-Cremer,N. (2008). Mechanisms by which the G333D mutation in human immunodeficiency virus type 1 Reverse transcriptase facilitates dual resistance to zidovudine and lamivudine. *Antimicrob. Agents Chemother.* *52*, 157-163.

Zhang,L., Diaz,R.S., Ho,D.D., Mosley,J.W., Busch,M.P., and Mayer,A. (1997a). Host-specific driving force in human immunodeficiency virus type 1 evolution in vivo. *J. Virol.* *71*, 2555-2561.

Zhang,Y.M., Imamichi,H., Imamichi,T., Lane,H.C., Falloon,J., Vasudevachari,M.B., and Salzman,N.P. (1997b). Drug resistance during indinavir therapy is caused by mutations in the protease gene and in its Gag substrate cleavage sites. *J. Virol.* 71, 6662-6670.

Zhu,T., Muthui,D., Holte,S., Nickle,D., Feng,F., Brodie,S., Hwangbo,Y., Mullins,J.I., and Corey,L. (2002). Evidence for human immunodeficiency virus type 1 replication in vivo in CD14(+) monocytes and its potential role as a source of virus in patients on highly active antiretroviral therapy. *J. Virol.* 76, 707-716.

Zhuang,J., Jetzt,A.E., Sun,G., Yu,H., Klarmann,G., Ron,Y., Preston,B.D., and Dougherty,J.P. (2002). Human immunodeficiency virus type 1 recombination: rate, fidelity, and putative hot spots. *J. Virol.* 76, 11273-11282.