Impact of Host Restriction Factors and Capsid Evolution on HIV Disease Progression



Michael Terence Boswell

Wolfson College

University of Oxford

Professors Sarah Rowland-Jones and Joakim Esbjörnsson

A thesis submitted for the degree of

Doctor of Philosophy

in

Clinical Medicine

Nuffield Department of Medicine

Trinity term 2019

Impact of Host Restriction Factors and Capsid Evolution on HIV Disease Progression

Michael Terence Boswell, Wolfson College

D. Phil. in Clinical Medicine, Trinity Term 2019

HIV-1 is responsible for the global HIV pandemic whereas HIV-2 has mostly been limited to West Africa. The key to their successful spread is their unparalleled ability to adapt to the host immune system. The contemporary HIV pandemics are the latest iteration in a millennia-long evolutionary arms race between retroviruses and their hosts. Two host restriction factors which have been important actors in this arms race are TRIM5α and its paralogue TRIM22. TRIM5α targets the retroviral capsid, whereas TRIM22 suppresses retroviral transcription indirectly.

The central hypothesis for the thesis is that the HIV capsid and TRIM restriction are associated with HIV disease progression. In this thesis we present findings from four projects. In chapter three we discuss the effect of *TRIM* genotype on disease outcomes in children and adolescents infected perinatally with HIV-1. We characterised *TRIM* genetic diversity in this cohort and found that *TRIM22* genotype is associated with $CD4^+$ T cell counts and, potentially, disease progression rates. Chapter four presents the outcomes of using $CD4^+$ T cell kinetic data to stratify HIV-2 infected adults into disease progressor groups. Forty-six percent of patients were classified as fast progressors and 54% as slow progressors. Slow progressors advanced to AIDS at approximately half the rate of fast progressors, had higher $CD4^+$ T cell levels and a slower $CD4^+$ T cell decline rate. In chapter five we discuss the effect of *TRIM* genotype on HIV-2 disease outcomes in adults. *TRIM22* genotype associated with significant effects on long-term $CD4^+$ levels, though these effects were modest. Finally, we report that intrahost evolutionary rates of HIV-2 *gag* are significantly higher in fast progressors, that the p24 region evolves under negative selection and that evolutionary rates correlate negatively with $CD4^+$ T cell percentage.

Statement of originality

I declare that the work presented within this thesis entitled “Impact of host restriction factors and capsid evolution on HIV disease progression” is my own. Contributions made by collaborators have been clearly acknowledged. I have not submitted this thesis, in whole, or in part for any other degree or qualification at the University of Oxford or elsewhere.

Publications

1. Boswell MT, Rossouw TM. Approach to acute kidney injury in HIV-infected patients in South Africa. South Afr J HIV Med. 2017 Nov 28;18(1):7 pages.
2. Weatherley DAV, Boswell MT, Rowland-Jones SL. Targeting TRIM5α in HIV Cure Strategies for the CRISPR-Cas9 Era. Front Immunol. 2017 Nov 22; 8:1616
3. Boswell MT, Rowland-Jones SL. Delayed disease progression in HIV-2: the importance of TRIM5α and the retroviral capsid. Clin Exp Immunol 2019;196:305–17
4. Boswell MT, Palm A, Karlson S, Månsson F, Norrgren H, Jansson M, et al. A17 The effect of intrahost evolution of HIV-2 capsid on disease progression. Virus Evol. 2019 Aug 1;5(Supplement_1) (**published abstract**)

Awards

1. Commonwealth Scholarship to fund doctoral studies at University of Oxford. A three-year stipendiary scholarship which covered all fees related to completing my D. Phil. degree. 2016 – 2019 (£161,053)
2. Wolfson College travel grants from 2017 – 2019 (£690)
3. Selected to attend the 23rd International Workshop on Virus Evolution and Molecular Epidemiology (VEME2018). Hosted by the Robert Koch Institute, Berlin, Germany 2018
4. British Society for Immunology international travel grant to attend the Southern African HIV Clinicians Society Conference 2018 (£1000)

Acknowledgments

I'm writing these words three years to the day since I left South Africa for Oxford. My time in Oxford and Lund have been life-changing in the most literal of senses. I am certain the change has been for the better.

As is customary I shall start at the beginning. This project happened because Sarah Rowland-Jones accepted me as a student, and there are no easy words to describe how I feel about that. I mention Joakim Esbjörnsson second only to keep this narrative's timeline consistent. Joakim's contribution to my development as a scientist is hard to overstate or exaggerate. My career has taken a completely different trajectory because we worked together. Sarah and Joakim – thank you for everything, there's simply too much to mention.

From here the acknowledgements are a little easier. Massive thank you to all the following people: Karen Poxon - you've made my time in the lab so much easier. Louis-Marie Yindom, Mimi Hou, Dan Hameiri Bowen, Bethany Charlton, Aljawharah Alrubayyi, Ester Gea-Mallorqui, Danielle Groves, Simon Brackenridge, The McKeating people and Brittany Chao – for being such wonderful people to work with. It's truly been a pleasure.

From 2018 – 2019 I spent several months at Lund University working on this project. Sara Karlson was instrumental in helping me to finish this thesis, chapter six would not have been possible without her. From teaching me lab techniques with patience, to finding the oldest and most obscure samples in the basement freezers of the BMC in Lund. Jamirah Nazziwa guided me through Bayesian phylogenetic analysis and was always willing to help me no matter how mundane the question. Emil Johanson, Sviat Sasinovich, Sara and Jamirah – thank you for making my time in the Lund the memorable experience that it was. Marianne Jansson, Patrik Medstrand and Fredrik Månsson provided much needed kind words and encouragement with my projects over this time.

I'm always grateful for all my friends and family – however, friends who help proofread your thesis deserve special mention. Niamh Appleby and Glynn Buchanan gave excellent feedback for this work. Ashley Jacobs deserves mention, not only for his tireless proofreading and feedback but also for being my emotional

support scientist over the past three years. Science is hard for clinicians, and I'm glad we went through a PhD together. Luckily, we also had Caron Jacobs with us.

To my friends and family: as I've said before, you make this life worth living. Thanks for being yourselves.

I may have started the acknowledgements with Sarah, but timeline-wise that's not strictly correct. Nadri said yes to me applying to come to Oxford long before then. She did so without hesitation and gave me the support I needed to finish this degree - I'll always remember that. This thesis is dedicated to her. So baie lief vir jou

Lastly – I'm immensely grateful to all the participants who contributed samples and information to this study.



"The microscopic and sub-microscopic parasites can evolve so much more rapidly than their hosts that the latter have little chance of evolving complete immunity to them… The most that the average species can achieve is to dodge its minute enemies by constantly producing new genotypes."

- J.B.S. Haldane, 1949

Contents

Introduction 1

1.1 Emergence of HIV-1 and HIV-2 2

1.2 HIV structure and genome 6

1.3 HIV replication cycle 10

1.4 HIV disease: from infection to AIDS 12

1.5 Viral factors associated with disease progression 15

1.6 The HIV-2 capsid's impact on disease progression 18

1.7 Host restriction factors and HIV 20

1.8 Host genetics and HIV-2 disease outcomes 25

1.9 T cell responses and HIV-2 disease outcomes 27

1.10 Immune activation and HIV disease 29

1.11 Clinical factors associated with HIV-2 disease progression 30

1.12 Antiretroviral therapy 32

1.13 Phylogenetic inference of HIV evolution 33

1.14 A model for slower disease progression in HIV-2 40

1.15 References for chapter 1 41

Materials and methods 57

2.1. Methods for chapter 3 57
2.1.1. Patient recruitment 57
2.1.2. $CD4^+$ T cell count analysis 58
2.1.3. DNA extraction and storage 58
2.1.4. *TRIM5* and *TRIM22* PCR amplification 59
2.1.5. Post-PCR product visualisation and purification 62
2.1.6. Sanger pre-sequencing reaction 63
2.1.7. Post-sequencing product purification 64
2.1.8. Sequence analysis 64

2.1.9. Quantification of plasma viral loads for CMV and HIV 66

2.2. Methods for chapter 4 68
2.2.1. Patient recruitment 68
2.2.2. $CD4^+$ T cell count analysis 69
2.2.3. Stratification of HIV-2 infected individuals into fast and slow disease progression groups 69

2.3. Methods for chapter 5 73
2.3.1. Patient recruitment 73
2.3.2. DNA extraction and storage 73
2.3.3. *TRIM5* and *TRIM22* PCR amplification 74

2.4. Methods for chapter 6 77
2.4.1 Study participants 77
2.4.2 RNA extraction 77
2.4.3 HIV-2 *gag* PCR 78
2.4.4 Cloning and sequencing 82

2.5 References for chapter 2 84

The impact of *TRIM22* genotype on disease outcomes of HIV-1 infected children and adolescents 86

3.1 Introduction 86
3.1.1 Specific aims 89

3.2 Methods 90
3.2.1 HIV-1 infected patients 90
3.2.2 HIV-1 negative controls 91
3.2.3 $CD4^+$ T cell measurement 91
3.2.4 Classification of progressor status 92
3.2.5 DNA extraction and storage 92
3.2.6 Amplification and sequencing of *TRIM5* and *TRIM22* 92
3.2.7 Quantification of plasma viral load for CMV and HIV 93
3.2.8 Sequence analysis 93
3.2.9 Statistical analysis 94

3.3 Results 95
3.3.1 Summary of genotype data for *TRIM5* and *TRIM22* 96
3.3.2 *TRIM22* SNP associations with clinical endpoints in HIV-1 infection 96
3.3.3 Association of age adjusted $CD4^+$ count by *TRIM22* genotype 102
3.3.4 Association of *TRIM22* genotype with AIDS at enrolment 105
3.3.5 Genotype frequency comparison 106

3.4 Discussion 108

3.5 Section summary 111

3.6 References for chapter 3 112

Appendix 3 116

Stratification of HIV-2 disease progression using CD4$^+$ T cell kinetic data 118

4.1 Introduction 118
4.1.1 Specific aims 120

4.2 Methods 121
4.2.1 Summary statistics for the Guinea-Bissau Police cohort clinical data 121
4.2.2 Patient enrolment and sample collection 121
4.2.3 CD4% measurements 122
4.2.4 Classification of HIV-2 infected individuals as relative fast and slow disease progressors 122
4.2.5 Statistical analysis 123

4.3 Results 124
4.3.1 Time to event analyses 126
4.3.2 Multivariate analysis of time to AIDS and mortality 128

4.4 Discussion 133

4.5 Section summary 136

4.6 References for chapter 4 137

Appendix 4 141

***TRIM22* genotype is associated with CD4$^+$ T cell kinetics in HIV-2 infected adults** 143

5.1 Introduction 143
5.1.1 Specific aims 145

5.2 Methods 146
5.2.1 Patients included in analysis 146
5.2.2 CD4% measurements 146
5.2.3 Stratification of patients into progressor groups 146
5.2.4 DNA extraction and storage 147
5.2.5 Amplification and sequencing of *TRIM5* and *TRIM22* 147
5.2.6 Sequence analysis 147

5.2.7 Statistical analysis 148

5.3 Results 149
5.3.1 Summary of genotype data for *TRIM5* and *TRIM22* 149
5.3.2 Association of *TRIM* genotype with midpoint CD4% and CD4% change rate 153
5.3.3 Change in longitudinal CD4% by *TRIM* genotype 155
5.3.4 Time to AIDS by *TRIM* genotype 155

5.4 Discussion 160

5.5 Section summary 163

5.6 References for chapter 5 164

Appendix 5 167

Intrahost evolutionary rates of HIV-2 *gag* correlate with progression to AIDS 169

6.1 Introduction 169
6.1.1 Specific aims 171

6.2 Methods 172
6.2.1 Study participants 172
6.2.2 Stratification of patients into progressor groups 172
6.2.3 RNA extraction and PCR amplification 173
6.2.4 Cloning and sequencing 173
6.2.5 Sequence analysis 174
6.2.6 Bayesian phylogenetic analysis 174
6.2.7 Variant identification and fingerprinting 176
6.2.8 Statistical analysis 177

6.3 Results 178
6.3.1 Variant identification and fingerprinting in HIV-2 p24 181
6.3.2 Estimation of site-specific selection pressures in the p24 region of *gag* 184
6.3.3 Comparison of *gag* evolutionary rates by disease progressor status using patient-specific relaxed molecular clock models 185
6.3.4 Correlation of CD4% kinetic data with *gag* evolutionary rates estimated by Hierarchical Phylogenetic Models 187

6.4 Discussion 189

6.5 Section summary 192

6.6 References for chapter 6 193

Appendix 6 197

Conclusions and future perspectives 204

7.1 Single nucleotide polymorphism association studies of *TRIM5* and *TRIM22* with immunological, viral and clinical outcomes of HIV infection 205
7.2 Reporting a HIV-2 disease progressor stratification system based on $CD4^+$ T cell kinetic data over long-term follow-up 207
7.3 Comparing intrahost evolution of HIV-2 *gag* between relative fast and slow disease progressors. 208
7.4 References for chapter 7 210

List of figures

Chapter 1

Figure 1.1	Phylogenetic tree of HIV and SIV sequences	3
Figure 1.2	Timeline of recent history of Guinea-Bissau	6
Figure 1.3	HIV-2 virion	7
Figure 1.4	HIV-1 and HIV-2 gene maps	8
Figure 1.5	HIV replication cycle	11
Figure 1.6	HIV disease progression from infection to AIDS	14
Figure 1.7	Genomic layout of *TRIM5* and *TRIM22*	20
Figure 1.8	Model of HIV-1 capsid	22
Figure 1.9	Mechanism of retroviral restriction by TRIM5α	24
Figure 1.10	Evolution of the viral quasispecies	35
Figure 1.11	Bayesian inference	39
Figure 1.12	Slower disease progression in HIV-2 infection	40

Chapter 2

Figure 2.1	Stratification of HIV-2 infected adults in fast and slow disease progressor groups	72
Figure 2.2	Primer design for HIV-2 *gag* PCR	80
Figure 2.3	Cloning process for HIV-2 gag final PCR products	82

Chapter 3

Figure 3.1	Linkage map of variants in *TRIM5* and *TRIM22*	98
Figure 3.2	Comparison of $CD4^+$ count and HIV viral load by clinical factors	99
Figure 3.3	Comparison of $CD4^+$ count and HIV viral load by *TRIM22* genotype	100
Figure 3.4	HIV viral load and $CD4^+$ count by progressor status	102
Figure 3.5	Multivariate linear regression of $CD4^+$ count by age and *TRIM22* genotype	104
Figure 3.6	Comparison of genotype frequencies in *TRIM22*	107

Chapter 4

Figure 4.1	Midpoint CD4% and CD4% change rate by progressor groups defined by the combined coefficient	125
Figure 4.2	CD4% change over follow-up between progressor groups	126
Figure 4.3	Kaplan-Meier curves for time to AIDS for progressor groups	127
Figure 4.4	Age at enrolment by sex and age at estimated date of infection	129
Figure 4.5	Cox regression adjusted survival curves	132

Chapter 5

Figure 5.1	Genotype frequencies of non-synonymous variants in *TRIM5* and *TRIM22*	151
Figure 5.2	Linkage map of variants in *TRIM5* and *TRIM22*	152
Figure 5.3	Comparison of midpoint CD4% and CD4% change rate by *TRIM5* genotype	153
Figure 5.4	Comparison of midpoint CD4% and CD4% change rate by *TRIM22* genotype	154
Figure 5.5	CD4% decline from HIV-2 detection, compared across genotypes for *TRIM22* rs7935564 and rs10838543	157
Figure 5.6	Time to event for AIDS by *TRIM22* genotype	158
Figure 5.7	Genotype frequencies for *TRIM5* and *TRIM22* genotype across fast and slow progressor groups with HIV-2 infection	159

Chapter 6

Figure 6.1	Cladogram of all HIV-2 *gag* sequences included in phylogenetic analysis	180
Figure 6.2	Sequence diversity in HIV-2 p24 by progressor status	182
Figure 6.3	Fingerprinting of identified major variants	183
Figure 6.4	Per-site selection pressures in p24 by progressor groups	184
Figure 6.5	Joint, synonymous and nonsynonymous substitution rates for progressor groups	186
Figure 6.6	Synonymous and nonsynonymous divergence over time by progressor groups	187
Figure 6.7	Correlation of median molecular clock rates derived from HPM with CD4% decline rates and midpoint CD4%	188
Figure S6.1	Phylogenetic tree of subtyping analysis for HIV-2 *gag* sequences	198
Figure S6.2	Nonsynonymous divergence over time	201

List of tables

Chapter 1

Table 1.1	HIV genes, proteins and their respective functions in the lifecycle	9
Table 1.2	Protein domain of *TRIM5α* and *TRIM22*	23
Table 1.3	Antiretroviral therapy agents currently in use	33
Table 1.4	Nucleotide substitution models	36

Chapter 2

Table 2.1	Primers for amplification and sequencing of *TRIM5*	60
Table 2.2	Primers for amplifying and sequencing *TRIM22*	61
Table 2.3	Thermocycler settings for touchdown PCR of *TRIM5* and *TRIM22*	62
Table 2.4	Patient stratification values	71
Table 2.5	Primers for amplification and sequencing of *TRIM5*	74
Table 2.6	Primers for amplifying and sequencing *TRIM22*	75
Table 2.7	Thermocycler settings for touchdown PCR of *TRIM5* and *TRIM22*	76
Table 2.8	HIV-2 *gag* sequences used in primer design	79
Table 2.9	Primers used in the HIV-2 *gag* PCR	80
Table 2.10	Reagents per reaction for HIV-2 *gag* PCR	81
Table 2.11	Cycling conditions for HIV-2 *gag* PCR	81

Chapter 3

Table 3.1	Summary statistics for the ZENITH cohort patients	95
Table 3.2	Summary of genotype data for *TRIM5* and *TRIM22*	97
Table 3.3	Results of regression modelling to determine associations with AIDS	106
Table S3.1	Results of model testing for testing host factors against dependent variable square root CD4+ count	116
Table S3.2	Summary for linear regression model 3	116
Table S3.3	Sensitivity analysis for linear model 3	116
Table S3.4	Model fitting procedures for multivariate binary logistic regression to determine associations with AIDS	117

Chapter 4

Table 4.1	Police cohort summary statistics	124
Table 4.2	Summary of time to AIDS	130
Table 4.3	Summary of time to death	130
Table S4.1	Summary of time to AIDS for Police cohort seroconverters	141
Table S4.2	Time to AIDS for patients with longer CD4% follow up	142

Chapter 5

Table 5.1	Police cohort summary statistics for genotyped patients	149
Table 5.2	Summary of genotype data for *TRIM5* and *TRIM22*	150
Table S5.1	Comparisons of random slope/intercept models for CD4%	167
Table S5.2	Cox proportional hazard model summary	168
Table S5.3	Cox proportional hazard model 2 output	168

Chapter 6

Table 6.1	Primers used in the HIV-2 *gag* PCR	173
Table 6.2	Summary clinical data for included patients	178
Table 6.3	VESPA output summary for progressor groups	182
Table S6.1	Accession numbers for sequences used in HIV-2 subtyping analysis	197
Table S6.2	Most recent common ancestor reconstruction phylogenetic model parameter summaries	198
Table S6.3	Summary epidemiological data for included patients	199
Table S6.4	Summary sample data for included patients	200
Table S6.5	Relaxed molecular clock estimates	201
Table S6.6	Hierarchical Phylogenetic Model parameters used to estimate molecular clock rates	202
Table S6.7	Strict molecular clock estimates from the Hierarchical Phylogenetic Model	202
Table S6.8	Results of Hierarchical Phylogenetic Model sensitivity analysis	203

List of abbreviations

AIC	Akaike information criterion
AIDS	Acquired Immunodeficiency Syndrome
ANRS	National Agency for AIDS Research (French)
ART	antiretroviral therapy
AZT	Zidovudine
BDT	Big Dye Terminator
BEAGLE	Broad-platform Evolutionary Analysis General Likelihood Evaluator
BEAST	Bayesian Evolutionary Analysis Sampling Trees
BEAUti	Bayesian Evolutionary Analysis Utility
BIC	Bayesian information criteria
CA	capsid
CADD	Combined Annotation Dependent Depletion
CC	Coiled-coil
CCR5	C-C chemokine receptor type 5
CD	cluster of differentiation
CGAS-STING	cyclic GMP–AMP synthase and stimulator of interferon genes
CIITA	class II transactivator
CMV	cytomegalovirus
CRF	circulating recombinant form
CSW	commercial sex worker
CTL	cytotoxic lymphocyte
CypA	cyclophilin A
CXCR4	C-X-C Motif Chemokine Receptor 4
DC-SIGN	dendritic cell–specific intercellular adhesion molecular 3-grabbing non-integrin
dN	nonsynonymous substitution rate
DNA	deoxyribonucleic acid
dNTP	deoxyribonucleotide triphosphate
dS	synonymous substitution rate
EDTA	Ethylenediaminetetraacetic acid
ELISpot	enzyme-linked immune absorbent spot
eQTL	Expression quantitative trait loci
ESS	effective sample size
F81	Felsenstein 1981
FET	Fisher's Exact Test
GTR	General Time Reversible
HBV	Hepatitis B virus
HIV	Human Immunodeficiency Virus
HKY85	Hasegawa, Kishino and Yano 1985
HLA	Human Leucocyte Antigen
HPM	hierarchical phylogenetic model
HTLV	Human T Cell Lymphotropic Virus

HWE	Hardy-Weinberg Equilibrium
IFN	Interferon
IN	integrase
INHALE	INvestigation of Heart And Lung disEase
JC69	Jukes Cantor 1969
K80	Kimura 1980
KIR	Killer-cell immunoglobulin-like receptors
LANL	Los Alamos National Laboratory
LAV	Lymphadenopathy Associated Virus
LD	linkage disequilibrium
LTFU	lost to follow-up
LTNP	long term non-progressors
LTR	long terminal repeat
MA	matrix
MCC	maximum clade credibility
MCMC	Markov Chain Monte Carlo
ME	minimum evolution
MRCA	most recent common ancestor
MSA	multiple sequence alignment
MW	Mann Whitney-U
NC	nucleocapsid
NCBI	National Center for Biotechnology Information
NFKB	nuclear factor kappa-light-chain-enhancer of activated B cells
NNRTI	non-nucleoside reverse transcriptase inhibitor
NOD-2	Nucleotide-binding oligomerization domain-containing protein 2
NONO	Non-POU domain-containing octamer-binding protein
NPC	nuclear pore complex
NRTI	nucleoside reverse transcriptase inhibitor
ORF	open reading frame
PBMC	peripheral blood mononuclear cell
PCR	polymerase chain reaction
PD1-PDL1	Programmed cell death protein 1 and its ligand
PIC	pre-integration complex
PML	promyelocytic leukaemia protein
PR	protease
pVL	plasma viral load
qPCR	quantitative polymerase chain reaction
RC	replicative capacity
RING	Really Interesting New Gene
RNA	ribonucleic acid
RT	reverse transcriptase
RT-PCR	reverse transcription polymerase chain reaction
SAAP	single amino acid polymorphisms
SAMHD1	SAM domain and HD domain-containing protein 1
SIV	Simian Immunodeficiency Virus

SNP	single nucleotide polymorphism
Sp1	specificity protein 1
SPVL	setpoint viral load
TAK	Transforming growth factor beta-activated kinase 1
TRIM	Tripartite interaction motif
VESPA	Viral Epidemiology Signature Pattern Analysis
WHO	World Health Organisation
ZENITH	Zimbabwe study for Enhancing Testing and Improving Treatment of HIV in Children

Chapter 1

Introduction

"Yet AIDS is not only an emergency. It is a massive and long-term problem – one of the make-or-break issues of our century" – Peter Piot as UNAIDS Executive Director, Statement at the UN General Assembly High Level Meeting on AIDS. New York, 31st May 2006.

Since the first cases of HIV-related illness were reported in 1981 the disease has morphed from an obscure illness affecting a socially marginalised population to a massive global pandemic (1). In 2017 the estimated global prevalence of HIV-1 was 37.9 million people, with 1.7 million new infections and 770 000 deaths per year (2).The burden of disease is greatest in countries which would be described as resource-limited settings, predominantly in Sub-Saharan Africa.

Unlike HIV-1, HIV-2 has not led to a global pandemic and recent data indicate that HIV-2 prevalence is declining in some West African states where it was formerly endemic. Although many early isolates of HIV-2 were derived from patients presenting with AIDS-defining illnesses, many HIV-2-infected adults showed evidence of slow disease progression. Many of the patient with lower viral loads showed limited evidence of T cell loss. Despite this, HIV-2 does progress to clinical AIDS without treatment in most patients – albeit at a slower rate than HIV-1. In addition, successful treatment with antiretroviral therapy (ART) is more challenging than for HIV-1.

HIV-2 is significantly more sensitive than HIV-1 to restriction by host restriction factor TRIM5α, and this difference in sensitivity is linked to differences in capsid structure. In this Introduction we will focus particularly on the potential impact of TRIM5α and HIV-2 capsid on disease progression.

1.1 Emergence of HIV-1 and HIV-2

The Human Immunodeficiency Viruses (HIV) are a group of zoonotic retroviruses which originated from Simian Immunodeficiency Viruses (SIV) in a wide range of African primates (3,4). There are four groups of HIV-1: M, N, O and P. Within the pandemic group HIV-1 M there are a further 14 subtypes and several circulating recombinant forms (CRFs) (5). HIV-1 group M and N originated in SIVcpz (chimpanzee), which was a probable recombinant of several precursor SIVs circulating in other primates that chimpanzees hunt (6). HIV-1 groups O and P emerged from SIVgor (gorilla), which itself originated in SIVcpz (7)(**figure 1.1**).

HIV-2 has nine known groups (A - I), and only two, A and B, have spread consistently in humans (8). Each group of HIV-2 represents an independent sylvatic transmission of SIVsmm from sooty mangabeys which range throughout West Africa (9). A single HIV-2 AB CRF has been described in a small infection cluster in Japan, and its prevalence in West Africa is unknown (10). Most SIV infections do not usually cause disease in their primate hosts. However, SIVcpz infection has been associated with an AIDS-like illness and premature death in wild chimpanzees (11). SIVsmm infection in sooty mangabeys is usually asymptomatic despite high concentrations of circulating viral RNA (12). In rare cases SIVsmm can cause an AIDS-like disease, but the time to AIDS is probably longer

than the lifespan of sooty mangabeys in the wild (13). For the rest of this discussion HIV-1 refers to HIV-1 group M and HIV-2 refers to HIV-2 group A or B.

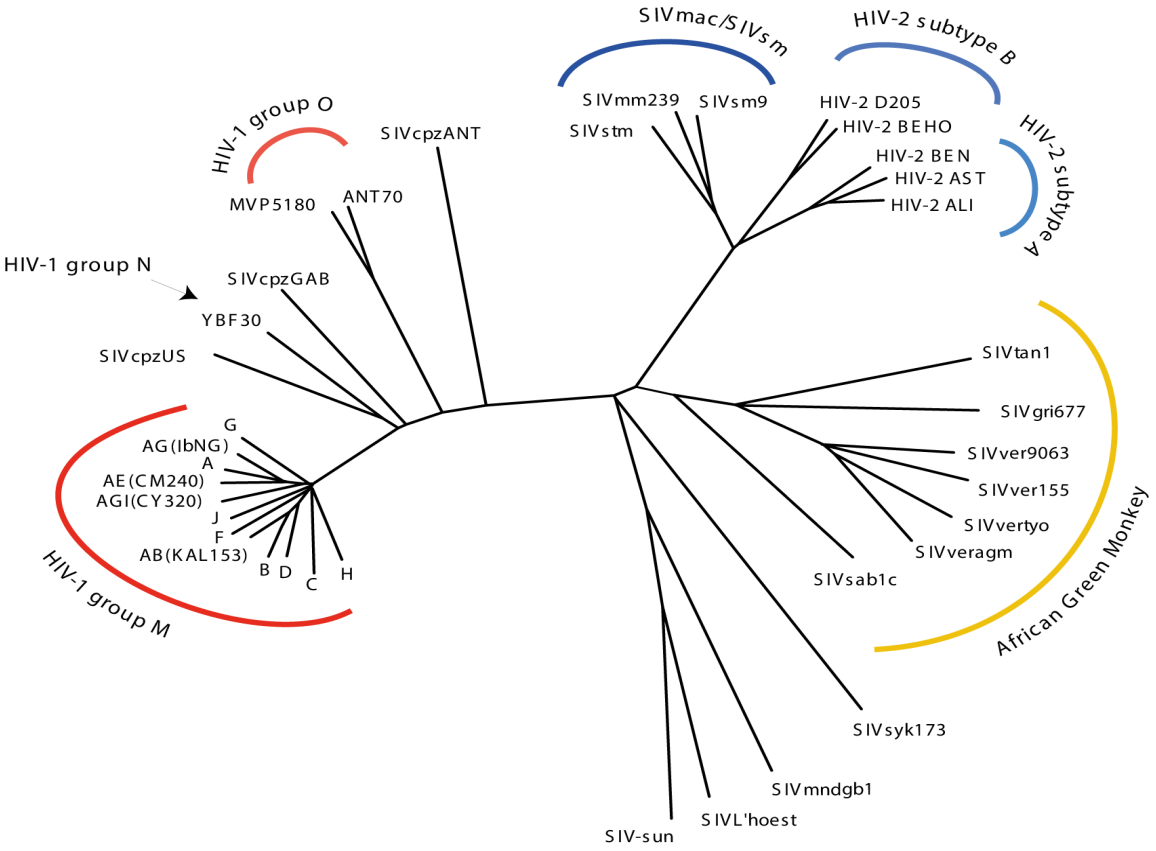


Figure 1.1. Phylogenetic tree of HIV and SIV sequences: HIV and SIV reference sequences are shown in an unrooted phylogenetic tree. HIV-1 groups cluster with SIVcpz sequences on the left (red) and HIV-2 groups cluster with the SIVsmm sequences top right (blue). SIV sequences from non-human primates are shown at the bottom right (yellow). Adapted with permission from (14)

HIV-1 and 2 are spread horizontally, that is sexually through bodily fluids, through blood transfusions or contaminated hypodermic needles and syringes. They are also transmitted vertically between woman and their children, either prenatally during pregnancy, perinatally during the delivery of the neonate or postnatally via breastfeeding. HIV-2 spreads far less efficiently than HIV-1 and is associated with lower concentration of viral ribonucleic acid (RNA) in bodily fluids (15).

HIV-1 is likely to have emerged in humans the early 1920s in Central Africa (16). Several factors coalesced in the surrounding decades which amplified the spread of HIV-1. Firstly, economic expansion under colonialism had brought distant communities closer together via roads and rail. This further facilitated greater movement of people and concentration of younger people in cities. During this time there was an ongoing epidemic of genital ulcer disease (GUD) in several West and Central African countries coupled with low male circumcision rates – this greatly increases the likelihood of onward transmission of HIV (17). In addition, mass vaccination campaigns and parenteral treatment of tropical diseases using unsterile procedures may have amplified spread of the virus (18). Historical records show that the number of cases of cryptococcal meningitis increased from the 1950s throughout the 20th century in the Congo River Basin – this was probably caused by the emerging AIDS epidemic (19). From Central Africa HIV-1 spread globally, with the first cases attributed to it in the 1980s in the United States of America.

HIV-2 was first isolated in 1986 from healthy commercial sex workers in Senegal and named human T-lymphotropic virus IV (HTLV-IV) (20). Shortly afterwards a similar virus (named Lymphadenopathy associated virus-II) was isolated from two West African patients with AIDS and renamed HIV-2. Subsequent molecular characterisation showed that HIV-2 was related to HIV-1 but was closer to SIV derived from macaques displaying an AIDS-like syndrome (20). HIV-2 group A probably entered the human population in approximately 1938 in Guinea-Bissau (21). Guinea-Bissau has a complicated recent history, starting the 20th century as a Portuguese colony, then fighting a protracted war for independence from 1963 – 1974, followed by several decades of political instability culminating in a civil war from 1998 - 1999 (**figure 1.2**). After its emergence in Caió, HIV-2 initially spread slowly in Guinea-Bissau then its incidence accelerated during the war for

independence (21). The accelerated spread was most likely to have been driven by an influx of Portuguese soldiers into Bissau and a concomitant increase in commercial sex-workers (CSW) – the majority of whom were from Caió (22). Evidence links the lower circumcision rates in the occupying soldiers with amplifying the spread of HIV-2 within Bissau and later to Portugal (23). Furthermore, around this time mass vaccination campaigns were undertaken and, through unsterile injections, may have accelerated HIV-2 spread (24,25). Blood transfusions and surgeries with unsterile procedures contributed to further spreading HIV-2 (26,27). Guinea-Bissau is where many of the long-term studies of HIV-2 have been conducted, and it is these studies which delineate the main differences between clinical outcomes of HIV-1 and HIV-2 infection.

An estimated 1-2 million people are infected with HIV-2 today (28). However, as HIV-2 is not regularly included in national testing strategies and requires specialist laboratory facilities for accurate diagnosis the current prevalence is unknown. The recent falls in national HIV-2 prevalence in some West African countries (29) have led to predictions that the epidemic will reach extinction around 2068 (**figure 1.2**) (30,31).

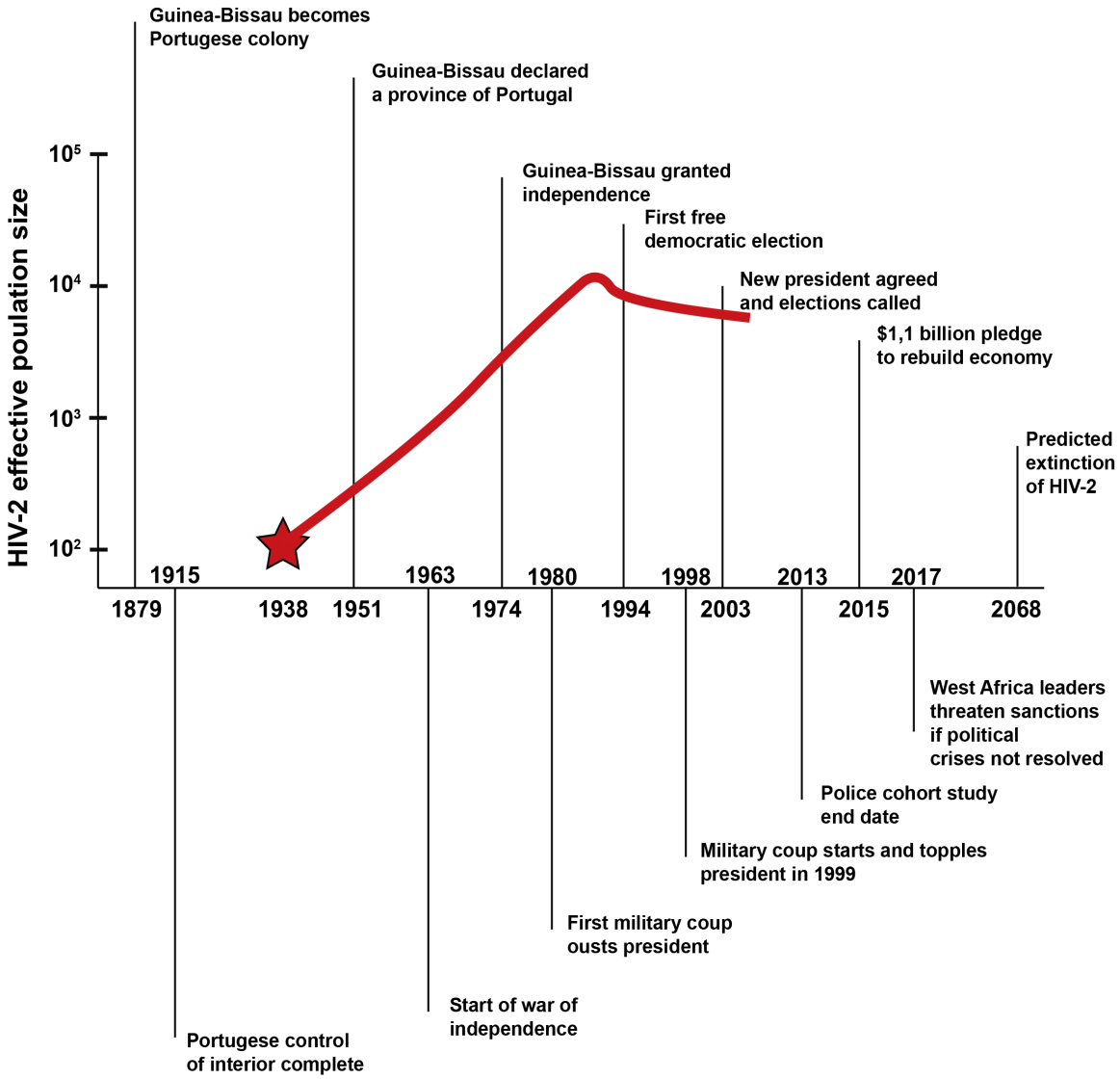


Figure 1.2. Timeline of recent history of Guinea-Bissau: Major political events of Guinea-Bissau's recent history are shown on the x axis from colonial settlement in 1879 until 2017 (32). The Guinea-Bissau Police cohort was recruited from 1990 till 2013, with five years of suspended work due to the civil war 1998 – 2003. The y-axis shows the HIV-2 effective population size from emergence around 1938 (shown by the star) (21).

1.2 HIV structure and genome

Human Immunodeficiency viruses are Retroviruses from the Lentivirus genus. Lentiviruses are characterised by their long incubation period from infection to disease (*lente* in Latin means "slow") and retroviruses are those which reverse transcribe their genomes (RNA to DNA). The HIV genome is diploid with two, positive sense, single-stranded RNA (ssRNA)

genomes packaged within each virion (**figure 1.3**). This diploid genome provides a source of genetic diversity *via* recombination, which occurs when the ssRNA genome is reverse transcribed into DNA.

The HIV virion is spherical in shape and is covered in an envelope with a lipid bilayer which is acquired from host cells when it is released. This outer envelope is stabilised by matrix proteins. The viral capsid is contained within the envelope and houses the viral genome and enzymes involved in reverse transcription, integration and replication. In addition, several different proteins which antagonise host defences (e.g. Vpx) are transported in the virion**.**

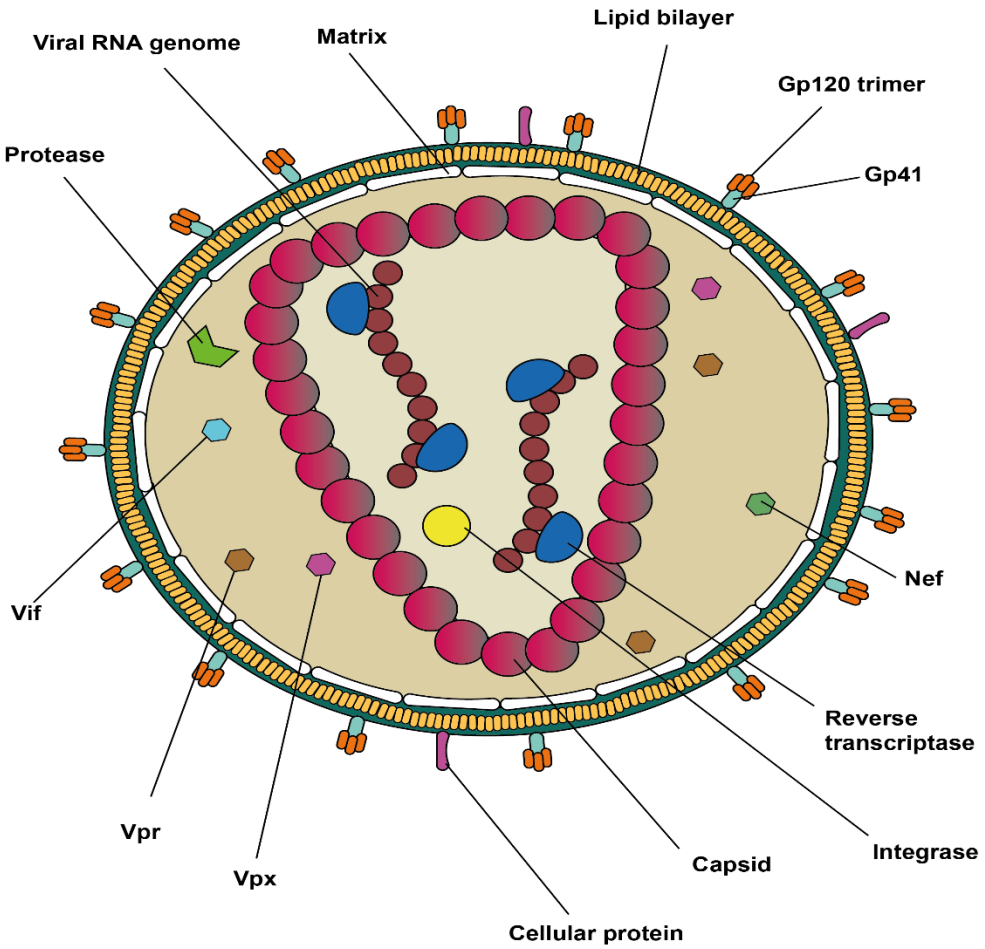


Figure 1.3. HIV-2 virion: HIV is a spherical virus covered in an envelope. Within the virion is the viral capsid which contains the viral genome and several different proteins and enzymes which are required for integration of the viral DNA into the host cell genome and further replication.

Each genome is an open reading frame (ORF) capped by a 5' and 3' long terminal repeat (LTR) region (**figure 1.4**). The genome contains three large genes which code for the major structural and functional proteins (*gag*, *pol* and *env*) and several smaller genes which code for accessory proteins. *gag* (group-specific antigen) codes for structural proteins which make up the matrix (MA), nucleocapsid (NC) and capsid (CA/p24) (**table 1.1**). The MA protein stabilises the spherical shape of the virion and outer envelope, the nucleocapsid is part of the RNA genome structure within the capsid. The p24 protein monomers organise into a hexagonal lattice in the shape of a fullerene cone made of 216 hexamers and 12 pentamers (33).

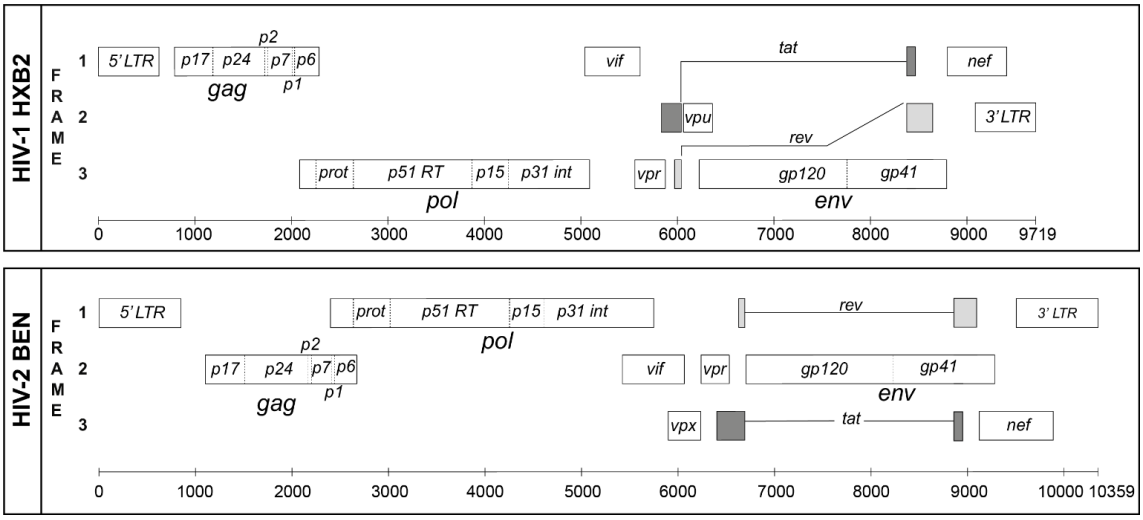


Figure 1.4. HIV-1 and HIV-2 gene maps: HIV genomes' organisation is shown as ORF read from 5' to 3'. These maps are for reference sequences HIV-1 HXB2 above and HIV-2 BEN below. Nucleotide positions for each gene are shown at the bottom of each map. Reproduced with permission from (34)

The *pol* gene codes for reverse transcriptase (RT), protease (PR) and integrase (IN). These enzymes are crucial to the replication cycle of HIV. Reverse transcriptase transcribes the ssRNA genomes into proviral double-stranded DNA (dsDNA). This proviral DNA genome is transported to the infected cell's nucleus in the capsid and integrated, by IN, into the host cell's genome. These enzymes are the chief targets of antiretroviral therapy (ART). After

the proviral genome is transcribed into RNA by the host cell, the precursor protein molecules are proteolytically cleaved by PR into their constituent proteins (**table 1.1**).

The envelope glycoproteins Gp120 and Gp41 are encoded by *env*. Glycoprotein 120 forms a trimer which is anchored to the virion by Gp41; together they attach the virion to a target cell's cluster of differentiation 4 (CD4) receptor and facilitate viral entry.

Table 1.1: HIV-2 genes, proteins and their respective functions in the lifecycle

Gene	**Size** [1]	**Protein**	**Function**
gag		Pr55Gag	precursor of the inner structural proteins
	p24	capsid (CA)	formation of conical capsid
	p17	matrix (MA)	Forms inner membrane layer
	p7	nucleoprotein (NC)	formation of the nucleoprotein/RNA complex
	p6		involved in virus particle release
pol		Pr160GagPol	precursor of the viral enzymes
	p10	protease (PR)	proteolytic cleavage of Gag and Gag-Pol precursor protein
	p51	reverse transcriptase (RT)	transcription of HIV RNA in proviral DNA
	p15 (66)	RNase H	degradation of viral RNA in the viral RNA/DNA replication complex
	p32	integrase (IN)	integration of proviral DNA into the host genome
env		PrGp160	precursor of the envelope Gp120 and Gp41, cleavage by cellular protease
	Gp120	surface glycoprotein	attachment of virus to the target cell CD4 receptor
	Gp41	transmembrane protein	anchorage of Gp120 to virion, fusion of viral and cell membrane
tat	p14	transactivator protein	activator of transcription of viral genes
rev	p19	RNA splicing regulator	regulates the export of non-spliced and partially spliced viral mRNA
nef	p27	negative regulating factor	influence on HIV replication, enhancement of infectivity of viral particles, downregulation of CD4 on target cells and HLA on target cells
vif	p23	viral infectivity protein	critical for infectious virus production *in vivo*
vpr	p15	virus protein r	component of virus particles, interaction with p6, facilitates virus infectivity, effect on the cell cycle
vpx	p15	virus protein x	interaction with p6 in virus particles, antagonises SAMHD1 [2]
tev	p26	tat/rev protein	Tat-Env-Rev fusion protein, regulates the activity of Tat and Rev in nucleus

[1] Numbers correspond to the size of the proteins (p) or glycoproteins (gp) in 1,000 Da. [2] SAMHDI = SAM domain and HD domain-containing protein 1, a cellular protein which inhibits replication of HIV. Reproduced with permission from (34)

1.3 HIV replication cycle

The HIV replication cycle begins with attachment of the envelope proteins to the host cell. First the virion is brought into close contact with the cell membrane by a variety of attachment factors (e.g. dendritic cell–specific intercellular adhesion molecular 3-grabbing non-integrin (DC-SIGN)) (35). Next the Gp120 trimer binds to the host CD4 receptor followed by binding to coreceptor proteins. To facilitate fusion HIV-1 uses host cell co-receptors CCR5 or CXCR4; HIV-2 is able to use a wider range of co-receptors for this step (36). After fusion of the virion and host cell membrane is complete, the viral capsid is released in the cytoplasm (**figure 1.5**).

After an HIV virion enters a cell the capsid houses the viral RNA and facilitates reverse transcription during transit to the nucleus – this pre-integration complex (PIC) consists of the replicating viral genome, RT, IN, Vpr and MA (37,38). HIV-1 capsids prevent both innate immune sensing of viral dsDNA by the CGAS-STING (cyclic GMP–AMP synthase and stimulator of interferon genes) pathway as well as DNA degradation and sensing by endogenous DNase enzymes such as TREX1 (three prime repair exonuclease 1) (39,40). After entry the capsid is targeted by a range of host restriction factors which aim to prevent further replication. These include tripartite motif containing protein, TRIM5α, as well as myxovirus resistance 2 (Mx2/MxB) (41,42). HIV capsids recruit cyclophilin A (CypA) at this time to evade TRIM5α and successfully enter the nucleus (43).

To pass from the cytoplasm into the nucleus the PIC crosses nuclear pore complexes (NPC) which are made up of a variety of nuclear pore proteins (Nups) (44). Two Nups which are relevant to HIV replication are Nup153 and Nup358 (43). These two proteins form a channel on the nuclear membrane and facilitate PIC entry and uncoating into the nucleus

in a CypA dependent manner (44). Experimentally blocking this specific pathway using cyclosporine A to deplete CypA causes HIV to use alternative nuclear entry pathways. This affects the choice of integration site for the viral genome, often integrating into less transcriptionally active sites (44).

After integration the proviral genome is transcribed into RNA by the host cells enzymes as HIV lacks RNA-polymerase II. Transcription is induced by transcription factors such as Sp1 (specificity protein one) and NFKB (nuclear factor kappa-light-chain-enhancer of activated B cells) binding to the LTR region (45). Viral accessory protein Tat also regulates proviral genome transcription (46). The *gag, pol* and *env* genes are translated into large precursor proteins which are then proteolytically cleaved by the protease enzyme into their constituents. These are then packaged into virions during the assembly and budding phase.

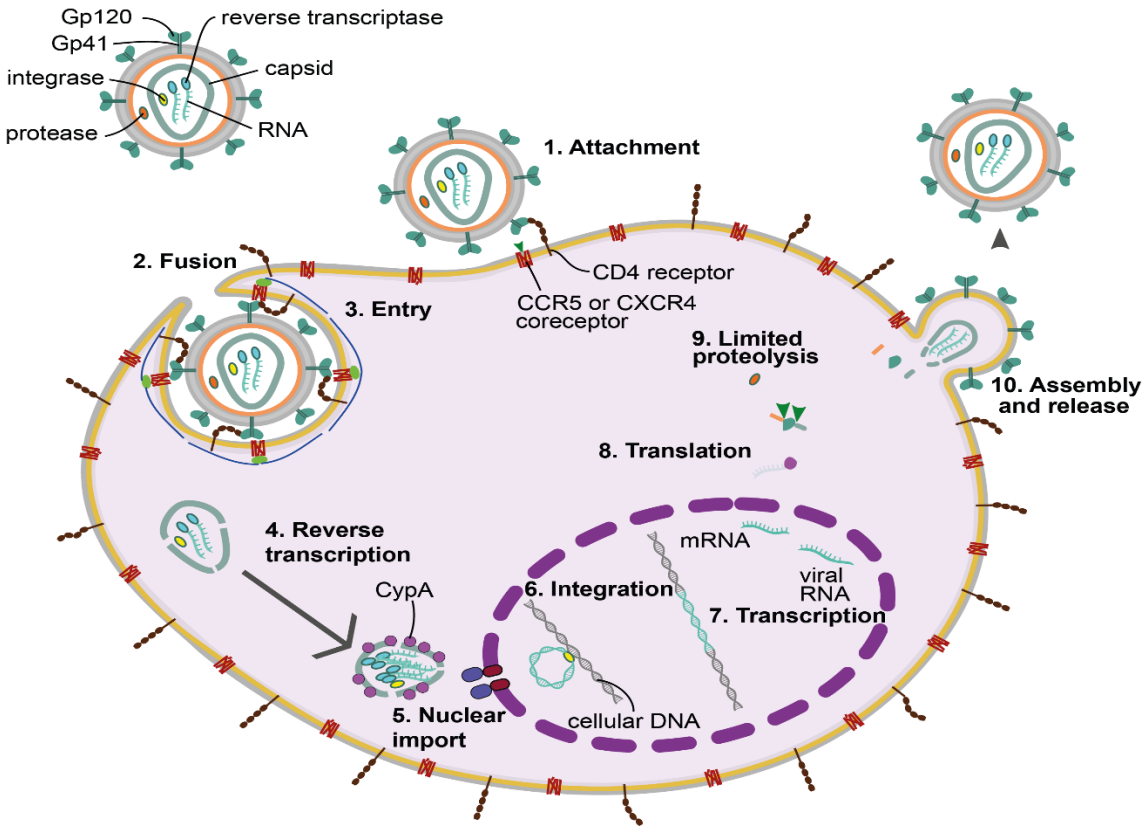


Figure 1.5. HIV replication cycle: The replication cycle of HIV-1 and HIV-2 are similar. Steps start at 1. Attachment and progress sequentially to 10. Assembly and release. Adapted with permission from (47)

1.4 HIV disease: from infection to AIDS

The authors of the 1981 Centre for Disease Control report which documented the first cases of AIDS in the USA made the following conclusion: "All the above observations suggest the possibility of a cellular-immune dysfunction related to a common exposure that predisposes individuals to opportunistic infections such as pneumocystosis and candidiasis". These prescient words summarise the hallmark of HIV disease. After infection with a (usually single) founder virus there is progressive spread to target cells (48). With ongoing viral replication there is associated immune activation and depletion of $CD4^+$ T lymphocytes in the mucosa and lymphoid organs which is reflected in a rapid decline in circulating $CD4^+$ T cells (**figure 1.6**). This eventually results in AIDS, and death as a result is a near certainty in the absence of ART. Progression to AIDS from infection is highly variable and occurs as follows:

The acute phase: After HIV infection the founder virus propagates and creates a population of recombinants and mutants – this is the viral quasispecies. Over the first three to six days there is further infection of localised $CD4^+$ T cells and the draining lymph nodes (49). In the following weeks there is ongoing systemic spread, this coincides with a profound decrease in gut-associated mucosal $CD4^+$ T cells, a peak in plasma viral load (measured as copies of viral RNA per millilitre of plasma) and a drop in peripheral $CD4^+$ T cells (50)(**figure 1.6**). This is also the time when the latent viral reservoir is established (51). The acute phase is often associated with a flu-like illness - fever, fatigue and muscle pain are the most common symptoms (52). The acute phase ends when an HIV-specific cytotoxic T lymphocyte (CTL) response is mounted, this leads to a fall in pVL and the selection of CTL escape variants (53). These escape variants then trigger new CTL

responses to different epitopes and this host-pathogen dynamic results in an eventual setpoint pVL (SPVL).

The asymptomatic phase: After the SPVL has been established the peripheral $CD4^+$ T cell population reconstitutes; immune activation remains high and slowly increases over time in tandem with the pVL (**figure 1.6**). The pathophysiology of the constant $CD4^+$ depletion is best described as "a slowly decaying unstable equilibrium between T cell activation, death, and renewal and the production and removal of virus" (54). During this asymptomatic phase synonymous and nonsynonymous viral divergence increases in a relatively linear way (55). Viral diversity accumulates with time, correlates positively with viral fitness and may in turn drive disease progression (56). The chronic immune activation associated with this phase causes end-organ damage of the kidneys, central nervous system and cardiovascular system, which is only partially reversed with ART (57). As the $CD4^+$ T cell count declines the risk of opportunistic infections increases exponentially (58).

AIDS: Once the $CD4^+$ T cell count falls past a certain threshold ($CD4^+$ count less than 200 cells/μL or CD4% less than 14%) or an individual develops an AIDS-defining illness, then the patient has developed AIDS (59). This is the end stage of HIV infection and is associated with low $CD4^+$ counts, an increasing HIV pVL and widespread immune dysfunction. The majority of patients who develop AIDS will die within two years if ART is not initiated (60).

The most striking difference between HIV-1 and HIV-2 is their disease progression rates. In a study of patients with estimated dates of infection the median time to AIDS after HIV-2 infection was approximately 14 years and after HIV-1 infection, six years (61). Mean CD4% in early infection was significantly higher in HIV-2 than HIV-1 (28% versus 22%) and the CD4% decline rate was 0.4/year for HIV-2 and 0.9/year for HIV-1. Intriguingly,

HIV-2 infected patients showed signs of clinical AIDS at a higher mean CD4% than HIV-1 infected patients.

The representations of disease progression shown in figure 1.6 are idealised models, these averages belie an extraordinary amount of inter-individual variability. In the following sections we discuss the viral, immunological and host factors which impact on HIV disease progression. We will focus on HIV-2 infection and how this differs from HIV-1.

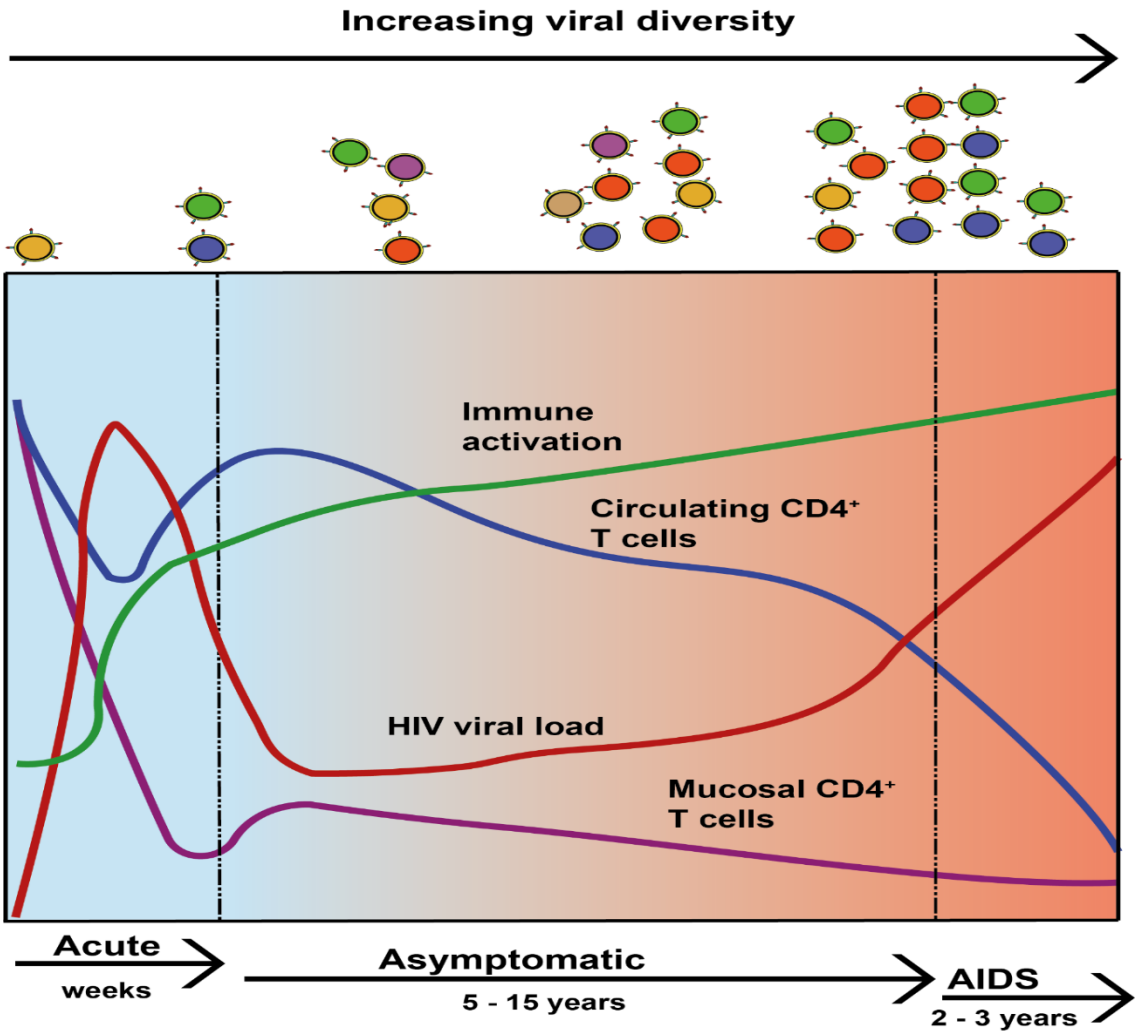


Figure 1.6. HIV disease progression from infection to AIDS: Representation of the average kinetics of HIV pVL (red), peripheral $CD4^+$ T cell count in cells/μL (blue) and mucosal $CD4^+$ density (purple) are shown. Overlying this is the average level of immune activation (green) which can be measured in concentrations of soluble biomarkers in plasma (e.g. β2-microglobulin or soluble CD4) or by T cell phenotyping to measure activation markers. From infection with a single founder virus, viral diversity increases in a relatively linear fashion (top)(55). Adapted with permission from (62)

1.5 Viral factors associated with disease progression

Viral load, replicative capacity, genetic diversity and intrahost evolution rates have all been linked to differences in disease outcomes in HIV infection and are discussed in turn.

Plasma viral load: In HIV-1 infection pVL correlates negatively with $CD4^+$ T cell counts and increasingly high pVLs predict faster time to AIDS and death (63,64). $CD4^+$ T cell decline rates vary widely between patients and a small amount of the variation in these rates is explained by pVL (65). HIV-1 pVLs in acute infection show greater effects on $CD4^+$ T cell decline rate at the extremes of the spectrum: patients with a pVL less than 1000 copies/mL have little evidence of $CD4^+$ T cell decline after 6 years and those with a pVL above 100 000 copies/mL all had a $CD4^+$ T cell count less than 300 cells/μL within three years of infection (66).

Across multiple studies, HIV-2 infection is associated with lower pVLs than seen in HIV-1 (67,68). HIV-2 pVL correlates negatively with CD4% - as in HIV-1 infection. The CD4% decline rate approximately doubles with each log10 increase in pVL and survival estimates diverge strongly at the extremes of pVL (69–71). A limitation of these studies is that the patients presented with an unknown date of infection, and therefore it is possible that these associations were biased by recruiting patients at various stages of disease progression. One study of HIV-2 infected patients (n = 29) with estimated seroconversion dates (seroconverters) showed that pVL correlates negatively with $CD4^+$ T cell count in the first two years of infection and was approximately 28-fold lower than in HIV-1 seroconverters (72). Patients who were recruited with unknown seroconversion dates (seroprevalent patients), who were expected to be at a later stage of infection, had a four-fold higher median pVL and lower $CD4^+$ T cell count when compared to seroconverters. There are

unfortunately no studies of HIV-2 where longitudinal sampling of pVL has been done in a sufficiently large cohort to resolve these limitations.

Viral replicative capacity: In HIV-1 infected patients, a greater viral replicative capacity (RC) in the Transmitted/founder (T/f) strain is associated with greater immune activation, higher SPVL and faster disease progression (66). In early infection RC has a stronger association with immune activation markers and T cell exhaustion than with pVL (66). Replicative capacity is a heritable trait of the virus which remains relatively constant during early infection, whereas the pVL is affected by the immune response (73). The RC of Gag-MJ4 chimeric viruses correlates closely with that of full-length T/f molecular clones. This suggests that the *gag* sequences of the T/f strains influence replicative capacity. In an epidemiologically-linked transmission cohort, CTL pressure on *gag* in transmitted viruses negatively influenced viral RC (74). Residue substitutions in all *gag* regions influence RC, though few occur in the p24 region (74). Later studies identified residues in *pol* which also associated with differences in RC (75).

An alternative approach to measuring replicative capacity is to use competition assays with different viral strains. Using this method most HIV-2 isolates have lower RC than HIV-1. Clinical HIV-2 isolates from aviraemic patients have a lower RC than those from viraemic patients (76,77). This indicates that viral RC is probably an important contributor to pVL in HIV-2 infection.

Viral genetic diversity and divergence: HIV-1 diversity shows a strong correlation with pVL, fitness and disease progression but less is known about how HIV-2 diversity and divergence influence disease outcomes (56,78). Comparisons of HIV-1 and HIV-2 intrahost evolutionary rate estimates have reported conflicting results. Skar *et al* reported that HIV-2 had a greater evolutionary rate than HIV-1, specifically in the Gp125 region

that contains the V3 loop (79). Mutations were more likely to be synonymous and therefore due to viral factors (negative selection) as opposed to immunological (positive selection) pressure - escape mutants to neutralising antibodies were notably rare. In contrast, findings from a Senegalese cohort showed limited divergence of *env* (80). One potential reason for the disparity in the results of these studies is that the two cohort profiles were different. The Senegalese cohort were ART-naïve, with limited disease progression, and had a long follow-up time. Skar *et al* reported on *env* divergence in patients with mostly progressive HIV-2 infection. Many of these patients had high pVLs and low $CD4^+$ T cell counts. HIV-1 synonymous substitution rates correlate positively with pVL and negatively with the $CD4^+$ T cell decline rate (55). HIV-2 *env* evolutionary rates associate with CD4% kinetics such that evolutionary rates in fast disease progressors (lower CD4% in follow-up) are approximately double that of slow-progressors (higher CD4%)(81). Furthermore, positive selection pressure on envelope surface-exposed residues which are conserved in SIVsmm and HIV-2 was associated with slower disease progression. The authors speculated that this may indicate that adaptation at these sites had deleterious effects on envelope structure or function, impairing viral fitness and slowing progression to AIDS.

The following pattern emerges: individuals with controlled HIV-2 infection have viruses with a low replicative rate, low pVLs and diversity. Those that develop progressive disease have higher pVLs and greater diversity, with prominent negative selection pressure.

1.6 The HIV-2 capsid's impact on disease progression

There are several significant differences in how HIV-1 and HIV-2 capsids influence disease outcomes. Capsid structures are highly conserved amongst retroviruses and mutations in the p24 sequence may either reduce viral infectivity or increase the virus' susceptibility to host immune responses (82,83). HIV-2 is able to replicate in dendritic cells and stimulates innate immune sensing via its capsid in a CypA-dependent manner, which is in contrast to HIV-1 (84). HIV-2 capsid binds more strongly than HIV-1 to the NONO (Non-POU domain-containing octamer-binding protein) protein in the nucleus. NONO-binding promotes DNA sensing by CGAS and stimulates a potent interferon response (85). The capsid residues involved in NONO binding are highly conserved in primate lentiviruses (85). As noted earlier, capsids determine the pathway for nuclear import and influence the choice of integration sites for viral DNA (44).

Polymorphisms in HIV-2 p24 sequences have been implicated in multiple aspects of disease progression. P24 sequences differ between patients with low and high HIV-2 pVLs, an important determinant of immune activation and disease progression. Specifically, patients with low pVLs often displayed prolines in capsid at positions 119, 159 and 178. Prolines in these positions had an additive effect on decreasing viraemia (86). In addition, prolines in these positions are associated with enhanced proteasomal processing of a $CD8^+$ T cell Gag epitope 165-DRFYKSLRA-173. This epitope is associated with greater Gag-specific T cell responses as determined by ELISpot and is a frequent target for T cell responses in patients with low pVLs (87). Therefore, it is possible that p24 sequences directly influence adaptive immune responses. HIV-2 p24 sequences with a proline at position 119 are more sensitive to restriction by the antiviral protein TRIM5α (rhesus macaque) than sequences with alternative amino acids at this position. This enhanced

sensitivity to TRIM5α may be linked to conformational changes in the p24 protein (88,89). HIV-1 NL4-3 chimeras which contain HIV-2 p24 sequences are more susceptible to restriction by human TRIM5α than the wildtype virus, though proline substitutions at position 119 do not correlate with TRIM5α restriction (90). Furthermore, HIV-2 p24 sequences derived from rapid progressors and long term non-progressors do not differ in their sensitivity to TRIM5α – though they are all more sensitive than HIV-1 (43).

Despite the increased sensitivity of the HIV-2 capsid to TRIM5α restriction, as well as the strong cytotoxic T cell response directed towards Gag epitopes, there is little evidence of positive selection pressure on *gag* divergence in a cross-sectional analysis of viral sequences from the Caió cohort (91). Negative selection predominates in interhost evolution, and this reflects the constrained evolution of this gene.

A study of *env* and *gag* sequences showed that pVLs in phylogenetically linked HIV-2 infection clusters were discordant (92). Therefore, viral genetic factors, at least in *gag* and *env*, probably do not account for the full spectrum of disease progression in HIV-2 infection. This finding strengthens the hypothesis that interactions between viral, host genetic and immunological factors are collective determinants of disease outcomes in HIV-2 infection.

1.7 Host restriction factors and HIV

TRIM proteins: Tripartite interaction motif (TRIM) E3 ligases are a group of structurally related host restriction factors which play an important role in antiviral defence in many organisms. Their structure consists of a conserved modular tripartite motif structure consisting of a RING (Really Interesting New Gene) domain, followed by one or two zinc binding areas named B-box and a predicted coiled-coil (CC) region (93). Collectively the RING, B-box and Coiled-coil region are designated as the RBCC domain, which is conserved in *TRIM* genes and functionally interchangeable between paralogues (**table 1.2**) (94). In a subset of TRIM proteins a PRY-SPRY domain is found on the C-terminal domain. The PRY-SPRY domain contains four variable regions (V1 – V4) and usually mediates protein-protein interactions and is essential in antiviral responses (93,95). *TRIM5* and its paralogues *TRIM6, TRIM34* and *TRIM22* are found on chromosome 11 in a cluster (**figure 1.7**). TRIM5α primarily targets retroviruses whereas TRIM22 is involved in host defence against a wider range of viruses (95–98).

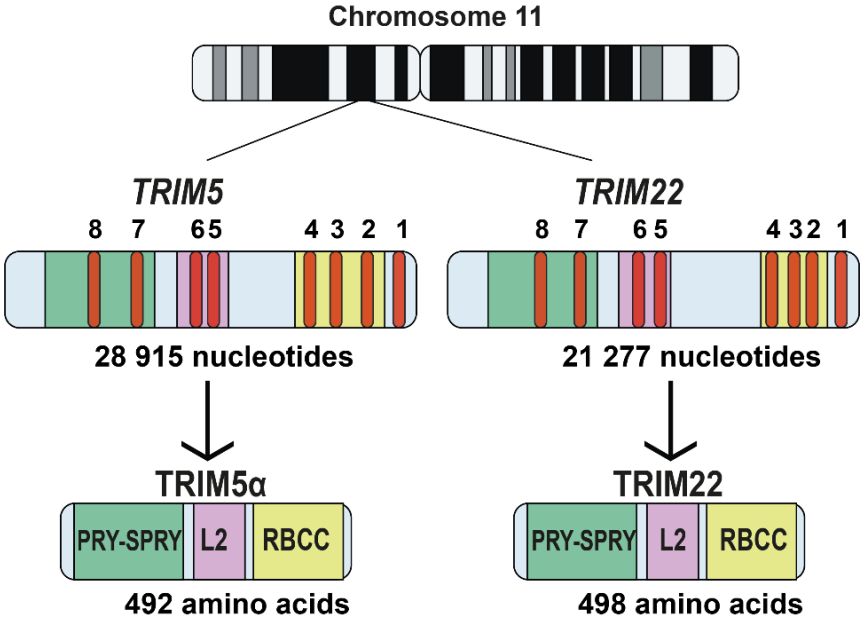


Figure 1.7: Genomic layout of *TRIM5* and *TRIM22*. Exons are shown in red within the gene's layout. Exon 1 is non-coding. Protein domains are coloured individually. L2 = linker 2 domain.

TRIM5α is an intracellular antiviral protein which disrupts the HIV life-cycle at multiple points (99). TRIM5α localises to the cytoplasm and its expression is induced by interferon (IFN) (99). *TRIM5* is transcribed into five isoforms and its antiviral activity is mediated by isoform TRIM5α. The other isoforms result in truncated proteins lacking a PRY-SPRY domain and inhibit TRIM5α activity, therefore the proportion of TRIM5α amongst total *TRIM5* transcripts may influence HIV restriction (100).

Host restriction factors play a key role in the prevention of cross-species movement of lentiviruses. As an example, the orthologue TRIM5α derived from rhesus macaques potently restricts HIV-1, whereas the human equivalent does not (101). *TRIM5* is one of the fastest evolving genes in mammals (102). It has evolved under positive selection pressure, probably due to various pathogens exerting selection pressure at different periods of primate to human evolution (103). Positive selection on residues in the V1 region of the PRY-SPRY domain is prominent in *TRIM5;* this V1 region determines viral capsid specificity and hence viral restriction (103).

TRIM5α's ability to restrict HIV-1 by targeting the viral capsid was reported in 2004 (96). The sensitivity of HIV-2 to TRIM5α is determined by a region of the capsid which corresponds to the cyclophilin binding loop in HIV-1 (104). Depletion of CypA enhances TRIM5α mediated restriction and reduces infectivity of HIV-2 (43). TRIM5α monomers bind at the interface of three capsid hexamers (**figure 1.8**). This binding site was shown for rhesus TRIM5α and HIV-1 and is associated with high affinity binding; it also determines species-specific restriction of TRIM5α orthologues (105).

The anti-retroviral mechanism of TRIM5α acts at the pre-integration phase of the HIV lifecycle (**figure 1.9**). TRIM5α in rhesus macaques is also capable of late viral restriction by incorporation into nascent HIV-2 virions, thereby decreasing the production of

functional new virus particles. This late restriction is a saturable process as increased virion production may overwhelm it (106). In feline cell lines permissive to HIV-2 infection, experimentally-induced expression of human TRIM5α restricts viral replication and TRIM5α -specific short interfering RNA rescues infectivity in resistant cells (104). Additional antiretroviral mechanisms include that TRIM5α binding to the incoming capsid interferes with nuclear entry *via* nucleoporin channels (107) and also stimulates TAK-1 phosphorylation, which activates the NFKB transcription pathway (108). This capsid-specific recognition then stimulates further downstream transcription of inflammatory cytokines, which enhances the antiviral response against HIV.

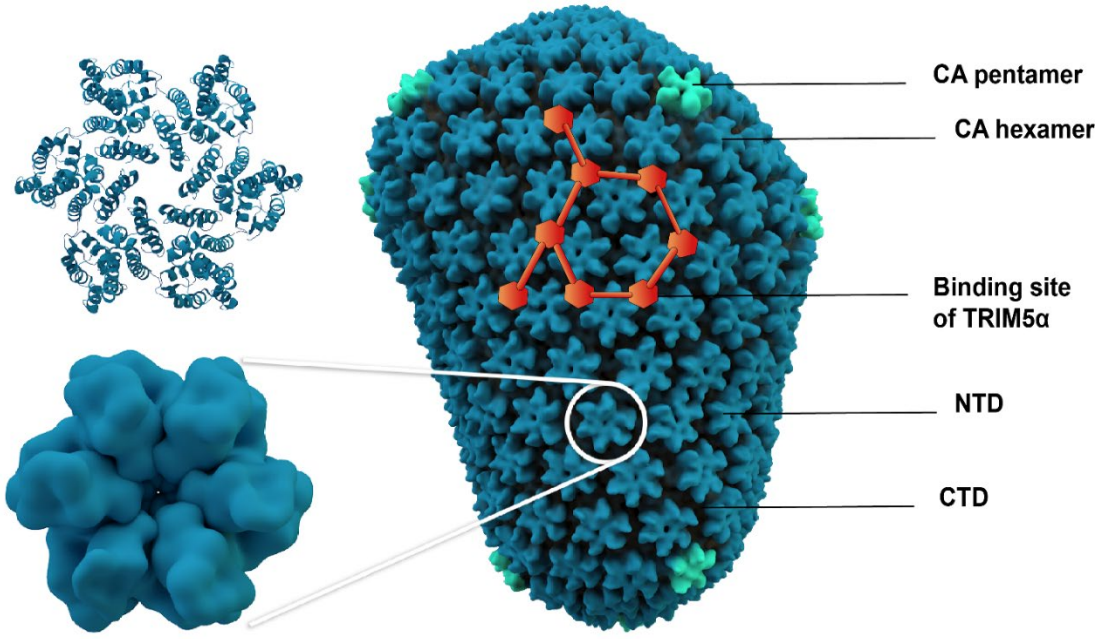


Figure 1.8. Model of HIV-1 capsid: p24 hexamers are shown on the left, with the larger fullerene cone structure on the right. HIV p24 pentamers are shown in turquoise and hexamers in blue. TRIM5α (red) binds at the p24 trimer interface then organises into a hexagonal matrix across the capsid. Adapted with permission from (109)

Despite high *TRIM* diversity, *TRIM5* genotype has minimal impact on HIV-1 disease progression (110). Moreover, TRIM5α does not appear to drive evolutionary change in HIV-1 p24 sequences from subtype B (111). HIV-infected patients who are homozygous

for the *TRIM5* H43Y substitution develop AIDS more quickly than heterozygotes or H43 homozygotes (112). The H43Y substitution impairs the restrictive capacity of TRIM5α (112). Furthermore, the appearance of putative TRIM5α viral escape variants has been found to precede rapid immunological decline in HIV-1 infection (113). Increased expression of TRIM5α in peripheral blood mononuclear cells is associated with a reduced risk of incident HIV-1 infection (114). A primate model of human HIV-2 infection using SIVsmm-infected Indian Macaques showed that *TRIM5* polymorphisms in the V1 loop of the PRY-SPRY domain were associated with increased memory $CD4^+$ cells and longer AIDS-free survival (115). This longitudinal study also demonstrated the development of escape mutants to TRIM5α, which could accelerate immunological decline, resulting in simian AIDS in the test subjects (115).

Table 1.2: Protein domains of TRIM5α and TRIM22

Domain	Exon	Residues	Function	Reference
TRIM5α				
RING	2	15 - 59	Self-ubiquitination via E3 ligase activity	(116,117)
B-box	2	90 - 132	B-box and coiled-coil domains cooperate to form higher order multimers	(118)
Coiled-coil	2 - 4	130 - 241		
Linker 2	5 - 6	241 - 281	Stabilising higher order multimers	(119)
PRYSPRY	6 - 8	281 - 493	Retroviral capsid recognition and binding.	(120)
TRIM22				
RING	2	15 - 60	Self-ubiquitination via E3 ligase activity	(121)
B-box	2	92 - 133	Higher order multimer formation. Polyubiquitination and activation of NOD-2	(122,123)
Coiled-coil	2 - 4	132 - 248		
Linker 2	4 - 6	249 - 282		
PRYSPRY	6 - 8	283 - 498	Protein binding; nuclear localisation	(124,125)

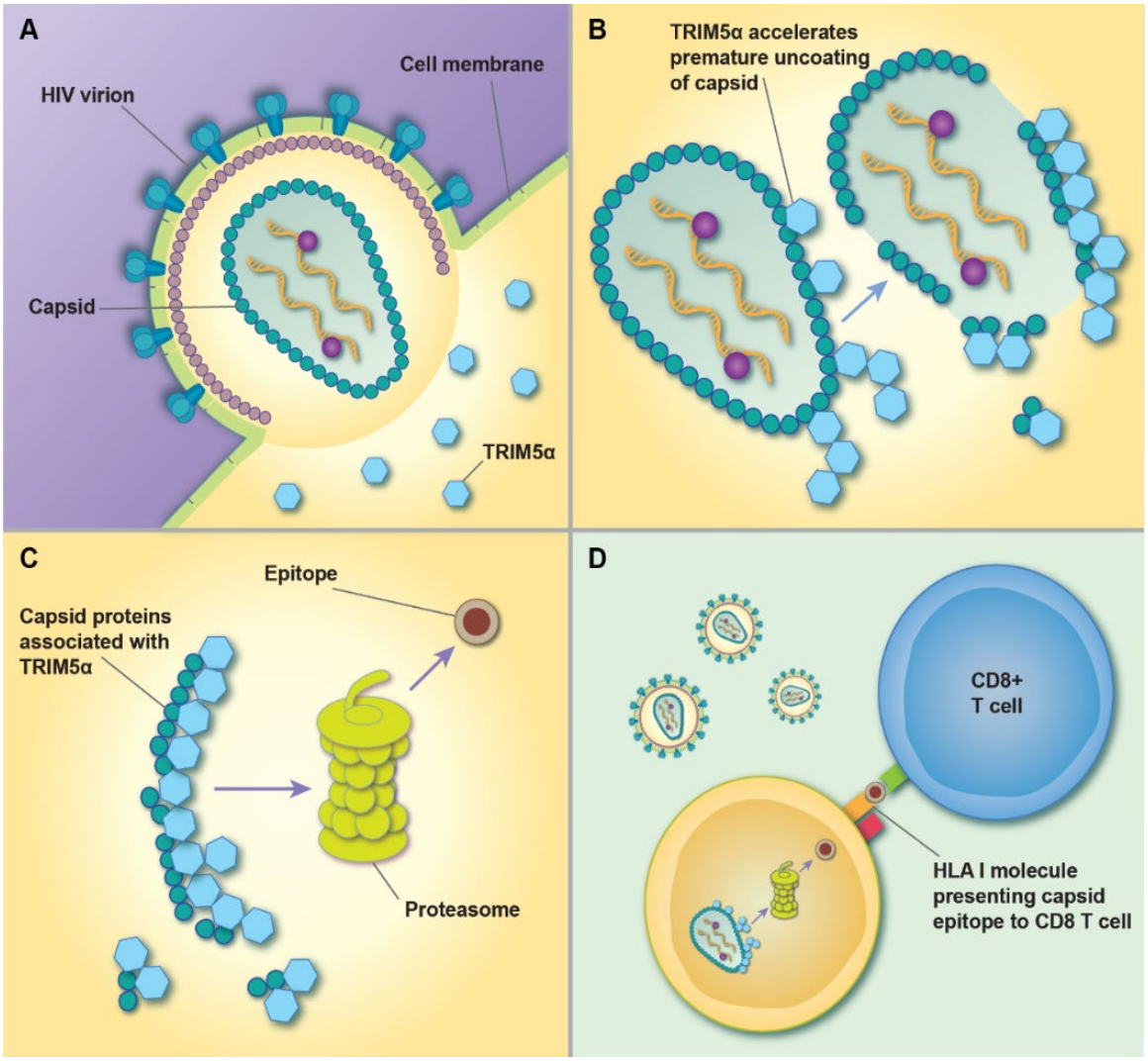


Figure 1.9. Mechanism of retroviral restriction by TRIM5α: A) after viral entry TRIM5α monomers binds to the capsid and oligomerise with other TRIM5α molecules. **B)** these are then poly-ubiquitinated via TRIM's E3 ligase activity. This prematurely uncoats the capsid, disrupting the pre-integration complex. **C)** The PIC is recruited to the cellular proteasome for degradation. **D)** The proteasomal processing selects for epitopes which are associated with protective Gag-specific $CD8^+$ T cell responses presented on HLA class I molecules (126,127). Reproduced with permission from (128)

TRIM22 is a paralogue of *TRIM5* and also an interferon-response gene (97,103). TRIM22 localises to the nucleus and inhibits post-integration replication of HIV-1 by forming a complex with two additional restriction factors, class II transactivator (CIITA) and promyelocytic leukaemia protein (PML), which blocks Sp1 driven transcription (97). This reduces downstream viral replication (129). *TRIM22* RNA expression correlates negatively with HIV-1 pVL in acute HIV-1 infection and correlates positively with type-1 IFN levels and $CD4^+$ T cell counts measured in the first year of infection (130). In the same patients these association were not observed for *TRIM5* expression. HIV-1 infected patients also had higher levels of *TRIM22* and lower levels of *TRIM5* expression when compared to uninfected controls (130).

TRIM22 is highly polymorphic and several single nucleotide polymorphisms (SNPs) alter its structure or function (131). Two SNPs in *TRIM22*, rs7935564 and rs1063303, have been associated with rapid progression to AIDS in HIV-1 infected adults (132). It is not known whether TRIM22 restricts HIV-2 or whether *TRIM22* genotype associates with disease outcomes in HIV-2 infected adults.

1.8 Host genetics and HIV-2 disease outcomes

The most robust immunogenetic associations with disease progression in HIV-1 are found in *HLA* and *KIR* genotypes and *CCR5* polymorphisms (133). Human leucocyte antigens (HLA) class I molecules are commonly distributed on all nucleated cell lines and provide an inhibitory signal to cytotoxic lymphocytes (CTL) in healthy cells by presenting self-antigens. When cells are infected with a virus the HLA molecules present viral epitopes to $CD8^+$ T cells which then destroy the infected cell. *KIR* genes code for Killer cell immunoglobulin-like receptors (KIR) which are found on Natural Killer (NK) cells. They are categorised as inhibitory or activatory, and constantly associate with HLA I molecules.

If an inhibitory KIR receptor receives sufficient stimulation from an HLA I molecule expressing self-antigens, then the NK cell will not activate. If HLA I molecules are downregulated this signal is lost and the activation threshold for the NK cell is lowered. This will cause the NK cell to activate and attack the offending cell.

Few studies have reported on the impact of HLA genotypes on HIV-2 infection outcomes. This reflects the paucity of well-characterised cohorts in West Africa, compounded by the variability of HLA repertoires present in different ethnic groups in the region (134). HLA class I associations with HIV-2 outcomes were strongest for HLA B*1503, which correlated with high pVLs and low $CD4^+$ T cell counts (135). Surprisingly HLA B*5703 did not correlate with low pVL, despite its strong association with long term non-progression in HIV-1 infection (135). In Senegalese female sex workers HLA-B35, together with the HLA-B35-Cw4 and HLA-A23-C7 haplotypes, were associated with increased risk of disease progression (136). Three additional HLA types have been associated with lower odds of having a detectable pVL in HIV-2 infection, HLA-B*58:01, HLA-DPB1*10:01 and HLA-DRB1*11:01 (137). However, these effects were not significant after correction for multiple comparisons.

Many associations between compound *HLA/KIR* genotypes and outcomes in HIV-1 infection have been reported, but in the Caió cohort there were no significant associations between individual *KIR* alleles or *HLA/KIR* genotypes and either the risk of infection with HIV-2 or HIV-2 disease progression (135).

An additional study from the Caió cohort reported that a haplotype for polymorphisms rs11575097-rs10849523 in the *CD4* gene was associated with high pVL in HIV-2 (138). In the same study, no correlations were found between polymorphisms in the *CD209* gene, encoding DC-SIGN, and HIV-2 susceptibility or progression.

1.9 T cell responses and HIV-2 disease outcomes

$CD4^+$ T helper cells: HIV-1-specific $CD4^+$ T cells are particularly susceptible to HIV-1 infection (139). HIV-2-specific $CD4^+$ T cells are as vulnerable to HIV-2 infection as their HIV-1-specific counterparts are to HIV-1 infection, however, the HIV-2-specific $CD4^+$ T cell response is relatively well-preserved in most patients despite this (139,140). The relative proportion of infected virus-specific $CD4^+$ T cells between HIV-2 progressor groups has not been investigated.

$CD4^+$ T cells from aviraemic HIV-2 infected patients show preserved proliferative and cytokine-releasing functions, IFN-γ and interleukin-2 being the dominant cytokines released (140,141). Aviraemic patients' $CD4^+$ T cells release a wider variety of cytokines (142), whereas in HIV-2 viraemic patients the functional profiles of HIV-specific $CD4^+$ T cells are similar to those seen in HIV-1 infection. Therefore, HIV-2 disease progression is associated with greater infection of HIV-2-specific $CD4^+$ populations, which may reduce T helper cell activity. A greater $CD4^+$ response against homologous Gag proteins correlates positively with immune activation and negatively with HIV-2 pVL. The host responses which promote immune activation may then also limit HIV-2 viral replication (143). Expanded T regulatory populations coupled with increasing PD1-PDL1 (Programmed cell death protein 1 and its ligand) expression on T cells have been observed in patients with advanced HIV-2 disease and low $CD4^+$ T cell counts, but these findings did not correlate with viraemia. This observation may indicate that aberrant immune tolerance mechanisms in the face of longstanding immune activation can drive HIV-2 disease progression (144–146).

Cytotoxic T cell responses: Many studies have implicated the HIV-1-specific $CD8^+$ CTL response as the major factor controlling viral replication from acute infection through the asymptomatic stage of infection (147). In HIV-1 the CTL response tends to be broad, targeting a wide variety of epitopes, but is largely undermined by escape mutants that evade CTL recognition (148). Despite the importance of the CTL response in HIV-1 control, the magnitude or breadth of their HIV-1-specific CTL response does not predict disease progression rates. Distinct characteristics of CTL found in long term non-progressors have been described, including restriction by "protective" HLA molecules such as HLA-B27 or B57, polyfunctionality, proliferative capacity, levels of cytolytic molecules and the ability to potently suppress HIV replication *in vitro* (149). In addition, greater targeting of HIV-1 Gag is associated with lower pVL (150).

The CTL response in HIV-2 is characterised as being narrow, consistently targeting a smaller number of epitopes than HIV-1 (151). CTL from HIV-2 infected subjects produce a wider range of cytokines, as well as greater quantities of individual cytokines per cell, compared to those from HIV-1 infected individuals (142). This is linked with the high functional avidity reported for some HIV-2-specific CTL responses (151). Using ELISpot to measure CTL responses to Gag peptides, aviraemic patients showed higher magnitude T cell responses than viraemic patients (152). Gag-specific CTL responses were shown to correlate strongly and inversely with pVL in HIV-2 infected subjects, with most viraemic patients lacking Gag-specific responses (which would be rare in HIV-1 infection) (153). $CD8^+$ T cells from aviraemic patients can potently suppress viral replication in $CD4^+$ T cells; these $CD8^+$ T cells have an early differentiation phenotype and are potent effectors (154).

Therefore, there are both quantitative and qualitative differences between CTL responses in HIV-2-infected subjects with and without detectable viraemia. The generation of a potent and effective CTL response may be central to limiting HIV-2 replication. Robust Gag-specific $CD8^+$ T cell responses are common in subjects with undetectable pVL, raising the question of how these CTL populations are maintained in the absence of detectable viraemia. HIV-2 replication is detectable in gut mucosa in patients who are otherwise aviraemic, and this may offer an explanation if CTL are stimulated through persistent HIV-2 replication in gut-associated lymphoid tissue (155).

1.10 Immune activation and HIV disease

Prolonged immune activation is thought to be the primary driver of HIV-1 disease. Immune activation also appears drive HIV-2 disease progression: the soluble activation marker beta-2 microglobulin (β2M) predicts HIV-2 disease progression as reliably as pVL in ART-naïve patients (156). T cell activation markers closely correlate with pVL in viraemic patients – however, there is a subgroup of patients with undetectable viraemia who exhibit elevated T cell activation (151). There has been speculation that the lower levels of immune activation in the many aviraemic HIV-2 infected patients constitute the key factor explaining their slower rates of disease progression (157). Levels of biomarkers associated with enhanced immune activation are equivalent in HIV-1 and HIV-2 infection and when pVL is taken into account, HIV-2 appears to elicit greater immune activation per unit of pVL than does HIV-1 (158). Moreover, immune activation in patients with HIV-1 and HIV-2 infection is equivalent when matched for $CD4^+$ T cell count (159).

Translocation into the blood of microbial products from the gut promotes immune activation in HIV-1-infected subjects - including lipopolysaccharide, which promotes

macrophage activation (160). Lipopolysaccharide levels were elevated in an African cohort with HIV-2 infection and correlated with disease progression (161). However, this was not observed in a European cohort, even though there was evidence of significant T cell and monocyte activation (162). In marked contrast to HIV-1 infection, mucosal integrity and gut-associated lymphocyte numbers are preserved in aviraemic HIV-2 infection (with or without ART), despite the presence of replicating virus in the gut mucosa (155). HIV-associated gut mucosal pathology may differ significantly between HIV-1 and HIV-2 infection, although how this might link with HIV-2 progression is not entirely clear.

1.11 Clinical factors associated with HIV-2 disease progression

Disease progression in HIV reflects a complex interplay between social factors and biological factors which influence how rapidly a person will develop AIDS after HIV infection.

Sex is an important factor in disease outcomes for both HIV-1 and HIV-2 and males tend to have worse outcomes then females (163). These findings may be explained, in part, by reduced health-seeking behaviour so they often present later in infection, together with poorer adherence to ART (164). In addition, baseline mortality rates due to non-infectious causes are higher in males and this contributes to the observed higher mortality rates in HIV-infected males (165).

Social factors do not fully account for the discrepancy in disease progression rates between males and females. Plasmacytoid dendritic cells from females produce higher transcription levels of several interferon-stimulated genes in response to IFNα than those from males

(166). This enhanced response may be beneficial in acute infection – limiting the set-point pVL (SPVL) and initial $CD4^+$ T cell count nadir (167). Females on average have lower SPVL coupled with an enhanced Type I IFN response and greater $CD8^+$ T cell activation – this may favour rapid $CD4^+$ T cell recovery, and improved outcomes after ART initiation (168,169). Statistical modelling of transmitted viral sequences indicates that transmission bottlenecks in heterosexuals exist, and that they favour the transmission of viruses with greater fitness from females to males (170). It has been postulated that because females have a large population of cells in the vaginal mucosa which are susceptible to HIV infection, they have a lower barrier to infection than males. This bottleneck selects viruses infecting males with greater predicted replicative capacity based on the presence of specific amino acid residues. The viruses which infect males may therefore predispose them to accelerated disease progression. HIV-1 infected males in Cameroon had an elevated risk of virological failure after ART initiation which was not linked to pre-existing viral resistance or poor adherence. Furthermore, females in this study showed greater $CD4^+$ T cell count recovery after two years of ART (171).

In HIV-1 advanced age at infection is a powerful co-factor for accelerated disease progression – which is probably due to immunological senescence (172). Males are often older than females when they initiate ART which may explain their diminished immunological reconstitution (173). HIV-2 prevalence in Guinea-Bissau was highest amongst females in the 45-60-year age group (174) and in the pre-ART era older HIV-2-infected adults had mortality rates equivalent to the HIV negative population, indicating that age was not a risk factor for progression in HIV-2 infection (175). In contrast, data from the ANRS cohort showed that age greater than 40 years was a significant risk factor for HIV-2 disease progression after ART initiation (176).

1.12 Antiretroviral therapy

The first antiretroviral agent used to treat HIV was zidovudine (AZT), a nucleoside reverse transcriptase inhibitor (NRTI), which was licenced in 1987 (177). This first successful treatment of HIV accelerated the development of further NRTIs to treat HIV. However, viral resistance to ART was reported shortly after more widespread use of AZT (178), which indicated that ART monotherapy was not a viable long-term solution. In 1996 the first report of triple combination ART (two NRTIs plus a protease inhibitor) showed significantly slower disease progression when compared to dual NRTI therapy (179). Triple therapy was later shown to suppress the emergence of viral resistance and maintain an undetectable pVL – however, replication was still detectable in cellular compartments (180). Triple therapy was not a cure as the integrated proviral DNA remains present in reservoir sites, and stopping ART results in a rapid rebound of plasma viraemia in most patients (181). Since 1987 there have been many new antiretroviral agents approved for treatment of HIV-1 (**table 1.3**).

HIV-2 is inherently resistant to NNRTIs, Enfuvirtide and several PIs (182–184). Extensive drug resistance has also been documented in ART-experienced patients in Senegal (185). The treatment of HIV-2 is complicated by the lack of large randomised controlled trials (RCT) to inform drug choice, though observational cohort studies indicate that triple NRTI regimens are inferior to PI-based regimens in both $CD4^+$ T cell reconstitution and pVL suppression (186). Two open-label trials have shown that single tablet combinations of integrase inhibitor based regimens are safe and as effective as PI-based regimens (187,188). The First-Line Treatment for HIV-2 (FIT-2) trial has completed recruitment of patients and will determine whether three different regimens treat HIV-2 effectively (189).

Table 1.3: Antiretroviral therapy agents currently in use

Class	Example	Mechanism of action
NRTI	Tenofovir disproxil fumarate	Incorporates into replicating viral genome and terminates further reverse transcription
NNRTI	Efavirenz	Binds to RT enzyme and inhibits further replication
PI	Atazanavir	Blocks activity of protease
Integrase inhibitor	Dolutegravir	Blocks activity of integrase
Fusion inhibitor	Enfuvirtide	Binds to Gp41 and disrupts the final stage of viral fusion with the host cell
CCR5 inhibitor	Maraviroc	Block CCR5 coreceptors and prevents entry of R5 tropic viruses
Post-attachment inhibitor	Ibalizumab	Block CD4 receptors and prevents entry
Pharmacokinetic enhancer	Cobicistat	Inhibits host cytochrome P450 enzymes which metabolise elvitegravir (EVG) and increases concentrations of EVG

NRTI = nucleoside reverse transcriptase inhibitor. NNRTI = non-nucleoside reverse transcriptase inhibitor. PI = protease inhibitor. RT = reverse transcriptase.

1.13 Phylogenetic inference of HIV evolution

RNA-based viruses are some of the fastest evolving organisms on earth, with evolutionary rates orders of magnitude greater than DNA-based organisms (190,191). Evolution is modelled as a bifurcating process from parent to offspring which produces a tree-like pattern of cumulative genetic distance (**figure 1.10**). Several factors drive the evolution of HIV:

i) HIV has **notoriously error-prone replication** with no inherent proof-reading mechanism acting during reverse transcription. HIV replicates close to the upper limit of a proposed "error threshold", beyond which the accumulation of harmful mutations leads to the collapse of subsequent populations (192).

ii) **Selection pressure** from the host's immune response and the availability of target cells for replication drive viral evolution down specific pathways (193).

iii) **Recombination** between viral genomes during the reverse transcription process – this not only corrects errors during replication, but gives rise to the rapid generation of new variants (194). Recombination violates the tree-like pattern of genetic accumulation which most phylogenetic models assume (**figure 1.10**).

iv) HIV has a short **generation time** of approximately three days (time from release of a virion until it infects another cell and releases a new generation of viral particles) (195). This leads to the generation of an extraordinarily **large circulating population of viruses** in an infected individual – approximately $1.3x10^{10}$ new virions produced on a daily basis (195). This circulating population, the viral quasispecies, provides a source of constant diversity which allows HIV to explore sequence space as it navigates the host environment. Viral quasispecies benefit from greater diversity as subpopulations are able to cooperate in networks to ensure successful replication (192,196).

Phylogenetic inference of HIV uses evolutionary models and statistical inference methods to study how viral populations change over time – either within an individual (intrahost) or between infected individuals (interhost). Phylogenetic inference is stepwise process: first a collection of homologous sequences are aligned to each other, then a tree model is selected to show how these sequences have evolved and, lastly, statistical support is provided to the tree estimates and parameters to test hypotheses about the observed sequence evolution.

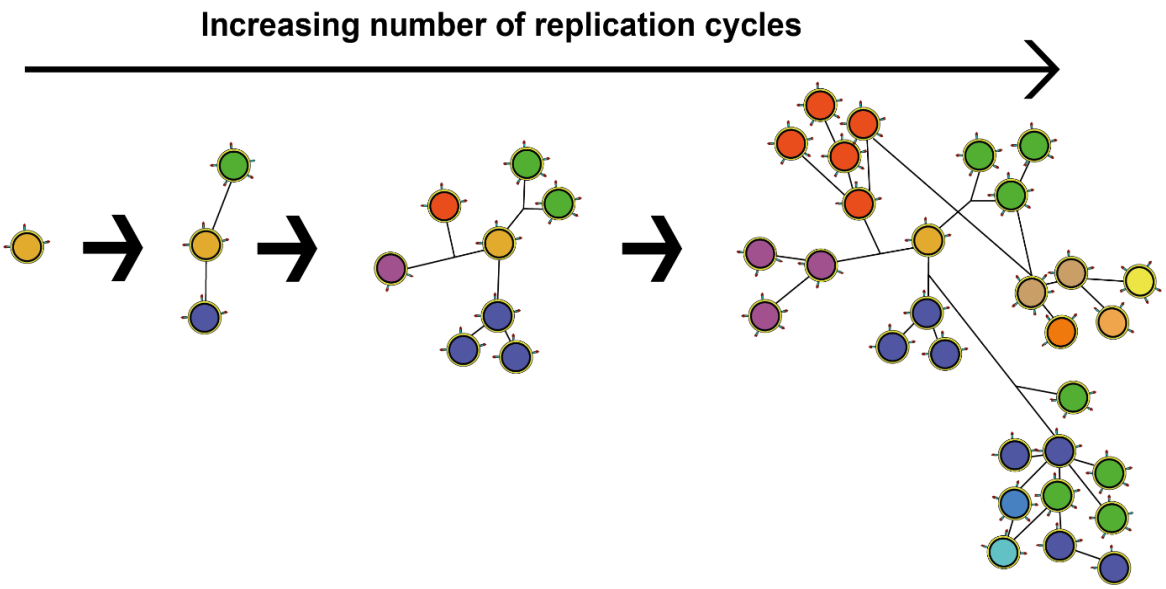


Figure 1.10: Evolution of the viral quasispecies which is usually modelled as a bifurcating, tree-like process (left and middle image). With each replication cycle new viral variants can be generated to found subpopulations in the quasispecies. This process is shaped by selection pressures. Recombination results in non-tree like patterns of evolution and the generation of new variants (far right image).

Multiple sequence alignment (MSA) is usually done via an algorithm which employs heuristic methods (e.g. MUSCLE or MAFFT) to create columns of nucleotides or amino acids from the sequence data matrix. Various alignment algorithms trade-off computational efficiency, accuracy and insertion of gaps into the alignment, and choosing a specific algorithm is dependent on the sequence length, relatedness of individual sequences, number of sequences and computational resources available. The goal of the MSA is to then determine how individual sequences have changed over time in relation to each other. The simplest way to do this is to quantify the genetic distance between sequences (the number of nucleotides that differ between a pair of sequences, iterated over an alignment). However, simple genetic distance estimates can be misleading. Individual sequences may have acquired a mutation (C → G) and then lost them again (G → C) – a distance-based tree-model would not account for this back mutation. Alternatively, two separate sequences

may acquire the same mutation in parallel; when their distance from each other is calculated they may appear closer than they are.

Nucleotide substation models are used to account for the variability in how mutations accumulate in an evolutionary model. These models vary in complexity and add free parameters to subsequent tree models (**table 1.4**). Nucleotide substitution models are time homogenous time-continuous stationary Markov models that use a Q matrix to specify the relative rates of change of each nucleotide along a sequence. Implicit in all substitution models are the following assumptions i) the rate of change from base *i* to base *j* is independent from the base that occupied the site prior to *i,* ii) substitution rates are constant over time and iii) the relative frequencies of A,C,T and G bases are in equilibrium.

Table 1.4: Nucleotide substitution models

Model	**Base frequencies**	**Substitution rates**	**Free parameters**
JC69	A = C = T = G	All rates equal	0
K80	A = C = T = G	Transition rate ≠ transversion rate	1
F81	A ≠ C ≠ T ≠ G	All rates equal	3
HKY85	A ≠ C ≠ T ≠ G	Transition rate ≠ transversion rate	4
GTR	A ≠ C ≠ T ≠ G	All rates unequal	8

JC69 = Jukes Cantor 1969; K80 = Kimura 1980; F81 = Felsenstein 1981; HKY85 = Hasegawa, Kishino and Yano 1985; GTR = General Time Reversible. Transitions are purine ↔ purine or pyrimidine ↔ pyrimidine substitutions; transversions are pyrimidine ↔ purine.

Once a MSA and substitution model have been selected the next step in phylogenetic inference is to produce an evolutionary model, that is a phylogenetic tree quantifying the relatedness of sequences in the MSA. Several methods to produce tree models have been proposed – reviewed in detail in (197).

Maximum Parsimony trees aim to select a tree where the fewest number of substitutions would have had to occur to explain the tree. This is done by assigning character states to

internal nodes on the tree. These models are simple and computationally inexpensive. However, their assumptions are not modifiable, they fail to account for multiple substitutions at the same site resulting in long branch attraction (198) and, most importantly, they are unable to incorporate additional prior knowledge about sequence evolution (e.g. assigning a molecular clock).

Distance methods use a matrix of nucleotide substitutions to minimise the difference between the calculated and expected distances or the sum of branch lengths. Examples of distance methods include least squares and minimum evolution (ME). Least squares methods use a linear regression of the branch lengths matrix, the tree with lowest score is the optimal tree. Minimum evolution chooses the tree with the lowest score after summing the branch lengths – this is done via two main algorithms. i) The unweighted pair group method with arithmetic mean (UPGMA) algorithm is a clustering method which uses ME to determine a phylogenetic tree. This algorithm assumes a strict molecular clock and results in an ultrametric tree where all lengths from root to tip are equal (199). ii) The Neighbour joining algorithm also uses agglomerative clustering to build a tree but does not assume a molecular clock model; it starts with an initial unresolved, star-like tree then iteratively adds pairs of sequences with the aim of minimising tree length (200). Limitations on distance methods include that highly divergent alignments with many gaps are difficult to resolve, NJ methods can return negative branch lengths and they cannot integrate uncertainty of the distance estimates (197).

Maximum likelihood uses likelihood functions to select a tree and its evolutionary model parameters (model) which has the highest probability of explaining the observed sequence evolution (data). The likelihood function asks the following question: **what is the probability of the data given the model parameters?** In maximum likelihood methods

the tree is the overarching model that is selected, and the branch lengths, substitutions models and molecular clock rates are specifiable parameters. The likelihood ratio test allows the comparison of many different model parameters to find the tree which fits the data the best. This method allows us to integrate many different known aspects of evolution into one model and simultaneously test specific hypotheses – e.g. which substitution model best explains the evolution of these sequences (197). This method is, however, computationally expensive and as the tree is the model itself it is difficult to use maximum likelihood to gauge the support for the topology.

Bayesian inference starts by asking the following question: **what is the probability of the model given the observed data?** Bayesian inference for model selection treats parameters and trees as variables with distributions and does not search for a single "best" parameter or tree but rather generates a large set of these, which is called the posterior distribution. Before the analysis prior distributions are assigned to these parameters which combine with the data to generate a posterior distribution.

Bayes theorem in phylogenetics is as follows (197):

$$P(T,\theta|D) = \frac{P(T,\theta)P(D|T,\theta)}{P(D)}$$

T, θ are the tree and model parameters with *D* the data and *P* is conditional probability. The equation translated into English is: the posterior probability is equal to the prior information (tree and parameters) plus the probability of the data given the prior information (which is equivalent to the likelihood) all divided by a normalising constant *P(D);* alternatively – the posterior distribution equals prior knowledge plus data. Markov Chain Monte Carlo algorithms are used to generate a posterior distribution and then iteratively explore this tree and parameter space. Bayesian inference uses the Metropolis-Hastings algorithm to explore

multidimensional space and draw random samples from the posterior distribution. This generates thousands of data points (estimated trees or molecular clock rates), which are used for statistical inference. In the first step of the algorithm a random set of parameters are drawn from the specified prior distribution and a random tree created. In the following iteration a new set is proposed and the probability of these parameters versus the previous iteration is determined (Bayesian inference uses the likelihood function as well). If the probability of the new state is higher it is automatically accepted as more accurate (**figure 1.11**). If the probability is lower, then the new state can be accepted, depending on how much worse it is. After many states (usually millions) have been tested the Markov chain may converge on a posterior distribution for the tree and parameters which is proportional to how often they were sampled. The credibility of the estimate is reported by the 95% highest posterior density (HPD), this is the interval which contains 95% of the sampled values.

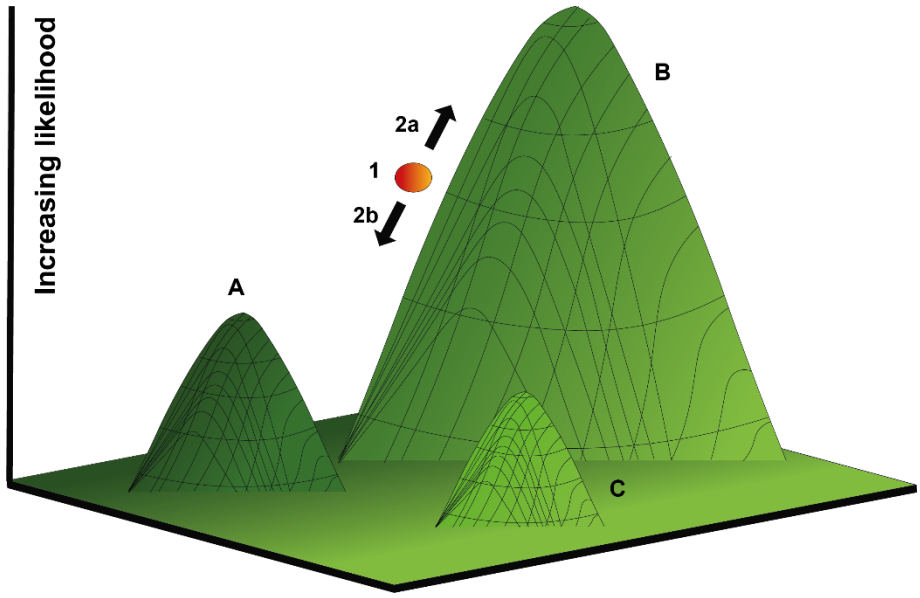


Figure 1.11: Bayesian inference uses likelihood function to search multidimensional tree and parameter space. A and C are local maxima where models can get "stuck" if prior limits don't allow movement outside these values. B represents the global maximum where the most probable model exists. 1 (red circle) shows the current state in the Markov chain, if the next state shows a higher probability over 1 then it is automatically accepted (2a). If the next state's probability is lower, then it is conditionally accepted (2b).

1.14 A model for slower disease progression in HIV-2

The preceding discussions show that HIV-2 disease progression to AIDS follows a similar trajectory to HIV-1, albeit at about half the rate (61). There are unfortunately few sufficiently powered studies to determine the events around acute infection and how these might influence longitudinal outcomes. What is agreed upon is that HIV-2 pVL are lower in all stages of disease than in HIV-1 infection, and that $CD4^+$ T cell decline rates are also slower. It would be reasonable to assume that levels of immune activation are also lower in most patients during the asymptomatic stages – given the positive correlations with HIV-2 pVL and disease progression rates (**figure 1.12**). In subsequent chapters we will describe the impact of *TRIM* genotype and intrahost evolution of p24 on HIV disease progression.

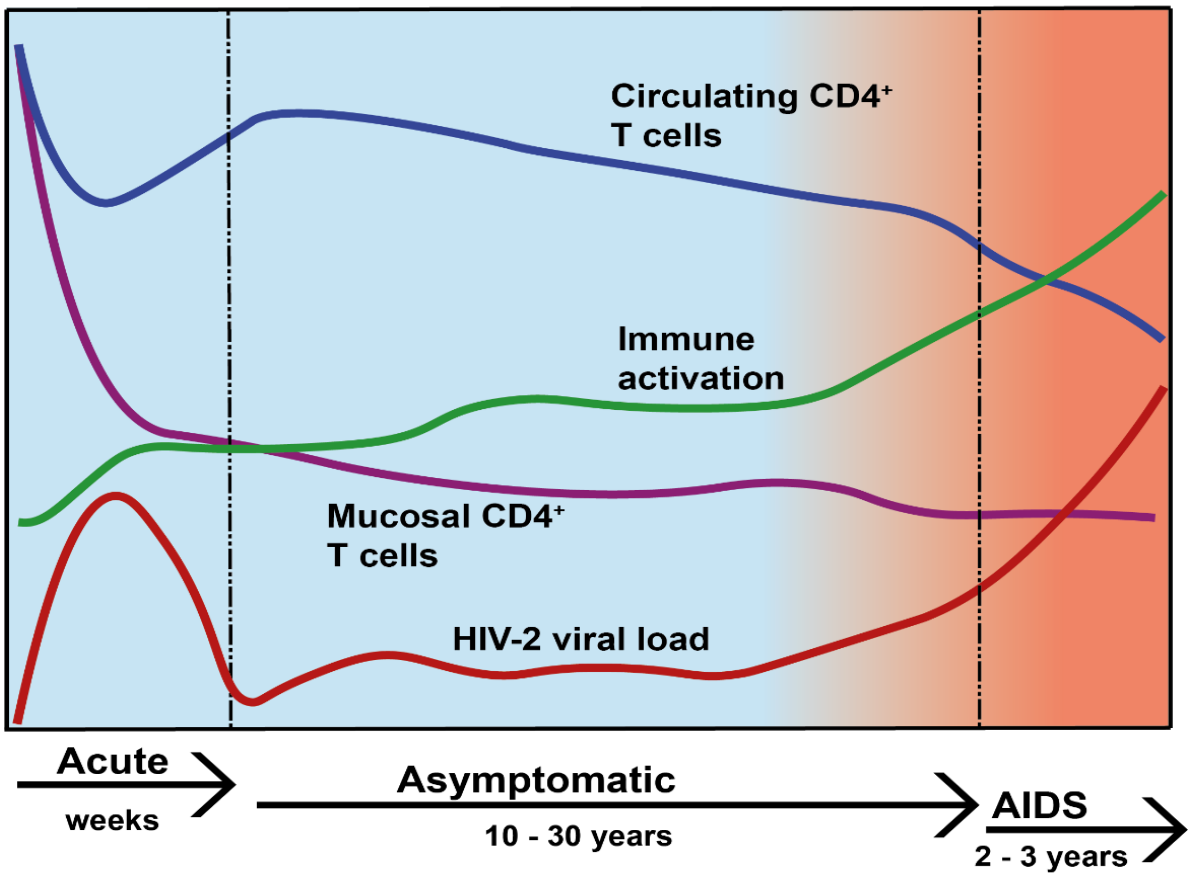


Figure 1.12. Slower disease progression in HIV-2 infection. Proposed longitudinal model of HIV-2 disease progression.

1.15 References for chapter 1

1. Centers for Disease Control (CDC). Pneumocystis pneumonia--Los Angeles. MMWR Morb Mortal Wkly Rep. 1981 Jun 5;30(21):250–2.
2. AIDSinfo | UNAIDS [Internet]. [cited 2019 Feb 1]. Available from: http://aidsinfo.unaids.org/
3. Gao F, Bailes E, Robertson DL, Chen Y, Rodenburg CM, Michael SF, et al. Origin of HIV-1 in the chimpanzee Pan troglodytes troglodytes. Nature. 1999 Feb 4;397(6718):436–41.
4. Santiago ML, Range F, Keele BF, Li Y, Bailes E, Bibollet-Ruche F, et al. Simian Immunodeficiency Virus Infection in Free-Ranging Sooty Mangabeys (Cercocebus atys atys) from the Tai Forest, Cote d'Ivoire: Implications for the Origin of Epidemic Human Immunodeficiency Virus Type 2. J Virol. 2005 Oct 1;79(19):12515–27.
5. Los Alamos National Laboratory. HIV and SIV Nomenclature [Internet]. http://www.hiv.lanl.gov/. [cited 2019 Aug 26]. Available from: https://www.hiv.lanl.gov/content/sequence/HelpDocs/subtypes-more.html#targetText=HIV-1 M Group&targetText=Subtypes and sub-subtypes of,%2C H%2C J and K.
6. Sauter D, Kirchhoff F. Key Viral Adaptations Preceding the AIDS Pandemic. Cell Host Microbe. 2019;25(1):27–38.
7. Takehisa J, Kraus MH, Ayouba A, Bailes E, Van Heuverswyn F, Decker JM, et al. Origin and biology of simian immunodeficiency virus in wild-living western gorillas. J Virol. 2009 Feb;83(4):1635–48.
8. Smith SM, Christian D, de Lame V, Shah U, Austin L, Gautam R, et al. Isolation of a new HIV-2 group in the US. Retrovirology. 2008 Nov 14;5.
9. Ayouba A, Akoua-Koffi C, Calvignac-Spencer S, Esteban A, Locatelli S, Li H, et al. Evidence for continuing cross-species transmission of SIVsmm to humans: Characterization of a new HIV-2 lineage in rural Côte d'Ivoire. Aids. 2013;27(15):2488–91.
10. Ibe S, Yokomaku Y, Shiino T, Tanaka R, Hattori J, Fujisaki S, et al. HIV-2 CRF01_AB: first circulating recombinant form of HIV-2. J Acquir Immune Defic Syndr. 2010 Jul;54(3):241–7.
11. Keele BF, Jones JH, Terio KA, Estes JD, Rudicell RS, Wilson ML, et al. Increased mortality and AIDS-like immunopathology in wild chimpanzees infected with SIVcpz. Nature. 2009 Jul 23;460(7254):515–9.
12. Rey-Cuillé MA, Hovanessian AG, Montagnier L, Chakrabarti LA, Berthier JL, Bomsel-Demontoy MC, et al. Simian immunodeficiency virus replicates to high levels in sooty mangabeys without inducing disease. J Virol. 1998;72(5):3872–86.

13. Ling B, Apetrei C, Pandrea I, Veazey RS, Lackner AA, Gormus B, et al. Classic AIDS in a Sooty Mangabey after an 18-Year Natural Infection. J Virol. 2004;78(16):8902–8.

14. Calef C, Mokili J, O'connor DH, Watkins DI, Korber B. Numbering Positions in SIV Numbering Positions in SIV Relative to SIVMM239.

15. Gottlieb GS, Hawes SE, Agne HD, Stern JE, Critchlow CW, Kiviat NB, et al. Lower levels of HIV RNA in semen in HIV-2 compared with HIV-1 infection: implications for differences in transmission. AIDS. 2006 Apr 4;20(6):895–900.

16. Faria NR, Rambaut A, Suchard MA, Baele G, Bedford T, Ward MJ, et al. HIV epidemiology. The early spread and epidemic ignition of HIV-1 in human populations. Science. 2014 Oct 3;346(6205):56–61.

17. de Sousa JD, Müller V, Lemey P, Vandamme A-M. High GUD incidence in the early 20 century created a particularly permissive time window for the origin and initial spread of epidemic HIV strains. PLoS One. 2010 Apr 1;5(4):e9936.

18. Pépin J, Labbé A-C. Noble goals, unforeseen consequences: control of tropical diseases in colonial Central Africa and the iatrogenic transmission of blood-borne viruses. Trop Med Int Health. 2008 Jun;13(6):744–53.

19. Molez JF. The historical question of acquired immunodeficiency syndrome in the 1960s in the Congo River basin area in relation to cryptococcal meningitis. Am J Trop Med Hyg. 1998;58(3):273–6.

20. Clavel F, Guétard D, Brun-Vézinet F, Chamaret S, Rey MA, Santos-Ferreira MO, et al. Isolation of a new human retrovirus from West African patients with AIDS. Science. 1986 Jul 18;233(4761):343–6.

21. Faria NR, Hodges-Mameletzis I, Silva JC, Rodés B, Erasmus S, Paolucci S, et al. Phylogeographical footprint of colonial history in the global dispersal of human immunodeficiency virus type 2 group A. J Gen Virol. 2012 Apr;93(Pt 4):889–99.

22. Buckner M. Village women as town prostitutes: cultural factors relevant to prostitution and HIV epidemiology in Guinea-Bissau. Vivre Penser le Sida en Afrique. 1999;389–404.

23. Sousa JD, Temudo MP, Hewlett BS, Camacho RJ, Müller V, Vandamme AM. Male Circumcision and the Epidemic Emergence of HIV-2 in West Africa. PLoS One. 2016;11(12):e0166805.

24. Venter A. Portugal's Guerrilla War: The Campaign for Africa. Malherbe J, editor. Cape Town South Africa. 1973.

25. Marx PA, Alcabes PG, Drucker E. Serial human passage of simian immunodeficiency virus by unsterile injections and the emergence of epidemic human immunodeficiency virus in Africa. In: Philosophical Transactions of the Royal Society B: Biological Sciences. Royal Society; 2001. p. 911–20.

26. Mota-Miranda A, Gomes H, Marques R, Serrão R, Lourenço H, Santos-Ferreira O, et al. HIV-2 infection with a long asymptomatic period. J Infect. 1995 Sep;31(2):163–4.

27. Pépin J, Plamondon M, Alves AC, Beaudet M, Labbé A-C. Parenteral transmission during excision and treatment of tuberculosis and trypanosomiasis may be responsible for the HIV-2 epidemic in Guinea-Bissau. AIDS. 2006 Jun 12;20(9):1303–11.

28. Visseaux B, Damond F, Matheron S, Descamps D, Charpentier C. Hiv-2 molecular epidemiology. Infect Genet Evol. 2016 Dec;46:233–40.

29. Forbi JC, Esona MD, Iperepolu HO, Adoga MP, Agwale SM. Absence of routine molecular testing and prevalence of HIV-2 infection in regions hardest-hit by HIV infection. J Infect Dev Ctries. 2012 Dec 15;6(12):854–9.

30. Fryer HR, Van Tienen C, Van Der Loeff MS, Aaby P, Da Silva ZJ, Whittle H, et al. Predicting the extinction of HIV-2 in rural Guinea-Bissau. AIDS. 2015 Nov 28;29(18):2479–86.

31. Olesen JS, Jespersen S, da Silva ZJ, Rodrigues A, Erikstrup C, Aaby P, et al. HIV-2 continues to decrease, whereas HIV-1 is stabilizing in Guinea-Bissau. AIDS. 2018 Jun 1;32(9):1193–8.

32. Guinea-Bissau profile - Timeline - BBC News [Internet]. [cited 2019 Aug 20]. Available from: www.bbc.co.uk/news/world-africa-13579838#targetText=Pre-15th century - The area,tributary kingdom known as Gabu.&targetText=1446-47 - First Portuguese arrive,important in the slave-trade.

33. Zhao G, Perilla JR, Yufenyuy EL, Meng X, Chen B, Ning J, et al. Mature HIV-1 capsid structure by cryo-electron microscopy and all-atom molecular dynamics. Nature. 2013 May 30;497(7451):643–6.

34. Division of AIDS NI of A and ID. HIV Sequence Compendium 2018. 2018;2.

35. Geijtenbeek TBH, van Kooyk Y. DC-SIGN: a novel HIV receptor on DCs that mediates HIV-1 transmission. Curr Top Microbiol Immunol. 2003;276:31–54.

36. Mörner A, Björndal A, Albert J, Kewalramani VN, Littman DR, Inoue R, et al. Primary human immunodeficiency virus type 2 (HIV-2) isolates, like HIV-1 isolates, frequently use CCR5 but show promiscuity in coreceptor usage. J Virol. 1999 Mar;73(3):2343–9.

37. Popov S, Rexach M, Zybarth G, Railing N, Lee MA, Ratner L, et al. Viral protein R regulates nuclear import of the HIV-1 pre-integration complex. EMBO J. 1998 Feb 4;17(4):909–17.

38. Levy N, Eiler S, Pradeau-Aubreton K, Crucifix C, Schaetzel A, Drillien R, et al. Structural and functional studies of the HIV-1 pre-integration complex. Retrovirology. 2013;10(Suppl 1).

39. Lahaye X, Satoh T, Gentili M, Cerboni S, Conrad C, Hurbain I, et al. The Capsids of HIV-1 and HIV-2 Determine Immune Detection of the Viral cDNA by the Innate Sensor cGAS in Dendritic Cells. Immunity. 2013 Dec 12;39(6):1132–42.

40. Kumar S, Morrison JH, Dingli D, Poeschla E. HIV-1 Activation of Innate Immunity Depends Strongly on the Intracellular Level of TREX1 and Sensing of Incomplete Reverse Transcription Products. J Virol. 2018 Aug 15;92(16):e00001-18.

41. Pertel T, Hausmann S, Morger D, Züger S, Guerra J, Lascano J, et al. TRIM5 is an innate immune sensor for the retrovirus capsid lattice. Nature. 2011 Apr 21;472(7343):361–5.

42. Bulli L, Apolonia L, Kutzner J, Pollpeter D, Goujon C, Herold N, et al. Complex Interplay between HIV-1 Capsid and MX2-Independent Alpha Interferon-Induced Antiviral Factors. J Virol. 2016;90(16):7469–80.

43. Mamede JI, Damond F, Bernardo A de, Matheron S, Descamps D, Battini J-L, et al. Cyclophilins and nucleoporins are required for infection mediated by capsids from circulating HIV-2 primary isolates. Sci Rep. 2017 Mar 27;7:45214.

44. Schaller T, Ocwieja KE, Rasaiyaah J, Price AJ, Brady TL, Roth SL, et al. HIV-1 Capsid-Cyclophilin Interactions Determine Nuclear Import Pathway, Integration Targeting and Replication Efficiency. Aiken C, editor. PLoS Pathog. 2011 Dec 8;7(12):e1002439.

45. Perkins ND, Edwards NL, Duckett CS, Agranoff AB, Schmid RM, Nabel GJ. A cooperative interaction between NF-kappa B and Sp1 is required for HIV-1 enhancer activation. EMBO J. 1993 Sep;12(9):3551–8.

46. Liu J, Perkins ND, Schmid RM, Nabel GJ. Specific NF-kappa B subunits act in concert with Tat to stimulate human immunodeficiency virus type 1 transcription. J Virol. 1992 Jun;66(6):3883–7.

47. Jmarchn. www.wikimedia.org [Internet]. 2019. Available from: https://commons.wikimedia.org/wiki/File:HIV-replication-cycle-en.svg#filehistory

48. Keele BF, Giorgi EE, Salazar-Gonzalez JF, Decker JM, Pham KT, Salazar MG, et al. Identification and characterization of transmitted and early founder virus envelopes in primary HIV-1 infection. Proc Natl Acad Sci U S A. 2008 May 27;105(21):7552–7.

49. Cohen MS, Shaw GM, McMichael AJ, Haynes BF. Acute HIV-1 infection. Vol. 364, New England Journal of Medicine. Massachussetts Medical Society; 2011. p. 1943–54.

50. Chun TW, Engel D, Berrey MM, Shea T, Corey L, Fauci AS. Early establishment of a pool of latently infected, resting CD4(+) T cells during primary HIV-1 infection. Proc Natl Acad Sci U S A. 1998 Jul 21;95(15):8869–73.

51. Schacker T, Little S, Connick E, Gebhard K, Zhang ZQ, Krieger J, et al. Productive infection of T cells in lymphoid tissues during primary and early human immunodeficiency virus infection. J Infect Dis. 2001 Feb 15;183(4):555–62.

52. Lin TC, Gianella S, Tenenbaum T, Little SJ, Hoenigl M. A simple symptom score for acute human immunodeficiency virus infection in a San Diego community-based screening program. Clin Infect Dis. 2018;67(1):105–11.

53. Goonetilleke N, Liu MKP, Salazar-Gonzalez JF, Ferrari G, Giorgi E, Ganusov V V, et al. The first T cell response to transmitted/founder virus contributes to the control of acute viremia in HIV-1 infection. J Exp Med. 2009 Jun 8;206(6):1253–72.

54. Douek DC, Picker LJ, Koup RA. T C ELL D YNAMICS IN HIV-1 I NFECTION . Annu Rev Immunol. 2003 Apr;21(1):265–304.

55. Lemey P, Kosakovsky Pond SL, Drummond AJ, Pybus OG, Shapiro B, Barroso H, et al. Synonymous substitution rates predict HIV disease progression as a result of underlying replication dynamics. PLoS Comput Biol. 2007 Feb 16;3(2):e29.

56. Troyer RM, Collins KR, Abraha A, Fraundorf E, Moore DM, Krizan RW, et al. Changes in Human Immunodeficiency Virus Type 1 Fitness and Genetic Diversity during Disease Progression. J Virol. 2005 Jul 15;79(14):9006–18.

57. Deeks SG, Tracy R, Douek DC. Systemic Effects of Inflammation on Health during Chronic HIV Infection. Vol. 39, Immunity. 2013. p. 633–45.

58. Ellis PK, Martin WJ, Dodd PJ. CD4 count and tuberculosis risk in HIV-positive adults not on ART: A systematic review and meta-analysis. PeerJ. 2017;2017(12).

59. WHO case definitions of hiv for surveillance and revised clinical staging and immunological classification of hiv-related disease in adults and children hiv/aids programme. 2007.

60. Poorolajal J, Hooshmand E, Mahjub H, Esmailnasab N, Jenabi E. Survival rate of AIDS disease and mortality in HIV-infected patients: a meta-analysis. Public Health. 2016 Oct;139:3–12.

61. Esbjörnsson J, Månsson F, Kvist A, da Silva ZJ, Andersson S, Fenyö EM, et al. Long-term follow-up of HIV-2-related AIDS and mortality in Guinea-Bissau: a prospective open cohort study. lancet HIV. 2018 Nov 1;

62. Mogensen TH, Melchjorsen J, Larsen CS, Paludan SR. Innate immune recognition and activation during HIV infection. Retrovirology. 2010;7:1–19.

63. Mellors JW, Muñoz A, Giorgi J V, Margolick JB, Tassoni CJ, Gupta P, et al. Plasma viral load and CD4+ lymphocytes as prognostic markers of HIV-1 infection. Ann Intern Med. 1997 Jun 15;126(12):946–54.

64. Lyles RH, Muñoz A, Yamashita TE, Bazmi H, Detels R, Rinaldo CR, et al. Natural History of Human Immunodeficiency Virus Type 1 Viremia after Seroconversion and Proximal to AIDS in a Large Cohort of Homosexual Men. J Infect Dis. 2000 Mar;181(3):872–80.

65. Rodríguez B, Sethi AK, Cheruvu VK, Mackay W, Bosch RJ, Kitahata M, et al. Predictive Value of Plasma HIV RNA Level on Rate of CD4 T-Cell Decline in Untreated HIV Infection. JAMA. 2006 Sep 27;296(12):1498.

66. Claiborne DT, Prince JL, Scully E, Macharia G, Micci L, Lawson B, et al. Replicative fitness of transmitted HIV-1 drives acute immune activation, proviral load in memory CD4+ T cells, and disease progression. Proc Natl Acad Sci U S A. 2015 Mar 24;112(12):E1480-9.

67. Alabi AS, Jaffar S, Ariyoshi K, Blanchard T, Schim Van Der Loeff M, Awasana AA, et al. Plasma viral load, CD4 cell percentage, HLA and survival of HIV-1, HIV-2, and dually infected Gambian patients. AIDS. 2003 Jul 4;17(10):1513–20.

68. Andersson S, Norrgren H, da Silva Z, Biague A, Bamba S, Kwok S, et al. Plasma viral load in HIV-1 and HIV-2 singly and dually infected individuals in Guinea-Bissau, West Africa: significantly lower plasma virus set point in HIV-2 infection than in HIV-1 infection. Arch Intern Med. 2000 Nov 27;160(21):3286–93.

69. Ariyoshi K, Jaffar S, Alabi a S, Berry N, Schim van der Loeff M, Sabally S, et al. Plasma RNA viral load predicts the rate of CD4 T cell decline and death in HIV-2-infected patients in West Africa. AIDS. 2000;14(4):339–44.

70. Berry N, Jaffar S, van der Loeff MS, Ariyoshi K, Harding E, N'Gom PT, et al. Low Level Viremia and High CD4% Predict Normal Survival in a Cohort of HIV Type-2-Infected Villagers. AIDS Res Hum Retroviruses. 2002 Nov 1;18(16):1167–73.

71. van der Loeff MFS, Larke N, Kaye S, Berry N, Ariyoshi K, Alabi A, et al. Undetectable plasma viral load predicts normal survival in HIV-2-infected people in a West African village. Retrovirology. 2010 May 19;7(1):46.

72. Andersson S, Norrgren H, da Silva Z, Biague A, Bamba S, Kwok S, et al. Plasma viral load in HIV-1 and HIV-2 singly and dually infected individuals in Guinea-Bissau, West Africa: significantly lower plasma virus set point in HIV-2 infection than in HIV-1 infection. Arch Intern Med. 2000 Nov 27;160(21):3286–93.

73. Williamson C, Swanstrom R. HIV-1 replication capacity: Setting the pace of disease. Proc Natl Acad Sci U S A. 2015 Mar 24;112(12):3591–2.

74. Prince JL, Claiborne DT, Carlson JM, Schaefer M, Yu T, Lahki S, et al. Role of transmitted Gag CTL polymorphisms in defining replicative capacity and early HIV-1 pathogenesis. PLoS Pathog. 2012;8(11):e1003041.

75. Selhorst P, Combrinck C, Ndabambi N, Ismail SD, Abrahams M-R, Lacerda M, et al. Replication Capacity of Viruses from Acute Infection Drives HIV-1 Disease Progression. Kirchhoff F, editor. J Virol. 2017 Apr 15;91(8):e01806-16.

76. Ariën KK, Abraha A, Quiñones-Mateu ME, Kestens L, Vanham G, Arts EJ. The replicative fitness of primary human immunodeficiency virus type 1 (HIV-1) group M, HIV-1 group O, and HIV-2 isolates. J Virol. 2005 Jul 15;79(14):8979–90.

77. Blaak H, van der Ende ME, Boers PHMM, Schuitemaker H, Osterhaus ADMEME. In vitro replication capacity of HIV-2 variants from long-term aviremic individuals. Virology. 2006 Sep 15;353(1):144–54.

78. Esbjörnsson J, Månsson F, Kvist A, Isberg P-E, Nowroozalizadeh S, Biague AJ, et al. Inhibition of HIV-1 disease progression by contemporaneous HIV-2 infection. N Engl J Med. 2012 Jul 19;367(3):224–32.

79. Skar H, Borrego P, Wallstrom TC, Mild M, Marcelino JM, Barroso H, et al. HIV-2 genetic evolution in patients with advanced disease is faster than that in matched HIV-1 patients. J Virol. 2010;84(14):7412–5.

80. MacNeil A, Sankalé J, Meloni ST, Sarr AD, Mboup S, Kanki P. Long-Term Intrapatient Viral Evolution during HIV-2 Infection. J Infect Dis. 2007;195(5):726–33.

81. Palm AA, Lemey P, Jansson M, Månsson F, Kvist A, Szojka Z, et al. Low Postseroconversion CD4 $^{+}$ T-cell Level Is Associated with Faster Disease Progression and Higher Viral Evolutionary Rate in HIV-2 Infection. Paraskevis D, editor. MBio. 2019 Jan 8;10(1).

82. Forshey BM, von Schwedler U, Sundquist WI, Aiken C. Formation of a human immunodeficiency virus type 1 core of optimal stability is crucial for viral replication. J Virol. 2002 Jun;76(11):5667–77.

83. Schommers P, Martrus G, Matschl U, Sirignano M, Lütgehetmann M, Richert L, et al. Changes in HIV-1 Capsid Stability Induced by Common Cytotoxic-T-Lymphocyte-Driven Viral Sequence Mutations. J Virol. 2016 Aug 15;90(16):7579–86.

84. Manel N, Hogstad B, Wang Y, Levy DE, Unutmaz D, Littman DR, et al. A cryptic sensor for HIV-1 activates antiviral innate immunity in dendritic cells. Nature. 2010 Sep 9;467(7312):214–7.

85. Lahaye X, Gentili M, Silvin A, Conrad C, Picard L, Jouve M, et al. NONO Detects the Nuclear HIV Capsid to Promote cGAS-Mediated Innate Immune Activation. Cell. 2018;

86. Onyango CO, Leligdowicz A, Yokoyama M, Sato H, Song H, Nakayama EE, et al. HIV-2 capsids distinguish high and low virus load patients in a West African community cohort. Vaccine. 2010 May;28:B60–7.

87. Jallow S, Leligdowicz A, Kramer HB, Onyango C, Cotten M, Wright C, et al. The presence of prolines in the flanking region of an immunodominant HIV-2 gag epitope influences the quality and quantity of the epitope generated. Eur J Immunol. 2015;45(8):2232–42.

88. Song H, Nakayama EE, Yokoyama M, Sato H, Levy JA, Shioda T. A Single Amino Acid of the Human Immunodeficiency Virus Type 2 Capsid Affects Its Replication in the Presence of Cynomolgus Monkey and Human TRIM5 s. J Virol. 2007 Jul 1;81(13):7280–5.

89. Miyamoto T, Yokoyama M, Kono K, Shioda T, Sato H, Nakayama EE. A Single Amino Acid of Human Immunodeficiency Virus Type 2 Capsid Protein Affects Conformation of Two External Loops and Viral Sensitivity to TRIM5α. Lee Y-M, editor. PLoS One. 2011 Jul 28;6(7):e22779.

90. Takeuchi JS, Perche B, Migraine J, Mercier-Delarue S, Ponscarme D, Simon F, et al. High level of susceptibility to human TRIM5α conferred by HIV-2 capsid sequences. Retrovirology. 2013;10(1):50.

91. de Silva TI, Leligdowicz A, Carlson J, Garcia-Knight M, Onyango C, Miller N, et al. HLA-associated polymorphisms in the HIV-2 capsid highlight key differences between HIV-1 and HIV-2 immune adaptation. AIDS. 2018 Mar 27;32(6):709–14.

92. de Silva TI, van Tienen C, Onyango C, Jabang A, Vincent T, Loeff MFS van der, et al. Population dynamics of HIV-2 in rural West Africa: comparison with HIV-1 and ongoing transmission at the heart of the epidemic. AIDS. 2013 Jan 2;27(1):125–34.

93. Uchil PD, Quinlan BD, Chan W-T, Luna JM, Mothes W. TRIM E3 ligases interfere with early and late stages of the retroviral life cycle. PLoS Pathog. 2008 Feb 8;4(2):e16.

94. Li X, Li Y, Stremlau M, Yuan W, Song B, Perron M, et al. Functional Replacement of the RING, B-Box 2, and Coiled-Coil Domains of Tripartite Motif 5 (TRIM5) by Heterologous TRIM Domains. J Virol. 2006 Jul 1;80(13):6198–206.

95. Biris N, Yang Y, Taylor AB, Tomashevski A, Guo M, Hart PJ, et al. Structure of the rhesus monkey TRIM5α PRYSPRY domain, the HIV capsid recognition module. Proc Natl Acad Sci U S A. 2012 Aug 14;109(33):13278–83.

96. Stremlau M, Owens CM, Perron MJ, Kiessling M, Autissier P, Sodroski J. The cytoplasmic body component TRIM5α restricts HIV-1 infection in Old World monkeys. Nature. 2004 Feb 26;427(6977):848–53.

97. Barr SD, Smiley JR, Bushman FD. The interferon response inhibits HIV particle production by induction of TRIM22. Hope TJ, editor. PLoS Pathog. 2008 Feb 29;4(2):e1000007.

98. Di Pietro A, Kajaste-Rudnitski A, Oteiza A, Nicora L, Towers GJ, Mechti N, et al. TRIM22 inhibits influenza A virus infection by targeting the viral nucleoprotein for degradation. J Virol. 2013 Apr 15;87(8):4523–33.

99. Doyle T, Goujon C, Malim MH. HIV-1 and interferons: who's interfering with whom? Nat Rev Micro. 2015 Apr 27;13(7):403–13.

100. Battivelli E, Migraine J, Lecossier D, Matsuoka S, Perez-Bercoff D, Saragosti S, et al. Modulation of TRIM5alpha activity in human cells by alternatively spliced TRIM5 isoforms. J Virol. 2011 Aug 1;85(15):7828–35.

101. Wilson SJ, Webb BLJ, Maplanka C, Newman RM, Verschoor EJ, Heeney JL, et al. Rhesus Macaque TRIM5 Alleles Have Divergent Antiretroviral Specificities. J Virol. 2008 Jul 15;82(14):7243–7.

102. Meyerson NR, Sawyer SL. Two-stepping through time: mammals and viruses. Trends Microbiol. 2011 Jun 1;19(6):286–94.

103. Sawyer SL, Emerman M, Malik HS, Finlay M, Barton G. Discordant Evolution of the Adjacent Antiretroviral Genes TRIM22 and TRIM5 in Mammals. PLoS Pathog. 2007;3(12):e197.

104. Ylinen LMJ, Keckesova Z, Wilson SJ, Ranasinghe S, Towers GJ. Differential Restriction of Human Immunodeficiency Virus Type 2 and Simian Immunodeficiency Virus SIVmac by TRIM5 Alleles. J Virol. 2005 Sep 15;79(18):11580–7.

105. Morger D, Zosel F, Bühlmann M, Züger S, Mittelviefhaus M, Schuler B, et al. The three-fold axis of the HIV-1 capsid lattice is the species-specific binding interface for TRIM5α. J Virol. 2017;(December):JVI.01541-17.

106. Ohmine S, Sakuma R, Sakuma T, Thatava T, Takeuchi H, Ikeda Y. The antiviral spectra of TRIM5α orthologues and human TRIM family proteins against lentiviral production. Harris RS, editor. PLoS One. 2011 Jan 14;6(1):e16121.

107. Veillette M, Bichel K, Pawlica P, Freund SM V, Plourde MB, Pham QT, et al. The V86M mutation in HIV-1 capsid confers resistance to TRIM5α by abrogation of cyclophilin a-dependent restriction and enhancement of viral nuclear import. Retrovirology. 2013 Feb 28;10(1):25.

108. Pertel T, Hausmann S, Morger D, Zuger S, Guerra J, Lascano J, et al. TRIM5 is an innate immune sensor for the retrovirus capsid lattice. Nature. 2011;472(7343):361–5.

109. Splettstoesser T. P24, HIV-capsid [Internet]. www.wikimedia.org. 2019. Available from: https://commons.wikimedia.org/wiki/File:P24_HIV-capsid.png

110. Goldschmidt V, Bleiber G, May M, Martinez R, Ortiz M, Telenti A, et al. Role of common human TRIM5alpha variants in HIV-1 disease progression. Retrovirology. 2006 Aug 22;3:54.

111. Rahm N, Gfeller D, Snoeck J, Martinez R, McLaren PJ, Ortiz M, et al. Susceptibility and adaptation to human TRIM5α alleles at positive selected sites in HIV-1 capsid. Virology. 2013 Jul 5;441(2):162–70.

112. van Manen D, Rits MAN, Beugeling C, van Dort K, Schuitemaker H, Kootstra NA, et al. The Effect of Trim5 Polymorphisms on the Clinical Course of HIV-1 Infection. PLoS Pathog. 2008;4(2):e18.

113. Kootstra NA, Navis M, Beugeling C, van Dort KA, Schuitemaker H. The presence of the Trim5α escape mutation H87Q in the capsid of late stage HIV-1 variants is preceded by a prolonged asymptomatic infection phase. AIDS. 2007 Oct 1;21(15):2015–23.

114. Sewram S, Singh R, Kormuth E, Werner L, Mlisana K, Karim SSA, et al. Human TRIM5α Expression Levels and Reduced Susceptibility to HIV-1 Infection. J Infect Dis. 2009 Jun 1;199(11):1657–63.

115. Wu F, Ourmanov I, Riddick N, Matsuda K, Whitted S, Plishka RJ, et al. TRIM5α Restriction Affects Clinical Outcome and Disease Progression in Simian Immunodeficiency Virus-Infected Rhesus Macaques. Kirchhoff F, editor. J Virol. 2015 Feb 15;89(4):2233–40.

116. Javanbakht H, Diaz-Griffero F, Stremlau M, Si Z, Sodroski J. The contribution of RING and B-box 2 domains to retroviral restriction mediated by monkey TRIM5α. J Biol Chem. 2005 Jul 22;280(29):26933–40.

117. Yamauchi K, Wada K, Tanji K, Tanaka M, Kamitani T. Ubiquitination of E3 ubiquitin ligase TRIM5 alpha and its potential role. FEBS J. 2008 Apr;275(7):1540–55.

118. Grütter MG, Luban J. TRIM5 structure, HIV-1 capsid recognition, and innate immune signaling. Vol. 2, Current Opinion in Virology. Elsevier B.V.; 2012. p. 142–50.

119. Kömürlü S, Bradley M, Smolin N, Imam S, Pauszek RF, Robia SL, et al. Defects in assembly explain reduced antiviral activity of the G249D polymorphism in human TRIM5α. Menéndez-Arias L, editor. PLoS One. 2019 Mar 19;14(3):e0212888.

120. Ohkura S, Yap MW, Sheldon T, Stoye JP. All Three Variable Regions of the TRIM5 B30.2 Domain Can Contribute to the Specificity of Retrovirus Restriction. J Virol. 2006 Sep 1;80(17):8554–65.

121. Duan Z, Gao B, Xu W, Xiong S. Identification of TRIM22 as a RING finger E3 ubiquitin ligase. Biochem Biophys Res Commun. 2008 Sep 26;374(3):502–6.

122. Li Q, Lee CH, Peters LA, Mastropaolo LA, Thoeni C, Elkadri A, et al. Variants in TRIM22 That Affect NOD2 Signaling Are Associated With Very-Early-Onset Inflammatory Bowel Disease. Gastroenterology. 2016 May;150(5):1196–207.

123. Esposito D, Koliopoulos MG, Rittinger K. Structural determinants of TRIM protein function. Biochem Soc Trans. 2017;45(1):183–91.

124. Gao B, Duan Z, Xu W, Xiong S. Tripartite motif-containing 22 inhibits the activity of hepatitis B virus core promoter, which is dependent on nuclear-located RING domain. Hepatology. 2009 Aug;50(2):424–33.

125. Yang C, Zhao X, Sun D, Yang L, Chong C, Pan Y, et al. Interferon alpha (IFNα)-induced TRIM22 interrupts HCV replication by ubiquitinating NS5A. Cell Mol Immunol. 2016 Jan;13(1):94–102.

126. Jimenez-Moyano E, Ruiz A, Kløverpris HN, Rodriguez-Plata MT, Peña R, Blondeau C, et al. Nonhuman TRIM5 Variants Enhance Recognition of HIV-1-Infected Cells by CD8 $^{+}$ T Cells. Silvestri G, editor. J Virol. 2016 Oct 1;90(19):8552–62.

127. Nakayama EE, Shioda T. Impact of TRIM5α in vivo. AIDS. 2015 Sep 10;29(14):1733–43.

128. Boswell MT, Rowland-Jones SL. Delayed disease progression in HIV-2: the importance of TRIM5α and the retroviral capsid. Clin Exp Immunol. 2019 Jun 21;196(3):305–17.

129. Forlani G, Accolla RS. Tripartite Motif 22 and Class II Transactivator Restriction Factors: Unveiling Their Concerted Action against Retroviruses. Front Immunol. 2017;8(October):1–7.

130. Singh R, Gaiha G, Werner L, McKim K, Mlisana K, Luban J, et al. Association of TRIM22 with the Type 1 Interferon Response and Viral Control during Primary HIV-1 Infection. J Virol. 2011 Jan 1;85(1):208–16.

131. Kelly JN, Barr SD. In Silico Analysis of Functional Single Nucleotide Polymorphisms in the Human TRIM22 Gene. Uversky VN, editor. PLoS One. 2014 Jul 1;9(7):e101436.

132. Ghezzi S, Galli L, Kajaste-Rudnitski A, Turrini F, Marelli S, Toniolo D, et al. Identification of TRIM22 single nucleotide polymorphisms associated with loss of inhibition of HIV-1 transcription and advanced HIV-1 disease. AIDS. 2013 Sep 24;27(15):2335–44.

133. Naranbhai V, Carrington M. Host genetic variation and HIV disease: from mapping to mechanism. Immunogenetics. 2017 Aug 10;69(8–9):489–98.

134. Allsopp CE, Harding RM, Taylor C, Bunce M, Kwiatkowski D, Anstey N, et al. Interethnic genetic differentiation in Africa: HLA class I antigens in The Gambia. Am J Hum Genet. 1992 Feb;50(2):411–21.

135. Yindom L-M, Leligdowicz A, Martin MP, Gao X, Qi Y, Zaman SM a, et al. Influence of HLA class I and HLA-KIR compound genotypes on HIV-2 infection and markers of disease progression in a Manjako community in West Africa. J Virol. 2010;84(16):8202–8.

136. Diouf K, Sarr AD, Eisen G, Popper S, Mboup S, Kanki P. Associations between MHC class I and susceptibility to HIV-2 disease progression. J Hum Virol. 5(1):1–7.

137. Thomsen D, Erikstrup C, Jespersen S, Medina C, Té D da S, Correira FG, et al. The influence of human leukocyte antigen-types on disease progression among HIV-2 infected patients in Guinea-Bissau. AIDS. 2018 Mar 27;32(6):721–8.

138. Hennig BJ, Velez-Edwards DR, Schim van der Loeff MF, Bisseye C, Edwards TL, Tacconelli A, et al. CD4 intragenic SNPs associate with HIV-2 plasma viral load and CD4 count in a community-based study from Guinea-Bissau, West Africa. J Acquir Immune Defic Syndr. 2011 Jan 1;56(1):1–8.

139. Douek DC, Brenchley JM, Betts MR, Ambrozak DR, Hill BJ, Okamoto Y, et al. HIV preferentially infects HIV-specific CD4+ T cells. Nature. 2002 May 2;417(6884):95–8.

140. Duvall MG, Jaye A, Dong T, Brenchley JM, Alabi AS, Jeffries DJ, et al. Maintenance of HIV-specific CD4+ T cell help distinguishes HIV-2 from HIV-1 infection. J Immunol. 2006 Jun 1;176(11):6973–81.

141. Alatrakchi N, Damond F, Matheron S, Beretta-Tempelhoff S, Campa P, Carcelain G, et al. Proliferative, IFNgamma and IL-2-producing T-cell responses to HIV-2 in untreated HIV-2 infection. AIDS. 2006 Jan 2;20(1):29–34.

142. Duvall MG, Precopio ML, Ambrozak DA, Jaye A, McMichael AJ, Whittle HC, et al. Polyfunctional T cell responses are a hallmark of HIV-2 infection. Eur J Immunol. 2008 Feb;38(2):350–63.

143. Foxall RB, Cortesão CS, Albuquerque AS, Soares RS, Victorino RMM, Sousa AE. Gag-specific CD4+ T-cell frequency is inversely correlated with proviral load and directly correlated with immune activation in infection with human immunodeficiency virus type 2 (HIV-2) but not HIV-1. J Virol. 2008 Oct;82(19):9795–9.

144. Francisco LM, Sage PT, Sharpe AH. The PD-1 pathway in tolerance and autoimmunity. Immunol Rev. 2010 Jul;236:219–42.

145. Soares RS, Tendeiro R, Foxall RB, Baptista AP, Cavaleiro R, Gomes P, et al. Cell-Associated Viral Burden Provides Evidence of Ongoing Viral Replication in Aviremic HIV-2-Infected Patients. J Virol. 2011 Mar 1;85(5):2429–38.

146. Tendeiro R, Foxall RB, Baptista AP, Pinto F, Soares RS, Cavaleiro R, et al. PD-1 and its ligand PD-L1 are progressively up-regulated on CD4 and CD8 T-cells in HIV-2 infection irrespective of the presence of viremia. AIDS. 2012 Jun;26(9):1065–71.

147. McMichael AJ, Rowland-Jones SL. Cellular immune responses to HIV. Nature. 2001 Apr 19;410(6831):980–7.

148. Phillips RE, Rowland-Jones S, Nixon DF, Gotch FM, Edwards JP, Ogunlesi AO, et al. Human immunodeficiency virus genetic variation that can escape cytotoxic T cell recognition. Nature. 1991 Dec 12;354(6353):453–9.

149. O'Connell KA, Bailey JR, Blankson JN. Elucidating the elite: mechanisms of control in HIV-1 infection. Trends Pharmacol Sci. 2009 Dec;30(12):631–7.

150. Kiepiela P, Ngumbela K, Thobakgale C, Ramduth D, Honeyborne I, Moodley E, et al. CD8+ T-cell responses to different HIV proteins have discordant associations with viral load. Nat Med. 2007 Jan 17;13(1):46–53.

151. Leligdowicz A, Onyango C, Yindom L-M, Peng Y, Cotten M, Jaye A, et al. Highly avid, oligoclonal, early-differentiated antigen-specific CD8+ T cells in chronic HIV-2 infection. Eur J Immunol. 2010 Jul;40(7):1963–72.

152. Leligdowicz A, Yindom L-M, Onyango C, Sarge-Njie R, Alabi A, Cotten M, et al. Robust Gag-specific T cell responses characterize viremia control in HIV-2 infection. J Clin Invest. 2007 Oct 1;117(10):3067–74.

153. de Silva TI, Peng Y, Leligdowicz A, Zaidi I, Li L, Griffin H, et al. Correlates of T-cell-mediated viral control and phenotype of CD8+ T cells in HIV-2, a naturally contained human retroviral infection. Blood. 2013 May 23;121(21):4330–9.

154. Angin M, Wong G, Papagno L, Versmisse P, David A, Bayard C, et al. Preservation of Lymphopoietic Potential and Virus Suppressive Capacity by CD8+ T Cells in HIV-2-Infected Controllers. J Immunol. 2016;1600693.

155. Fernandes SM, Pires AR, Matoso P, Ferreira C, Nunes-Cabaço H, Correia L, et al. HIV-2 infection is associated with preserved GALT homeostasis and epithelial integrity despite ongoing mucosal viral replication. Mucosal Immunol. 2017 May 17;mi201744.

156. Jaffar S, Grant AD, Whitworth J, Smith PG, Whittle H. The natural history of HIV-1 and HIV-2 infections in adults in Africa: A literature review. Bull World Health Organ. 2004;82(6):462–9.

157. Grossman Z, Meier-Schellersheim M, Sousa AE, Victorino RMM, Paul WE. CD4+ T-cell depletion in HIV infection: are we closer to understanding the cause? Nat Med. 2002 Apr 1;8(4):319–23.

158. Nyamweya S, Townend J, Zaman A, Steele SJ, Jeffries D, Rowland-Jones S, et al. Are plasma biomarkers of immune activation predictive of HIV progression: A longitudinal comparison and analyses in HIV-1 and HIV-2 infections? Gray CM, editor. PLoS One. 2012 Sep 10;7(9):e44411.

159. Sousa AE, Carneiro J, Meier-Schellersheim M, Grossman Z, Victorino RMM. CD4 T cell depletion is linked directly to immune activation in the pathogenesis of HIV-1 and HIV-2 but only indirectly to the viral load. J Immunol. 2002 Sep 15;169(6):3400–6.

160. Brenchley JM, Price DA, Schacker TW, Asher TE, Silvestri G, Rao S, et al. Microbial translocation is a cause of systemic immune activation in chronic HIV infection. Nat Med. 2006 Dec 19;12(12):1365–71.

161. Nowroozalizadeh S, Månsson F, da Silva Z, Repits J, Dabo B, Pereira C, et al. Microbial translocation correlates with the severity of both HIV-1 and HIV-2 infections. J Infect Dis. 2010 Apr 15;201(8):1150–4.

162. Cavaleiro R, Tendeiro R, Foxall RB, Soares RS, Baptista AP, Gomes P, et al. Monocyte and myeloid dendritic cell activation occurs throughout HIV type 2 infection, an attenuated form of HIV disease. J Infect Dis. 2013 Jun 1;207(11):1730–42.

163. Jespersen S, Hønge BL, Esbjörnsson J, Medina C, da Silva Té D, Correira FG, et al. Differential effects of sex in a West African cohort of HIV-1, HIV-2 and HIV-1/2 dually infected patients: men are worse off. Trop Med Int Heal. 2016 Feb;21(2):253–62.

164. Fonsah JY, Njamnshi AK, Kouanfack C, Qiu F, Njamnshi DM, Tagny CT, et al. Adherence to Antiretroviral Therapy (ART) in Yaoundé-Cameroon: Association with Opportunistic Infections, Depression, ART Regimen and Side Effects. Kumar A, editor. PLoS One. 2017 Jan 31;12(1):e0170893.

165. Cornell M, Schomaker M, Garone DB, Giddy J, Hoffmann CJ, Lessells R, et al. Gender Differences in Survival among Adult Patients Starting Antiretroviral Therapy in South Africa: A Multicentre Cohort Study. Celentano DD, editor. PLoS Med. 2012 Sep 4;9(9):e1001304.

166. Chang JJ, Woods M, Lindsay RJ, Doyle EH, Griesbeck M, Chan ES, et al. Higher expression of several interferon-stimulated genes in HIV-1-infected females after adjusting for the level of viral replication. J Infect Dis. 2013 Sep 1;208(5):830–8.

167. Addo MM, Altfeld M. Sex-based differences in HIV type 1 pathogenesis. J Infect Dis. 2014 Jul 15;209 Suppl 3(Suppl 3):S86-92.

168. Sterling TR, Vlahov D, Astemborski J, Hoover DR, Margolick JB, Quinn TC. Initial Plasma HIV-1 RNA Levels and Progression to AIDS in Women and Men. N Engl J Med. 2001 Mar 8;344(10):720–5.

169. Meier A, Chang JJ, Chan ES, Pollard RB, Sidhu HK, Kulkarni S, et al. Sex differences in the Toll-like receptor–mediated response of plasmacytoid dendritic cells to HIV-1. Nat Med. 2009 Aug 13;15(8):955–9.

170. Carlson JM, Schaefer M, Batorsky R, Claiborne DT, Prince J, Deymier MJ, et al. Selection bias at the heterosexual HIV-1 transmission bottleneck. 2015;345(6193).

171. Boullé C, Kouanfack C, Laborde-Balen G, Boyer S, Aghokeng AF, Carrieri MP, et al. Gender Differences in Adherence and Response to Antiretroviral Treatment in the Stratall Trial in Rural District Hospitals in Cameroon. JAIDS J Acquir Immune Defic Syndr. 2015 Jul 1;69(3):355–64.

172. Darby SC, Ewart DW, Giangrande PL, Spooner RJ, Rizza CR. Importance of age at infection with HIV-1 for survival and development of AIDS in UK haemophilia population. UK Haemophilia Centre Directors' Organisation. Lancet (London, England). 1996 Jun 8;347(9015):1573–9.

173. Balestre E, Eholié SP, Lokossue A, Sow PS, Charurat M, Minga A, et al. Effect of age on immunological response in the first year of antiretroviral therapy in HIV-1-infected adults in West Africa. AIDS. 2012 May 15;26(8):951–7.

174. Aaby P, Ariyoshi K, Buckner M, Jensen H, Berry N, Wilkins A, et al. Age of wife as a major determinant of male-to-female transmission of HIV-2 infection: a community study from rural West Africa. AIDS. 1996 Nov;10(13):1585–90.

175. Poulsen AG, Aaby P, Larsen O, Jensen H, Nauclér A, Lisse IM, et al. 9-year HIV-2-associated mortality in an urban community in Bissau, west Africa. Lancet (London, England). 1997 Mar 29;349(9056):911–4.

176. Matheron S, Pueyo S, Damond F, Simon F, Leprêtre A, Campa P, et al. Factors associated with clinical progression in HIV-2 infected-patients: the French ANRS cohort. AIDS. 2003 Dec 5;17(18):2593–601.

177. Broder S. The development of antiretroviral therapy and its impact on the HIV-1/AIDS pandemic. Vol. 85, Antiviral Research. 2010. p. 1–18.

178. Larder BA, Darby G, Richman DD. HIV with reduced sensitivity to zidovudine (AZT) isolated during prolonged therapy. Science. 1989 Mar 31;243(4899):1731–4.

179. Hammer SM, Squires KE, Hughes MD, Grimes JM, Demeter LM, Currier JS, et al. A controlled trial of two nucleoside analogues plus indinavir in persons with human immunodeficiency virus infection and CD4 cell counts of 200 per cubic millimeter or less. N Engl J Med. 1997 Sep 11;337(11):725–33.

180. Zhang L, Ramratnam B, Tenner-Racz K, He Y, Vesanen M, Lewin S, et al. Quantifying residual HIV-1 replication in patients receiving combination antiretroviral therapy. N Engl J Med. 1999 May 27;340(21):1605–13.

181. Chun TW, Davey RT, Engel D, Lane HC, Fauci AS. AIDS: Re-emergence of HIV after stopping therapy. Nature. 1999 Oct 28;401(6756):874–5.

182. Menéndez-Arias L, Álvarez M. Antiretroviral therapy and drug resistance in human immunodeficiency virus type 2 infection. Antiviral Res. 2014 Feb 1;102:70–86.

183. Ntemgwa M, Brenner BG, Oliveira M, Moisi D, Wainberg MA. Natural Polymorphisms in the Human Immunodeficiency Virus Type 2 Protease Can Accelerate Time to Development of Resistance to Protease Inhibitors. Antimicrob Agents Chemother. 2007 Feb 1;51(2):604–10.

184. Poveda E, Rodes B, Toro C, Soriano V. Are Fusion Inhibitors Active against All HIV Variants? AIDS Res Hum Retroviruses. 2004 Mar;20(3):347–8.

185. Gottlieb GS, Badiane NMD, Hawes SE, Fortes L, Toure M, Ndour CT, et al. Emergence of multiclass drug-resistance in HIV-2 in antiretroviral-treated individuals in Senegal: implications for HIV-2 treatment in resouce-limited West Africa. Clin Infect Dis. 2009 Feb 15;48(4):476–83.

186. Benard A, Sighem A van. Immunovirological response to triple nucleotide reverse-transcriptase inhibitors and ritonavir-boosted protease inhibitors in treatment-naive HIV-2–infected patients: Clin Infect Dis. 2011;52(1537-6591 (Electronic)):1257–66.

187. Ba S, Raugi DN, Smith RA, Sall F, Faye K, Hawes SE, et al. A Trial of a Single-tablet Regimen of Elvitegravir, Cobicistat, Emtricitabine, and Tenofovir Disoproxil Fumarate for the Initial Treatment of Human Immunodeficiency Virus Type 2 Infection in a Resource-limited Setting: 48-Week Results From Senegal, West Africa. Clin Infect Dis. 2018 Oct 30;67(10):1588–94.

188. Matheron S, Descamps D, Gallien S, Besseghir A, Sellier P, Blum L, et al. First-line Raltegravir/Emtricitabine/Tenofovir Combination in Human Immunodeficiency Virus Type 2 (HIV-2) Infection: A Phase 2, Noncomparative Trial (ANRS 159 HIV-2). Clin Infect Dis. 2018 Sep 28;67(8):1161–7.

189. Gottlieb GS, Raugi DN, Smith RA. 90-90-90 for HIV-2? Ending the HIV-2 epidemic by enhancing care and clinical management of patients infected with HIV-2. Vol. 5, The Lancet HIV. Elsevier Ltd; 2018. p. e390–9.

190. Duffy S, Shackelton LA, Holmes EC. Rates of evolutionary change in viruses: patterns and determinants. Nat Rev Genet. 2008 Apr;9(4):267–76.

191. Sanjuán R. From molecular genetics to phylodynamics: Evolutionary relevance of mutation rates across viruses. PLoS Pathog. 2012 May;8(5).

192. Lauring AS, Andino R. Quasispecies Theory and the Behavior of RNA Viruses. Manchester M, editor. PLoS Pathog. 2010 Jul 22;6(7):e1001005.

193. Overbaugh J, Bangham CRM. Selection forces and constraints on retroviral sequence variation. Vol. 292, Science. 2001. p. 1106–9.

194. Sanborn KB, Somasundaran M, Luzuriaga K, Leitner T. Recombination elevates the effective evolutionary rate and facilitates the establishment of HIV-1 infection in infants after mother-to-child transmission. Retrovirology. 2015 Dec 16;12(1):96.

195. Perelson AS, Neumann AU, Markowitz M, Leonard JM, Ho DD. HIV-1 dynamics in vivo: Virion clearance rate, infected cell life-span, and viral generation time. Science (80-). 1996 Mar 15;271(5255):1582–6.

196. Vignuzzi M, Stone JK, Arnold JJ, Cameron CE, Andino R. Quasispecies diversity determines pathogenesis through cooperative interactions in a viral population. Nature. 2006 Jan 19;439(7074):344–8.

197. Yang Z, Rannala B. Molecular phylogenetics: Principles and practice. Nat Rev Genet. 2012;13(5):303–14.

198. Felsenstein J. Cases in which Parsimony or Compatibility Methods Will be Positively Misleading. Syst Zool. 1978 Dec;27(4):401.

199. Robert R. Sokal and Charles Michener. A statistical method for evaluating systematic relationships. Vol. 38, University of Kansas Science Bulletin. 1958. 1409–1438 p.

200. Rzhetsky AM. A Simple Method for Estimating and Testing Minimum-Evolution Trees. Mol Biol Evol. 1992 Sep 1;9(5):945.

Chapter 2

Materials and Methods

This chapter provides supplementary details about the materials and methods used in the subsequent chapters.

2.1. Methods for chapter 3

2.1.1. Patient recruitment

Participants aged 6 – 15 years were recruited from seven public sector primary healthcare clinics in Harare, Zimbabwe between January 2013 and January 2015, following provider-initiated HIV testing and counselling. Those who tested HIV-positive were offered enrolment into a prospective cohort study. Exclusion criteria included living outside Harare or choosing not to seek HIV care at one of the study clinics. Participants were invited for follow-up visits at three monthly intervals for assessment of clinical status, up until 18 months post-enrolment. Within the cohort, participants who consented to home visits were enrolled into the ZENITH randomised controlled trial, reported elsewhere (1).

Informed consent (parent or guardian) and assent (participant) were obtained from all participants prior to enrolment. The Medical Research Council of Zimbabwe, the Biomedical Research Training Institute Institutional Review Board, and the Ethics Committees of Harare City Health Department, Harare Central Hospital, and the London School of Hygiene and Tropical Medicine approved the study.

2.1.2. $CD4^+$ T cell count analysis

The guideline in use in January 2013 stated that children should be initiated on ART at a $CD4^+$ T cell count below 350cells/μL or if they had World Health Organisation (WHO) stage 3 or 4 HIV disease (2). In March 2014, Zimbabwe adopted WHO guidelines (3), with a revised threshold for ART initiation of 500 $CD4^+$ T cells/μL.

$CD4^+$ T cell counts were measured at three- and six-months post-initiation, and otherwise six monthly, with an Alere PIMA CD4 (Waltham, Massachusetts, USA) machine. Participant data were recorded by a study nurse on paper clinical report forms and entered into a Microsoft Access database using Cardiff TELEFORM Intelligent Character Optical Mark Recognition Software (Version 10.9).

2.1.3. DNA extraction and storage

DNA extraction was done by laboratory personnel at the Biomedical Research and Training Institute, Harare, Zimbabwe.

Day 1

Red blood cell lysis: 10mL of whole blood was added to a 50mL falcon tube. The sample was vortexed and then 30mL of Tris 20Mm- Ethylenediaminetetraacetic acid (EDTA) 5mM (TE 20-5) was added up to volume of 40mL, the tube was then inverted several times and incubated on ice for 50min. Following this the sample was centrifuged at 3500rpm for 15min and the supernatant then removed. The remaining pellet was then vortexed. TE was then added to a total volume of 40mL and the sample then vortexed and centrifuged at 3500rpm for 15min. This step was repeated until the remaining pellet was white.

White blood cell lysis: After the red cells were adequately lysed the remaining white pellet was vortexed until no cell clumps remained. 2mL of TE 20-5 was added followed by 100µL of 20% solution of sodium dodecyl sulphate. To digest remaining proteins 20µL of proteinase-K (10mg/mL) was added and the sample placed in a shaking water bath for overnight incubation at 42ºC.

Day 2

Ethanol precipitation: After the overnight incubation the sample was removed from the water bath and cooled to room temperature. 2.5mL of 7.5M ammonium acetate was added and the sample placed on ice. Next, 10mL of chilled 100% ethanol was added followed by centrifugation at 5000rpm for 30min. The supernatant was then removed, and 4mL of sterile Tris 20mM-EDTA 5mM-NaCl 0.2M (TE 20-5 NaCl 0.2M) was added, followed by vortexing the sample. The sample was then incubated at 42ºC until the DNA was fully dissolved into the solution. The DNA was then re-precipitated in 8mL of 100% ethanol, followed by centrifugation at 5000rpm for 30min. The supernatant was then removed and allowed to stand until no fluid remained.

The DNA sample was then re-suspended in 150µL of Tris 20mM-EDTA 1 Mm (TE 20-1) and incubated at 58ºC to dissolve the sample completely. Samples were then stored in 1.5mL Eppendorf's at 4ºC. DNA concentration and quality were quantified by spectrophotometry (Nanodrop 2000, Thermo Fisher Scientific, Waltham, USA).

2.1.4. *TRIM5* and *TRIM22* PCR amplification

TRIM5 and *TRIM22* are large genes, at 28915 and 21277 base pairs respectively. Both genes contain a total of eight exons, with exons two to eight coding for protein domains. A

large intronic sequence separates exons four and five which makes sequencing the full coding DNA sequence difficult.

Amplification primers were designed to flank exons 2 - 4 and 5 - 8 using the *TRIM5* and *TRIM22* reference sequences NG_029122 and NC_000011 respectively from the NCBI database (**table 2.1**). Non-specific amplification producing multiple products was a significant problem during optimisation of these protocols. This can be attributed to the presence of multiple paralogues of *TRIM5* in the human genome. To rectify this, touchdown amplification protocols were developed for polymerase chain reaction (PCR). Single step, touchdown PCR reactions were optimised, and separate inner sequencing primers were designed to sequence individual exons within the larger amplicons (**table 2.2**). All primers were synthesised by Life Technologies (Thermo Fisher Scientific).

Table 2.1: Primers for amplification and sequencing of *TRIM5*

Primer name	**5' – 3' oligo sequence**	**Binding site** [1]
Primers used in PCR		
TRIM5 EX2_4F	GCTGAGAGAAGAGCAAATGTG	9463 - 9483
TRIM5 EX2_4R	ATTGCTCAGCTTAGCCAC	12210 - 12227
TRIM5 EX5_8F	CCGCGTAAACATCTTGAGC	21710 - 21728
TRIM5 EX5_8R	AACTCCTGTCTCTCCATATACC	25808 - 25829
Primers used in sequencing		
TRIM5A EXO2F	GCCCATTTTCTAATTGTGCACAATC	9786 - 9810
TRIM5A EXO2R	AACTGCCTTGTCATGATAGACC	10426 - 10447
T5E3F SEQ PRI	GCCATTCTCACAGTTTCTGCA	10817 - 10837
T5E3R SEQ PRI	TAGTCCTCTTCATCCTTTCC	11149 - 11168
T5E4F SEQ PRI	GGCAGCTTTCCTAAGACCTACT	11538 - 11559
T5E4R SEQ PRI	GAAGTAGCAAGAGATAATCCC	12061 - 12081
TRIM5A EXO5F	GCTATGTTGCCTAGGCTGTA	22023 - 22042
TRIM5A EXO5R	TAGACCAGATTTGGCCAGGT	22600 - 22619
T5E67F SEQ PRI	GCCTCAATGGGTTGAAATGA	23936 - 23955
T5E67R SEQ PRI	CAAGACACACATAACTTACCC	24573 - 24593
T5E8F SEQ PRI	GGTTCCTCCCAGTTTTCTCT	24615 - 24634
T5E8R SEQ PRI	CCATCTTTTCCTTGCTGTCTCC	25387 - 25408

[1] Binding site in reference sequence NG_029122. Reverse primers were ordered as reverse complement sequences of those listed in the table.

Table 2.2: Primers for amplifying and sequencing *TRIM22*

Primer name	**5' – 3' oligo sequence**	**Binding site** [1]
Primers for PCR		
T22 E2_4F_2	GGGTTACACGAAGCTCTTGG	6292 - 6311
T22 E2_4R_2	TGAATGGGAAGGGAAATTCC	9137 - 9157
T22 E5_8F_3	TCTGCAGTGGTCAGTCTGAA	16647 - 16666
T22 E5_8R_3	CATTTGCTAGGGCTTCCATAGC	20196 - 20217
Primers for sequencing		
T22E2-3F SEQ PRI	ACTGCAGGAGTTTGTGACCAAG	6582 - 6603
T22E2-3R SEQ PRI	CCTTCCTCTTTGCTGTGCTAG	7833 - 7853
T22E4F SEQ PRI	CAGGCTCATACAAAGCAG	8663 - 8680
T22E4R SEQ PRI	GACATGAAATGGTTCCTTTGGCC	9059 - 9081
T22E5F SEQ PRI	CCAGGGATAGTAAGAGTGTG	16722 - 16741
T22E5R SEQ PRI	GTCCCAAATTGCTAGTCACAG	17154 - 17174
T22E6-7F SEQ PRI	TTGGATGGAGAGAAATAGGG	18530 - 18549
T22E6-7R SEQ PRI	CCTTTCAGAATTGTGGTTAC	19048 - 19067
T22E8F SEQ PRI	GTTCCGTTTGATATCGACAC	19360 - 19379
T22E8R SEQ PRI	CTTCCTTTCTTTCCCCTTCT	20153 - 20172

[1] Binding site in reference sequence NC_000011. Reverse primers were ordered as reverse complement sequences of those listed in the table.

Master mix and cycling conditions for PCR

All PCRs were done using the Advantage 2 PCR kit (Takara, Kusatsu, Japan), all PCR, PCR purification and pre-sequencing reactions were performed on an Alpha Cycler 4 (PCRmax Limited, Staffordshire, United Kingdom)(**table 2.3**).

Master Mix per reaction:

16.5µL PCR grade H_20 (sterile, RNA and DNA free)

2.5µL of Advantage 10X Buffer SA

2.0µL of forward primer at 10ng/µL

2.0µL of reverse primer at 10ng/µL

2.0µL of 2.5Mm mixed deoxyribonucleotide triphosphate (dNTP) (Takara)

0.5 µL of 50X Advantage 2 polymerase mix (Takara)

Then add 1.0µL of 50ng/µL gDNA sample for a total of 25µL per reaction

Table 2.3: Thermocycler settings for touchdown PCR of *TRIM5* and *TRIM22*

PCR step	***T5* exon2-4** [2]	***T5* exon5-8**	***T22* exon2-4** [3]	***T22* exon5-8**
1. Denaturation	95 ºC – 60s	95 ºC – 60s	95 ºC – 60s	95 ºC – 60s
2.1 Denaturation	95 ºC – 30s	95 ºC – 30s	95 ºC – 30s	95 ºC – 30s
2.2 Annealing	65.5 – 55.5 ºC – 30s	65.5 – 55.5 ºC – 30s	67 – 57 ºC – 30s	70 – 60 ºC – 30s
2.3 Elongation	68 ºC – 3min	68 ºC – 4min30s	68 ºC – 3min30s	68 ºC – 4min30s
3.1 Denaturation	95 ºC – 30s	95 ºC – 30s	95 ºC – 30s	95 ºC – 30s
3.2 Annealing	55.5 ºC – 30s	55.5 ºC – 30s	57 ºC – 30s	60 ºC – 30s
3.3 Elongation	68 ºC – 3min	68 ºC – 4min30s	68 ºC – 3min30s	68 ºC – 4min30s
4. Final elongation	68 ºC – 5min	68 ºC – 5min	68 ºC – 5min	68 ºC – 5 min

[1] For each stage of the PCR temperature is shown in degree Celsius with the time for each step reported in minutes and seconds.[2] *T5* = *TRIM5*. [3] *T22* = *TRIM22*.

Steps 2.1 to 2.3 were repeated for 10 cycles, decreasing the annealing temperature by 1 ºC per cycle. Following this, steps 3.1 to 3.3 were repeated for 25 cycles with the same annealing temperature for a total of 35 cycles of amplification.

2.1.5. Post-PCR product visualisation and purification

After PCR, products were separated via gel electrophoresis on a 1% agarose, Tris-acetate-EDTA (TEA) gel stained with ethidium bromide for 25min at 110V. Products were then visualised under ultraviolet light in a Gel-doc against a 1kb hyperladder (Bioline). PCR products were then purified for sequencing using an

Exo-Sap master mix per reaction:

1. Recombinant exonuclease I (New England Biolabs, Ipswich, USA) - 0.1µL per reaction

2. Recombinant shrimp alkaline phosphatase (New England Biolabs) - 0.5µL

3. PCR grade H_20 - 0.4µL

1.0µL of the EXO-SAP mix was then added to each PCR product and incubated on a thermocycler as follows:

1 cycle 37 ºC for 45min

1 cycle 80 ºC for 25min

2.1.6. Sanger pre-sequencing reaction

After post-PCR processing PCR products were pre-sequenced via Sanger sequencing using the Big Dye Terminator (BDT) v3.1 Cycle sequencing kit (Thermofisher). Template negative controls were carried through from PCR through product purification and sequencing to ensure that contamination between wells did not occur in a sequencing plate.

The BDT master mix per reaction:

1. BDT 5X buffer - 2.08µL
2. BDT v1.1 V3.1 - 0.25µL
3. PCR grade H_20 - 5.36µL
4. Primer (10 µM) - 0.32µL
5. PCR product - 2.0µL

Pre-Sequencing reactions were performed on a thermocycler as follows:

1. 96ºC – 1min
2. 96ºC – 10s
3. 50ºC – 5s
4. 60ºC – 4min

2.1.7. Post-sequencing product purification

Excess oligonucleotides were removed from pre-sequenced PCR products using ethanol precipitation as follows:

5µL of 125mM EDTA was added to each sample well containing 2µL of sequenced PCR products and incubated at room temperature for 5min. 40µL of 100% ethanol was added to each well and incubated at -20ºC for 1 hour. Following incubation, DNA salts were centrifuged at 3000g for 30min at 4ºC on a Heraeus 40 centrifuge (Thermofisher). After centrifugation, uncovered sequencing plates containing the precipitated DNA salts were inverted and briefly centrifuged at 100g. 40µL of 70% ethanol was added into each well and centrifuged at 3000g for 15min at 4ºC. The uncovered sequencing plates were again inverted and centrifuged at 200g for 1min. The precipitated DNA salts were resuspended in 10µL of Hi-Di Formamide. Samples were sequenced the same day or stored at -20ºC. Sequencing was performed on a 3730XL ABI sequence analyser at the Medical Research Council Weatherall Institute for Molecular Medicine (University of Oxford, United Kingdom).

2.1.8. Sequence analysis

Sequence electropherograms for *TRIM5* and *TRIM22* were analysed in Geneious 2019.2 (4). Forward and reverse reads were collected and assembled into contiguous patient sequences. These assembled contigs were aligned to the relevant reference sequence for *TRIM5* and *TRIM22*. Sequences with HQ scores of less than 80% were not included in further analysis. This score indicates that 80% of bases are high quality (likelihood of a false positive base reading of 1:10 000). Single nucleotide polymorphisms were identified

at a 35% heterozygosity threshold and identified in relation to the dbSNP database administered by the National Center for Biotechnology Information (NCBI).

Linkage statistics were calculated in Haploview and linkage disequilibrium (LD) maps generated (5). Linkage disequilibrium describes the correlation between alleles at different loci within a population. Central to the measurement of LD is the following relationship (6).

$$DAB = pAB - pApB$$

DAB is the coefficient of LD between two alleles A and B. pAB is the frequency of haplotype frequency of these alleles with pA and pB the alleles' frequencies. Linkage disequilibrium is routinely reported using D' which is the ratio of D to its maximum possible absolute value, given the allele frequencies. Perfect LD is reported as D' = 1.0 (6). D' scores can bias toward high scores when an allele exists at a low frequency and here r^2 is preferred for measuring LD

$$r^2 = \frac{D^2}{pA\,(1-pA)pB(1-pB)}$$

D is affected by several evolutionary forces acting at the population level including, random mutation, genetic drift, gene flow and selective pressure. Measurements of heterozygosity and testing of Hardy-Weinberg Equilibrium (HWE) were also done in Haploview. If a locus is in HWE it indicates that alleles on each chromosome are independently associated. Significant deviations from HWE can indicate inbreeding in the population, strong positive selection or genotyping errors (6).

Combined Annotation Dependent Depletion (CADD) scores and population level genotype frequency data were extracted from the Ensembl Project (7,8). The CADD score for a variant integrates data from multiple sources via a machine-learning algorithm to

determine the probability that the variant in question is pathogenic. The data which is used to generate this score includes different predictive scores for determining the probable pathogenicity of a variant (e.g. Polyphen or SIFT) as well as the likelihood that it has been tolerated in human evolution. A scaled CADD score indicates which percentile of pathogenic mutations a variant falls in. A score of 20 – 29 indicates that the mutation is in the top 1% of pathogenic mutations, a score of >30 indicates that it is in the top 0.1%.

Expression quantitative trait loci (eQTL) statistics of gene expression for *TRIM22* in human tissue was extracted from the Ensembl Project (8). A significant negative eQTL score for a variant indicates that gene expression of a variant is decreased in relation to the reference allele.

2.1.9. Quantification of plasma viral loads for CMV and HIV

These experiments were done by Dan Bowen (DB) and Louis-Marie Yindom (L-MY), using similar methods for both cytomegalovirus (CMV) and HIV.

Extraction of CMV DNA From Plasma: Total viral nucleic acids were extracted at the Biomedical Research and Training Institute laboratories in Harare from 200µL plasma using the QIAamp MinElute Virus Spin Kit (Qiagen, Hilden, Germany). Then, 60µL of total viral nucleic acids including CMV DNA were eluted and immediately stored at −80°C for subsequent testing.

CMV Detection and Quantification: CMV DNAaemia detection and quantification was performed by quantitative polymerase chain reaction (qPCR) using the RealStar CMV PCR kit v1.0 (Altona Diagnostics, Hamburg, Germany). CMV DNA was quantified with the QuantStudio 3 Real-Time PCR System (Applied Biosystems, CA). Kits were purchased as a single lot and were used following the manufacturer's instructions. Each qPCR reaction

contained a heterologous internal control to validate PCR efficiency. Each patient sample was tested in duplicates to minimise false negative results and any sample with discordant results was repeated. To minimise other sources of variation, all CMV DNA measurements were performed in the same laboratory using the same instrument by the same person (L-MY). Each experimental batch contained four quantification standards as positive controls, with known CMV DNA concentrations ranging from 10 - 10^4 IU/μL, which were used to generate standard curves. The analytical sensitivity (Limit of Detection - LoD) of the RealStar CMV kit is defined as the concentration of CMV DNA molecules that can be detected with a positivity rate of 95%. A dilution series of CMV DNA was analysed in eight replicates. Using the WHO International Standard for Human Cytomegalovirus and the Applied Biosystems real-time PCR instruments the LoD was 0.67IU/μL (CI ≥ 95%: 0.32 – 2.26 IU/μL). The conversion factor for DNA values to copies/mL was 0.6. Of note, longitudinal samples (baseline and follow-up) were tested in the same plate, at the same time, with the same reagents, by the same operator using the same qPCR instrument as described above.

2.2. Methods for chapter 4

2.2.1. Patient recruitment

The Police cohort is a prospective open cohort study of police officers recruited from urban and rural areas in Guinea-Bissau. All police officers were invited to join the study and 98% agreed to do so. The cohort started recruiting patients on 6th February 1990 and continued recruitment until 28th September 2009. A total of 4817 people were enrolled and 872 tested HIV positive during the study, 408 HIV-1 positive and 464 HIV-2 positive. HIV-infected participants were followed up until 28th September 2013. Subsets of HIV-infected participants have been followed up after this date, but unfortunately not systematically for the entire cohort. Our analysis therefore ends at the 28th of September 2013.

Blood samples for serology and $CD4^+$ T cell counts were collected at enrolment and at scheduled follow-up visits every 12 – 18 months. HIV testing was done at the National Public Health Laboratory in the capital Bissau. HIV disease was classified according to the WHO staging criteria (9). All participants received detailed information about the study before inclusion, as well as pre- and post-test counselling. If a participant died mortality reports from the Ministry of Interior were crosschecked by health post staff at the main police station in Bissau. Reported symptoms before death were used to classify cause of death. ART was introduced in Guinea-Bissau in 2005 through a national treatment programme. The Police cohort was included in the programme in early 2006. Individuals receiving ART were right-censored from the analysis from the timepoint of ART initiation. Date of HIV infection was estimated as the mid-timepoint between the last HIV-seronegative and the first HIV-seropositive sample. Patients with estimated dates of infection are referred to as **seroconverters** and those HIV-2 infected at enrolment as

seroprevalent patients. Patient data for HIV-2 infection was right-censored in the event of HIV-1 superinfection.

Ethics approval and consent to participate. The study was approved by the ethics committees of the government of Guinea-Bissau; the University of Lund, Sweden; and the Karolinska Institute, Stockholm. Study participants were counselled and provided informed oral consent

2.2.2. $CD4^+$ T cell count analysis

Both absolute $CD4^+$ T cell counts and $CD4^+$ T cell percentage (CD4%) are reliable immunological markers of HIV disease progression. In resource-limited settings, the CD4% has lower variability and sensitivity to specimen handling, patient age, or time of sampling (10). T-lymphocyte subsets were determined at the National Public Health Laboratory (LNSP), by conventional flow cytometry (Until 2005: FACStrak; Becton Dickinson, San Jose, CA, and from 2006 and onwards: CyFlow, Partec, Münster, Germany). Leucocyte counts were performed with a cell counter until 2005 (Coulter Counter CBC5; Coulter Electronics Ltd, Luton, England), and by CyFlow from 2006 and onwards.

2.2.3. Stratification of HIV-2 infected individuals into fast and slow disease progression groups

Longitudinal follow-up of ART naïve HIV-1 infected individuals has suggested that time to AIDS and death can be explained by $CD4^+$ T cell absolute counts and decline rate as well as viral load (11–14).

For patients with two or more CD4% measurements we analysed per patient linear regression models in SPSS to estimate the midpoint CD4% in follow-up time (**figure 2.1**)(15). We used the coefficient of the regression model as the CD4% change rate, measured as CD4% change/day. CD4% increased, stayed the same and decreased for different patients over follow-up; we therefore described CD4% as change rate and not decline rate. We have indicated a negative change rate for a decreasing CD4% with time and a positive change rate for increasing CD4%. This procedure then gave two statistics per patient, a midpoint CD4% level and CD4% change rate. These data were then transformed into proportional values, with each data point expressed as a proportion of the highest value which was 1.0 (highest midpoint CD4% or change rate divided by itself). The transformed data points now fell on a positive scale between 0 and 1. Patients were then ranked from highest CD4% level to lowest, and highest change rate to lowest (they were all on a positive scale at this point). Those with CD4% levels and change rates greater than the mean value were classified as **slow progressors**, those below the mean as **fast progressors**. An example of patients with their values used in the transformation is shown in **table 2.4**.

Table 2.4: Patient stratification values

Police	Midpoint CD4% [1]		CD4% change rate [1]		Combined coefficient	Midpoint CD4%	CD4% change rate	Combined coefficient
DL1751	31.10	0.95	-0.005	0.04	0.49	slow	fast	fast
DL2086	25.75	0.78	-0.004	0.08	0.43	slow	fast	fast
DL2224	32.91	1.00	0.002	0.78	0.89	slow	slow	slow
DL2233	24.83	0.75	0.003	0.89	0.82	fast	slow	slow
DL2948	29.33	0.89	0.003	0.88	0.88	slow	slow	slow
DL2973	18.00	0.55	0.000	0.54	0.55	fast	slow	fast
DL2986	32.80	1.00	0.000	0.55	0.77	slow	slow	slow
DL3105	32.25	0.98	0.004	1.00	0.99	slow	slow	slow
DL3242	31.11	0.95	-0.003	0.28	0.61	slow	fast	fast
DL3346	18.89	0.57	0.000	0.58	0.58	fast	slow	fast
DL3440	16.33	0.50	0.000	0.55	0.52	fast	slow	fast
DL3461	15.00	0.46	-0.001	0.46	0.46	fast	fast	fast
DL3641	18.00	0.55	-0.005	0.00	0.27	fast	fast	fast

[1] Midpoint CD4% and CD4% change rates per day are shown in their left columns with the transformed proportional values alongside them. Stratifications into fast and slow progressor groups are shown in the three right-most columns

Finally, to account for variability in disease stages at enrolment we used a combined coefficient derived from an equally weighted estimate using both CD4% level and change rate. To do this, the ranked midpoint CD4% values and CD4% change rates were added together and divided by two. As an example, DL1751's transformed midpoint CD4% value of 0.95 (slow progressor) was added to the transformed CD4% 0.04 (fast progressor) and averaged to give a combined coefficient of 0.49 (fast progressor). This is referred to as the combined coefficient. Using the combined coefficient to stratify patients we divided the cohort into relatively equal sized groups, 101 fast progressors and 118 slow progressors. Patients were consistently classified as fast or slow progressors across all three measures in 108/219 (49%) cases. The coefficients plotted per patient showed a normal distribution with some overlap between the plots, this indicates clearly differentiated groups when stratified according to a combined estimate of CD4% level and change rate (**figure 2.1**).

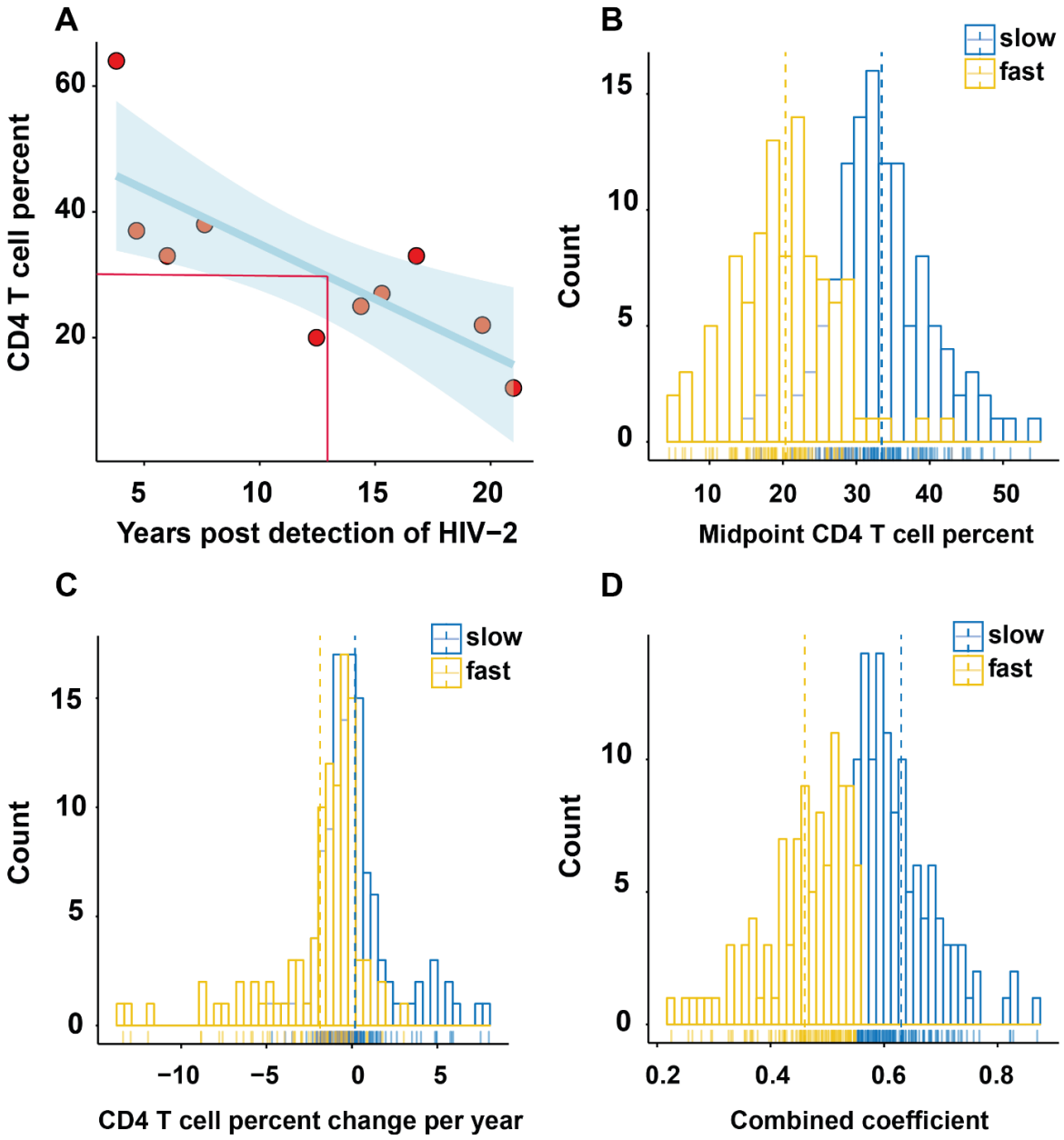


Figure 2.1: Stratification of HIV-2 infected adults in fast and slow disease progressor groups: **A)** Patient DL1751's CD4% over time is shown, there was a high CD4% and classified as a slow progressor based on the midpoint CD4% but the decline rate was also rapid (fast progressor). This patient was stratified as a fast disease progressor on the combined coefficient given the magnitude of the CD4% decline rate. **B)** Distribution of midpoint CD4%, prior to transformation, by the combined coefficient stratification. **C)** Distribution of CD4% change rates, prior to transformation, by the combined coefficient. **D)** Distribution of combined coefficient. Dashed lines for panels B – D indicate the mean x-axis values for each group.

2.3. Methods for chapter 5

2.3.1. Patient recruitment

Patients were recruited from the Police cohort described in 2.2.1

2.3.2. DNA extraction and storage

DNA was extracted from frozen blood pellets using the QIAamp DNA blood MIDI extraction kit (Qiagen, Hilden, Germany). 100µl of Qiagen Protease was pipetted into the bottom of a sterile 15mL centrifuge tube. After this 500µL of whole blood or 1mL of uncounted PBMC was added and mixed thoroughly. After this 1.2mL of Buffer AL was added, and mixed by inverting the tube 15 times, followed by vortexing for at least 1min. 10µL of carrier RNA was then added and incubated at 70°C for 10min in a water bath. 1mL of ethanol (96–100%) was added to the sample and mixed by inverting the tube 10 times, followed by additional vortexing. The sample was then added to a QIAamp Midi column placed in a 15mL centrifuge tube, without moistening the rim, and centrifuged at 1850xg (3000rpm) for 3min in a Megastar 1.6R centrifuge (VWR, Radnor, USA). The filtrate was then discarded and then 2mL Buffer AW1 added to the QIAamp Midi column. The sample was then centrifuged at 4100xg (5000rpm) for 3min. After this 2mL Buffer AW2 was added to the QIAamp Midi column. The sample was then centrifuged at 4100xg (5000rpm) for 15min. The QIAamp Midi column was then placed in a clean 15mL centrifuge tube and the filtrate discarded. 200µL Buffer AE, equilibrated to room temperature (15 – 25°C), was pipetted directly onto the membrane of the QIAamp Midi column. The sample was then incubated at room temperature for 5 min, and centrifuge at 4100xg (5000rpm) for 5min. To increase the DNA concentration in the final sample the eluate was reloaded on the QIAamp

Midi Column, incubated at room temperature for 5min. Finally, the sample was centrifuged at 4100xg (5000rpm) for 5min. DNA concentration and quality were quantified by spectrophotometry (Nanodrop 2000, Thermo Fisher Scientific, Hemel Hempstead, United Kingdom).

2.3.3. *TRIM5* and *TRIM22* PCR amplification

Touch down polymerase chain reactions were used to amplify *TRIM5* and *TRIM22* using similar methods as described for chapter three above. However, primers which amplified *TRIM5* and *TRIM22* in the ZENITH cohort were often unsuccessful in the Police cohort. We determined that this was likely to be due to polymorphisms in these primer binding sites and we therefore redesigned amplification primers for the Police cohort (**table 2.5 and 2.6**). All other steps were unchanged.

Table 2.5: Primers for amplification and sequencing of *TRIM5*

Primer name	**5' – 3' oligo sequence**	**Binding site** [1]
Primers used in PCR		
TRIM5A EXO2F	GCCCATTTTCTAATTGTGCACAATC	9786 - 9810
T5E4R SEQ PRI	GAAGTAGCAAGAGATAATCCC	12061 - 12081
TRIM5 EX5-8F	CCGCGTAAACATCTTGAGC	21710 - 21728
TRIM5 EX5-8F	AACTCCTGTCTCTCCATATACC	25808 - 25829
Primers used in sequencing		
T5E2F SEQ PRI	CCTTTCTTACTTGGTCCC	9830 - 9847
T5E2R SEQ PRI	GTCATGATAGACCTGAGACCCGGG	10435 - 10458
T5E3F SEQ PRI	GCCATTCTCACAGTTTCTGCA	10817 - 10837
T5E3R SEQ PRI	TAGTCCTCTTCATCCTTTCC	11149 - 11168
T5E4F SEQ PRI	GGCAGCTTTCCTAAGACCTACT	11538 - 11559
T5E4R SEQ PRI	GAAGTAGCAAGAGATAATCCC	12061 - 12081
TRIM5A EXO5F	GCTATGTTGCCTAGGCTGTA	22023-22042
TRIM5A EXO5R	TAGACCAGATTTGGCCAGGT	22600-22619
T5E67F SEQ PRI 2	CCCCAAGAAGAGAGTTCAAGTGCC	23915 - 23938
T5E67R SEQ PRI 2	CCAAGACACACATAACTTACCCC	24572 - 24594
T5E8F SEQ PRI	GGTTCCTCCCAGTTTTCTCT	24615 - 24634
T5E8R SEQ PRI	CCATCTTTTCCTTGCTGTCTCC	25387 - 25408

[1] Binding site in reference sequence NG_029122. Reverse primers were ordered as reverse complement sequences of those listed in the table.

Table 2.6: Primers for amplifying and sequencing *TRIM22*

Primer name	**5' – 3' oligo sequence**	**Binding site** [1]
Primers for PCR		
TRIM22 EX2_3F	CCTGTCCCCTTCAACATTCT	6558 - 6577
TRIM22 EX4R	GTATGAATGGGAAGGGAAATTCC	9134 - 9156
TRIM22 E5_8F_3	TCTGCAGTGGTCAGTCTGAA	16647 - 16666
TRIM22 E5_8R_3	CATTTGCTAGGGCTTCCATAGC	20196 - 20217
Primers for sequencing		
T22E2-3F SEQ PRI	ACTGCAGGAGTTTGTGACCAAG	6582 - 6603
T22E2R SEQ PRI	GCCATAGGTTCTAATCACCAG	7523 - 7543
T22E3F SEQ PRI	CTCTTTGCTTAGTTTCTGATGCC	7201 - 7223
T22E2_3R SEQ PRI	CCTTCCTCTTTGCTGTGCTAG	7833 - 7853
T22E4F SEQ PRI	CAGGCTCATACAAAGCAG	8663 - 8680
T22E4R SEQ PRI 3	ACAGAGGCCGATTTTCCTTCCC	9006 - 9027
T22E5F SEQ PRI	CCAGGGATAGTAAGAGTGTG	16722 - 16741
T22E5R SEQ PRI 2	GTCCTGAGAGATATCTTCCTTCA	17081 - 17103
T22E6-7F SEQ PRI	TTGGATGGAGAGAAATAGGG	18530 - 18549
T22E6-7R SEQ PRI 2	CTGAGTCTCATGTGTTCCTCCC	19084 - 19105
T22E8F SEQ PRI	GTTCCGTTTGATATCGACAC	19360 - 19379
T22E8R SEQ PRI	CTTCCTTTCTTTCCCCTTCT	20153 - 20172

[1] Binding site in reference sequence NC_000011. Reverse primers were ordered as reverse complement sequences of those listed in the table.

Master mix and cycling conditions for PCR

All PCRs were done using the Advantage 2 PCR kit (Takara), all PCR, PCR purifications and pre-sequencing reactions were performed on an Alpha Cycler 4 (PCRmax Limited, Staffordshire, United Kingdom, **table 2.7**).

Master Mix per reaction:

16.5µL PCR grade H_20 (sterile, RNA and DNA free)

2.5µL of Advantage 10X Buffer SA

2.0µL of forward primer at 10ng/µL

2.0µL of reverse primer at 10ng/µL

2.0µL of 2.5Mm mixed deoxyribonucleotide triphosphate (dNTP) (Takara)

0.5µL of 50X Advantage 2 polymerase mix (Takara)

Then add 1.0µL of 50ng/µL gDNA sample for a total of 25µL per reaction

Table 2.7: Thermocycler settings for touchdown PCR of *TRIM5* and *TRIM22*

PCR step	***T5* exon2-4** [2]	***T5* exon5-8**	***T22* exon2-4** [3]	***T22* exon5-8**
1. Denaturation	95ºC – 60s	95ºC – 60s	95ºC – 60s	95ºC – 60s
2.1 Denaturation	95ºC – 30s	95ºC – 30s	95ºC – 30s	95ºC – 30s
2.2 Annealing	65.5 – 55.5ºC – 30s	65.5 – 55.5ºC – 30s	67 – 57ºC – 30s	70 – 60ºC – 30s
2.3 Elongation	68ºC – 3min	68ºC – 4min30s	68ºC – 3min30s	68ºC – 4min30s
3.1 Denaturation	95ºC – 30s	95ºC – 30s	95ºC – 30s	95ºC – 30s
3.2 Annealing	55.5ºC – 30s	55.5ºC – 30s	57ºC – 30s	60ºC – 30s
3.3 Elongation	68ºC – 3min	68ºC – 4min30s	68ºC – 3min30s	68ºC – 4min30s
4. Final elongation	68ºC – 5min	68ºC – 5min	68ºC – 5min	68ºC – 5 min

[1] For each stage of the PCR temperature is shown in degree Celsius with the time for each step reported in minutes and seconds.[2] *T5 = TRIM5*. [3] *T22 = TRIM22*.

Steps 2.1 to 2.3 were repeated for 10 cycles, decreasing the annealing temperature by 1ºC per cycle. Following this, steps 3.1 to 3.3 were repeated for 25 cycles with the same annealing temperature for a total of 35 cycles of amplification.

Post-PCR product visualisation and purification, Sanger pre-sequencing reaction, post sequencing product purification and sequence analysis were done as described for chapter three.

2.4. Methods for chapter 6

2.4.1 Study participants

Study participants were drawn from the Guinea-Bissau Police cohort which was described in 2.2.1. Twelve HIV-2 infected patients were selected based on availability of plasma samples and meeting the following inclusion criteria:

i) HIV-2 mono-infected

ii) Antiretroviral therapy naïve at the time of plasma sample collection

iii) Longitudinal CD4% data available to classify patients as relative fast or slow disease progressors as described previously

2.4.2 RNA extraction

RNA extractions were performed by Angelica Palm and Sara Karlson at Lund University. Total RNA was extracted from plasma samples which had been stored at -80ºC, using the Qiagen RNeasy Lipid Tissue Mini Kit with minor modifications to the manufacturers' instructions. 100µL of plasma was used for each extraction. 1mL of Qiazol was added to each sample, mixed thoroughly and then incubated at room temperature for 10min. 5µg of carrier RNA was then added followed by 200µL of chloroform. The sample was mixed, then incubated for 3min at room temperature and then centrifuged at 12000xg for 15min at 4ºC in a Megastar 1.6R centrifuge (VWR, Radnor, USA). The aqueous phase was removed and a further 5µg of carrier RNA added to it, followed by an equivalent volume of 70% ethanol. 700µL of the aqueous phase was added to a provided spin column and then centrifuged at 8000xg for 15s at room temperature. The flow through was then removed,

and the rest of the sample added to the spin column for centrifugation. After this 350µL RW1 was added to the sample which was then centrifuged at 8000xg for 15s at room temperature. 10µL of DNase I was added to 70µL of Buffer RDD and mixed. This 80µL solution was then added to the sample and incubated for 15min at room temperature. Next 350µL RW1 was added and the sample centrifuged at 8000xg for 15s at room temperature. 500µL of RPE was added and the sample centrifuged at 8000xg for 15s at room temperature. Next a further 500µL RPE was added and the sample centrifuged at 8000xg for 2min at room temperature RT. The spin column was then placed in a new collection tube, centrifuged at 12000xg for 1min at room temperature. Finally, the spin column was placed in a new 1.5mL tube, 40µL of PCR grade water was added to the membrane and incubated for 10min at room temperature. The sample was then centrifuged at 8000xg for 1min at room temperature.

2.4.3 HIV-2 *gag* PCR

Primer design: HIV displays significant diversity between patients and also within the quasispecies circulating in a patient at any given timepoint (16). To design a PCR which would amplify the p24 region of *gag* consistently across patients, and across timepoints, it is necessary to select primers which included conserved sites between patients. We downloaded 356 sequences from the Los Alamos National Laboratory (LANL) sequence database spanning HIV-2 *gag* from position 1103 – 2668 on the HIV-2 BEN.M30502 reference sequence. We then removed short sequences, clonal sequences and all non-group A sequences - the Police cohort is very unlikely to be infected with other HIV-2 groups (17). This left 27 sequences (**table 2.8**) which were then aligned in Geneious 2019.2 using the Muscle alignment algorithm (4).

Table 2.8: HIV-2 *gag* sequences used in primer design

Number	Accession number [1]
1	A.GW.1987.CAM2CG.D00835:
2	A.NL.2003.RH2.21.MF595865:
3	A.FR.2000.LA38.KY025539:
4	A.FR.1998.LA41.KY025542:
5	A.DE.-.PEI2_KR_KRCG.U22047:
6	A.JP.2008.NMC786_clone_41.AB731742:
7	A.DE.-.BEN.M30502:
8	A.FR.1993.LA37.KY025538:
9	A.PT.-.ALI.AF082339:
10	A.IN.2007.NNVA.EU980602:
11	A.FR.1996.LA40.KY025541:
12	A.CI.1988.UC2.U38293:
13	A.FR.2001.LA42.KY025543:
14	A.FR.2002.LA36GomM.KU168287:
15	A.FR.1998.LA39.KY025540:
16	A.IN.1995.CRIK_147.DQ307022:
17	A.GH.-.GH1.M30895:
18	A.GW.1986.FG_clone_NIHZ.J03654:
19	A.GW.-.MDS.Z48731:
20	A.GM.1987.D194.X52223:
21	A.GM.-.MCR35.AY509260:
22	A.SN.1986.ST_JSP4_27.M31113:
23	A.SN.1985.ROD.M15390:
24	A.GM.-.ISY_SBL_6669_85.J04498:
25	A.GM.-.MCN13.AY509259:
26	A.GW.1987.CAM2.KU179861:
27	A.IN.-.NIM_8.DQ973520:

[1] Accession numbers for sequences on the LANL database.

We designed primers for a nested PCR to flank the p24 region of HIV-2 BEN.M30502. Eight forward and six reverse degenerate primers were selected and tested in a variety of primer pairs (**figure 2.2**). We selected one pair of outer primers and one pair of inner primers (**table 2.9**). The nested PCR was able to amplify p24 consistently at an HIV-2 RNA concentration of 200 copies/mL. Below this level amplification became inconsistent. All PCRs were run on a LifeTouch thermal cycler (Bioer, Technology, Tokyo, Japan)

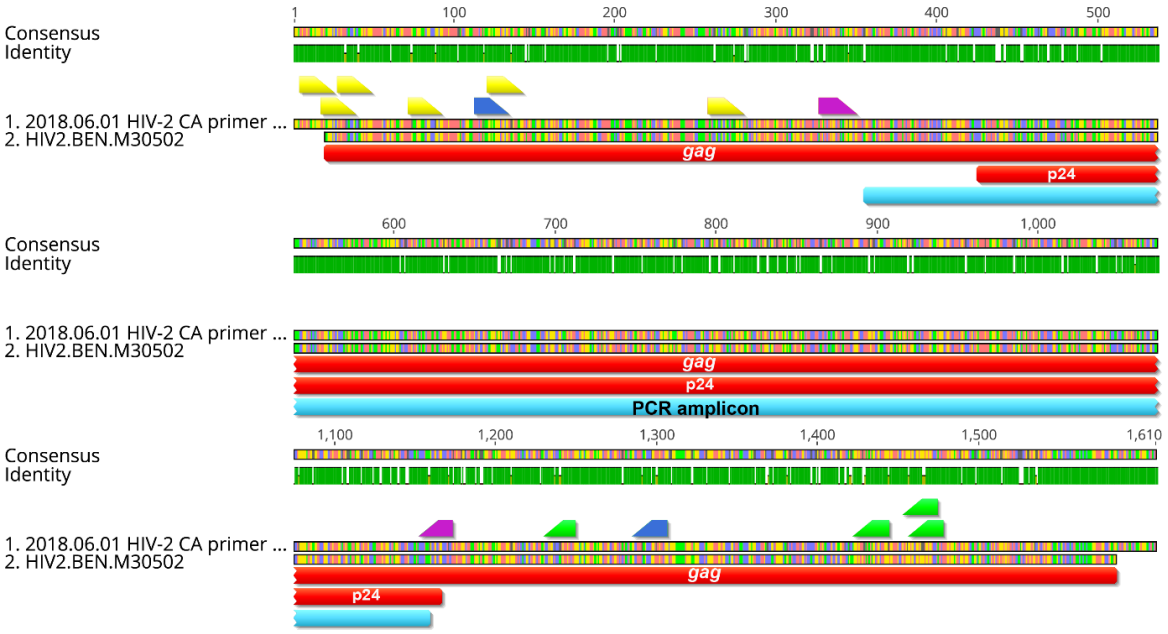


Figure 2.2: Primer design for HIV-2 *gag* PCR. The HIV-2 *gag* sequences alignment consensus sequence (2018.06.01 HIV-2 CA primer) is shown mapped to HIV-2 BEN.M30502. Identity agreement at the nucleotide level is shown with green indicating 100% agreement. The consensus of this pairwise alignment is also shown. Forward primers are highlighted in yellow, reverse in green. Outer primers used in the nested reaction are shown in blue, inner primers in purple. Annotations for the maximum inner amplicon length and the p24 region are shown.

Table 2.9: Primers used in the HIV-2 *gag* PCR

Primers	**Oligo sequence**	**Binding site** [1]	**Amplicon size** [2]
Outer primers			
HIV2CA_IF4	AAACATRTTGTGTGGGCAGCGA	1197 - 1218	1193
HIV2CA_IR2	CTGTCTWTCTGGGCARTTTGCC	2369 - 2390	
Inner primers			
HIV2CA_IF2	TAGTACAGAGACATCTAGYGGCAG	1411 - 1434	846
HIV2CA_IR3	TGYTGGGCTGCTGCRAATGGG	2237 - 2257	

[1] Binding site in reference to HIV-2 BEN.M30502 from 5' to 3'. [2] Amplicon length shown here includes the primer sequences. Forward primers are shown 5' to 3' and reverse primers are shown 3' to 5'. Degenerate primers were selected to allow amplification on sequences with variable bases.

Nested PCRs include two stages of amplification. The first step involved a one-step reverse transcription PCR (**RT-PCR**) using the SuperScript IV One-Step RT-PCR System with Platinum Taq DNA Polymerase (Thermofisher) (**table 2.10**). RT-PCR's start with reverse transcription: the RT-PCR master mix includes an RT enzyme. 5µL of extracted RNA was

added to 15.5µL of master mix which was then incubated at 50ºC for 30min followed by the PCR with the outer primers (**table 2.11**). The outer set of primers amplified a 1193 nucleotide base pair (bp) fragment from nucleotide 1197 – 2390 of HIV-2 BEN.M30502 (**figure 2.2**). After the RT-PCR 2µL of the amplified outer product was used as a template for the **nested reaction**. The nested reaction with the inner set of primers was performed using the Dream Taq PCR kit (Thermofisher). A negative control was carried through from RNA extraction to RT-PCR and the final nested reaction .The final sequenced PCR product represented a single open reading frame 735 nucleotides long and mapped to nucleotide 1460 – 2195 on HIV-2 BEN.M30502. PCR products were purified using the in-house EXO-SAP protocol described above and stored at -20ºC until further downstream applications.

Table 2.10: Reagents per reaction for HIV-2 *gag* PCR

RT-PCR	**Nested reaction**
12.5µL 2X buffer	16.875µL PCR-grade H_20
1.0µL forward primer (100ng/L)	2.5µL 10X Dream Taq buffer
1.0µL reverse primer (100ng/L)	2.5µL dNTP (2mM)
1.0µL enzyme mix (RT + Taq)	1µL forward primer (100ng/L)
	1µL reverse primer (100ng/L)
	0.125µL Dream Taq DNA polymerase (5U/L)

Table 2.11: Cycling conditions for HIV-2 *gag* PCR

PCR step	**RT-PCR**	**Nested PCR**
1. Denaturation	94ºC - 2min	94ºC - 2min
2.1 Denaturation	94ºC - 15s	94ºC - 30s
2.2 Annealing	53ºC - 15s	55ºC - 30s
2.3 Elongation	68ºC - 60s	72ºC - 90s
3. Final elongation	68ºC - 5min	72ºC - 5min

2.4.4 Cloning and sequencing

Twenty-five separate PCR products were amplified from the 12 patients, nine of the patients successfully amplified multiple PCR products from different timepoints and three had only a single timepoint successfully amplified. Sara Karlson performed several PCRs on samples in Lund which I cloned, amplified and sequenced in Oxford. The PCR products were cloned using a TOPO-TA cloning kit (Invitrogen, Carlsbad, CA, USA)(**figure 2.3**).

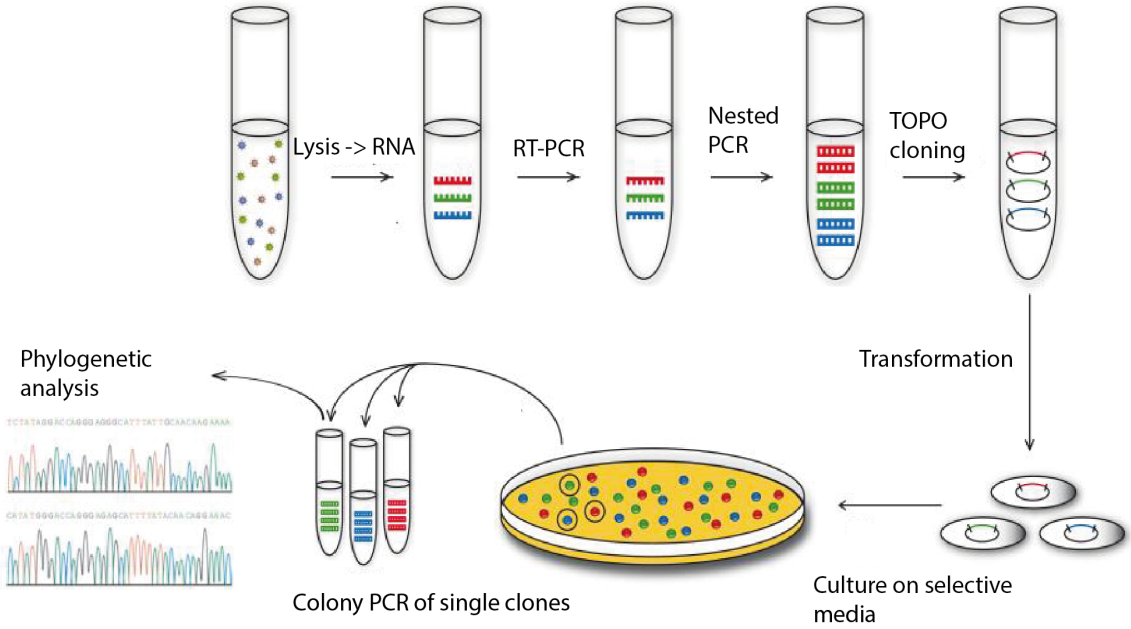


Figure 2.3: Cloning process for HIV-2 *gag* final PCR products. Steps involved in cloning and DNA sequencing of HIV-2 *gag* amplicons. Image courtesy of Joakim Esbjörnsson.

TOPO-TA cloning was performed in three separate steps: ligation, transformation and colony PCR.

Ligation: 4µL of purified HIV-2 *gag* PCR product was aliquoted into a clean 1.5mL Eppendorf tube. 1µL of salt solution and 1µL of pCRTM4-TOPO plasmid stock was added

and gently mixed by tapping the tube. This was then incubated for 30min at room temperature.

Transformation: For each vector (PCR product ligated to plasmid) one vial of chemically competent E. coli TOP10 cells was thawed from -80ºC on ice. 5µL of vector was added to one vial, gently mixed and incubated on ice for 30min. The cells were then heat-shocked at 42ºC for 30s in a water bath. Cells were then rested on ice for 2min. 250µL of sterile Super Optimal broth with Catabolite repression Media was added to each vial. Cells were then incubated at 37ºC for 1h in a shaking incubator rotating at 225rpm. After incubation 75µL of each vial was added to kanamycin (50µg/mL) selective Luria-Bertani media growth plates and incubated for 16h at 37ºC. Vector only negative controls were added to each experiment to exclude contamination between plates.

23 clones were picked routinely and amplified in a colony PCR using the Advantage 2 PCR kit (Takara). The colony PCR used the same primers as the nested reaction described above and gave an amplicon of identical length to the insert (735bp). The colony PCR products were then sequenced with the inner amplification primers with Macrogen Europe (Amsterdam, Netherlands) using Sanger sequencing. Clonal sequences from each timepoint represent the circulating quasispecies.

2.5 References for chapter 2

1. Ferrand RA, Simms V, Dauya E, Bandason T, Mchugh G, Mujuru H, et al. The effect of community-based support for caregivers on the risk of virological failure in children and adolescents with HIV in Harare, Zimbabwe (ZENITH): an open-label, randomised controlled trial. Lancet Child Adolesc Heal. 2017 Nov;4642(17):1–9.
2. World Health Organization. Antiretroviral therapy for HIV infection in infants and children: Towards Universal Access. Recommendations for a public health approach. 2010 Revision. World Health. 2010.
3. World Health Organization (WHO). Consolidated guideline on the use of antiretroviral drugs for treating and preventing HIV. Consolidated Guidelines on the Use of Antiretroviral Drugs for Treating and Preventing HIV Infection: Recommendations for a Public Health Approach. 2013.
4. Kearse M, Moir R, Wilson A, Stones-Havas S, Cheung M, Sturrock S, et al. Geneious Basic: An integrated and extendable desktop software platform for the organization and analysis of sequence data. Bioinformatics. 2012;
5. Barrett JC, Fry B, Maller J, Daly MJ. Haploview: analysis and visualization of LD and haplotype maps. Bioinformatics. 2005 Jan 15;21(2):263–5.
6. Slatkin M. Linkage disequilibrium-understanding the genetic past and mapping the medical future. Nat Rev Genet. 2016;9(6):477–85.
7. Kircher M, Witten DM, Jain P, O'roak BJ, Cooper GM, Shendure J. A general framework for estimating the relative pathogenicity of human genetic variants. Nat Genet. 2014;46(3):310–5.
8. Zerbino DR, Achuthan P, Akanni W, Amode MR, Barrell D, Bhai J, et al. Ensembl 2018. Nucleic Acids Res. 2018 Jan 4;46(D1):D754–61.
9. WHO case definitions of HIV for surveillance and revised clinical staging and immunological classification of HIV-related disease in adults and children HIV/AIDS programme. 2007.
10. van der Loeff MFS, Larke N, Kaye S, Berry N, Ariyoshi K, Alabi A, et al. Undetectable plasma viral load predicts normal survival in HIV-2-infected people in a West African village. Retrovirology. 2010 May 19;7(1):46.
11. Phillips AN, Staszewski S, Weber R, Kirk O, Francioli P, Miller V, et al. HIV Viral Load Response to Antiretroviral Therapy According to the Baseline CD4 Cell Count and Viral Load. JAMA. 2001 Nov 28;286(20):2560.
12. Mellors JW, Margolick JB, Phair JP, Rinaldo CR, Detels R, Jacobson LP, et al. Prognostic Value of HIV-1 RNA, CD4 Cell Count, and CD4 Cell Count Slope for Progression to AIDS and Death in Untreated HIV-1 Infection. JAMA. 2007 Jun 6;297(21):2346.

13. Rodríguez B, Sethi AK, Cheruvu VK, Mackay W, Bosch RJ, Kitahata M, et al. Predictive Value of Plasma HIV RNA Level on Rate of CD4 T-Cell Decline in Untreated HIV Infection. JAMA. 2006 Sep 27;296(12):1498.

14. Cori A, Pickles M, van Sighem A, Gras L, Bezemer D, Reiss P, et al. CD4+ cell dynamics in untreated HIV-1 infection. AIDS. 2015 Nov;29(18):2435–46.

15. IBM SPSS Inc. SPSS Statistics for Windows. IBM Corp Released 2012. 2016;

16. Lemey P, Rambaut A, Pybus OG. HIV evolutionary dynamics within and among hosts. Vol. 8, AIDS Reviews. 2006. p. 125–40.

17. Palm AA, Lemey P, Jansson M, Månsson F, Kvist A, Szojka Z, et al. Low Postseroconversion CD4 $^{+}$ T-cell Level Is Associated with Faster Disease Progression and Higher Viral Evolutionary Rate in HIV-2 Infection. Paraskevis D, editor. MBio. 2019 Jan 8;10(1).

Chapter 3

The impact of *TRIM22* genotype on disease outcomes of HIV-1 infected children and adolescents

3.1 Introduction

In 2018 there were approximately 37.9 million people living with HIV-1 across the globe with 1.7 million new infections and 770 000 deaths that year. In children and adolescents global prevalence was estimated at 1.7 million, with 160 000 incident infections and 100 000 deaths (1). Children and adolescents therefore account for 5% of the global HIV-1 infected population but 13% of HIV-1 related deaths. Children and adolescents who were perinatally infected with HIV-1 display distinct differences in disease progression and outcomes when compared to adults (2). The most striking difference is that time to developing AIDS and subsequently dying is, on average, far shorter with approximately 50% of these individuals progressing to AIDS at one year of age and death by two years (3). Children with perinatal HIV-1 infection have been stratified into slow progressors and fast progressors based on those who reach a CD4% of less than 25% by 24 months of age (4). Median plasma HIV-1 viral load (pVL) is significantly higher in fast progressors than slow progressors (1.1 million copies/mL versus 30 000 copies/mL). In addition, slow progressors maintain low levels of immune activation in the presence of a pVL (5). This

suppression of immune activation is maintained by an expanded population of suppressive regulatory T cells (5).

Congenital Cytomegalovirus (CMV) infection is associated with a greatly increased risk of HIV transmission in HIV-exposed neonates (6). In infants, HIV-1 infection earlier in life correlates with higher mortality (3). Coinfection with CMV has been associated with negative effects on growth and T cell activation in HIV-exposed, but uninfected, infants (7). In older HIV-1 infected children, CMV viraemia is associated with lower $CD4^+$ T cell counts and stunting (8). In pathogenesis studies of HIV-1 infected children it is therefore important to control for CMV viraemia, if possible.

In adults, immunogenetic factors which have been consistently associated with slower disease progression include HLA and KIR as well as deletions in *CCR5* (9). In contrast, HLA types do not associate with significant differences in viral replicative capacity (RC), $CD4^+$ T cell counts or pVL levels in children – nor do they affect disease progression rates (4). This weakened effect of HLA types on disease outcomes in perinatally infected children is thought to be caused by pre-adaptation to the mother's HLA type by the transmitted virus (10). Evolutionary bottlenecks which characterise vertical transmission of HIV-1 select for consensus viral sequences infecting children which have significantly lower RC (11). Higher RC has consistently been shown to correlate with fast disease progression rates in this population, and *vice versa*. Further mechanisms shaping disease phenotypes in perinatally infected children appear to be, as-yet, undefined and HLA-independent. We therefore aimed to study the effect of polymorphisms in host restriction factors *TRIM5* and *TRIM22* on disease progression in HIV-1 infected children and adolescents.

TRIM5α is a host restriction factor, one of a group of intracellular antiviral proteins which act as the effector arm of the Interferon response to disrupt the HIV life cycle at multiple points (12–16). This capsid-specific recognition then stimulates further downstream transcription of inflammatory cytokines, which enhances the antiviral response against HIV. HIV-1 displays high levels of resistance to human TRIM5α mediated restriction, likely to be due to longstanding adaptation of HIV lineages to closely related primates immune systems (17). However, HIV-1 cytotoxic lymphocyte (CTL) escape variants which have accumulated compensatory mutations in *gag* display increased sensitivity to restriction by TRIM5α (18). Therefore, single nucleotide polymorphisms (SNP) in *TRIM5* which result in loss or gain of function mutations may have a significant effect on HIV-1 disease outcomes in perinatally infected children due, in-part, to the accumulation of CTL compensatory mutations in transmitted viruses.

TRIM22 blocks transcription factor Sp1 from binding to the HIV-1 LTR region, thereby decreasing viral transcription. TRIM22 also prevents viral particle release through its association with gag proteins (19,20). Increased *TRIM22* expression in peripheral blood mononuclear cell (PBMC) is associated with higher $CD4^+$ counts and lower HIV-1 pVL (21). *TRIM22* is highly polymorphic and several single nucleotide polymorphisms (SNPs) alter its structure or function (22). Two SNPs in *TRIM22*, rs7935564 and rs1063303, have been associated with rapid progression to AIDS in adults (23). Furthermore, TRIM22 is an essential component of the Nucleotide-binding oligomerization domain-containing protein 2 (NOD-2) signalling pathway, and pathological loss of function mutations in *TRIM22* have been associated with early-onset inflammatory bowel syndrome (24). Defective NOD-2 signalling may affect downstream activation of the NFKB and interferon pathways, thus

impairing local cellular anti-viral defence systems. TRIM22 therefore represents an HLA-independent candidate gene for further association studies with HIV-1 disease outcomes.

In this chapter we test the hypothesis that *TRIM5* and *TRIM22* genotype, specifically non-synonymous SNPs, are associated with differences in disease outcomes in HIV-1 infected children and adolescents.

3.1.1 Specific aims

1. Characterise genetic diversity, defined as the number and frequencies of SNPs, in *TRIM5* and *TRIM22* in HIV-1 infected children and adolescents from Zimbabwean cohorts.

2. Perform SNP association studies with clinical outcomes of HIV-1 infection in this population.

3.2 Methods

3.2.1 HIV-1 infected patients

Samples were collected from 385 perinatally HIV-1 infected children and adolescents as part of the **Z**imbabwe study for **En**hancing Testing and **I**mproving **T**reatment of **H**IV in Children (ZENITH) cohort. The ZENITH study, supported provision of provider-initiated testing and counselling for older children and adolescents in Harare's primary care clinics to determine if this intervention improved adherence to antiretroviral therapy (ART) (25). These patients were predominantly vertically infected with HIV-1 (97.5% vertically infected and 2.5% horizontally infected) and were not on ART at the time of recruitment into the study.

From the 385 patients we selected 150 patients who had both adequate DNA samples in terms of volume, concentration and quality for sequencing of the full-length coding regions of *TRIM5* and *TRIM22.* An additional 125 patients were included, where only *TRIM22* exon 3 and 4 were amplified and sequenced. We selected this region for an expanded study due to previously published data indicating that there are important associations for SNPs in this region with HIV-1 disease outcomes in adults. This was done to increase our sample size and power for statistical modelling.

Our total sample of HIV-1 infected children and adolescents for SNP association studies was 275 individuals and here we focused on SNPs which met the following criteria:

i) Non-synonymous substitution

ii) A minority variant genotype frequency of greater than 5%

3.2.2 HIV-1 negative controls

The INHALE (**IN**vestigation of **H**eart **A**nd **L**ung dis**E**ase among older children in Harare, Zimbabwe) cohort was recruited for a study to evaluate the clinical and functional features of chronic lung and cardiac disease among vertically infected older children, stable on antiretroviral therapy (ART). Two hundred vertically HIV-1 infected children who were on ART for at least six months were recruited. In addition, 200 HIV-negative age and sex-matched peers were recruited as controls.

Of these HIV-1 negative patients we identified 104 individuals with suitable samples for sequencing of *TRIM22* exon three and four. This was done to provide age and ethnically matched control genotype frequencies for comparison to the ZENITH cohort.

3.2.3 $CD4^+$ T cell measurement

$CD4^+$ T cell counts ($CD4^+$ counts) were measured at three- and six-months post-initiation, and otherwise six monthly, with an Alere PIMA CD4 (Waltham, Massachusetts, USA) machine. Participant data were recorded by a study nurse on paper clinical report forms and entered into a Microsoft Access database using Cardiff TELEFORM Intelligent Character Optical Mark Recognition Software (Version 10.9).

3.2.4 Classification of progressor status

AIDS was defined according the WHO staging criteria as either (26):

i) Diagnosis of a WHO HIV stage IV defining illness

ii) Having a $CD4^+$ count less than 200 cells/μL

For the purposes of this study we classified long term non-progressors (LTNP) as patients who met the following criteria:

i) Age greater than 10 years **AND**

iii) An ART-naive $CD4^+$ count of greater than 500 cells/μL **AND**

iii) WHO clinical stage one to two

We classified all other individuals as average progressors, given that paediatric rapid progressors are unlikely to survive past age five without initiating ART. This distinction is arbitrary and was used to determine if there are significant differences between those individuals who have survived for prolonged periods of time with minimal disease progression and those with clearly progressive HIV infection as determined by decreasing $CD4^+$ counts.

3.2.5 DNA extraction and storage

DNA was extracted from whole blood using an in-house protocol.

3.2.6 Amplification and sequencing of *TRIM5* and *TRIM22*

Briefly, amplification primers were designed to flank exons 2 - 4 and 5 - 8 using the *TRIM5* and *TRIM22* reference sequences NG_029122 and NC_000011 respectively from the NCBI

database. Touchdown amplification protocols were developed for PCR. Separate inner sequencing primers were designed to sequence individual exons within the larger amplicons in a primer-walking fashion. PCR products were purified using an in-house EXOSAP protocol. Sanger sequencing was done using the Big Dye Terminator system according the manufacturer's instructions. Post-sequencing purification was done using ethanol precipitation and purified sequenced products were resuspended in 10uL of Hi-Di Formamide. Sequencing was performed on 3730XL ABI sequence analyser at the Medical Research Council Weatherall Institute for Molecular Medicine (Oxford, United Kingdom).

3.2.7 Quantification of plasma viral load for CMV and HIV

Plasma levels of CMV viral DNA and HIV-1 viral RNA were quantified using commercially available kits - RealStar CMV PCR kit 1.0 and RealStar HIV RT-PCR kit 1.0 respectively (Altona diagnostics). CMV viral loads were measured for a previous publication and that data, were available, was used in our analysis (8).

3.2.8 Sequence analysis

Sequence electropherograms for *TRIM5* and *TRIM22* were analysed in Geneious 2019.2 (27). Forward and reverse reads were collected and assembled into contiguous patient sequences. These assembled contigs were aligned to the relevant reference sequence for *TRIM5* and *TRIM22*. Sequences with HQ scores of less than 80% were not included in further analysis. This score indicates that 80% of bases are high quality (likelihood of a false positive base reading of 1:10 000). Single nucleotide polymorphisms were identified at a 35% heterozygosity threshold and identified in relation to the dbSNP database administered by the NCBI. Linkage statistics were calculated in Haploview and LD maps

generated (28). Measurements of heterozygosity and testing of Hardy-Weinberg Equilibrium were also done in Haploview.

3.2.9 Statistical analysis

All further statistical analysis was done in R, version 3.4.0 (29). Pairwise comparisons of continuous numeric variables were done using the Students' T test and Mann Whitney-U (MW) test for parametric and non-parametric distributions respectively. Multiple group comparisons were done using a one-way ANOVA and Kruskall-Wallis test for parametric and non-parametric distributions respectively. Univariate odds ratios to test associations between categorical variables was calculated using Fishers' exact test. Testing of differences in genotype frequencies between groups was done using chi-squared tests. Linear and logistic regression modelling was done using R. Where appropriate, corrections for multiple comparisons was done using the Holm-Bonferroni method. Pirate plots were prepared using the YARRR package and all figures were prepared using the ggpubr package (30,31).

3.3 Results

Of the 275 ZENITH patients most patients were vertically infected, with 3% being horizontally infected (1 parenteral and 6 through probable sexual transmission). The patient sample was balanced for sex, median age was 11.37 years (**table 3.1**). Prior to starting ART 15% patients had an undetectable HIV-1 pVL and 25% of patients had a $CD4^+$ below 200 cells/μL. For 28% (76/275) individuals there were multiple ART-naive timepoints with $CD4^+$ and HIV pVL data available for analysis. These data points were collected into per-patient regression models and midpoint $CD4^+$ and HIV pVL values extracted for each. This provided a pre-ART average estimate for these measurements which were used in cross-sectional analyses. CMV viraemia data was available for 67% (185/275) of patients, with 77 testing positive and 108 negative.

Table 3.1: Summary statistics for the ZENITH cohort patients

	Number (%)	**Missing (%)**	**Mean/ Median**	**IQR**
HIV-1 infected patients	275 (100)			
Male	134 (49.08)	2 (0.73)		
Female	139 (50.92)	2 (0.73)		
Stunting [1]	44 (16.12)	2 (0.73)		
AIDS at enrolment	69 (25.09)	0 (0.00)		
HIV pVL undetectable	35 (14.71)	37 (13.45)		
LTNP [2]	49 (18.21)	6 (2.55)		
CMV viraemia positive [3]	77 (41.62)	90 (32.73)		
Age (years) [4]			11.37	8.93 – 13.47
$CD4^+$ count (cells/μL)			388	215 - 607
HIV viral load log10(IU/mL) [5]			4.58	4.13 – 4.95
CMV viral load log10(IU/mL) [5]			3.07	2.82 – 3.50

[1] Stunting was defined as height-for-age more than two standard deviations below the WHO Child Growth Standards median. [2] LTNP = long term non-progressors. [3] Positive qPCR of CMV DNA in plasma. [4] Age at the time a sample was taken. [5] These values relate only to the samples with detectable viraemia. Samples with undetectable viraemia were treated as categorical outcomes.

3.3.1 Summary of genotype data for *TRIM5* and *TRIM22*

We identified five SNPs in *TRIM5* all of which had been previously reported. The sequence coding for the V1 loop of the PRY-SPRY domain was conserved in all patients. In *TRIM22* we identified 12 SNPs, all previously reported (**table 3.2**). Linkage disequilibrium (LD) values between SNPs and tests for deviation from Hardy Weinberg Equilibrium (HWE) were calculated (**figure 3.1**). *TRIM22* SNPs rs1063303 and rs2291842 were excluded from LD analysis as they were not in Hardy Weinberg Equilibrium. The observed heterozygosity scores in these SNPs was lower than predicted - rs1063303 (rs1063303 observed heterozygosity: 0.31 and predicted heterozygosity: 0.49, $P = 1.49 \times 10^{-5}$). There was no evidence of LD between SNPs in *TRIM5* and *TRIM22*. rs10838543 and rs1063303 in *TRIM22* had strong evidence of LD (D' = 0.91, $r^2 = 0.72$, LOD = 39.43).

3.3.2 *TRIM22* SNP associations with clinical endpoints in HIV-1 infection

We included *TRIM22* SNPs rs7935564, rs1063303 and rs61735273 in further association studies with HIV-1 clinical outcomes. The SNPs which we identified in *TRIM5* either did not meet our inclusion criteria or did not occur in TRIM5 domains where substitutions are likely to affect its antiviral activity (32). The primary endpoints which we included for SNP association studies were $CD4^+$ count, HIV pVL and CMV pVL (**figure 3.2**).

$CD4^+$ counts did not vary in the patients based on sex or presence of stunting (**figure 3.2**). Patients who were stunted had higher mean HIV pVL than those who were within normal height-for-age ranges (T test: 4.9 vs 4.5 log10 HIV pVL; 95% CI: -0.64 to -0.15; $P = 0.002$). CMV viraemia was associated with lower $CD4^+$ counts (MW: difference in median $CD4^+$

counts = 126 cells/μL, 95% CI: 33 - 225; $P = 0.007$) and higher HIV pVL (T test: 4.7 vs 4.5 log10 HIV pVL; 95% CI: -0.41 to -0.03; $P = 0.02$, **figure 3.2**). $CD4^+$ counts, HIV pVL and CMV pVL did not vary by *TRIM22* genotype (**figure 3.3 –** CMV pVL comparisons not shown).

Table 3.2: Summary of genotype data for *TRIM5* and *TRIM22*

***TRIM5* exon**	**dbSNP ID** [1]	**CADD score** [2]	**AA change** [3]	**Genotype frequencies in the ZENITH cohort**		
2	rs10838525	12	R136Q	G/G: 0.87	G/A: 0.12	A/A: 0.01
3	rs3740995			G/G: 0.54	G/A: 0.37	A/A: 0.09
4	rs58187248			A/A: 0.78	A/G: 0.22	A/A: 0
5	rs11038628	4	G249D	G/G: 0.52	G/A: 0.41	A/A: 0.07
8	rs7104422	21	P479L	C/C: 0.87	C/T: 0.13	T/T: 0
***TRIM22* exon**	**dbSNP ID**	**CADD score**	**AA change**	**Genotype frequencies in the ZENITH cohort**		
2	rs10838543			T/T: 0.32	T/C: 0.39	C/C: 0.29
3	rs200924168	2	R150T	G/G: 0.99	G/C: 0.01	C/C: 0
3	**rs7935564**	14	D155N	G/G: 0.29	G/A: 0.39	A/A: 0.32
4	rs78484876			T/T: 0.88	T/C: 0.09	C/C: 0.03
4	rs2291842			T/T: 0.67	T/C: 0.20	C/C: 0.13
4	**rs1063303**	13	R238T	G/G: 0.41	G/C: 0.31	C/C: 0.28
4	**rs61735273**	22	S244L	C/C: 0.72	C/T: 0.23	T/T: 0.05
7	rs73404240	4	T294K	C/C: 0.86	C/A: 0.14	A/A: 0
8	rs140700337			G/G: 0.98	G/A: 0.02	A/A: 0
8	rs200243523	21	V359I	G/G: 0.92	G/A: 0.06	A/A: 0.01
8	rs150095329	5	T415I	C/C: 0.97	C/T: 0.03	T/T: 0
8	rs138529937	17	C472S	T/T: 0.98	T/A: 0.02	A/A: 0

These genotype frequencies were calculated for the 150 patients who had *TRIM5* and *TRIM22* sequenced. [1] rs number refers to Reference SNP Cluster ID and is used as an accession number on the dbSNP database which is administered by the NCBI. SNPs used for association studies are shown in **bold text.** [2] Combined Annotation Dependent Depletion (CADD) scores are a continuous predictor of the likelihood that a SNP is deleterious. [3] AA: Amino acid.

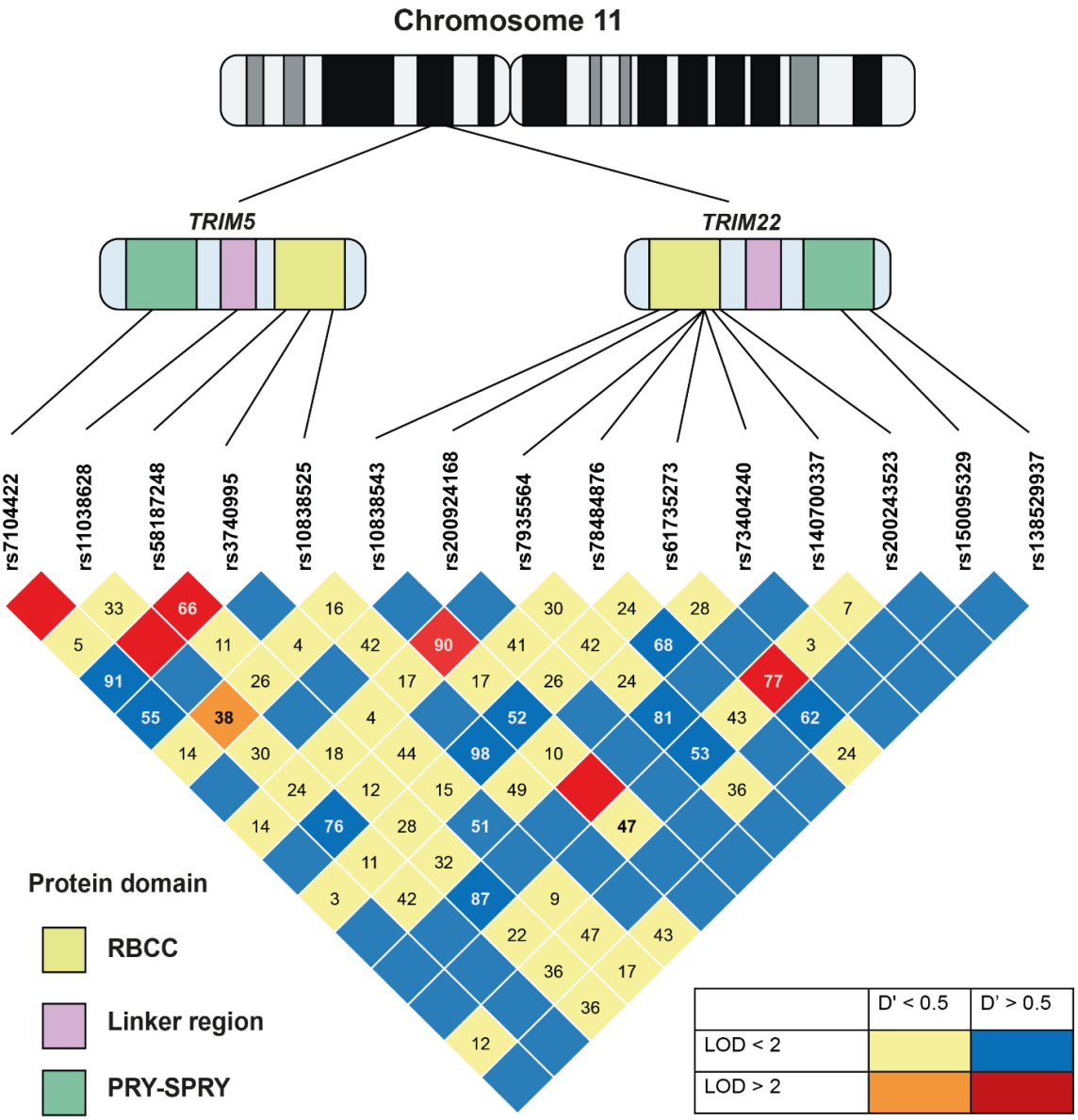


Figure 3.1: **Linkage map of variants in *TRIM5* and *TRIM22***. *TRIM5* and *TRIM22* are in band 12 of the short arm of chromosome 11. Protein coding domains key are shown in the bottom left corner. In the linkage map, numbers in blocks denote D′ x 100 with D' values = 100 not shown. Representations of *TRIM5* and *TRIM22* are shown with their respective SNPs mapped to their coding regions. Bottom right key for interpreting LD scores using D': D' scores range from 0 (no linkage) to 1 (complete linkage), Logarithm (base 10) of odds (LOD) scores indicates the likelihood that a pair of SNPs are in linkage, LOD of 3 = 1000:1 odds that observed linkage occurred by chance. Red blocks indicate strong LD with high likelihood.

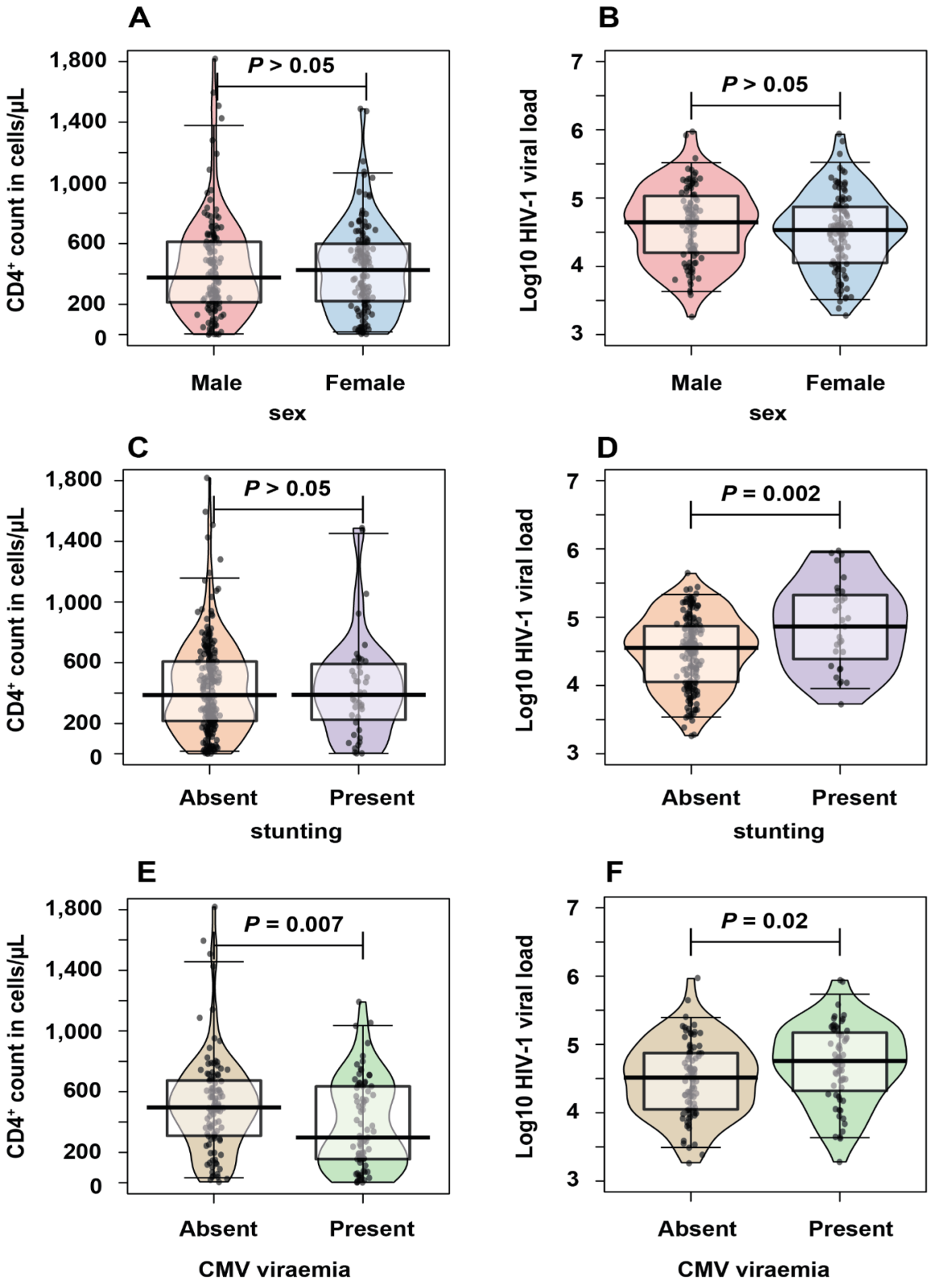


Figure 3.2: **Comparison of CD4⁺ count and HIV viral load by clinical factors.** Pairwise comparisons of CD4⁺ count and HIV pVL by sex (**panels A and B**), stunting (**panels C and D**) and CMV viraemia (**panels E and F**). Y-axes show the continuous variables and x-axes the groupings for pairwise comparisons. *P* values are shown for pairwise MW tests for CD4⁺ and T tests for HIV pVL.

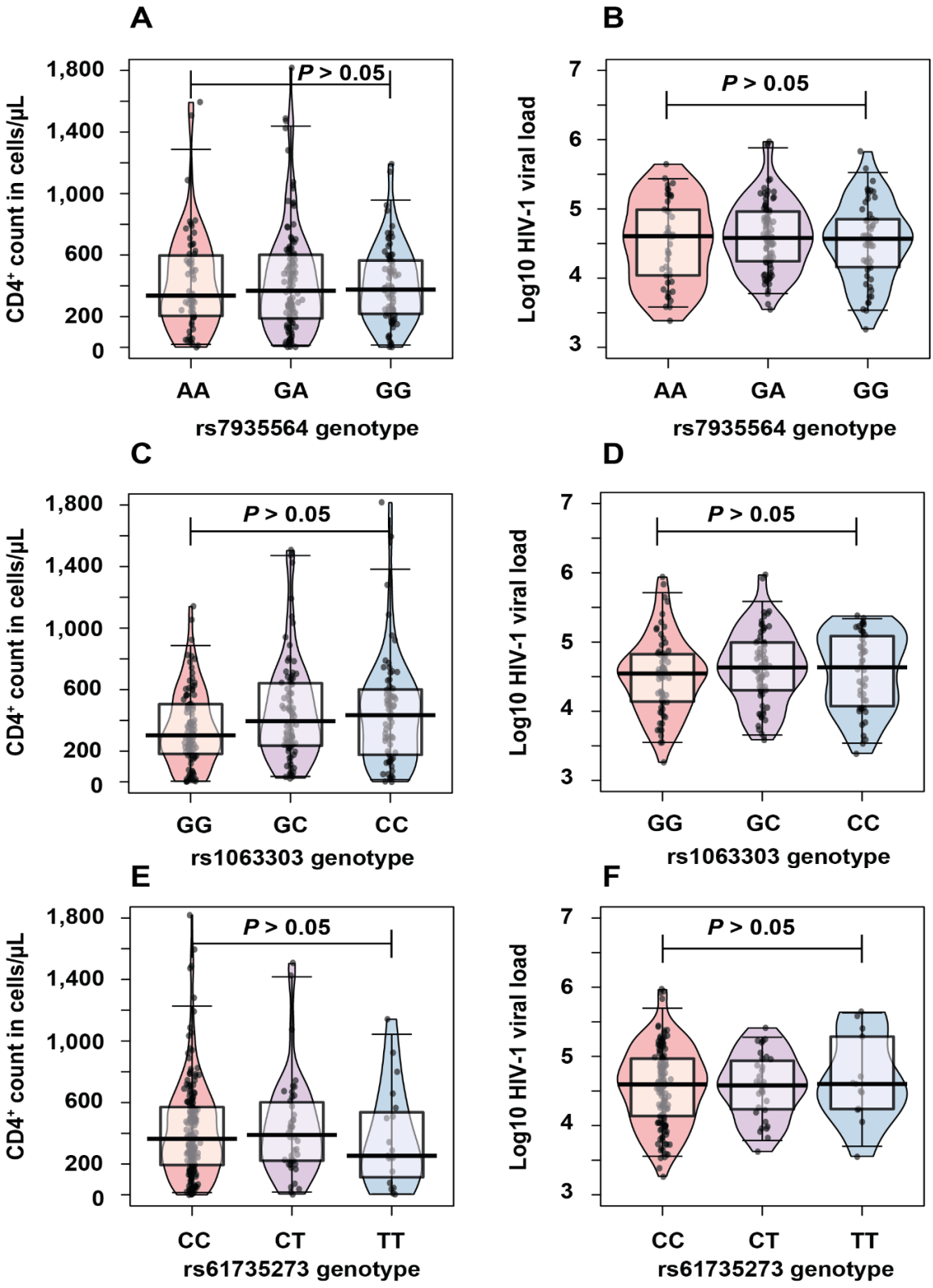


Figure 3.3: **Comparison of CD4⁺ count and HIV viral load by *TRIM22* genotype.** CD4⁺ count and HIV pVL did not vary significantly by genotype in pairwise comparisons or across genotypes at rs7935564 – **panel A** and **B**, rs1063303 – **panel C** and **D** and rs61735273 – **panel E** and **F**. *P* values are shown for pairwise MW tests for CD4⁺ and T tests for HIV pVL.

There was no significant correlation between CMV pVL and HIV pVL ($\rho = 0.23$, $P = 0.06$), $CD4^+$ count ($\rho = -0.07$, $P = 0.55$) or age ($\rho = 0.16$, $P = 0.16$). Correlation between HIV pVL and age was not significant (spearman correlation: $\rho = 0.01$, $P = 0.85$), and this did not differ between normal progressors and LTNP (**figure 3.4**). HIV pVL did not correlate with $CD4^+$ count (Pearson correlation: $r^2 = 0.04$, $P = 0.6$) and this did not change when the sample was stratified according to progressor status. There was no significant difference in HIV pVL between patients who were classified as LTNP (mean log10 HIV pVL = 4.4) and normal progressors (mean log10 HIV pVL = 4.6) (T test: 95% CI: -0.05, 0.37; $P = 0.13$, **figure 3.4**).

There was a negative correlation between $CD4^+$ count and age of patient (Spearman correlation: $\rho = -0.32$, $P = <0.001$). This reflects a combination of normal T cell kinetics in children as well as probable HIV disease progression where age gives a reasonable estimate of time of infection in this cohort (33).

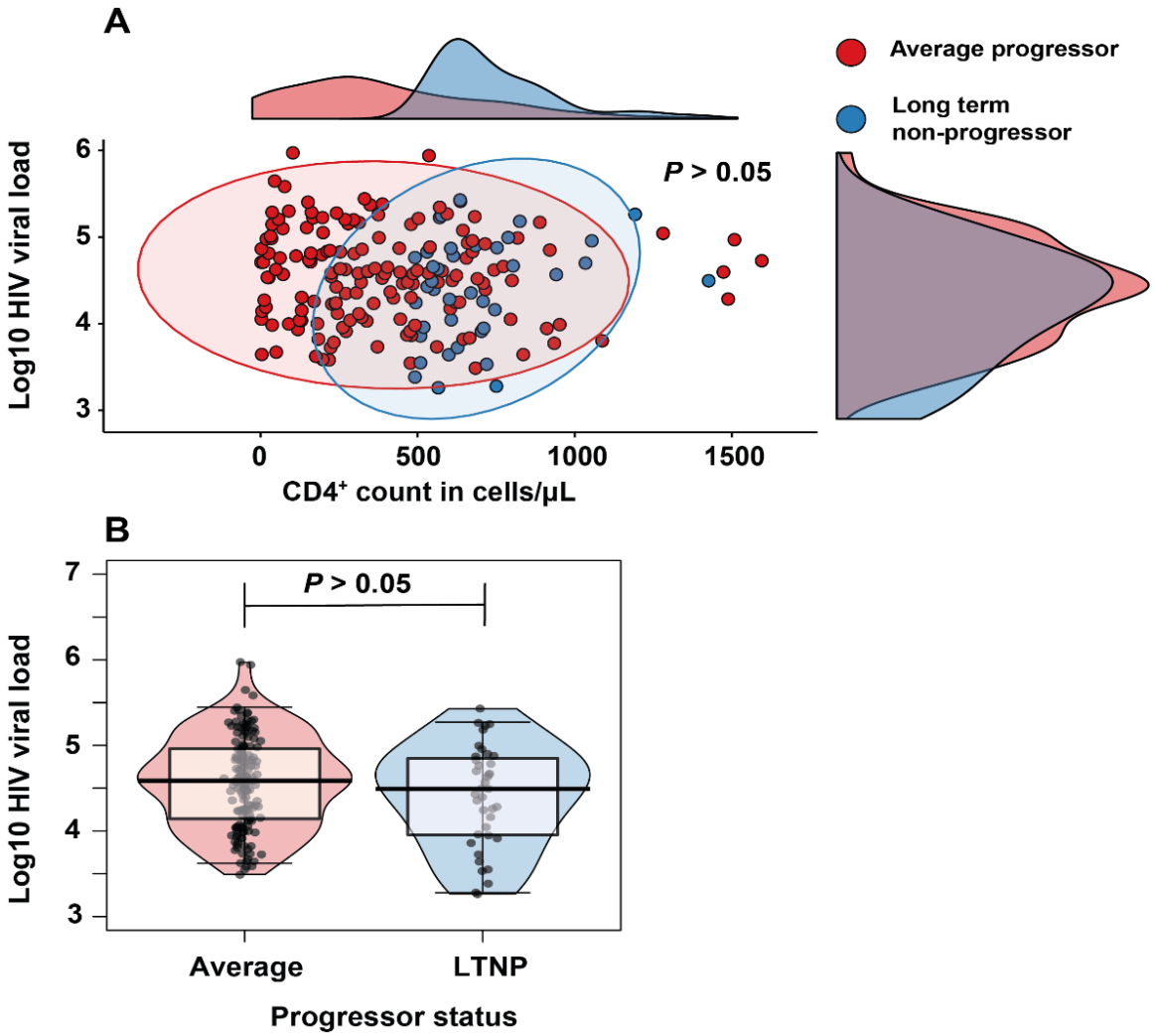


Figure 3.4: **HIV viral load and CD4⁺ count by progressor status**. **A)** Log10 HIV pVL correlated with CD4⁺ count, density plots of the y-axis show the overlapping distributions of pVL when coloured by progressor status. Density plot on the x-axis shows the distribution of CD4⁺ by progressor status. *P* value shown is for the result of a Pearson correlation test. **B)** Pirate plots of log10 HIV pVL by progressor group. *P* value shown is for the result of a T test.

3.3.3 Association of age adjusted CD4⁺ count by *TRIM22* genotype

We tested a series of multivariate models to determine if *TRIM22* genotype contributed to variation in CD4⁺ counts when controlled for host factors and covariates (**figure 3.5**).

To meet least squares regression model heteroscedasticity assumptions criteria CD4⁺ counts were transformed using square roots of the values. Host factors and covariates were added to the first model and *P* values calculated for their coefficients. Factors and covariates

were removed from the analysis in subsequent models if the *P* value for their coefficient was > 0.1 (backward step approach). Following this individual SNPs were added to the model to i) determine if the genotype at this SNP was associated with significant differences in $CD4^+$ count and ii) whether the addition of this SNP improved model fit as measured by adjusted r^2. Using this approach model three which included rs1063303 genotype as an additional factor to the model fitted the data best ($r^2 = 0.13$, **table S3.1**). Results of model fitting are shown in (**table S3.2**). Age was associated with decreasing $CD4^+$ counts (-0.71 square root $CD4^+$, SE = 0.24, $P = 0.03$). Detectable CMV viraemia was associated with lower $CD4^+$ counts (-3.92 square root $CD4^+$ if CMV DNA detected, SE = 1.21, $P = 0.002$). rs1063303(GG) genotype was associated with lower $CD4^+$ counts (-3.81 square root $CD4^+$ for GG genotype vs CC genotype, SE 1.21, $P = 0.01$). This model did not violate the assumptions of ordinary least squares regression. However, rs1063303 genotype did not affect $CD4^+$ counts in an analysis which did not control for CMV viraemia (GG genotype vs. GC/CC: -1.91 square root $CD4^+$, SE = 1.21, $P = 0.12$). A post-hoc power analysis showed that this model had a power of 0.63 to detect a small effect size (r^2 increase of 0.03).

We next performed a sensitivity analysis to determine whether our 76 midpoint derived $CD4^+$ count and HIV pVL data points biased the model's results. In addition, the 7 horizontally infected patients were excluded in this sensitivity analysis. Therefore, we included 196 individuals in the sensitivity analysis. Only rs1063303(GG) genotype were associated with significantly lower $CD4^+$ counts in the sensitivity analysis, though the effect sizes of age and detectable CMV DNA were quite similar with the larger sample size model (**table S3.3**).

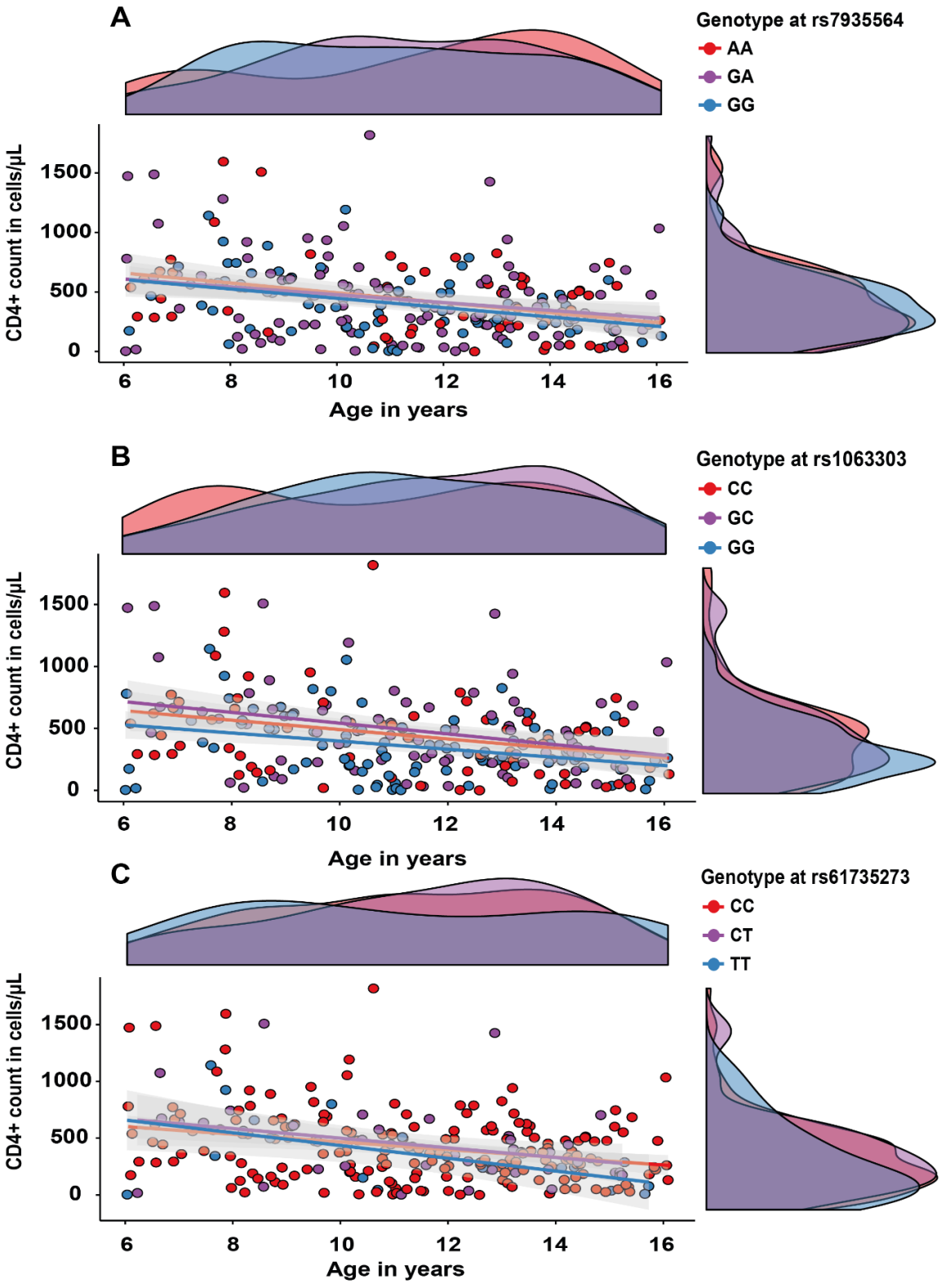


Figure 3.5: **Multivariate linear regression of CD4$^+$ count by age and *TRIM22* genotype**. CD4$^+$ decline was plotted against age of participants and further divided according to genotype at *TRIM22* SNPs rs7935564, rs1063303 and rs61735273 in **panels A**, **B** and **C** respectively. Density plots on the x and y-axes show the distribution of age and CD4$^+$ respectively.

3.3.4 Association of *TRIM22* genotype with AIDS at enrolment

To determine whether *TRIM22* genotype was associated with categorical outcomes univariate odds ratios were calculated for the following outcomes i) Stunting ii) Chronic diarrhoea iii) Having an ART-naïve undetectable HIV-1 pVL iv) Presence of detectable CMV viraemia. rs61735273(TT) genotype was associated with increased odds of having chronic diarrhoea, however this result was not significant after correction for multiple comparisons. This outcome was specifically tested as this SNP has been associated with increased risk for developing early-onset inflammatory bowel syndrome and this disease may manifest as chronic diarrhoea (24). *TRIM22* genotype did not associate with any of the above-mentioned categorical outcomes.

Multivariate logistic regression models were used to determine whether *TRIM22* genotype associated with clinical stage, specifically AIDS at time of inclusion. Clinical factors and covariates were added and if the *P* value on the coefficient for that variable was >0.1 it was excluded from subsequent models in a backward step approach. The significant host variables were then included in subsequent models where *TRIM22* SNPs were added individually. Model fit was compared using AIC (**table S3.4**). Model two fit the data best with an AIC of 191.84. CMV viraemia and rs1063303(GG) genotype were each associated with increased odds of progression to AIDS prior to ART initiation (**table 3.3**). However, the result for rs1063303(GG) genotype was of borderline significance at $P = 0.048$ (rounded to 0.05).

Table 3.3: Results of regression modelling to determine associations with AIDS

	Univariate			**Multivariate**		
	OR	*P* value	SE	OR	*P* value	SE
HIV viral load [1]	1.48	0.19	0.30			
CMV viraemia	2.77	<0.001	0.35	2.83	0.01	0.37
Age [2]	1.16	0.01	0.05	1.14	0.08	0.07
rs1063303(CC)	1			1		
rs1063303(GG)	0.99	0.97	0.35	2.45	0.05	0.45
rs7935564(AA)	1			1		
rs7935564(GG)	0.92	0.83	0.40	1.40	0.50	0.50
rs61735273(CC)	1			1		
rs61735273(TT)	1.32	0.60	0.52	1.65	0.40	0.62

Uni- and multivariate OR were calculated with homozygous genotypes and clinical outcomes. SE = standard error. [1] HIV viral load was treated as a continuous covariate in these models. [2] Age at the time of the sample.

3.3.5 Genotype frequency comparison

Genotype frequencies for *TRIM22* exons three and four in the HIV-1 positive patients and negative controls were calculated and compared with global and African genotype estimate frequencies available on the 1000 Genomes database (34). Genotype frequencies at synonymous substitutions rs78484876 and rs2291842 in exon four appeared to be quite similar between global and African estimates (**figure 3.6**). This was also true for nonsynonymous substitutions rs7935564, rs1063303 and rs61735273. Genotype frequencies for rs7935564 did not differ significantly between HIV positive cases and HIV negative controls ($X^2 = 3.39$, $P = 0.18$). Genotype frequencies differed significantly at rs1063303 between HIV positive cases and HIV negative controls. Specifically, the GG genotype occurred at a higher frequency in HIV negative controls than HIV positive cases (chi-squared test: 54% versus 39%, $X^2 = 6.74$, $P = 0.03$).

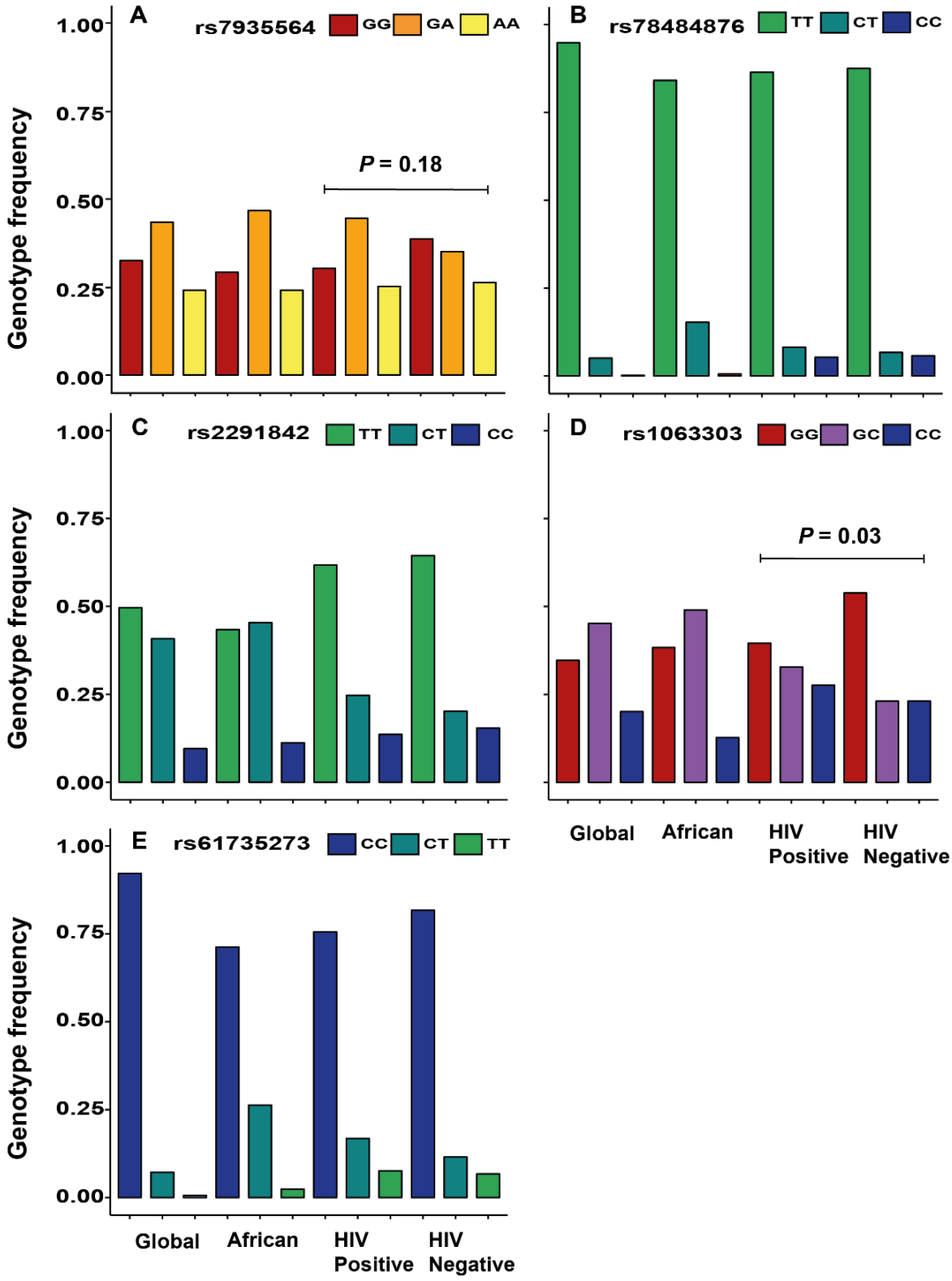


Figure 3.6: Comparison of genotype frequencies in *TRIM22*. Genotype frequencies (y-axis) were grouped by global, African, HIV positive ZENITH patients and HIV negative INHALE controls (x-axis). Genotype colours and SNP numbers are shown above each bar plot group. *P* values for X^2 tests are shown for rs7935564 and rs1063303.

3.4 Discussion

HIV-1 pVL did not contribute significantly to differences in $CD4^+$ count. The observed median pVL for the HIV-1 infected patients correlates closely with previous estimates for paediatric slow progressors (35). CMV co-infection in HIV-1 infected infants has previously been reported to lead to faster disease progression and this is supported by our finding (36).

We did not identify any novel variants by sequencing the coding regions of *TRIM5* and *TRIM22* in 150 of the ZENITH cohort patients. There was no evidence of LD between *TRIM5* and *TRIM22* which is in keeping with the hypothesis that selection pressure in these genes has been driven at different times in primate evolution, possibly by different pathogens (37). TRIM5α restriction is specific to the viral capsid of the infecting retrovirus. TRIM22 has not been shown to bind SIV or HIV viral proteins despite it being a paralogue of *TRIM5* and encoding for a PRY-SPRY domain. However, TRIM22 does bind Hepatitis C viral protein NS5A, and this interaction is mediated by the PRY-SPRY domain (38). This further supports the hypothesis that related *TRIM* paralogues may have evolved in mammals to defend hosts against diverse pathogens.

rs7935564 genotype did not associate with any significant outcomes in our study. Patients with the rs1063303(GG) genotype had significantly lower $CD4^+$ counts than heterozygotes and (CC) homozygotes at this position when controlling for age and CMV viraemia. The rs1063303(GG) genotype was also associated with increased odds of progression to AIDS before enrolment. Both results had weak statistical support and may represent false positive findings.

HIV-1 pVL did not differ by *TRIM22* genotype despite previous publications indicating that TRIM22 expression correlates negatively with HIV-1 pVL (20). A possible explanation for this finding is that vertically infected children who were rapid progressors, who are likely to have much higher pVL, were not captured by the sampling strategy used to recruit patients to the ZENITH cohort. This cohort is likely to consist of slower disease progressors who are unlikely to have large differences in HIV-1 pVL (5). In support of this assertion we found that genotype frequencies at this SNP differed significantly with between HIV-1 infected children and matched HIV-1 negative controls. The rs1063303(GG) genotype occurred at a significantly higher frequency in controls than the HIV-1 infected patients. It is therefore reasonable to conclude that this genotype may have been more frequent in rapid progressors who died, and had higher pVL, than slow progressors.

Transfection of plasmids bearing the rs1063303(CC) genotype into immortalised cell lines results in markedly increased expression of TRIM22 protein levels; however, in these cell models a paradoxical reduction of HIV-1 antiviral activity was observed with increasing *TRIM22* expression (39). These findings are not incompatible with our results – as noted earlier, increased expression of *TRIM22* in PBMC correlated positively with increasing $CD4^+$ counts and negatively with HIV-1 pVL, and that study recruited patients of Southern African ancestry (40). Therefore, if we accept that the rs1063303(CC) genotype results in increased *TRIM22* expression in human T cells, it would explain why the GG genotype is associated with worse outcomes in perinatal HIV-1 infection. The paradoxical *in vitro* results reported by Kelly *et al* might reflect the limitations of translating results of infection assays on immortalised cell lines to *in vivo* data (39).

TRIM22 suppresses Hepatitis B (HBV) replication, and mutations in the RING domain can disrupt this anti-HBV activity (41). The rs10838543 CC genotype has been associated with chronic HBV infection in Chinese adults (42). Prevalence of active HBV infection in HIV-1 infected children and adolescents in Sub-Saharan Africa ranges from 7 – 10% (43–45). We were unable to control for HBV coinfection in this study, and this may be a confounder in our results.

Despite rs61735273 being previously categorised as a pathological loss of function mutation there were no significant associations of this substitution with any HIV-1 disease outcomes in our study (22,24). Genotype frequencies did not differ from HIV-1 negative controls. In HIV-1 infection it appears that expression of *TRIM22* is central to its antiviral activity, and this might not necessarily depend on whether NOD-2 signalling is activated by TRIM22. Limitations of this study include that we were unable to assess a balanced spectrum of HIV-1 infection outcomes as our sample predominantly consists of slower progressors. This was mitigated somewhat by the inclusion of matched HIV-1 negative controls. A further limitation is that this control group sample size was not matched to our patients, however genotype frequencies did not differ significantly from the initial smaller sample of HIV-1 infected patients versus the larger sample where we only sequenced *TRIM22* exon 3 and 4.

3.5 Section summary

1. rs1063303(GG) homozygotes had lower $CD4^+$ counts than heterozygotes and (CC) homozygotes in a multivariate analysis controlling for age and CMV viraemia.
2. rs1063303(GG) was associated with increased odds of having a $CD4^+$ less than 200 cells/μL at enrolment into the trial.
3. The rs1063303(GG) genotype was less common HIV-1 infected children and adolescents when compared to HIV-1 negative, age and ethnically matched controls.

3.6 References for chapter 3

1. AIDSinfo | UNAIDS [Internet]. [cited 2019 Feb 1]. Available from: http://aidsinfo.unaids.org/

2. Goulder PJ, Lewin SR, Leitman EM. Paediatric HIV infection: the potential for cure. Nat Rev Immunol. 2016 Mar 14;16(4):259–71.

3. Newell M-L, Coovadia H, Cortina-Borja M, Rollins N, Gaillard P, Dabis F, et al. Mortality of infected and uninfected infants born to HIV-infected mothers in Africa: a pooled analysis. Lancet. 2004 Oct;364(9441):1236–43.

4. Adland E, Paioni P, Thobakgale C, Laker L, Mori L, Muenchhoff M, et al. Discordant Impact of HLA on Viral Replicative Capacity and Disease Progression in Pediatric and Adult HIV Infection. Douek DC, editor. PLOS Pathog. 2015 Jun 15;11(6):e1004954.

5. Roider J, Ngoepe A, Muenchhoff M, Adland E, Groll A, Ndung'u T, et al. Slow progression in HIV infected children is associated with increased regulatory T cell activity and enhanced T-cell homeostatic signaling. Front Immunol. 2019;10:213.

6. Pathirana J, Groome M, Dorfman J, Kwatra G, Boppana S, Cutland C, et al. Prevalence of Congenital Cytomegalovirus Infection and Associated Risk of In Utero Human Immunodeficiency Virus (HIV) Acquisition in a High-HIV Prevalence Setting, South Africa. Clin Infect Dis. 2019 Jan 7;

7. Garcia-Knight MA, Nduati E, Hassan AS, Nkumama I, Etyang TJ, Hajj NJ, et al. Cytomegalovirus viraemia is associated with poor growth and T-cell activation with an increased burden in HIV-exposed uninfected infants. AIDS. 2017 Aug 24;31(13):1809–18.

8. Yindom L-M, Simms V, Majonga ED, McHugh G, Dauya E, Bandason T, et al. Unexpectedly High Prevalence of Cytomegalovirus DNAemia in Older Children and Adolescents With Perinatally Acquired Human Immunodeficiency Virus Infection. Clin Infect Dis. 2019 Aug 1;69(4):580–7.

9. Naranbhai V, Carrington M. Host genetic variation and HIV disease: from mapping to mechanism. Immunogenetics. 2017 Aug 10;69(8–9):489–98.

10. Tsai M-H, Muenchhoff M, Adland E, Carlqvist A, Roider J, Cole DK, et al. Paediatric non-progression following grandmother-to-child HIV transmission. Retrovirology. 2016 Dec 8;13(1):65.

11. Naidoo VL, Mann JK, Noble C, Adland E, Carlson JM, Thomas J, et al. Mother-to-Child HIV Transmission Bottleneck Selects for Consensus Virus with Lower Gag-Protease-Driven Replication Capacity. Silvestri G, editor. J Virol. 2017 Sep 1;91(17).

12. Doyle T, Goujon C, Malim MH. HIV-1 and interferons: who's interfering with whom? Nat Rev Micro. 2015 Apr 27;13(7):403–13.

13. Lahaye X, Satoh T, Gentili M, Cerboni S, Conrad C, Hurbain I, et al. The Capsids of HIV-1 and HIV-2 Determine Immune Detection of the Viral cDNA by the Innate Sensor cGAS in Dendritic Cells. Immunity. 2013 Dec 12;39(6):1132–42.

14. Kumar S, Morrison JH, Dingli D, Poeschla E. HIV-1 Activation of Innate Immunity Depends Strongly on the Intracellular Level of TREX1 and Sensing of Incomplete Reverse Transcription Products. J Virol. 2018 Aug 15;92(16):e00001-18.

15. Veillette M, Bichel K, Pawlica P, Freund SM V, Plourde MB, Pham QT, et al. The V86M mutation in HIV-1 capsid confers resistance to TRIM5α by abrogation of cyclophilin a-dependent restriction and enhancement of viral nuclear import. Retrovirology. 2013 Feb 28;10(1):25.

16. Pertel T, Hausmann S, Morger D, Zuger S, Guerra J, Lascano J, et al. TRIM5 is an innate immune sensor for the retrovirus capsid lattice. Nature. 2011;472(7343):361–5.

17. Mamede JI, Sitbon M, Battini J-L, Courgnaud V. Heterogeneous susceptibility of circulating SIV isolate capsids to HIV-interacting factors. Retrovirology. 2013 Jul 24;10:77.

18. Merindol N, El-Far M, Sylla M, Masroori N, Dufour C, Li J, et al. HIV-1 capsids from B27/B57+ elite controllers escape Mx2 but are targeted by TRIM5α, leading to the induction of an antiviral state. Emerman M, editor. PLOS Pathog. 2018 Nov 12;14(11):e1007398.

19. Barr SD, Smiley JR, Bushman FD. The interferon response inhibits HIV particle production by induction of TRIM22. Hope TJ, editor. PLoS Pathog. 2008 Feb 29;4(2):e1000007.

20. Turrini F, Marelli S, Kajaste-Rudnitski A, Lusic M, Van Lint C, Das AT, et al. HIV-1 transcriptional silencing caused by TRIM22 inhibition of Sp1 binding to the viral promoter. Retrovirology. 2015 Dec 18;12(1):104.

21. Singh R, Gaiha G, Werner L, McKim K, Mlisana K, Luban J, et al. Association of TRIM22 with the Type 1 Interferon Response and Viral Control during Primary HIV-1 Infection. J Virol. 2011 Jan 1;85(1):208–16.

22. Kelly JN, Barr SD. In Silico Analysis of Functional Single Nucleotide Polymorphisms in the Human TRIM22 Gene. Uversky VN, editor. PLoS One. 2014 Jul 1;9(7):e101436.

23. Ghezzi S, Galli L, Kajaste-Rudnitski A, Turrini F, Marelli S, Toniolo D, et al. Identification of TRIM22 single nucleotide polymorphisms associated with loss of inhibition of HIV-1 transcription and advanced HIV-1 disease. AIDS. 2013 Sep 24;27(15):2335–44.

24. Li Q, Lee CH, Peters LA, Mastropaolo LA, Thoeni C, Elkadri A, et al. Variants in TRIM22 That Affect NOD2 Signaling Are Associated With Very-Early-Onset Inflammatory Bowel Disease. Gastroenterology. 2016 May;150(5):1196–207.

25. Ferrand RA, Simms V, Dauya E, Bandason T, Mchugh G, Mujuru H, et al. The effect of community-based support for caregivers on the risk of virological failure in children and adolescents with HIV in Harare, Zimbabwe (ZENITH): an open-label, randomised controlled trial. Lancet Child Adolesc Heal. 2017 Nov;4642(17):1–9.

26. World Health Organization. WHO case definitions of HIV for surveillance and revised clinical staging and immunological classification of HIV-related disease in adults and children. World Health. 2007.

27. Kearse M, Moir R, Wilson A, Stones-Havas S, Cheung M, Sturrock S, et al. Geneious Basic: An integrated and extendable desktop software platform for the organization and analysis of sequence data. Bioinformatics. 2012;

28. Barrett JC, Fry B, Maller J, Daly MJ. Haploview: analysis and visualization of LD and haplotype maps. Bioinformatics. 2005 Jan 15;21(2):263–5.

29. RStudio team, RStudio. RStudio: Integrated development environment for R. J Wildl Manage. 2015;

30. Kassambara A. ggpubr: "ggplot2" Based Publication Ready Plots. R Package. 2018.

31. Kampstra P. Beanplot: A Boxplot Alternative for Visual Comparison of Distributions. J Stat Softw. 2008 Nov 10;28(Code Snippet 1):1–9.

32. Goldschmidt V, Bleiber G, May M, Martinez R, Ortiz M, Telenti A, et al. Role of common human TRIM5alpha variants in HIV-1 disease progression. Retrovirology. 2006 Aug 22;3:54.

33. Shearer WT, Rosenblatt HM, Gelman RS, Oyomopito R, Plaeger S, Stiehm ER, et al. Lymphocyte subsets in healthy children from birth through 18 years of age: The Pediatric AIDS Clinical Trials Group P1009 study. J Allergy Clin Immunol. 2003;112(5):973–80.

34. Gibbs RA, Boerwinkle E, Doddapaneni H, Han Y, Korchina V, Kovar C, et al. A global reference for human genetic variation. Nature. 2015 Oct 1;526(7571):68–74.

35. Roider JM, Muenchhoff M, Goulder PJR. Immune activation and paediatric HIV-1 disease outcome. Curr Opin HIV AIDS. 2016 Mar;11(2):146–55.

36. Ellington SR, Clarke KEN, Kourtis AP. Cytomegalovirus Infection in Human Immunodeficiency Virus (HIV)-Exposed and HIV-Infected Infants: A Systematic Review. J Infect Dis. 2016 Mar 15;213(6):891–900.

37. Sawyer SL, Emerman M, Malik HS, Finlay M, Barton G. Discordant Evolution of the Adjacent Antiretroviral Genes TRIM22 and TRIM5 in Mammals. PLoS Pathog. 2007;3(12):e197.

38. Yang C, Zhao X, Sun D, Yang L, Chong C, Pan Y, et al. Interferon alpha (IFNα)-induced TRIM22 interrupts HCV replication by ubiquitinating NS5A. Cell Mol Immunol. 2016 Jan;13(1):94–102.

39. Kelly JN, Woods MW, Xhiku S, Barr SD. Ancient and Recent Adaptive Evolution in the Antiviral *TRIM22* Gene: Identification of a Single-Nucleotide Polymorphism That Impacts TRIM22 Function. Hum Mutat. 2014 Sep;35(9):1072–81.

40. Singh R, Gaiha G, Werner L, McKim K, Mlisana K, Luban J, et al. Association of TRIM22 with the Type 1 Interferon Response and Viral Control during Primary HIV-1 Infection. J Virol. 2011 Jan 1;85(1):208–16.

41. Gao B, Duan Z, Xu W, Xiong S. Tripartite motif-containing 22 inhibits the activity of hepatitis B virus core promoter, which is dependent on nuclear-located RING domain. Hepatology. 2009 Aug;50(2):424–33.

42. Zhao N, Wang X-L, Gu Q-H, Huang F, Zheng W, Li Z-W. Tripartite Motif-Containing 22 Gene -364T/C Polymorphism Associated With Hepatitis B Virus Infection in Chinese Han Population. Hepat Mon. 2014 Jan;14(1):e12110.

43. Mutwa PR, Boer KR, Rusine JB, Muganga N, Tuyishimire D, Reiss P, et al. Hepatitis B virus prevalence and vaccine response in HIV-infected children and adolescents on combination antiretroviral therapy in Kigali, Rwanda. Pediatr Infect Dis J. 2013 Mar;32(3):246–51.

44. Anigilaje EA, Olutola A. Prevalence and Clinical and Immunoviralogical Profile of Human Immunodeficiency Virus-Hepatitis B Coinfection among Children in an Antiretroviral Therapy Programme in Benue State, Nigeria. ISRN Pediatr. 2013;2013:1–7.

45. Peebles K, Nchimba L, Chilengi R, Moore CB, Mubiana-Mbewe M, Vinikoor MJ. Pediatric HIV-HBV coinfection in Lusaka, Zambia: Prevalence and short-term treatment outcomes. J Trop Pediatr. 2015 Dec 1;61(6):464–7.

Appendix 3

Table S3.1: Results of model testing for testing host factors against dependent variable square root $CD4^+$ count

Model	**Covariates** [1]	**Factors** [2]	**r^2**	**Intercept**	***P* value**
Model 1	Age / HIVVL	CMV + Stunting	0.07	29.78	0.003
Model 2	Age	CMV +	0.10	30.10	$2.29x10^{-5}$
Model 3	**Age**	**CMV + rs1063303**	**0.13**	**30.50**	**2.02**$x10^{-5}$
Model 4	Age	CMV + rs7935564	0.09	30.20	0.001
Model 5	Age	CMV + rs61735273	0.09	29.68	0.001

[1] "HIVVL" = HIV-1 pVL. *P* value reported of the model. [2] "CMV +" indicates that a patient had a positive plasma CMV qPCR.

Table S3.2: Summary for linear regression model 3

Factors [1,2]	**Estimate**	**SE**	**t value**	***P* value**
Age	-0.71	0.24	-3.00	0.003
rs1063303(CC)	1.00			
rs1063303(GC)	0.32	1.56	0.21	0.83
rs1063303(GG)	-3.81	1.48	-2.57	0.01
CMV viraemia -	1.00			
CMV viraemia +	-3.92	1.21	-3.23	0.002

[1] "CMV viraemia +" indicates that a patient had a positive plasma CMV qPCR. [2] Age at the time a sample was taken. *P* value reported for the factor or covariate effect.

Table S3.3: Sensitivity analysis for linear model 3

Factors [1,2]	**Estimate**	**SE**	**t value**	***P* value**
Age	-0.53	0.29	-1.85	0.07
rs1063303(CC)	1.00			
rs1063303(GC)	-0.80	1.93	-0.41	0.67
rs1063303(GG)	-4.88	1.85	-2.64	0.01
CMV viraemia -	1.00			
CMV viraemia +	-2.75	1.50	-1.84	0.07

[1] "CMV viraemia +" indicates that a patient had a positive plasma CMV qPCR. [2] Age at the time a sample was taken. *P* value reported for the factor or covariate effect.

Table S3.4: Model fitting procedures for multivariate binary logistic regression to determine associations with AIDS

Model name	**Dependent variable** [1]	**Factors** [2]	**Covariates** [3]	**AIC**
Model 1	CD4 < 200	CMV positive	Age	203.3
Model 2	CD4 < 200	CMV positive + rs1066303	Age	191.84
Model 3	CD4 < 200	CMV positive + rs7935564	Age	196.92
Model 4	CD4 < 200	CMV positive + rs61735273	Age	197.5

[1] CD4 < 200 = $CD4^+$ count less than 200 cells/μL at enrolment into the trial. [2] CMV positive indicates that a patient had a positive plasma CMV qPCR. [3] age at time of sample being taken. AIC = Akaike Information Criterion.

Chapter 4

Stratification of HIV-2 disease progression using $CD4^+$ T cell kinetic data

4.1 Introduction

In comparison to HIV-1, HIV-2 infection is characterised by a relatively slower rate of $CD4^+$ T cell decline, lower plasma viral loads (pVL) and slower disease progression (1,2). Early publications with short follow-up times (less than 10 years) reported that people infected with HIV-2 were unlikely to develop AIDS within five years of infection (3). Plasma viral loads and $CD4^+$ T cell percentage (CD4%) independently predict 10-year survival rates (4). Survival time is strongly tied to HIV-2 pVL, where survival time correlates negatively with HIV-2 pVL. Indeed, 37% of individuals with undetectable pVLs had reported survival times equivalent with HIV-negative participants (5). However, the authors urged caution when interpreting their findings as many individuals had only a single pVL measurement which was taken at various times in disease course between individuals where time of infection was unknown. This may have skewed the higher pVL toward later stage infections when pVLs are expected to increase and survival will be shorter. A 2016 systematic review summarised the factors associated with increased mortality after HIV-2 infection as: late stage presentation, male sex, $CD4^+$ T cell count less than 500 cell/μL, high pVL, anaemia (haemoglobin concentration of less than 8g/dL) and a low body mass index (6).

Researchers from the ANRS CO5 cohort reported that 6% of patients were classified as long term non-progressors (LTNP - 10 year follow-up with minimal clinical progression and $CD4^+$ T cell nadir more than 500 cells/μL) and 9% as HIV controllers (10 year follow-up with minimal clinical progression and 90% of HIV-2 pVL more than 500 copies/mL). There was considerable overlap between these disease progressor groups, with 55% of HIV controllers classified as LTNP (7). However, this report indicates that 45% of HIV controllers will have evidence of disease progression despite low/undetectable pVL. The ANRS cohort was clinically recruited and this may introduce a recruitment bias towards patients with symptomatic HIV-2 infection.

One of the shortcomings with many estimates of HIV-2 disease progression rates and survival times are uncertain dates of infection in the study participants. Findings from the Guinea-Bissau Police cohort which included people with estimated seroconversion dates showed that HIV-2 infected people had a high probability of developing AIDS, and dying as a result, when studied for a long period of time (23 years) (8). Median time to AIDS and death after HIV-1 infection was approximately six and eight years respectively. Comparable figures for HIV-2 were approximately double that, at 14 and 15 years respectively (8).

In HIV-1 research identifying extreme disease progression groups has proved useful in understanding biological mechanisms responsible for variation in viral control and disease progression rates. A 2014 systematic review reported 714 definitions of HIV extreme disease progression groups in 501 eligible studies. The reviewers suggested commonalities between these various definitions in an aim to promote consensus across studies and reduce variability in reported outcomes (9). Identifying distinct progressor groups of HIV-2 infection and disease will be useful in studying host and viral genetic determinants of

disease outcomes. However, given the relatively small sample sizes available in HIV-2 cohorts the "extremeness" of the studied phenotype should be balanced against numbers of patients included in an analysis.

Using CD4% kinetic data Palm *et al* showed that stratifying patients into relative fast and slow progressor groups could identify differences in evolutionary rates of the HIV-2 *env* gene (10). This stratification procedure resulted in similar sized groups which could be scaled continuously, allowing for flexibility in the definition. In this chapter we aim to expand on this stratification system as a method for identifying different disease progressor groups for host and viral genetic comparisons in later chapters.

We hypothesised that CD4% kinetics would provide a useful means of stratifying patients into disease progressors groups. Furthermore, that these progressor groups would have significant differences in immunological and clinical outcomes.

4.1.1 Specific aims

1. Stratify patients with HIV-2 infection into fast and slow disease progressors groups based on CD4% kinetic data

2. Compare immunological and clinical outcomes across the progressor groups

4.2 Methods

4.2.1 Summary statistics for the Guinea-Bissau Police cohort clinical data

The Police cohort is a prospective open cohort study of police officers recruited from urban and rural areas in Guinea-Bissau. All police officers were invited to join the study and 98% agreed to do so. The cohort started recruiting patients on 6th February 1990 and continued recruitment till 28th September 2009. HIV-infected participants were followed up until 28th September 2013. Subsets of HIV-infected participants have been followed-up after this date, but unfortunately not systematically for the entire cohort. Our analysis therefore ends at the 28th of September 2013. Our analysis excluded HIV-1 infected patients who were superinfected with HIV-2 and those that had inadequate clinical and immunological follow up.

4.2.2 Patient enrolment and sample collection

Blood samples were collected at enrolment and at scheduled annual follow-up visits. HIV testing was done at the National Public Health Laboratory in the capital Bissau. HIV disease was classified according to the WHO staging criteria (11). Mortality reports from the Ministry of Interior were crosschecked by health post staff at the main police station in Bissau. Reported symptoms before death were used to classify cause of death. Cause of death was classified as HIV-related, not HIV-related and accidental. Individuals who died from accidental causes were right-censored from the survival analysis.

ART was introduced in Guinea-Bissau in 2005 through a national treatment programme. The Police cohort was included in the programme in early 2006. Individuals receiving ART

were right-censored from the analysis from the timepoint of ART initiation. Date of HIV infection was estimated as the mid-timepoint between the last HIV-seronegative and the first HIV-seropositive sample. Patients with estimated dates of infection are referred to as **seroconverters** and those HIV-2 infected at enrolment as **seroprevalent** patients. Patient data for HIV-2 infection was right-censored in the event of HIV-1 superinfection.

4.2.3 CD4% measurements

Both absolute $CD4^+$ T cell counts ($CD4^+$ counts) and CD4% are reliable immunological markers of HIV disease progression. In resource-limited settings, the CD4% has lower variability and sensitivity to specimen handling, patient age, or time of sampling (5). T-lymphocyte subsets were determined at the National Public Health Laboratory (LNSP), by conventional flow cytometry (Until 2005: FACStrak; Becton Dickinson, San Jose, CA, and from 2006 and onwards: CyFlow, Partec, Münster, Germany). Leucocyte counts were performed with a cell counter until 2005 (Coulter Counter CBC5; Coulter Electronics Ltd, Luton, England), and by CyFlow from 2006 and onwards.

4.2.4 Classification of HIV-2 infected individuals as relative fast and slow disease progressors

Longitudinal follow-up of ART naïve HIV-1 infected individuals has suggested that time to AIDS and death can be explained by $CD4^+$ counts and decline rate as well as pVL (12–15). For patients with two or more CD4% measurements we used per-patient linear regression models to estimate the midpoint CD4% in follow-up time. We used the coefficient of the regression model as the CD4% change rate, measured as CD4% change/year. CD4% increased, stayed the same and decreased for different patients over

follow-up; we therefore described CD4% as change rate and not decline rate. We have indicated a negative change rate for a decreasing CD4% and a positive change rate for increasing CD4%. This procedure then gave two statistics per patient, a midpoint CD4% level and CD4% change rate. These data were then transformed into proportional values on a positive scale, between zero and one. Patients were then ranked from highest CD4% level to lowest, and highest change rate to lowest (they were all on a positive scale at this point). Those with CD4% levels and change rates greater than the mean value were classified as **slow progressors**, those below the mean as **fast progressors**. Finally, to account for variability in disease stages at enrolment we used a combined coefficient derived from an equally weighted estimate using both CD4% level and change rate. This is referred to as the combined coefficient.

4.2.5 Statistical analysis

Statistical analysis was done in R Studio, version 3.4.0 (16,17). Pairwise comparisons of continuous numeric variables were done using the Student's T test and Mann Whitney-U test for parametric and non-parametric distributions respectively. Where appropriate, corrections for multiple comparisons was done using the Holm-Bonferroni method. Linear mixed effects models were used to analyse longitudinal CD4% data using follow-up time as a fixed, continuous covariate in SPSS (IBM, version 25.0.1) (18). Univariate survival analysis was done by Kaplan-Meier survival curves and hypotheses tested using the logrank test. Multivariate survival analysis was done using Cox regression models. Proportional hazard assumptions were tested by examining Schoenfeld residuals for variance with time. All survival analysis was done in R using the survival and survminer packages (19,20). Pirate plots were prepared using the YARRR package and all other figures with the ggpubr package in R (21,22)

4.3 Results

A total of 389 HIV-2 infected individuals were included, of these 219 had two or more CD4% measurements during follow-up and were stratified into progressor groups (**table 4.1**). Using the combined coefficient to stratify patients we divided the cohort into relatively equal sized groups, 101 fast progressors and 118 slow progressors. The stratification procedure agreed on the progressor group using both the midpoint CD4% and change rate in 49% of patients (108/219).

Females had a higher average CD4% over time (female CD4% increase relative to males of 6.31 percent, 95% CI: 3.28 – 9.34. $P = 5.8 \times 10^{-5}$). CD4% change over time follows a linear pattern with an average change rate of -0.42/year (SE = 0.05). Males and females had similar CD4% change rates, -0.43/year (SE = 0.05) and -0.39/year (SE = 0.1) respectively. These finding were unrelated to age which did not correlate with midpoint CD4% or CD4% change rate (Pearson correlation $P = 0.75$ and $P = 0.37$ respectively).

Table 4.1: Police cohort summary statistics

	Number (%)	**Missing/ Censored (%)**	**Mean**	**IQR**
HIV-2 infected individuals	389			
Male	317 (81)			
Female	72 (19)			
Fast progressor [1]	101 (26)	170 (44)		
Slow progressor [1]	118 (30)			
Died during study [2]	171 (44)	218 (56)		
Diagnosed with AIDS during study	168 (43)	221 (57)		
HIV-related deaths	110 (28)	279 (72)		
Estimated date of infection [3]	108 (28)	281 (72)		
Age at enrolment (years)	388 (99.99)	1 (0.001)	37.5	31 - 44

[1] Progressor groups determined by the combined coefficient. [2] This variable excluded all accidental deaths. [3] Patients with an estimated date of seroconversion based on the serosurvey.

Slow progressors had significantly higher midpoint CD4% compared to fast progressors (34.40 vs 21.10, 95% CI of difference: 11.65 - 15.78, T test $P = 2.45 \times 10^{-12}$, **figure 4.1**). CD4% change rate per year was negative in fast progressors (-1.87 vs +0.18, MW $P = 4.92 \times 10^{-8}$, **figure 4.1**).

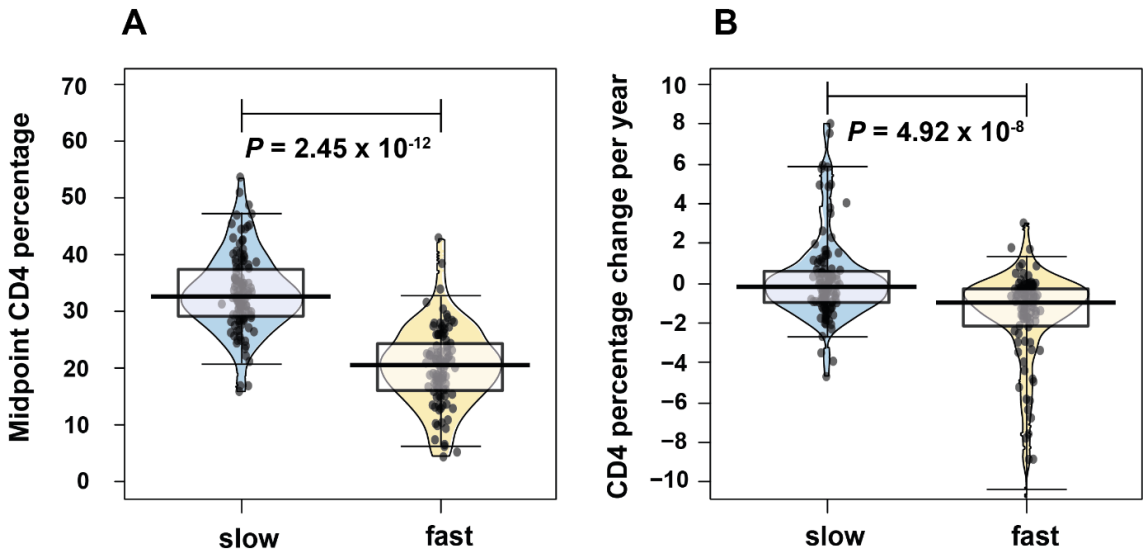


Figure 4.1: Midpoint CD4% and CD4% change rate by progressor groups defined by the combined coefficient. A) Pirate plot comparing midpoint CD4% for fast and slow progressors. **B)** Pirate plot comparing CD4% change rate per year for progressor groups. Results of the Students T test are shown in each panel.

A total of 219 patients contributed 897 CD4% data points, 509 (56.74%) were from slow progressors and 388 (43.23%) were from fast progressors. The majority of CD4% data clustered before 10 years of follow-up and were evenly distributed across the two groups in time (**figure 4.2**).

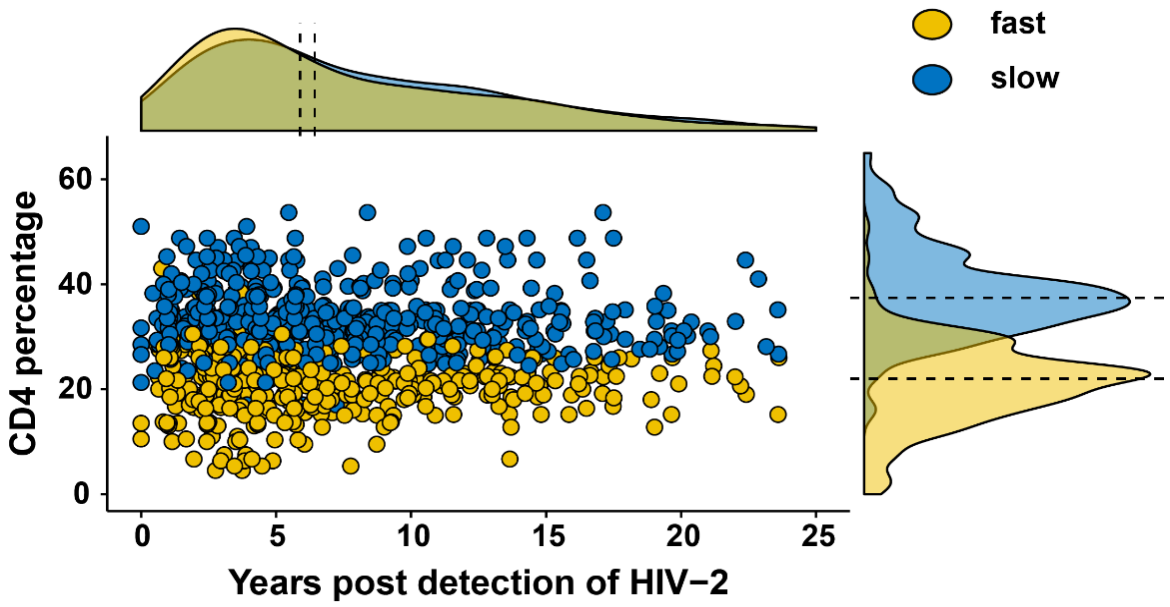


Figure 4.2: CD4% change over follow-up between progressor groups. Scatterplot of longitudinal CD4% change by progressor group with marginal density plots. The density plot of the x-axis shows that data points were evenly distributed in follow-up time between the groups. The y-axis density plots show the difference in CD4% distribution between the progressor groups, the dashed lines indicate the median years post HIV-2 detection and CD4% for the x and y density plots respectively.

4.3.1 Time to event analyses

We next determined the effect of progressor stratification by midpoint CD4%, CD4% change rate and the combined coefficient on time to AIDS (**figure 4.3**). Median time to AIDS for slow versus fast progressors stratified by midpoint CD4% was 18.80 years and 9.35 (logrank $P = 1.09\text{x}10^{-7}$). When stratifying by the CD4% change rate, slow progressors progressed to AIDS at a median of 16.70 years and fast progressors at 10.30 years (logrank $P = 0.001$). The combined coefficient's effect was larger - slow progressors time to AIDS was 20.59 years and fast progressors 9.35 years (logrank $P = 1.97\text{x}10^{-10}$).

Many patients (72%) had unknown dates of HIV-2 infection, therefore heterogeneity in outcomes might be accounted for by differences in disease stage at inclusion into the study. To assess whether the date of HIV-2 infection changed our estimates for time to AIDS and

death we performed a sensitivity analysis where we estimated median time to event from the estimated seroconversion date. Time to AIDS and death did not vary significantly for seroconverters versus seroprevalent patients (**table S4.1**). A further source of potential bias is that some patients had a greater number of CD4% measurements over a longer period of follow-up time. To address this, we performed a second sensitivity analysis where we excluded patients with increasingly longer duration of follow-up CD4% data (**table S4.2**).

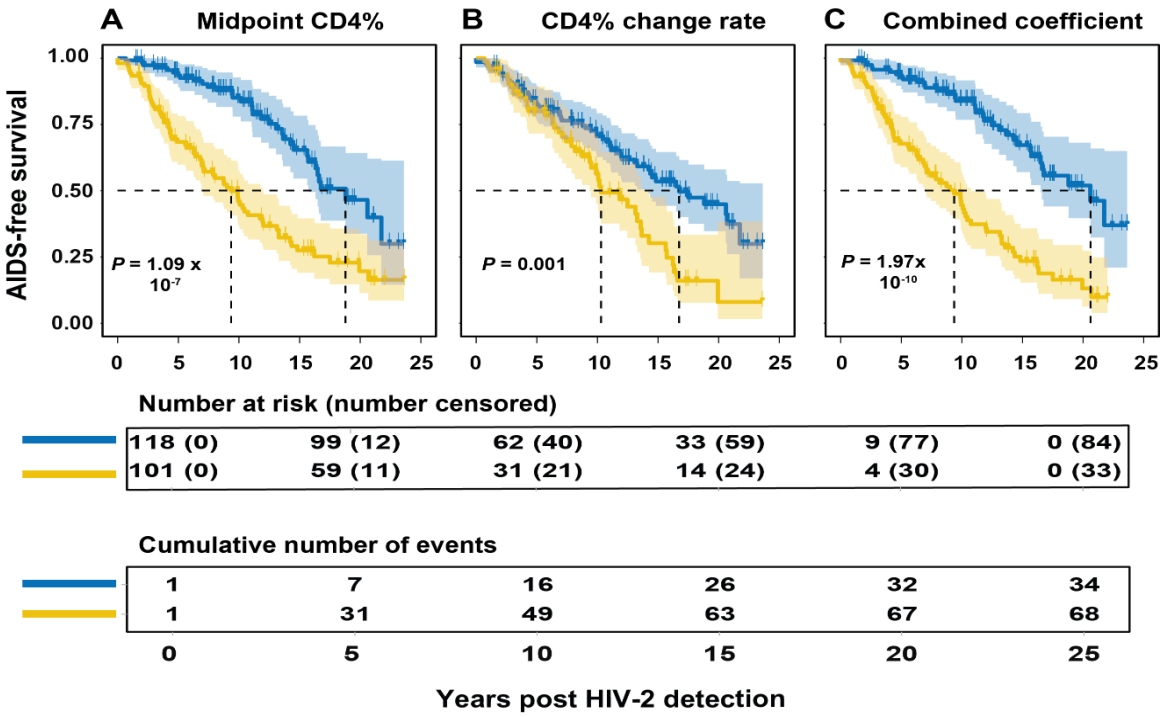


Figure 4.3: Kaplan-Meier curves for time to AIDS for progressor groups. Different stratification systems are shown in panels A – C. Proportion of patients who do not have AIDS are shown on the y-axis and follow-up time after HIV-2 detection on the x-axis. Fast progressors are coloured yellow and slow progressors in blue with shaded confidence intervals around the curves. **A)** Patients stratified only by midpoint CD4% **B)** Patients stratified by CD4% change rate **C)** Patients stratified by the combined coefficient. Number at risk and number censored for the combined coefficient are shown below the curves with the cumulative number of events at five-year intervals at the bottom.

4.3.2 Multivariate analysis of time to AIDS and mortality

The univariate survival estimates shown above for progressor groups might be confounded by other patient factors. For example, females were over-represented in the slow progressor group – 71% of females and 50% of males were slow progressors using the combined coefficient. Females were younger than males at enrolment (age at enrolment: females = 33.50 years versus males = 38.80 years; T test $P = 2.43x10^{-6}$) and infection (age at infection: females = 33.50 years versus males = 38.90 years; T test $P = 2.00x10^{-3}$, **figure 4.4**). Fast and slow progressors had similar ages at enrolment and their distributions overlapped (T test $P = 0.97$). To determine the specific effects of CD4% kinetics on survival estimates we used multivariate cox regression models to estimate relative hazard ratios for progressor status, sex and age.

Median time to AIDS for all patients was 12.50 years (95% CI: 10.56 – 14.34) and median time to death was 16.40 years (95% CI: 14.50 – 18.28, **figure 4.5)**. Treated as continuous covariates in regression models adjusted for age and sex, each unit increase in midpoint CD4% was associated with a relative risk reduction of 11% for progression to AIDS (HR = 0.89; 95% CI: 0.87 – 0.92. $P = 7.77x10^{-15}$). Each unit increase in CD4% change rates was associated with a 9% decreased relative risk of AIDS (HR = 0.91; 95% CI: 0.82 – 0.99. $P = 0.04$). These covariates' effects had a non-significant interaction (interaction term $P = 0.7$). We therefore decided to use the progressor status stratification based on the combined coefficient to determine the effect of CD4% kinetics on time to AIDS and death.

Age was treated as a categorical variable, divided at 40 years as this cut-off has been previously used to for survival analyses of HIV-2 and was close to our cohort's median age (23). Age was also analysed as a continuous covariate to determine cumulative risk of AIDS and death with increasing age at enrolment.

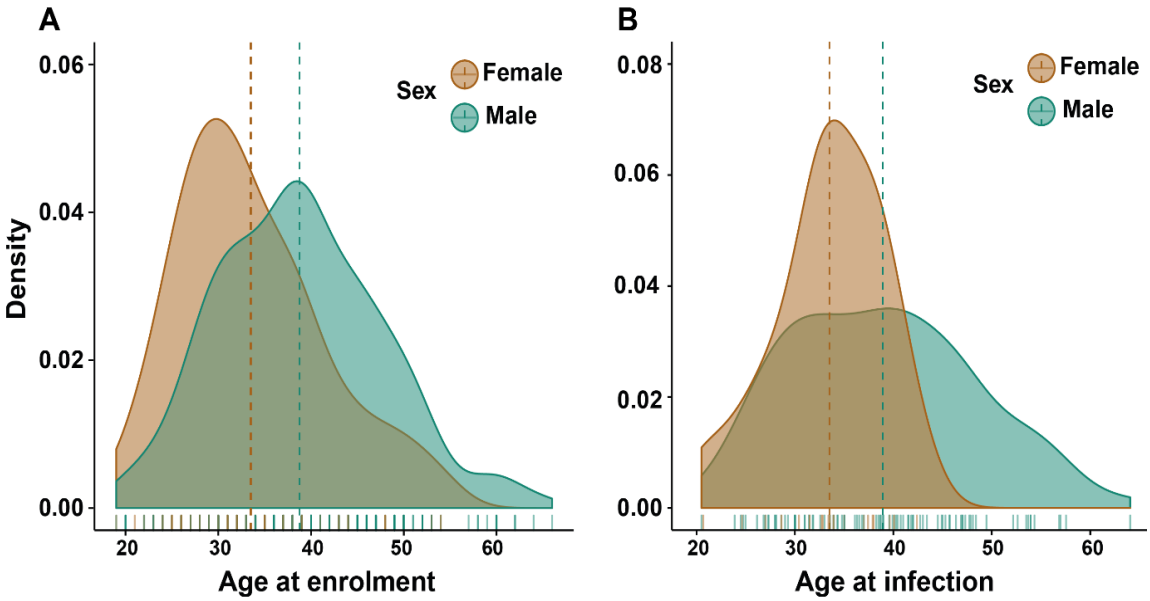


Figure 4.4: Age at enrolment by sex and age at estimated date of infection. A) Density plots for age at enrolment by sex. **B)** Density plots for age at the estimated date of infection for seroconverters by sex. Dashed lines indicate mean estimates for age in each group and the carpets on the x-axis shows the number of data points at each time point.

Age at enrolment was not associated with an increased risk of progressing to AIDS in follow-up (**table 4.2**). Males had a greater risk of progression to AIDS (HR = 1.93, 95% CI: 0.99 - 3.74, P = 0.05). Fast disease progressors had an approximately three-fold increased risk of progression to AIDS when compared to slow progressors (HR = 3.29, 95% CI: 2.18 - 5.07, $P = 2.25 \times 10^{-8}$).

After adjusting for sex and age, fast progressors had weak evidence of an increased relative risk of death during follow-up when compared to slow progressors (HR = 1.60, 95% CI: 0.98 – 2.63, $P = 0.06$, **table 4.3**). Females had a lower risk of death relative to men, however this estimate was not statistically significant (HR of male sex = 2.19, 95% CI: 0.94 – 5.10, $P = 0.07$).

Table 4.2: Summary of time to AIDS in years

Univariate estimates						
Factor	**Group**	**n**	**Events**	**Time to AIDS**	**95% CI**	***P* value**
Progressor [1]	Slow	118	34	20.59	16.49 - NA	$1.97x10^{-10}$
	Fast	101	68	9.35	7.05 - 10.70	
Sex	Female	72	19	18.80	15.70 - NA	0.01
	Male	317	149	11.10	10.10 - 13.30	
Age [2]	< 40 years	252	119	12.00	10.39 - 14.80	0.66
	> 40 years	137	49	13.60	8.73 - 16.50	
Multivariate estimates						
Factor	**Group**	**n**	**Events**	**Hazard ratio**	**95% CI**	***P* value**
Progressor [1]	Slow	118	34			
	Fast	101	68	3.29	2.18 - 5.07	$2.25x10^{-8}$
Sex	Female	44				
	Male	175		1.93	0.99 - 3.74	0.05
Age [2]	< 40 years	154				
	> 40 years	65		1.43	0.92 - 3.74	0.11

[1] Progressor groups determined by the combined coefficient. [2] Age at enrolment. 95% confidence intervals and *P* values are shown for a logrank test. n = number of patients per group

Table 4.3: Summary of time to death in years

Univariate estimates						
Factor	**Group**	**n**	**Events**	**Time to death**	**95% CI**	***P* value**
Progressor [1]	Slow	118	31	21.00	17.40 – NA	0.06
	Fast	101	34	17.80	15.90 - NA	
Sex	Female	72	14	NA	18.80 - NA	0.003
	Male	317	126	15.90	13.10 – 17.40	
Age [2]	< 40 years	252	88	17.80	16.60 – 21.00	$1.27x10^{-4}$
	> 40 years	137	52	12.30	9.80 – 16.00	
Multivariate estimates						
Factor	**Group**	**n**	**Events**	**Hazard ratio**	**95% CI**	***P* value**
Progressor [1]	Slow	118	31			
	Fast	101	34	1.60	0.98 – 2.63	0.06
Sex	Female	44				
	Male	175		2.19	0.94 – 5.10	0.07
Age [2]	< 40 years	154				
	> 40 years	65		3.10	1.82 – 5.27	$2.99x10^{-5}$

[1] Progressor groups determined by the combined coefficient. [2] Age at enrolment. 95% confidence intervals and *P* values are shown for a logrank test. n = number of patients per group

Each additional year of age at enrolment was associated with an increased risk of death of 7% during follow-up and age above 40 years at enrolment was associated with a large increase in risk of death in follow-up (HR = 3.10, 95% CI: 1.83 – 5.27, *P* = 2.99x10-5, **table 4.3**).

One hundred and sixty-eight patients were diagnosed with AIDS during the study, 67% of fast progressors and 29% of slow progressors. More men were diagnosed with AIDS than woman (47% versus 26%) and a greater number of younger patients (age less than 40) were diagnosed with AIDS than older patients – 47% versus 36% respectively.

There were 171 recorded deaths during the 24 years of follow-up and 64% were HIV-related (110/171). A greater proportion of males than females died during follow-up (40% versus 19%). There were more deaths in the fast progressor group (Fast versus slow progressors: 34% versus 26%) and more patients older than 40 years of age at enrolment died during the study (Age > 40 versus age < 40 years: 38% versus 35%).

The median survival time after progression to AIDS was 27.70 months (95% CI 11.80 – 45.60). After progression to AIDS, 57% fast progressors and 35% of slow progressors were censored. Of the 51 censored events three were for HIV-1 superinfection, two for ART initiation and two for accidental death - the remainder (86%) of censoring after AIDS is presumed as lost to follow-up (LTFU). Proportional hazards assumptions were not violated for the time to AIDS analysis (**table 4.2**). Progressor status in the time to death analysis violated the proportional hazards assumptions (Schoenfeld *P* = 0.03, **table 4.3**). Sex and age at enrolment did not violate these assumptions, nor did the global Schoenfeld test.

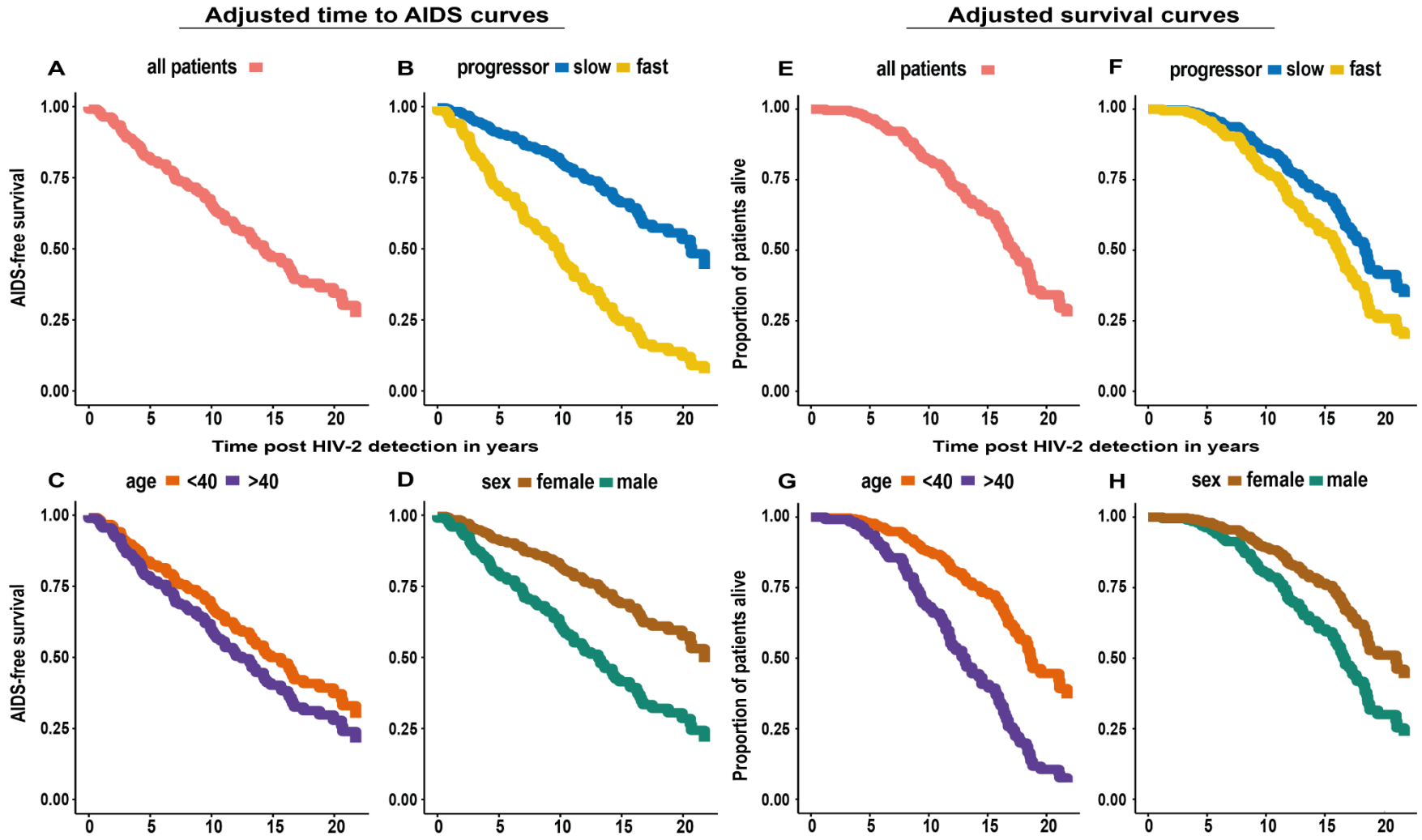


Figure 4.5: Cox regression adjusted survival curves. A - D) Survival curves for time to AIDS for all patients, progressors stratified using the combined coefficient, age and sex. **E - H)** Survival curves for time to death for all patients, progressors stratified using the combined coefficient, age and sex. Y-axes show the proportion of patients without AIDS and alive. X-axes show time post HIV-2 detection measured in years.

4.4 Discussion

Fast progressors had a significantly lower midpoint CD4% during follow-up and a faster CD4% decline rate. Consequently, fast progressors time to AIDS was approximately half that of slow progressors. In a multivariate model, which adjusted for age and sex, fast progressors had a large relative increase in risk of progressing to AIDS when compared to slow progressors. This did not translate into an increased risk of death or significantly shorter survival times. This finding is surprising given previous literature (4,24–26). The median survival times in HIV-2 infected adults which we reported are within previously published ranges (8,27,28).

Possible explanations for the discrepancy in survival times for fast and slow progressors include the following. Firstly, the Police cohort reported 45% of patients as LTFU, defined as individuals censored three years before the study end (8). Secondly, we additionally censored patients for HIV-1 superinfection, ART initiation and accidental death. After progression to AIDS a greater number of fast than slow progressors were censored, with most probably LTFU. In Sub-Saharan African HIV cohorts many patients who are LTFU will have died, the estimated mortality rate of patients LTFU ranges from 12 - 87% (29,30). Mortality estimates and LTFU rates are inversely correlated, and a high LTFU rate significantly biases true mortality rates in cohort studies (29,31). Mortality in LTFU is associated with male sex and advanced HIV disease. In contrast, LTFU in females and those with earlier stage HIV disease is associated with transfer to other health centres for continued care (32).Therefore it is likely that we have substantially underestimated the mortality rate in the fast progressor group.

Our estimates for time to AIDS for fast and slow progressors were similar to those of Palm *et al* who used the first CD4% measurement as their starting point for time to AIDS (10). We have also confirmed their finding that midpoint CD4% and the change rate associated independently with disease progression. The notably stronger effect size of midpoint CD4% versus CD4% change rate on disease progression suggests that events around the acute infection period (e.g. post-seroconversion CD4% or "immune activation set point") have a greater influence on long-term outcomes than events in the later stages (33,34).

Previous findings from the Police cohort showed that, in seroconverters, CD4% decline rate did not differ between males and females – which our findings support (8). Of further importance, females at the estimated date of infection had higher CD4% than males for both HIV-1 and HIV-2 infection (8). A greater proportion of females in our study were slow progressors than males. Females had significantly higher CD4% values in follow-up with similar CD4% decline rates to males. This suggests that the difference in CD4% based on sex may be explained by events in the acute infection period. Females produce a more robust inflammatory response after HIV-1 seroconversion, with lower pVL and higher median $CD4^+$ counts in the absence of ART (35). In HIV-1 infection this is associated with a faster progression to AIDS. Perhaps in HIV-2 infection an initial heightened inflammatory response in females increases the probability of inducing a protective CTL response, which may be protective in the long-term (36,37).

An alternative explanation is that founder viruses from males may have greater replicative capacity than those infecting females (38). The hypothesis for this heterosexual evolutionary bottleneck is that the vaginal mucosa is enriched with HIV-specific target cells. This would create an environment where less fit viruses could establish successful populations. In males, the dearth of cells to infect in the penile mucosa would select for

founder viruses with higher viral replicative capacity. After infection is established this may then lead to accelerated disease progression in males versus females (38).

Many studies which examined factors associated with extreme groups of disease progression have focused on HIV-2 pVL and clinical stage (36,39–41). Clinical factors which are associated with pVL may not necessarily associate with $CD4^+$ T cell kinetics or disease progression rates (7). Plasma viral load explains little of the variation in $CD4^+$ T cell decline rates in HIV-1 infection (13,14). Immune activation and pVL independently predict $CD4^+$ T cell decline and mortality in HIV-1 and HIV-2 infection (25,34,42). The dependence between immune activation and HIV-2 pVL has proved controversial in previous reports (43,44). One possible explanation for this discrepancy is that patients in these studies had unknown dates of infection, and therefore unknown stages of disease progression. Soluble immune activation markers correlate negatively with CD4% in the years preceding death in untreated HIV-1 and -2 infection and increase nonlinearly closer to death (28). Therefore, the strength of an association between pVL and immune activation may not be constant across different stages of HIV infection. Average age for both studies was greater than 45 years and patients were predominantly female which may also limit generalisability (43,44). The Police cohort is predominantly male and differences between our findings and those from other studies may be explained by this bias.

Studies correlating $CD4^+$ counts, decline rates, pVL and immune activation usually report weak to moderate dependence between these variables. This implies that there are other factors which may add random variation to disease outcomes. The presence of additional random variation in SPVL and $CD4^+$ T cell decline rates has been shown in a linked analysis of transmitted/founder viruses. In this analysis heritability (correlation of continuous variables between donor and recipient in a transmission pair) of SPVL and

$CD4^+$ decline rates were 29% and 17% respectively. Viral pathogenicity (regression coefficient of $CD4^+$ decline by SPVL adjusted for age) was independently heritable and linked to polymorphisms in *pol* which have been associated with viral replicative capacity (45).

An ideal disease stratification system would make use of viral and host genetic data, alongside $CD4^+$ T cell kinetic data and clinical progression rates as these variables are all likely to capture distinct aspects of HIV disease – perhaps independently of each other. Stratifying patients into disease progressor groups based on few aspects of HIV disease is an attempt to simplify a complex phenomenon. Using CD4% kinetic data we were able to divide patients into groups with clear differences in disease progression rates and mortality. This stratification system also associates with significant differences in synonymous substitutions rates in HIV-2 *env* (10). It is tempting to speculate that synonymous substitution rates in HIV-2 *env* would correlate strongly with immune activation or viral replicative capacity (46). Therefore, our stratification system may capture extraneous variables which explain additional variation in HIV disease outcomes.

4.5 Section summary

1. Forty-six percent of patients with longitudinal CD4% data were classified as fast progressors with a median time to AIDS of 9.35 years. Fifty-four percent were slow progressors with a median time to AIDS of 20.59 years.

2. Females were more likely to have slow disease progression after HIV-2 infection.

4.6 References for chapter 4

1. Gottlieb GS, Sow PS, Hawes SE, Ndoye I, Redman M, Coll-Seck AM, et al. Equal Plasma Viral Loads Predict a Similar Rate of CD4 $^{+}$ T Cell Decline in Human Immunodeficiency Virus (HIV) Type 1– and HIV-2–Infected Individuals from Senegal, West Africa. J Infect Dis. 2002 Apr;185(7):905–14.

2. Popper SJ, Sarr AD, Travers KU, Guèye-Ndiaye A, Mboup S, Essex ME, et al. Lower human immunodeficiency virus (HIV) type 2 viral load reflects the difference in pathogenicity of HIV-1 and HIV-2. J Infect Dis. 1999 Oct;180(4):1116–21.

3. Marlink R, Kanki P, Thior I, Travers K, Eisen G, Siby T, et al. Reduced rate of disease development after HIV-2 infection as compared to HIV-1. Science. 1994 Sep 9;265(5178):1587–90.

4. Berry N, Jaffar S, van der Loeff MS, Ariyoshi K, Harding E, N'Gom PT, et al. Low Level Viremia and High CD4% Predict Normal Survival in a Cohort of HIV Type-2-Infected Villagers. AIDS Res Hum Retroviruses. 2002 Nov 1;18(16):1167–73.

5. van der Loeff MFS, Larke N, Kaye S, Berry N, Ariyoshi K, Alabi A, et al. Undetectable plasma viral load predicts normal survival in HIV-2-infected people in a West African village. Retrovirology. 2010 May 19;7(1):46.

6. Tchounga B, Ekouevi DK, Balestre E, Dabis F. Mortality and survival patterns of people living with HIV-2. Curr Opin HIV AIDS. 2016 Sep;11(5):537–44.

7. Thiébaut R, Matheron S, Taieb A, Brun-Vezinet F, Chêne G, Autran B, et al. Long-term nonprogressors and elite controllers in the ANRS CO5 HIV-2 cohort. AIDS. 2011 Mar 27;25(6):865–7.

8. Esbjörnsson J, Månsson F, Kvist A, da Silva ZJ, Andersson S, Fenyö EM, et al. Long-term follow-up of HIV-2-related AIDS and mortality in Guinea-Bissau: a prospective open cohort study. lancet HIV. 2018 Nov 1;

9. Gurdasani D, Iles L, Dillon DG, Young EH, Olson AD, Naranbhai V, et al. A systematic review of definitions of extreme phenotypes of HIV control and progression. AIDS. 2014 Jan 14;28(2):149.

10. Palm AA, Lemey P, Jansson M, Månsson F, Kvist A, Szojka Z, et al. Low Postseroconversion CD4 $^{+}$ T-cell Level Is Associated with Faster Disease Progression and Higher Viral Evolutionary Rate in HIV-2 Infection. Paraskevis D, editor. MBio. 2019 Jan 8;10(1).

11. World Health Organization. WHO case definitions of HIV for surveillance and revised clinical staging and immunological classification of HIV-related disease in adults and children. World Health. 2007.

12. Phillips AN, Staszewski S, Weber R, Kirk O, Francioli P, Miller V, et al. HIV Viral Load Response to Antiretroviral Therapy According to the Baseline CD4 Cell Count and Viral Load. JAMA. 2001 Nov 28;286(20):2560.

13. Mellors JW, Margolick JB, Phair JP, Rinaldo CR, Detels R, Jacobson LP, et al. Prognostic Value of HIV-1 RNA, CD4 Cell Count, and CD4 Cell Count Slope for Progression to AIDS and Death in Untreated HIV-1 Infection. JAMA. 2007 Jun 6;297(21):2346.

14. Rodríguez B, Sethi AK, Cheruvu VK, Mackay W, Bosch RJ, Kitahata M, et al. Predictive Value of Plasma HIV RNA Level on Rate of CD4 T-Cell Decline in Untreated HIV Infection. JAMA. 2006 Sep 27;296(12):1498.

15. Cori A, Pickles M, van Sighem A, Gras L, Bezemer D, Reiss P, et al. CD4+ cell dynamics in untreated HIV-1 infection. AIDS. 2015 Nov;29(18):2435–46.

16. R Core Team. R Language Definition V. 3.1.1. Https://WwwR-ProjectOrg/. 2014;

17. RStudio team, RStudio. RStudio: Integrated development environment for R. J Wildl Manage. 2015;

18. IBM SPSS Inc. SPSS Statistics for Windows. IBM Corp Released 2012. 2016;

19. Therneau T, Crowson CS, Atkinson EJ. Adjusted Survival Curves. Study R. 2015;

20. Kassambara A, Kosinski M, Biecek P, Fabian S. survminer: Drawing Survival Curves using "ggplot2". R package. version 0.4.3. 2018.

21. Kassambara A. ggpubr: "ggplot2" Based Publication Ready Plots. R Package. 2018.

22. Kampstra P. Beanplot: A Boxplot Alternative for Visual Comparison of Distributions. J Stat Softw. 2008 Nov 10;28(Code Snippet 1):1–9.

23. Matheron S, Pueyo S, Damond F, Simon F, Leprêtre A, Campa P, et al. Factors associated with clinical progression in HIV-2 infected-patients: the French ANRS cohort. AIDS. 2003 Dec 5;17(18):2593–601.

24. Hansmann A, Schim van der Loeff MF, Kaye S, Awasana AA, Sarge-Njie R, O'Donovan D, et al. Baseline plasma viral load and CD4 cell percentage predict survival in HIV-1- and HIV-2-infected women in a community-based cohort in The Gambia. J Acquir Immune Defic Syndr. 2005 Mar 1;38(3):335–41.

25. Jaffar S, Van der Loeff MS, Eugen-Olsen J, Vincent T, Sarje-Njie R, Ngom P, et al. Immunological predictors of survival in HIV type 2-infected rural villagers in Guinea-Bissau. AIDS Res Hum Retroviruses. 2005 Jun;21(6):560–4.

26. Jaffar S, Wilkins A, Ngom PT, Sabally S, Corrah T, Bangali JE, et al. Rate of decline of percentage CD4+ cells is faster in HIV-1 than in HIV-2 infection. J Acquir Immune Defic Syndr Hum Retrovirol. 1997 Dec 15;16(5):327–32.

27. Martinez-Steele E, Awasana AA, Corrah T, Sabally S, van der Sande M, Jaye A, et al. Is HIV-2- induced AIDS different from HIV-1-associated AIDS? Data from a West African clinic. AIDS. 2007 Jan 30;21(3):317–24.

28. Nyamweya S, Townend J, Zaman A, Steele SJ, Jeffries D, Rowland-Jones S, et al. Are plasma biomarkers of immune activation predictive of HIV progression: A longitudinal comparison and analyses in HIV-1 and HIV-2 infections? Gray CM, editor. PLoS One. 2012 Sep 10;7(9):e44411.

29. Brinkhof MWG, Pujades-Rodriguez M, Egger M. Mortality of patients lost to follow-up in antiretroviral treatment programmes in resource-limited settings: systematic review and meta-analysis. PLoS One. 2009 Jun 4;4(6):e5790.

30. Fox MP, Rosen S. Retention of Adult Patients on Antiretroviral Therapy in Low- and Middle-Income Countries: Systematic Review and Meta-analysis 2008-2013. J Acquir Immune Defic Syndr. 2015 May 1;69(1):98–108.

31. Egger M, Spycher BD, Sidle J, Weigel R, Geng EH, Fox MP, et al. Correcting mortality for loss to follow-up: a nomogram applied to antiretroviral treatment programmes in sub-Saharan Africa. PLoS Med. 2011 Jan 18;8(1):e1000390.

32. Chammartin F, Zürcher K, Keiser O, Weigel R, Chu K, Kiragga AN, et al. Outcomes of Patients Lost to Follow-up in African Antiretroviral Therapy Programs: Individual Patient Data Meta-analysis. Clin Infect Dis. 2018 Nov 13;67(11):1643–52.

33. Audigé A, Taffé P, Rickenbach M, Battegay M, Vernazza P, Nadal D, et al. Low Postseroconversion CD4 Count and Rapid Decrease of CD4 Density Identify HIV $^{+}$ Fast Progressors. AIDS Res Hum Retroviruses. 2010 Sep;26(9):997–1005.

34. Deeks SG, Kitchen CMR, Liu L, Guo H, Gascon R, Narváez AB, et al. Immune activation set point during early HIV infection predicts subsequent CD4+ T-cell changes independent of viral load. Blood. 2004 Apr 22;104(4):942–7.

35. Addo MM, Altfeld M. Sex-based differences in HIV type 1 pathogenesis. J Infect Dis. 2014 Jul 15;209 Suppl 3(Suppl 3):S86-92.

36. Leligdowicz A, Yindom L-M, Onyango C, Sarge-Njie R, Alabi A, Cotten M, et al. Robust Gag-specific T cell responses characterize viremia control in HIV-2 infection. J Clin Invest. 2007 Oct 1;117(10):3067–74.

37. Angin M, Wong G, Papagno L, Versmisse P, David A, Bayard C, et al. Preservation of Lymphopoietic Potential and Virus Suppressive Capacity by CD8+ T Cells in HIV-2-Infected Controllers. J Immunol. 2016;1600693.

38. Carlson JM, Schaefer M, Monaco DC, Batorsky R, Claiborne DT, Prince J, et al. Selection bias at the heterosexual HIV-1 transmission bottleneck. Science (80-). 2014 Jul 11;345(6193):1254031–1254031.

39. Mamede JI, Damond F, Bernardo A de, Matheron S, Descamps D, Battini J-L, et al. Cyclophilins and nucleoporins are required for infection mediated by capsids from circulating HIV-2 primary isolates. Sci Rep. 2017 Mar 27;7:45214.

40. Yindom L-M, Leligdowicz A, Martin MP, Gao X, Qi Y, Zaman SM a, et al. Influence of HLA class I and HLA-KIR compound genotypes on HIV-2 infection and markers of disease progression in a Manjako community in West Africa. J Virol. 2010;84(16):8202–8.

41. de Silva TI, Peng Y, Leligdowicz A, Zaidi I, Li L, Griffin H, et al. Correlates of T-cell-mediated viral control and phenotype of CD8+ T cells in HIV-2, a naturally contained human retroviral infection. Blood. 2013 May 23;121(21):4330–9.

42. Ariyoshi K, Jaffar S, Alabi a S, Berry N, Schim van der Loeff M, Sabally S, et al. Plasma RNA viral load predicts the rate of CD4 T cell decline and death in HIV-2-infected patients in West Africa. AIDS. 2000;14(4):339–44.

43. Sousa AE, Carneiro J, Meier-Schellersheim M, Grossman Z, Victorino RMM. CD4 T cell depletion is linked directly to immune activation in the pathogenesis of HIV-1 and HIV-2 but only indirectly to the viral load. J Immunol. 2002 Sep 15;169(6):3400–6.

44. Leligdowicz A, Feldmann J, Jaye A, Cotten M, Dong T, McMichael A, et al. Direct relationship between virus load and systemic immune activation in HIV-2 infection. J Infect Dis. 2010;201(1):114–22.

45. Bertels F, Marzel A, Leventhal G, Mitov V, Fellay J, Günthard HF, et al. Dissecting HIV Virulence: Heritability of Setpoint Viral Load, CD41 T-Cell Decline, and Per-Parasite Pathogenicity. Mol Biol Evol. 2017;35(1):27.

46. Lemey P, Kosakovsky Pond SL, Drummond AJ, Pybus OG, Shapiro B, Barroso H, et al. Synonymous substitution rates predict HIV disease progression as a result of underlying replication dynamics. PLoS Comput Biol. 2007 Feb 16;3(2):e29.

Appendix 4

Sensitivity analysis 1

We analysed univariate time to AIDS and for all patients versus those with estimated seroconversion dates. Fast progressors had a significantly shorter time to AIDS when compared to slow progressors, 9.35 vs 20.59 years (logrank $P = 1.97x10^{-10}$). Females progressed AIDS later after infection than males, 18.80 years vs 11.10 years (logrank $P = 0.002$) and age did not appear to have a strong influence on progression to AIDS. These results were largely similar for seroconverters and seroprevalent individuals. Seroconverters time to AIDS had lower levels of certainty given the small group sizes after stratification for different factors. Particularly for the progressor groups stratified by the combined coefficient we found no notable difference in time to AIDS.

Table S4.1: Summary of time to AIDS for Police cohort seroconverters

Group	n	Events	Time to AIDS	95% CI	*P* value
Estimates for all patients					
Slow [1]	118	34	20.59	16.49 - NA	$1.97x10^{-10}$
Fast [1]	101	68	9.35	7.05 - 10.70	
Female	72	19	18.80	15.70 - NA	0.002
Male	317	149	11.10	10.10 - 13.30	
Age < 40	252	119	12.00	10.39 - 14.80	0.66
Age > 40	137	49	13.60	8.73 - 16.50	
Estimates for patients with seroconversion dates [2]					
Slow [1]	28	9	20.59	11.13 - NA	0.14
Fast [1]	30	15	9.88	6.28 - NA	
Female	19	2	18.8	18.84 - NA	0.01
Male	89	38	11.00	8.76 - 18.20	
Age < 40	83	32	14.30	10.36 - NA	0.7
Age > 40	25	8	14.90	8.15 - NA	

[1] Progressor groups determined by the combined coefficient. [2] Patients with an estimated date of seroconversion based on the serosurvey. 95% confidence intervals and *P* values are shown for a logrank test. n = number of patients per group

Sensitivity analysis 2

We next performed a sensitivity analysis where we excluded patients with shorter periods of longitudinal CD4% data. The concern was that patients who had two or more CD4% data points clustered close together in time would be biased toward being classified as slow progressors. We first identified patients with less than three years follow-up time between their first and last sample, then five years. We iteratively excluded them and determined if time to AIDS changed for fast and slow progressors stratified by the combined coefficient. Excluding patients with shorter length of follow-up time increased the time to AIDS for fast and slow progressors. This was likely due to selecting out slower progressing patients.

Table S4.2: Time to AIDS for patients with longer CD4% follow-up

	n	**Events**	**Time to AIDS**	**95% CI**	***P* value**
All patients [1]	219				
- fast	101	68	9.35	7.05 - 10.70	1.97×10^{-10}
- slow	118	34	20.59	16.49 – NA	
>3 years CD4% data	171				
- fast	72	51	10.00	9.01 - 13.30	8.29×10^{-9}
- slow	99	25	21.70	16.74 – NA	
>5 years CD4% data	125				
- fast	50	37	12.00	9.94 - 14.80	6.75×10^{-7}
- slow	75	21	21.70	16.74 - NA	

[1] Progressor groups determined by the combined coefficient. Separate analyses are shown for patients with more than three and five years of CD4% data compared to the larger cohort. 95% confidence intervals and *P* values are shown for a logrank test. n = number of patients per group.

Chapter 5

TRIM22 genotype is associated with $CD4^+$ T cell kinetics in HIV-2 infected adults

5.1 Introduction

The most robust immunogenetic associations with disease outcomes in HIV-1 are found in *HLA* alleles and *CCR5* polymorphisms (1). Evidence indicates that *HLA* class I genotype explains less variation in disease outcomes in HIV-2 than HIV-1 (2,3). HIV-2 is more sensitive to host restriction factor TRIM5α than HIV-1 (4). The sensitivity of HIV-2 to TRIM5α is determined by a region of the capsid which corresponds to the cyclophilin binding loop in HIV-1 (5).

Multiple lines of evidence report that the capsid of HIV-2 is a key factor in determining disease outcomes. P24 sequences differ between patients with low and high HIV-2 viral loads. Prolines at positions 119, 159 and 178 are associated with low viral loads and enhanced proteasomal processing of a T cell epitope 165-DRFYKSLRA-173 *in vitro* (6,7). This epitope is associated with greater Gag-specific cytotoxic lymphocyte (CTL) responses, *in vivo*, as determined by ELISpot which are associated with long term non-progression (8). Together these findings suggest that p24 sequences directly influence adaptive immune responses.

HIV-2 p24 sequences with a proline at position 119 are more sensitive to restriction by cmTRIM5α (cynomolgus macaque orthologue) than sequences with alternative amino

acids at this position. This enhanced sensitivity to cmTRIM5α is associated with conformational changes in the capsid (9,10). HIV-1 NL4-3 chimeras which contain HIV-2 p24 sequences are more susceptible to restriction by human TRIM5α than the wildtype virus (11). Furthermore, HIV-2 p24 sequences derived from rapid progressors and long term non-progressors do not differ in their sensitivity to TRIM5α – though they are all more sensitive than HIV-1 (4).

Despite high *TRIM* diversity, *TRIM5* genotype has minimal impact on HIV-1 disease progression (12). Moreover, TRIM5α does not appear to drive evolutionary change in HIV-1 p24 sequences from subtype B (13). However, HIV-1 infected patients who are homozygous for the H43Y substitution develop AIDS more quickly than heterozygotes or H43 homozygotes (14) and this substitution modestly impairs the restrictive capacity of TRIM5α (14). Furthermore, the appearance of putative TRIM5α escape variants has been found to precede rapid immunological decline in HIV-1 infection (15).

TRIM22 is a paralogue of *TRIM5* and polymorphisms in *TRIM22* have been linked to differences in diseases outcomes in HIV-1 and Hepatitis C infections (16–18). TRIM22 blocks transcription factor Sp1 from binding to the HIV long terminal repeat region and reduces downstream viral replication (19). *TRIM22* RNA expression correlates negatively with HIV-1 viral load in acute HIV-1 infection and correlates positively with type 1 interferon levels and $CD4^+$ counts measured in the first year of infection (20). In the same patients these association were not observed for *TRIM5* expression. HIV-1 infected patients also had higher levels of *TRIM22* and lower levels of *TRIM5* expression when compared to uninfected controls (20).

Our hypothesis is that polymorphisms in *TRIM5* and *TRIM22* affect the structure or expression of these genes and thereby influence disease outcomes in HIV-2 infection.

5.1.1 Specific aims

1. Determine *TRIM5* and *TRIM22* genotype by sequencing their coding regions and identify which SNPs are present in HIV-2 infected patients from the Police Cohort.

2. Determine if SNPs in *TRIM5* and *TRIM22* associate with relevant clinical outcomes using both cross-sectional and longitudinal analyses.

3. Determine if genotype frequencies of the identified SNPs vary between fast and slow disease progressors as determined by the combined coefficient.

5.2 Methods

5.2.1 Patients included in analysis

We included HIV-2 infected patients from the Guinea-Bissau Police cohort described in the chapter four.

5.2.2 CD4% measurements

Both absolute $CD4^+$ cell counts and CD4% are reliable immunological markers of HIV disease progression. In resource-limited settings, the CD4% has lower variability and sensitivity to specimen handling, patient age, or time of sampling (21). T-lymphocyte subsets were determined at the National Public Health Laboratory (LNSP), by conventional flow cytometry (Until 2005: FACStrak; Becton Dickinson, San Jose, CA, and from 2006 and onwards: CyFlow, Partec, Münster, Germany). Leukocyte counts were performed with a cell counter until 2005 (Coulter Counter CBC5; Coulter Electronics Ltd, Luton, England), and by CyFlow from 2006 and onwards.

5.2.3 Stratification of patients into progressor groups

Briefly – patients longitudinal $CD4^+$ T cell percentage (CD4%) data were analysed in per-patient linear regression models. The CD4% at the midpoint in follow-up and the CD4% change rate (slope of the regression line) were extracted from these models. Patients were ranked from highest midpoint CD4% to the lowest, those above the mean were classified as slow progressors and those below the mean as fast progressors. Separately, patients were ranked from highest positive change rate to the lowest negative change rate and classified again. The midpoint CD4% and CD4% change rate were then transformed into proportional

values, added together and averaged for each patient. This gave a combined coefficient for each patient which was then used to rank patients and stratify them into progressor groups for further analysis.

5.2.4 DNA extraction and storage

Genomic DNA was extracted from whole blood or PBMC using the QIAamp Blood Midi Kit according to the manufacturer's instructions with minor modifications. Extracted DNA was stored at -20˚C.

5.2.5 Amplification and sequencing of *TRIM5* and *TRIM22*

Briefly - amplification primers were designed to flank exons 2 - 4 and 5 - 8 using the *TRIM5* and *TRIM22* reference sequences NG_029122 and NC_000011.10 respectively from the National Centre for Biotechnology Information (NCBI) database. Single step, touchdown PCR reactions were optimised, and separate inner sequencing primers were designed to sequence individual exons within the larger amplicons in a primer-walking fashion. PCR products were purified using an in-house EXOSAP protocol. Sanger sequencing was done using the Big Dye Terminator system according the manufacturer's instructions with minor modifications. Post-sequencing purification was done using ethanol precipitation and purified sequenced products were resuspended in 10µL of Hi-Di Formamide. Sequencing was performed on 3730XL ABI sequence analyser at the Medical Research Council Weatherall Institute for Molecular Medicine (University of Oxford, United Kingdom).

5.2.6 Sequence analysis

Sequence electropherograms for *TRIM5* and *TRIM22* were analysed in Geneious 2019.2 (22). Forward and reverse reads for each exon were collected and assembled into a

contiguous sequences per patient. The patients' assembled contigs were aligned to the relevant reference sequence for *TRIM5* and *TRIM22*. Sequences with HQ scores of less than 80% were not included in further analysis. Single nucleotide polymorphisms were identified at a 35% heterozygosity threshold and identified in relation to the dbSNP database administered by the NCBI. Linkage statistics were calculated in Haploview and linkage disequilibrium (LD) maps generated (23). Measurements of heterozygosity and testing of Hardy-Weinberg Equilibrium (HWE) were done in Haploview. Combined Annotation Dependent Depletion (CADD) scores and population level genotype frequency data were extracted from the Ensembl Project (24). eQTL statistics of gene expression for *TRIM22* in human tissue was extracted from the Ensembl Project (24).

5.2.7 Statistical analysis

Statistical analysis was done in R Studio, version 3.4.0 (25,26). Pairwise comparisons of continuous numeric variables were done using the Students' T test and Mann Whitney-U test for parametric and non-parametric distributions respectively. Multiple group comparisons were done using a one-way ANOVA and Kruskall-Wallis test for parametric and non-parametric distributions respectively. Testing of differences in genotype frequencies between groups was done using chi-squared (X^2) tests. Where appropriate, corrections for multiple comparisons was done using the Holm-Bonferroni method. Pirate plots were prepared using the YARRR package and other figures were prepared using the R package ggpubr (27,28). Uni and multivariate survival analysis was done in R using the survival and survminer packages (29,30). Longitudinal CD4% data was analysed using full-factorial, linear mixed effects models in SPSS (31). Statistical significance was assessed at $P < 0.05$.

5.3 Results

A total of 125 patients from the Police cohort were included, 97 males and 28 females (**table 5.1**). Mean age at enrolment was 36.78 years. Longitudinal CD4% data was available for 99 patients, these patients were stratified into progressor groups as previously described.

Table 5.1: Police cohort summary statistics for genotyped patients

	Number (%)	**Missing (%)**	**Mean/ median**	**IQR**
HIV-2 infected patients	125 (100)			
Male	97 (78)	0		
Female	28 (22)			
Age	125 (100)	0	36.78	29 - 43
Fast progressor [1]	43 (34)	26 (21)		
Slow progressor [1]	56 (45)			
Midpoint CD4% [2]	99 (79)	26 (21)	28.15	21.6 - 33.2
CD4% change rate per year [2]	99 (79)	26 (21)	-0.33	-0.92 to +0.22
Median time to AIDS (years) [3]	43 (34)	57 (46)	16.5	

[1] Progressor group was determined by the combined coefficient. [2] These statistics were extracted from per-patient regression models described in experimental chapter two. [3] Estimates from Kaplan-Meier curves of time to AIDS and survival time for this sample of the Guinea-Bissau Police cohort.

5.3.1 Summary of genotype data for *TRIM5* and *TRIM22*

TRIM5 and *TRIM22* are highly polymorphic, with SNPs identified in each exon. We identified 17 SNPs of which 10 were nonsynonymous substitutions (**table 5.2**). For genotype association analysis we included SNPs which had a minority genotype frequency of greater than 5% in our sample. The nonsynonymous substitutions which we included had low CADD scores (less than 20) and were then unlikely to be pathological (**table 5.2**). Genotype frequencies across selected African populations for non-synonymous SNPs which have been linked to HIV-1 disease outcomes did not vary much across population groups, apart from rs1063303 (**figure 5.1**). rs11038628 in *TRIM5* and rs78484876 in

TRIM22 were the only SNPs in linkage disequilibrium (LD) between *TRIM5* and *TRIM22*. Several SNPs within TRIM5 and TRIM22 are in LD when using D' scores (**figure 5.2**). rs1063303 was not in HWE - observed heterozygosity of 0.33 and predicted heterozygosity 0.49 (HWE $P = 3.2x10^{-5}$). D' measurements can be biased by low allele frequency resulting in inflated linkage statistics between rare SNPs. Only two SNPs were in LD based on their correlation coefficient which corrects for allele frequency – rs10838543 and rs7935564 in *TRIM22* ($r^2 = 0.6$).

Table 5.2: Summary of genotype data for *TRIM5* and *TRIM22*

***TRIM5* exon**	**dbSNP ID** [1]	**CADD Score** [2]	**AA change** [3]	**Genotype frequencies in the Police cohort**		
2	rs59896509	24	G31S	GG: 0.98	GA: 0.02	AA: 0.01
2	rs3740996	13	H43Y	CC: 0.84	CT: 0.15	TT: 0.01
2	rs10838525	12	R136Q	GG: 0.77	GA: 0.20	AA: 0.03
3	**rs3740995**			GG: 0.46	GA: 0.44	AA: 0.10
4	rs58187248			AA: 0.77	AG: 0.20	AA: 0.03
5	**rs11038628**	4	G249D	GG: 0.50	GA: 0.42	AA: 0.08
8	rs7104422	21	P479L	CC: 0.90	CT: 0.08	TT: 0.02
***TRIM22* exon**	**dbSNP ID**	**CADD score**	**AA change**	**Genotype frequencies in the Police cohort**		
2	rs200668710	15	E83K	GG: 0.95	GA: 0.05	AA: 0
2	**rs10838543**			TT: 0.45	TC: 0.39	CC: 0.15
2	rs61735271			CC: 0.95	CT: 0.05	CC: 0
3	**rs7935564**	14	D155N	GG: 0.34	GA: 0.42	AA: 0.24
4	rs78484876			TT: 0.84	TC: 0.12	CC: 0.04
4	**rs2291842**			TT: 0.58	CT: 0.30	CC: 0.12
4	**rs1063303**	13	R238T	GG: 0.40	GC: 0.33	CC: 0.27
4	rs61735273	22	S244L	CC: 0.85	CT: 0.12	TT: 0.03
7	rs73404240	4	T294K	CC: 0.79	CA: 0.20	AA: 0.01
8	rs140700337			GG: 0.97	GA: 0.02	AA: 0.01

[1] rs number refers to Reference SNP Cluster ID and is used as an accession number on the dbSNP database which is administered by the NCBI. SNPs used for association studies are shown in **bold text.** [2] Combined Annotation Dependent Depletion (CADD) scores are a continuous predictor of the likelihood that a SNP is deleterious. CADD scores < 30 indicate that a SNP is likely to be benign. [3]AA: Amino acid.

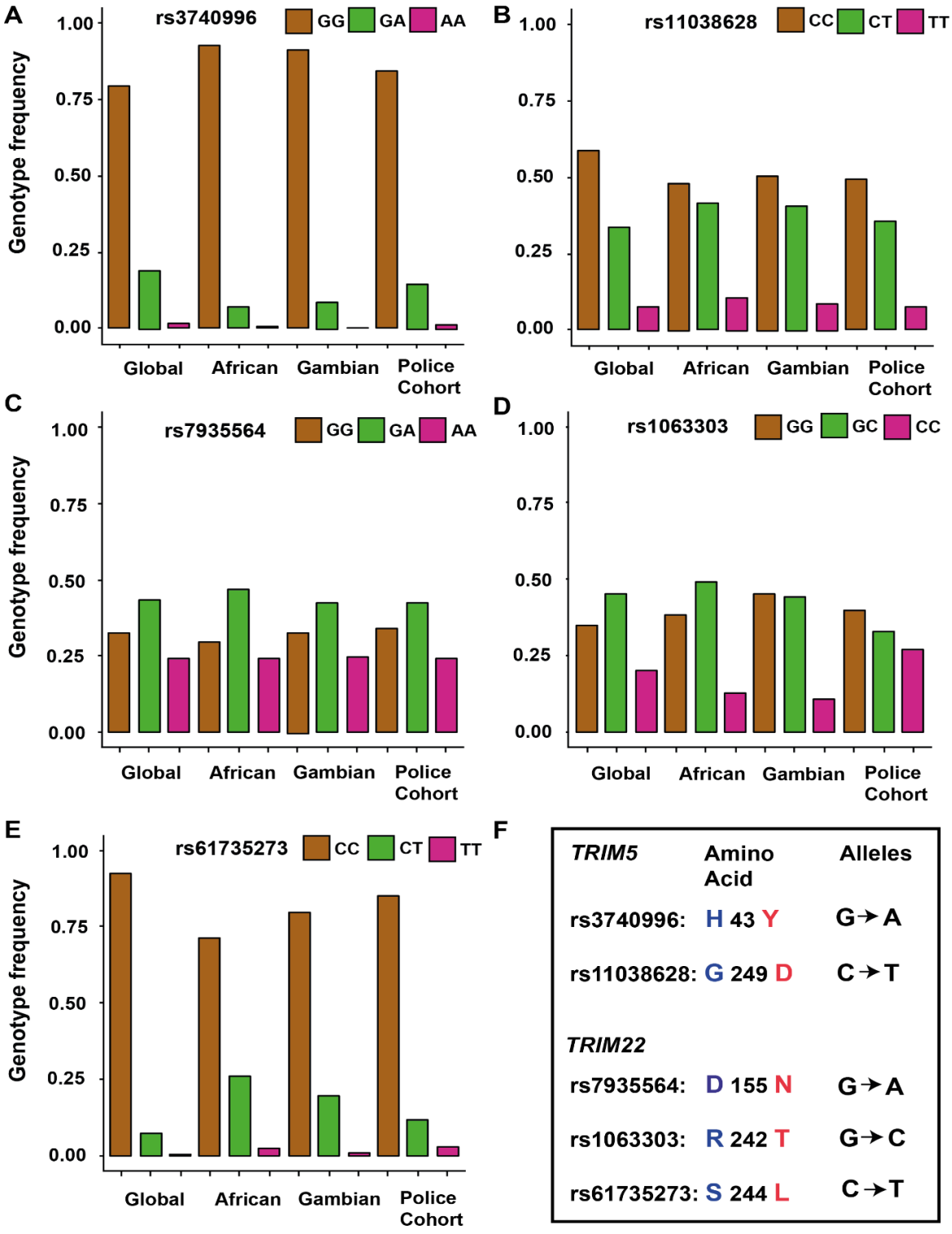


Figure 5.1: Genotype frequencies of non-synonymous variants in *TRIM5* and *TRIM22*. Genotype frequencies (y-axis) for Global, African and Gambian West Division categories (x-axis) were compared against Police genotype frequencies for non-synonymous SNPs in *TRIM5* (**panels A and B**) and *TRIM22* (**panels C – E**). Panel F shows the dbSNP reference sequence numbers with the associated amino acid substitutions, blue indicates the reference amino acid and red the alternative.

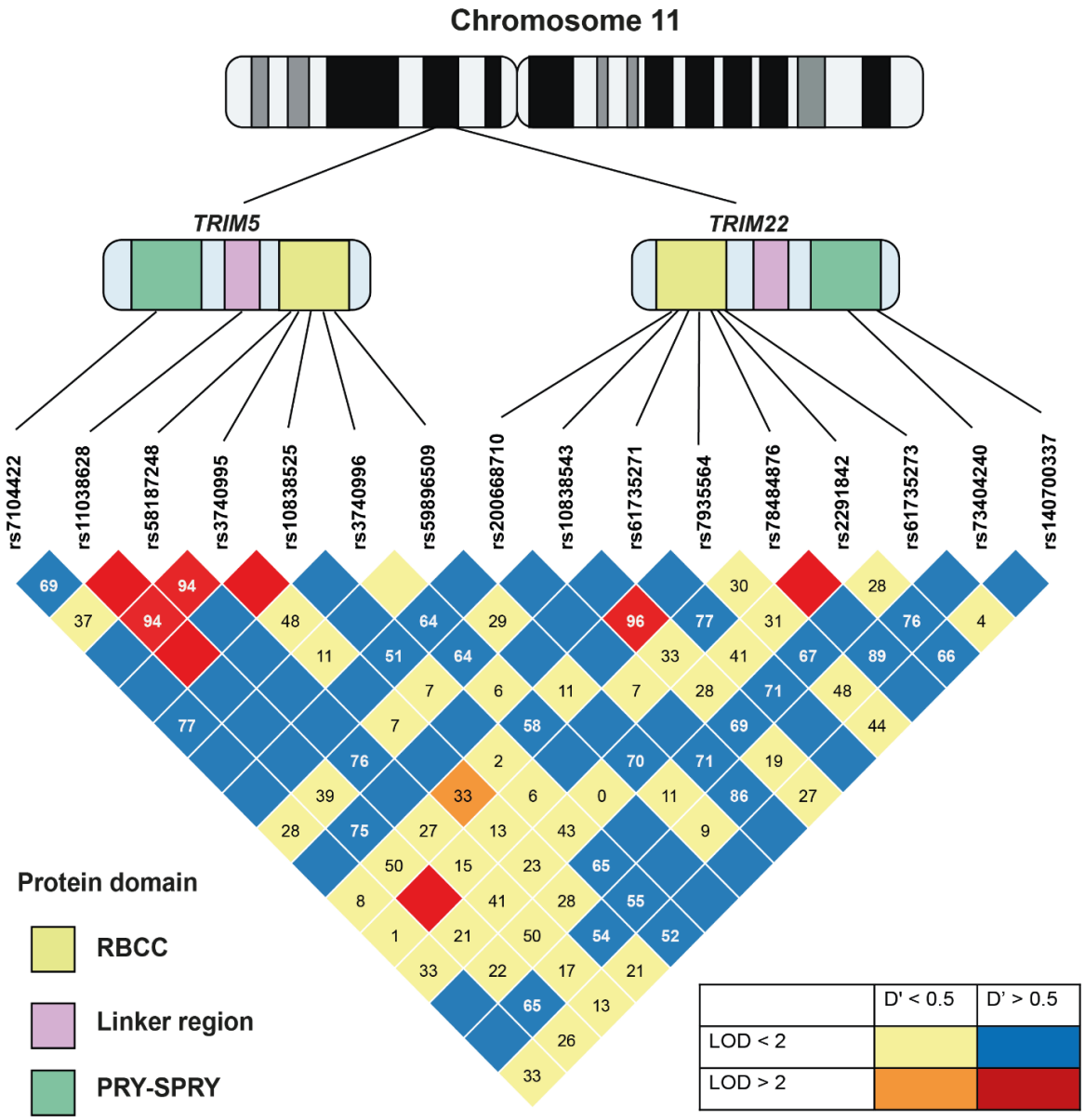


Figure 5.2: Linkage map of variants in *TRIM5* and *TRIM22*. *TRIM5* and *TRIM22* are in band 12 of the short arm of chromosome 11. Protein coding domains key are shown in the bottom left corner. In the linkage map, numbers in blocks denote D′ x 100 with D' values = 100 not shown. Representations of *TRIM5* and *TRIM22* are shown with their respective SNPs mapped to their coding regions. Bottom right key for interpreting LD scores using D': D' scores range from 0 (no linkage) to 1 (complete linkage), Logarithm (base 10) of odds (LOD) scores indicates the likelihood that a pair of SNPs are in linkage, LOD of 3 = 1000:1 odds that observed linkage occurred by chance. Red blocks indicate strong LD with high likelihood.

5.3.2 Association of *TRIM* genotype with midpoint CD4% and CD4% change rate

In univariate analyses *TRIM5* and *TRIM22* genotype did not have significant effects on midpoint CD4% and CD4% change rate (**figure 5.3 and figure 5.4**).

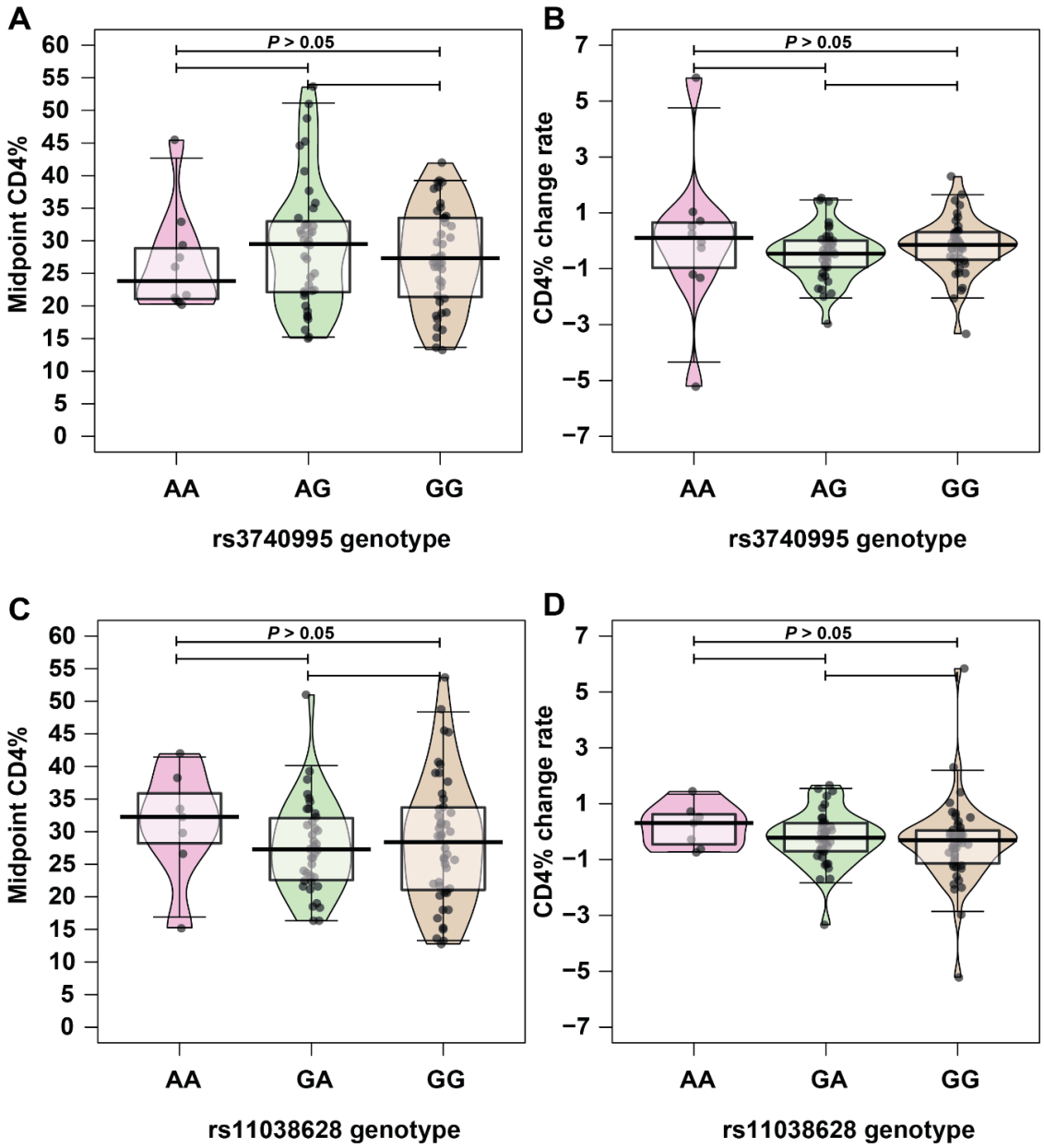


Figure 5.3: Comparison of midpoint CD4% and CD4% change rate by *TRIM5* genotype. Pairwise comparisons of midpoint CD4% and CD4 decline rate by rs3740995 genotype are shown in **panels A and B** and by rs11038628 genotype are shown in **panels C and D**. Midpoint CD4% comparisons were made using Students T test and CD4 decline rate with a Mann-Whitney U test. Continuous variables are shown on the y-axes and genotype groups on the x-axes.

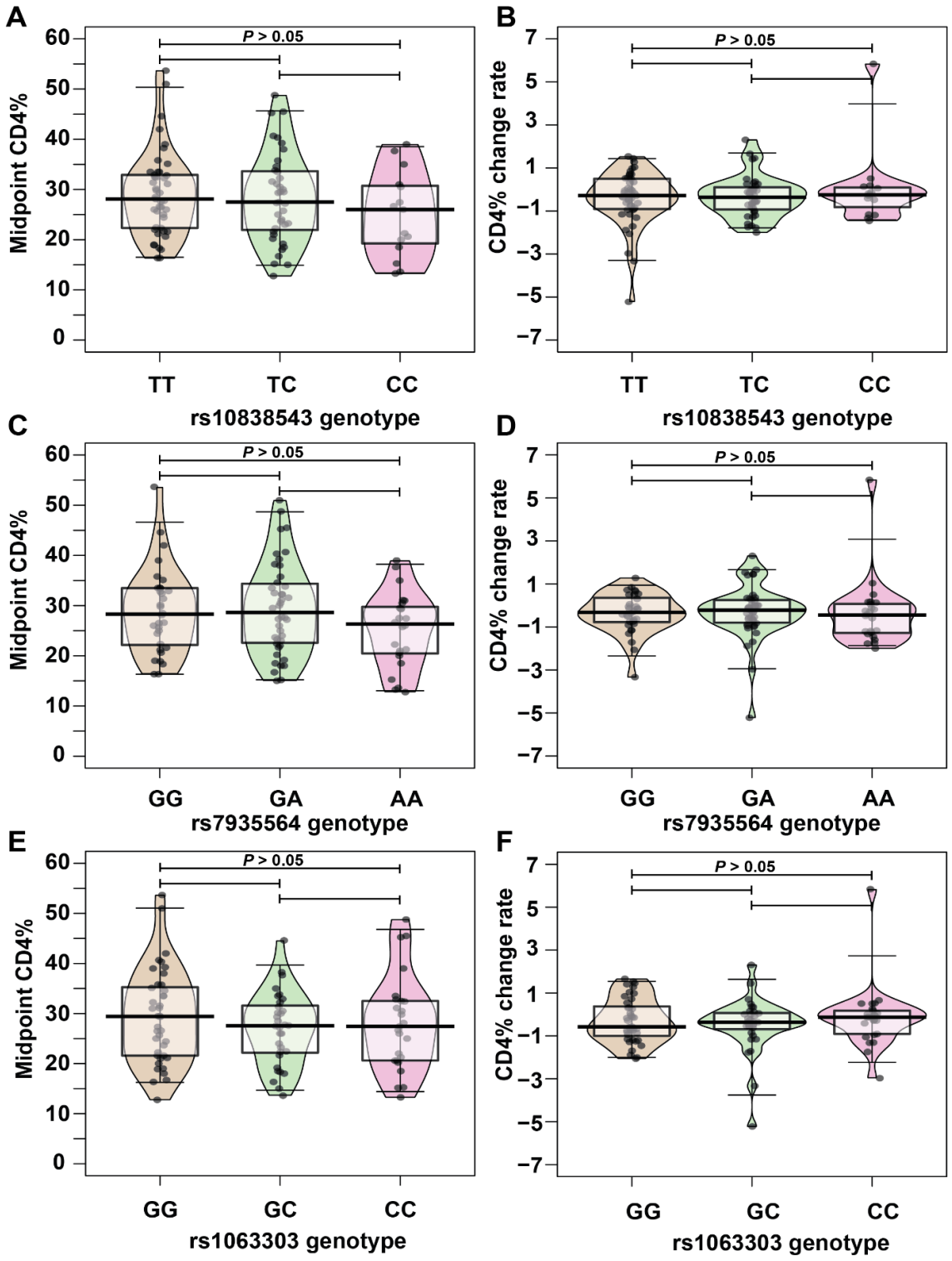


Figure 5.4: Comparison of midpoint CD4% and CD4% change rate by *TRIM22* genotype. Pairwise comparisons of midpoint CD4% and CD4% change rate by rs10838543 genotype are shown in **panels A and B**, by rs7935564 genotype in **panels C and D** and rs1063303 in **panels E and F** - plots for rs2291842 are not shown. Midpoint CD4% comparisons were made using Students T test and decline rate with a Mann-Whitney U test. Continuous variables are shown on the y-axes and genotype groups on the x-axes

5.3.3 Change in longitudinal CD4% by *TRIM* genotype

Ninety-nine patients contributed 511 CD4% data points. We first analysed patients as random intercepts with years as repeated measurements, then alternated random slopes and intercepts in iterative models. Model fit was assessed using AIC and BIC values (**table S5.1**). A random intercept only model provided the best fit for the data. Being female was consistently associated with a higher CD4% (female vs male increase of 4.49 CD4%, 95% CI: 0.41 - 8.57. $P = 0.03$). The CD4% change rate was -0.36/year (SE = 0.06, $P = 7.40 \times 10^{-10}$). Age at enrolment was not a significant determinant of CD4% during follow-up (age $P = 0.24$) and was excluded in later models. *TRIM* SNPs were then tested individually with an adjustment for sex (**table S5.1**). *TRIM5* genotype did not have a significant effect on longitudinal CD4%. In a multivariate analysis *TRIM22* rs7935564(AA) genotype was associated with a lower CD4% when compared to (GG) homozygotes (AA vs GG genotype = -4.53 CD4%, 95% CI: -9.01 to -0.04. $P = 0.048$). *TRIM22* rs10838543 genotype was associated with a larger difference in CD4% (CC vs TT/TC = -5.15 CD4%, 95% CI: -10.14 to -0.16, $P = 0.043$, **figure 5.5**). The effects of *TRIM22* were not significant when corrected for multiple comparisons across models.

5.3.4 Time to AIDS by *TRIM* genotype

We next investigated whether time to AIDS varied by *TRIM* genotype. Univariate associations of *TRIM* SNPs were not statistically significance. Cox proportional hazard models were used to evaluate whether *TRIM*22 rs7935564 genotype contributed to time to AIDS when adjusted for additional clinical factors. Age at enrolment did not influence time to AIDS in a multivariate analysis and was not included in the next models which tested whether *TRIM* genotype was associated with differences in time to AIDS when controlling

for sex. Male sex was associated with a significant and substantial increase in risk of developing AIDS during follow-up. The addition of *TRIM22* genotype improved model fit (**table S5.2**). rs10838543 did not influence time to AIDS nor risk of developing AIDS (X^2 P = 0.86, **figure 5.6**). rs7935564 AA genotype associated with an increased hazard ratio to develop AIDS when compared to GG homozygotes and heterozygotes, though this was not statistically significant (AA genotype adjusted HR = 2.2, 95% CI 0.99 – 4.84. P = 0.05, **table S5.3**). In addition, there were more AIDS cases in this group than either the GG or GA genotype groups, though this was not statistically significant (X^2 P = 0.06).

Lastly, we compared genotype frequencies for the six included *TRIM* SNPs across progressor groups. Differences in genotype frequency were small between fast and slow progressor groups and were not statistically significant after correction for multiple comparisons (X^2 for each SNP category $P > 0.05$, **figure 5.7**).

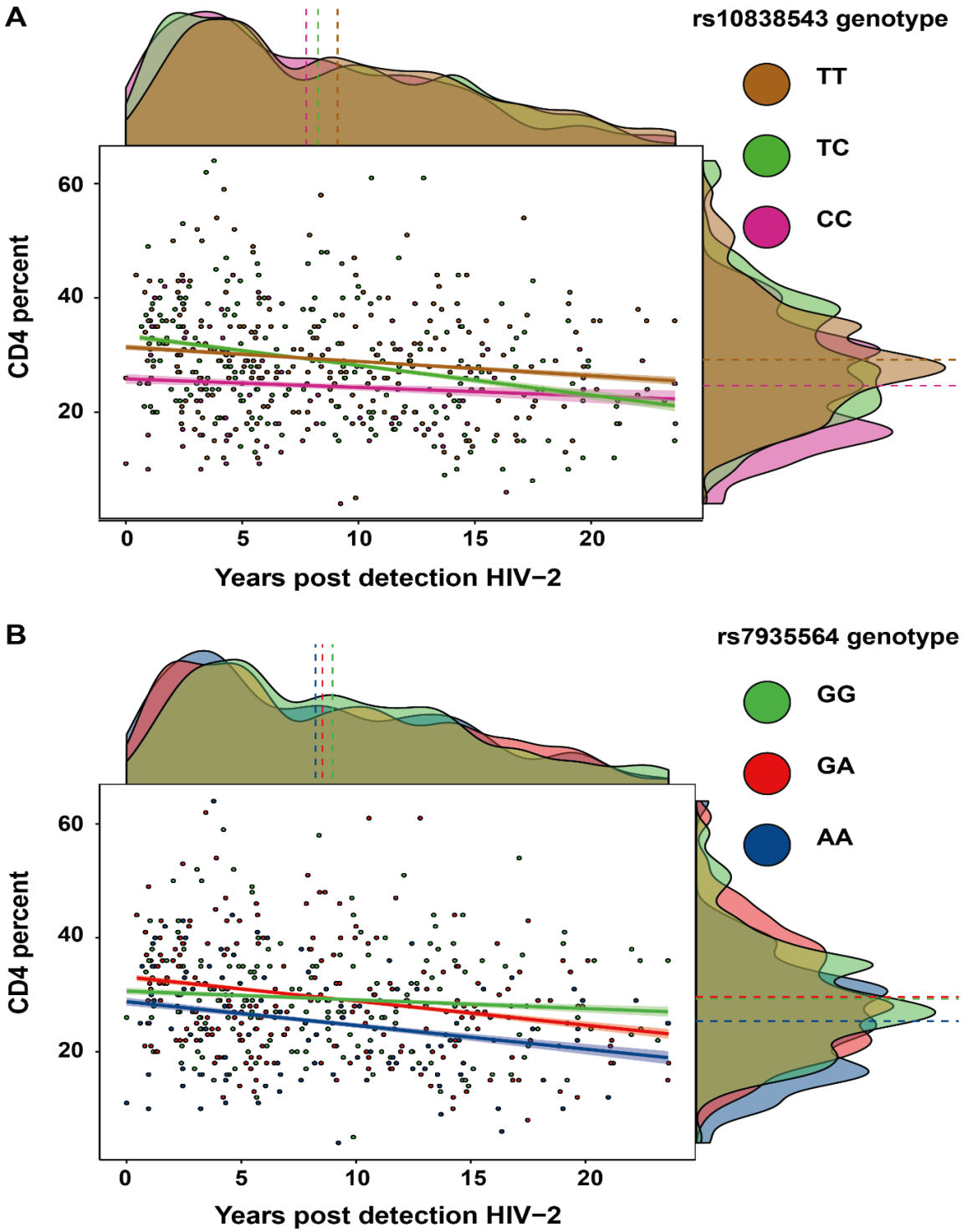


Figure 5.5: CD4% decline from HIV-2 detection, compared across genotypes for *TRIM22* rs7935564 and rs10838543. A bootstrapped sample of the CD4% datapoints was generated (1000 random samples with replacement). Linear regression lines were used to visualise CD4% change over time grouped by *TRIM22* genotype at rs7935564 and rs10838543, shown in **panels A and B** respectively. Marginal density plots indicate the distributions of the CD4% (y-axes) and years post HIV-2 detection (x-axes). Mean values for years post detection of HIV-2 by genotype are shown by the dashed lines on the x-axis and for CD4% by genotype on the y-axis.

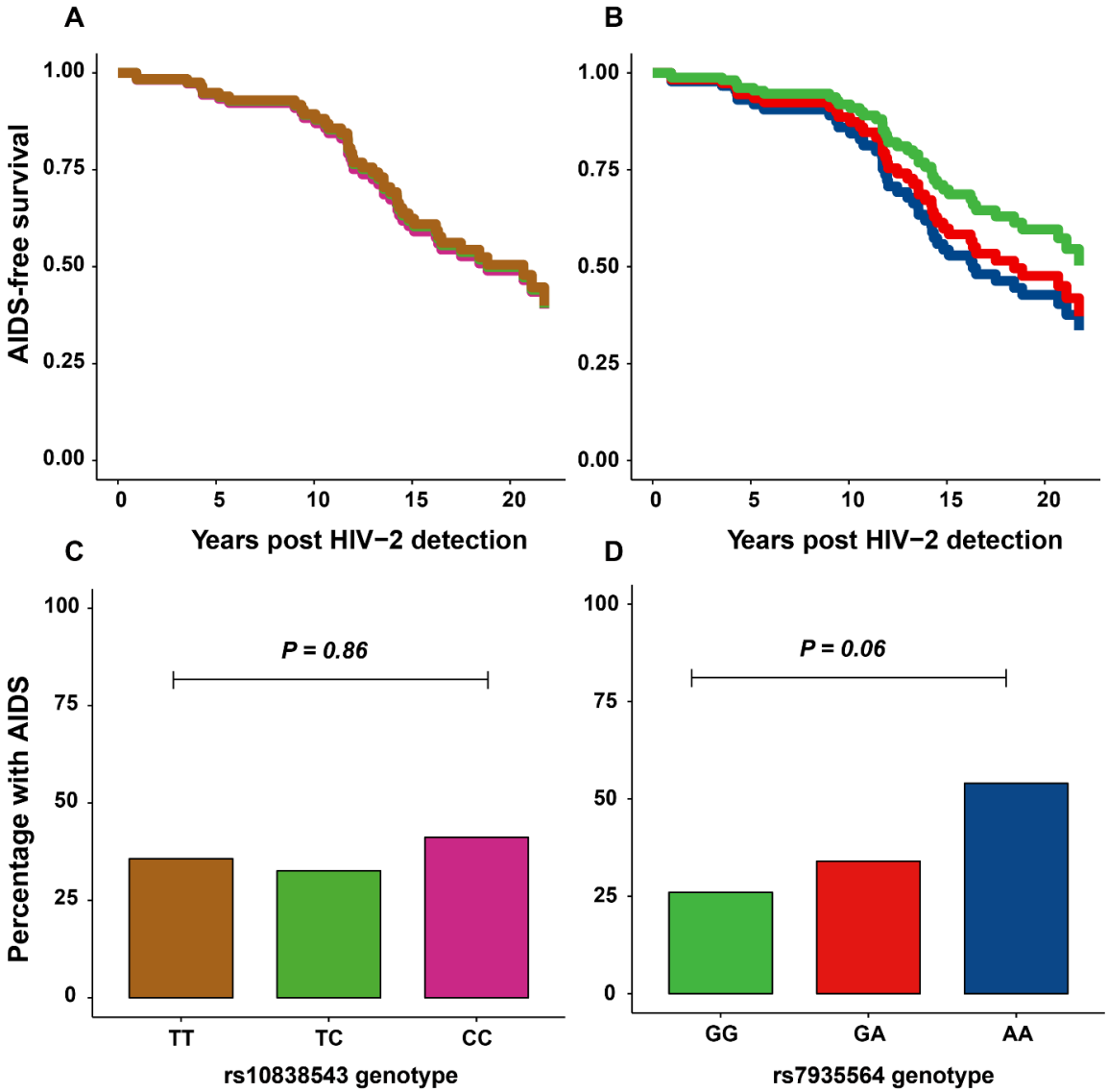


Figure 5.6: Time to event for AIDS by *TRIM22* genotype. Cox adjusted survival curve for time to AIDS by rs10838543 (**A**) and rs7935564 (**B**) genotype. Percentage of patients who developed AIDS by rs10838543 (**C**) and r7935564 (**D**) genotypes, *P* values are shown for X^2 tests comparing frequency of AIDS diagnoses by genotype. Survival curves are coloured for the genotypes shown in panels **C** and **D.**

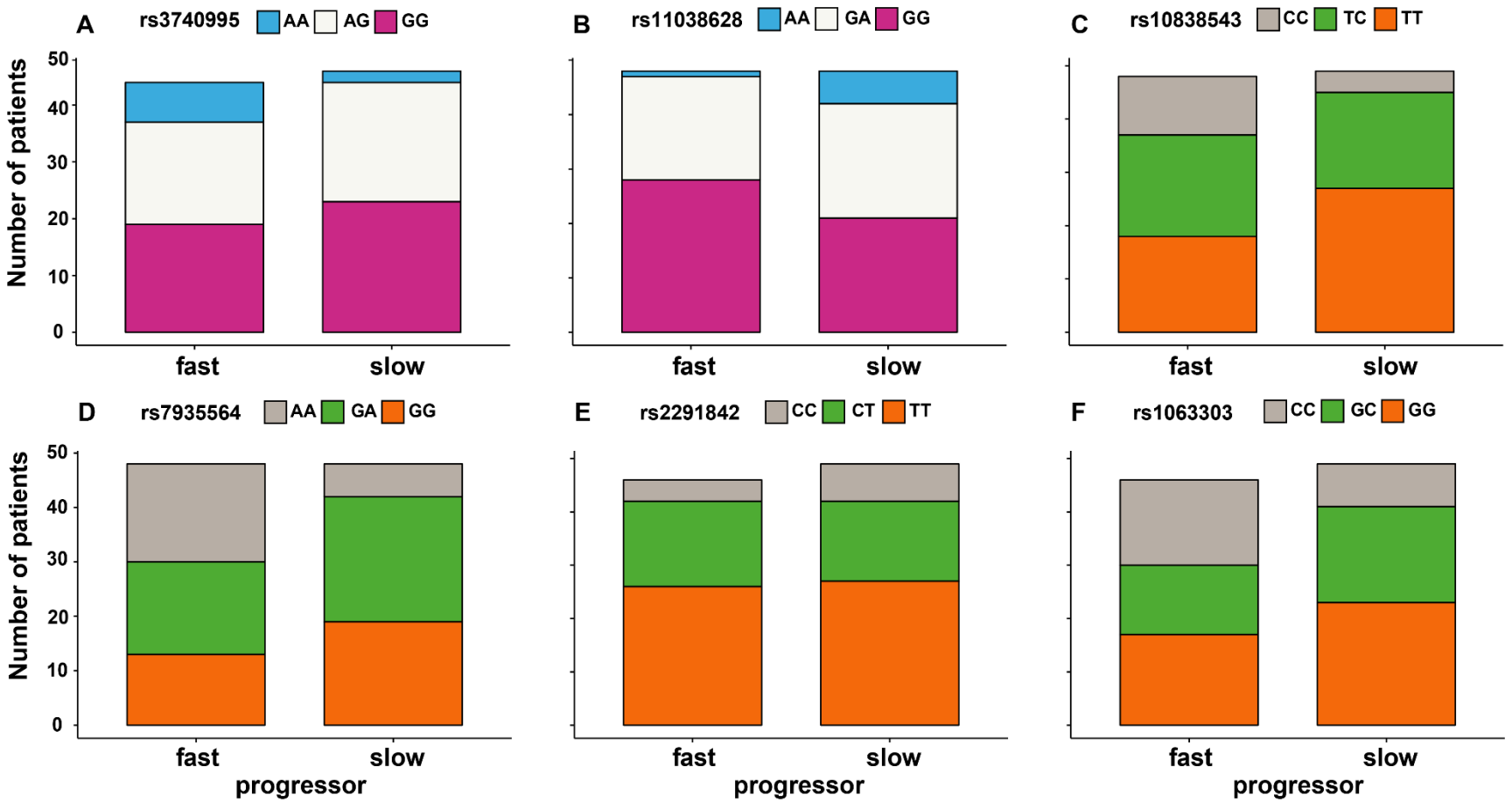


Figure 5.7: Genotype frequencies for *TRIM5* and *TRIM22* by disease progressor group. Panels **A and B** show *TRIM5* SNPs and panels **C – F** show *TRIM22* SNPs. Stacked bar plots show number of patients with specific *TRIM5* and *TRIM22* genotype frequencies clustered by progressor group as determined by the combined coefficient. Numbers of patients by genotype are shown on the y-axis, x-axis shows progressor groups.

5.4 Discussion

We successfully sequenced *TRIM5* and *TRIM22*'s coding regions in a group of HIV-2 infected individuals with long term follow-up. We identified 17 SNPs, 10 of which resulted in non-synonymous substitutions - the majority which were unlikely to be pathological. Indeed, when these nonsynonymous SNPs population level frequency was compared against global, African and a geographically close population in Gambia only minor differences were noted in the Police cohort. *TRIM22* rs1063303 is at a site which has evolved under positive selection in mammalian evolution and has been shown to vary by up to 10-fold between ethnicities (32). It was therefore unsurprising that this SNP showed the greatest variation in genotype frequencies across the selected population groups we analysed. The PRY-SPRY domain V1 loop in *TRIM5* was conserved throughout the cohort. This implies that TRIM5α specificity for retroviral capsids was unlikely to vary between patients – which supports the finding that HIV-2 capsids from rapid and long term non-progressors have similar sensitivity to TRIM5α (4).

Polymorphisms in *TRIM5* which are associated with impaired retroviral restriction (e.g. rs3740996) had a low genotype frequency which made analysis of disease outcomes difficult. We did not observe a significant impact of the rs11038628 substitution on a range of disease outcomes, which was surprising as the published literature indicates that this substitution negatively affects higher order assembly of cytoplasmic TRIM5α bodies (33,34). However, the HIV-2 restrictive activity of TRIM5α with the G249D substitution is largely unaffected *in vitro* (33). The G249D substitution has also been associated with increased risk of HIV-1 infection in the Indian population and this may reflect the differences in restrictive capacity of TRIM5α in different cellular compartments (33,35). The ability of TRIM5α to restrict HIV-1 replication is enhanced in Langerhans cells which

line the vaginal and penile mucosa. The mechanism involves HIV-1 binding Langerin (a C-type lectin receptor) on entry to the cytoplasm and this triggers an autophagy-based pathway for HIV-1 degradation (35). Therefore, SNPs in *TRIM5* may have strong effects on risk of HIV-1 infection but not disease progression afterwards. We were not able to test the hypothesis that *TRIM* genotype affects risk of HIV-2 infection.

rs7935564 and rs1063303 which code for the D155N and R242T substitutions in *TRIM22* have been linked to HIV-1 and Hepatitis C disease outcomes, often with contradictory results (16–18). Patients with the rs7935564 AA genotype had lower CD4% in follow-up relative to GG homozygotes and heterozygotes. Previous reports indicate that rs7935564 and rs1063303 GG homozygosity were more frequent in HIV-1 infected patients who developed AIDS versus long term non-progressors (17). Given that these SNPs are predicted to have little impact on the proteins function based on multiple scoring systems available on the Ensembl database, it is unclear why this mutation would influence disease outcomes or viral replication (24). rs1063303 CC homozygosity, *in vitro*, has been shown to simultaneously increase TRIM22 expression in immortalised cell lines while decreasing its restrictive capacity against HIV-1 – which would conflict with those of Singh *et al* (32). The retroviral "target" of TRIM22, the LTR region where the Sp1 transcription factor binds, varies by up to 50% between HIV-1 and HIV-2 LTR (36). The HIV-1 LTR has a stronger response to transcription factors and tumor necrosis factor alpha than HIV-2 (36,37). TRIM22 may therefore have different effects on HIV-2 replication, and studies assessing HIV-1 disease outcomes may not extrapolate to HIV-2.

We have shown that rs7935564 shows strong evidence of linkage with another *TRIM22* SNP, rs10838543. Similar linkage results have been observed in a cohort of Zimbabwean HIV-1 infected children and adolescents (shown in chapter three). rs10838543 is a

synonymous substitution located in a transcription factor binding site in exon two of *TRIM22* and is expected to influence downstream regulation of *TRIM22* expression (24). The CC genotype was associated with lower CD4% in longitudinal follow-up in a multivariate analysis which accounted for sex. This may be due to reduced RNA expression of *TRIM22*. The CD4% change rate did not vary based on genotype at this SNP. This may indicate that events early in infection alter the viral set point and CD4% nadir. *TRIM22* RNA expression correlates negatively with HIV-1 viral load in acute infection, and this may support our finding (20). Cox proportional hazard models showed that rs7935564, but not rs10838543, may be associated with an increased risk of developing AIDS during long term follow-up. These results had weak statistical support and probably represent chance variation.

Furthermore, TRIM22 has been shown to be important in regulating nuclear factor kappa-light-chain-enhancer of activated B cells (NFKB) signalling via nucleotide-binding oligomerization domain (NOD) activation (38). rs61735273 has been associated with reduced NFKB activation by disruption of NOD signalling, resulting in an increased risk of early inflammatory bowel syndrome. Changes in TRIM22 function might therefore also influence HIV disease outcomes by a) reducing viral transcription via NFKB stimulation in the LTR region b) reducing the stimulation of a downstream interferon response which might limit further viral spread and infection of new cells. rs61735273 did not have any significant effects on HIV-2 disease outcomes in the Police cohort, though this is probably the result of an underpowered analysis as the minority genotype had a low frequency in this cohort.

5.5 Section summary

1. *TRIM5* and *TRIM22* have high genetic diversity in this population and there was no evidence of linkage between the two genes

2. *TRIM5* genotype did not explain variation in HIV-2 disease outcomes

3. Single nucleotide polymorphisms in *TRIM22* accounted for variation in longitudinal CD4% kinetics – though these effects were modest, and not statistically significant if corrected for multiple comparisons.

4. rs7935564 may influence disease progression rates, measured in time to AIDS.

5. *TRIM5* and *TRIM22* genotype did not differ significantly between fast and slow disease progressor groups.

5.6 References for chapter 5

1. Naranbhai V, Carrington M. Host genetic variation and HIV disease: from mapping to mechanism. Immunogenetics. 2017 Aug 10;69(8–9):489–98.

2. Diouf K, Sarr AD, Eisen G, Popper S, Mboup S, Kanki P. Associations between MHC class I and susceptibility to HIV-2 disease progression. J Hum Virol. 5(1):1–7.

3. Thomsen D, Erikstrup C, Jespersen S, Medina C, Té D da S, Correira FG, et al. The influence of human leukocyte antigen-types on disease progression among HIV-2 infected patients in Guinea-Bissau. AIDS. 2018 Mar 27;32(6):721–8.

4. Mamede JI, Damond F, Bernardo A de, Matheron S, Descamps D, Battini J-L, et al. Cyclophilins and nucleoporins are required for infection mediated by capsids from circulating HIV-2 primary isolates. Sci Rep. 2017 Mar 27;7:45214.

5. Ylinen LMJ, Keckesova Z, Wilson SJ, Ranasinghe S, Towers GJ. Differential Restriction of Human Immunodeficiency Virus Type 2 and Simian Immunodeficiency Virus SIVmac by TRIM5 Alleles. J Virol. 2005 Sep 15;79(18):11580–7.

6. Onyango CO, Leligdowicz A, Yokoyama M, Sato H, Song H, Nakayama EE, et al. HIV-2 capsids distinguish high and low virus load patients in a West African community cohort. Vaccine. 2010 May;28:B60–7.

7. Jallow S, Leligdowicz A, Kramer HB, Onyango C, Cotten M, Wright C, et al. The presence of prolines in the flanking region of an immunodominant HIV-2 gag epitope influences the quality and quantity of the epitope generated. Eur J Immunol. 2015;45(8):2232–42.

8. Leligdowicz A, Yindom L-M, Onyango C, Sarge-Njie R, Alabi A, Cotten M, et al. Robust Gag-specific T cell responses characterize viremia control in HIV-2 infection. J Clin Invest. 2007 Oct 1;117(10):3067–74.

9. Song H, Nakayama EE, Yokoyama M, Sato H, Levy JA, Shioda T. A Single Amino Acid of the Human Immunodeficiency Virus Type 2 Capsid Affects Its Replication in the Presence of Cynomolgus Monkey and Human TRIM5 s. J Virol. 2007 Jul 1;81(13):7280–5.

10. Miyamoto T, Yokoyama M, Kono K, Shioda T, Sato H, Nakayama EE. A Single Amino Acid of Human Immunodeficiency Virus Type 2 Capsid Protein Affects Conformation of Two External Loops and Viral Sensitivity to TRIM5α. Lee Y-M, editor. PLoS One. 2011 Jul 28;6(7):e22779.

11. Takeuchi JS, Perche B, Migraine J, Mercier-Delarue S, Ponscarme D, Simon F, et al. High level of susceptibility to human TRIM5α conferred by HIV-2 capsid sequences. Retrovirology. 2013;10(1):50.

12. Goldschmidt V, Bleiber G, May M, Martinez R, Ortiz M, Telenti A, et al. Role of common human TRIM5alpha variants in HIV-1 disease progression. Retrovirology. 2006 Aug 22;3:54.

13. Rahm N, Gfeller D, Snoeck J, Martinez R, McLaren PJ, Ortiz M, et al. Susceptibility and adaptation to human TRIM5α alleles at positive selected sites in HIV-1 capsid. Virology. 2013 Jul 5;441(2):162–70.

14. van Manen D, Rits MAN, Beugeling C, van Dort K, Schuitemaker H, Kootstra NA, et al. The Effect of Trim5 Polymorphisms on the Clinical Course of HIV-1 Infection. PLoS Pathog. 2008;4(2):e18.

15. Kootstra NA, Navis M, Beugeling C, van Dort KA, Schuitemaker H. The presence of the Trim5α escape mutation H87Q in the capsid of late stage HIV-1 variants is preceded by a prolonged asymptomatic infection phase. AIDS. 2007 Oct 1;21(15):2015–23.

16. Moura Rodrigues R, Plana M, Garcia F, Zupin L, Kuhn L, Crovella S. Genome-wide scan in two groups of HIV-infected patients treated with dendritic cell-based immunotherapy. Immunol Res. 2016 Dec 4;64(5–6):1207–15.

17. Ghezzi S, Galli L, Kajaste-Rudnitski A, Turrini F, Marelli S, Toniolo D, et al. Identification of TRIM22 single nucleotide polymorphisms associated with loss of inhibition of HIV-1 transcription and advanced HIV-1 disease. AIDS. 2013 Sep 24;27(15):2335–44.

18. Medrano LM, Rallón N, Berenguer J, Jiménez-Sousa MA, Soriano V, Aldámiz-Echevarria T, et al. Relationship of TRIM5 and TRIM22 polymorphisms with liver disease and HCV clearance after antiviral therapy in HIV/HCV coinfected patients. J Transl Med. 2016 Dec 2;14(1):257.

19. Forlani G, Accolla RS. Tripartite Motif 22 and Class II Transactivator Restriction Factors: Unveiling Their Concerted Action against Retroviruses. Front Immunol. 2017;8(October):1–7.

20. Singh R, Gaiha G, Werner L, McKim K, Mlisana K, Luban J, et al. Association of TRIM22 with the Type 1 Interferon Response and Viral Control during Primary HIV-1 Infection. J Virol. 2011 Jan 1;85(1):208–16.

21. van der Loeff MFS, Larke N, Kaye S, Berry N, Ariyoshi K, Alabi A, et al. Undetectable plasma viral load predicts normal survival in HIV-2-infected people in a West African village. Retrovirology. 2010 May 19;7(1):46.

22. Kearse M, Moir R, Wilson A, Stones-Havas S, Cheung M, Sturrock S, et al. Geneious Basic: An integrated and extendable desktop software platform for the organization and analysis of sequence data. Bioinformatics. 2012;

23. Barrett JC, Fry B, Maller J, Daly MJ. Haploview: analysis and visualization of LD and haplotype maps. Bioinformatics. 2005 Jan 15;21(2):263–5.

24. Zerbino DR, Achuthan P, Akanni W, Amode MR, Barrell D, Bhai J, et al. Ensembl 2018. Nucleic Acids Res. 2018 Jan 4;46(D1):D754–61.

25. R Core Team. R Language Definition V. 3.1.1. Https://WwwR-ProjectOrg/. 2014;

26. RStudio team, RStudio. RStudio: Integrated development environment for R. J Wildl Manage. 2015;

27. Kassambara A. ggpubr: “ggplot2” Based Publication Ready Plots. R Package. 2018.

28. Kampstra P. Beanplot: A Boxplot Alternative for Visual Comparison of Distributions. J Stat Softw. 2008 Nov 10;28(Code Snippet 1):1–9.

29. Therneau T, Crowson CS, Atkinson EJ. Adjusted Survival Curves. Study R. 2015;

30. Kassambara A, Kosinski M, Biecek P, Fabian S. survminer: Drawing Survival Curves using “ggplot2”. R package. version 0.4.3. 2018.

31. IBM SPSS Inc. SPSS Statistics for Windows. IBM Corp Released 2012. 2016;

32. Kelly JN, Woods MW, Xhiku S, Barr SD. Ancient and Recent Adaptive Evolution in the Antiviral *TRIM22* Gene: Identification of a Single-Nucleotide Polymorphism That Impacts TRIM22 Function. Hum Mutat. 2014 Sep;35(9):1072–81.

33. Nakayama EE, Nakajima T, Kaur G, Mimaya J, Terunuma H, Mehra N, et al. A Naturally Occurring Single Amino Acid Substitution in Human TRIM5α Linker Region Affects Its Anti-HIV Type 1 Activity and Susceptibility to HIV Type 1 Infection. AIDS Res Hum Retroviruses. 2013 Jun;29(6):919–24.

34. Kömürlü S, Bradley M, Smolin N, Imam S, Pauszek RF, Robia SL, et al. Defects in assembly explain reduced antiviral activity of the G249D polymorphism in human TRIM5α. Menéndez-Arias L, editor. PLoS One. 2019 Mar 19;14(3):e0212888.

35. Ribeiro CMS, Sarrami-Forooshani R, Setiawan LC, Zijlstra-Willems EM, van Hamme JL, Tigchelaar W, et al. Receptor usage dictates HIV-1 restriction by human TRIM5α in dendritic cell subsets. Nature. 2016 Dec 7;540(7633):448–52.

36. Tong-Starksen SE, Welsh TM, Peterlin BM. Differences in transcriptional enhancers of HIV-1 and HIV-2. Response to T cell activation signals. J Immunol. 1990 Dec 15;145(12):4348–54.

37. Hannibal MC, Markovitz DM, Clark N, Nabel GJ. Differential activation of human immunodeficiency virus type 1 and 2 transcription by specific T-cell activation signals. J Virol. 1993 Aug;67(8):5035–40.

38. Li Q, Lee CH, Peters LA, Mastropaolo LA, Thoeni C, Elkadri A, et al. Variants in TRIM22 That Affect NOD2 Signaling Are Associated With Very-Early-Onset Inflammatory Bowel Disease. Gastroenterology. 2016 May;150(5):1196–207.

Appendix 5

Linear mixed effects models were used to analyse longitudinal CD4% by *TRIM* genotype. Models which treated years of follow-up after HIV-2 detection as a random factor failed to converge. Adding sex as a fixed factor improved model fit and *TRIM* genotype improved model fit further, though with minor difference between SNPs.

Table S5.1: Comparisons of random slope/intercept models for CD4%

Model	**Fixed factors** [1]	**Random factors**	**Dependent variable**	**AIC**	**BIC**	**C** [2]
Police.mixed 1	sex	patient + years follow-up	CD4%			No
Police.mixed 2	years follow-up	patient	CD4%	3543.69	3552.16	Yes
Police.mixed 3	sex + years follow-up	patient	CD4%	3535.71	3544.17	Yes
Police.mixed 4	sex + years follow-up + rs3740995	patient	CD4%	3522.49	3530.94	Yes
Police.mixed 5	sex + years follow-up + rs11038628	patient	CD4%	3518.09	3526.54	Yes
Police.mixed 6	sex + years follow-up + rs10838543	patient	CD4%	3524.89	3533.34	Yes
Police.mixed 7	sex + years follow-up + rs7935564	patient	CD4%	3513.04	3525.48	Yes
Police.mixed 8	sex + years follow-up + rs2291842	patient	CD4%	3522.14	3530.58	Yes
Police.mixed 9	sex + years follow-up + rs1063303	patient	CD4%	3523.27	3531.72	Yes

[1] Years of follow-up was calculated from time of HIV-2 detection, this did not differentiate between seroprevalent and seroconverter patients. AIC = Akaike's Information Criterion. BIC = Schwarz's Bayesian Criterion. The smaller the information criteria the better the model fit. [2] C = Converge.

TRIM did not have a significant effect on time to AIDS in univariate survival analysis. We tested whether *TRIM22* SNPs which had significant effects on CD4% kinetics would also have effects on time to AIDS. Age at enrolment had no effect on time to AIDS, whereas being male was associated with a significantly increased risk of AIDS. rs10838543 had no discernible effect on time to AIDS in a multivariate analysis which adjusted for sex. rs7935564 improved model fit and was associated with an increased risk of AIDS in follow-up (**tables S5.2 and S5.3**).

Table S5.2: Cox proportional hazard model summary

Model name	**Factors**	**Outcome**	**r^2**	**n**
Model 1	age + sex	time to AIDS	0.07	119
Model 2	rs7935564 + sex	time to AIDS	0.10	117

[1] model fit was assessed using r^2, increasing values indicate better model fit. Time to AIDS was calculated from HIV-2 detection. n = number of patients in a model.

Table S5.3: Cox proportional hazard model 2 output

Factor	**Category**	**Hazard Ratio**	**95% CI**	***P***
Sex	Female	1.00		0.02
	Male	4.24	1.30 - 13.80	
rs7935564	GG	1.00		
	GA	1.33	0.62 – 2.84	
	AA	2.20	0.99 – 4.84	0.05

Chapter 6

Intrahost evolutionary rates of HIV-2 *gag* correlate with progression to AIDS

6.1 Introduction

An extensive body of literature has connected intrahost evolutionary processes of HIV-1 to disease outcomes (1). Measuring evolutionary rates, synonymous and nonsynonymous nucleotide substitution rate, and linking them to host immune responses and disease outcomes can further our understanding of how shifting viral population dynamics shape disease outcomes. Synonymous substitution rates in HIV-1 *env* correlate negatively with $CD4^+$ T cell change over time and positively (albeit weakly) with viral load and this is thought to reflect greater viral replication in the presence of heightened immune activation (2). HIV-1 viral replicative capacity (RC) of founder viruses has been linked to significant variation in T cell activation, proliferation, and $CD8^+$ T cell exhaustion (3). The association of RC with disease progression is independent of host HLA type or set point viral load (4,5).

In HIV-2 infection the relationship between immune activation and progressive $CD4^+$ T cell depletion is similar to that in HIV-1 at equivalent degrees of immune suppression (6). HIV-2 evolutionary rates have been compared to HIV-1 infected individuals with conflicting results (7,8). One plausible reason for the discrepancy in results is that the two cohort profiles which were used to measure HIV-2 evolutionary rates were quite different. The Senegalese cohort was ART-naïve, with limited disease progression, and had a long

follow-up time. Skar *et al* reported on *env* divergence in patients with mostly advanced HIV-2 infection. Many of these patients had high viral loads and low $CD4^+$ T cell counts. These results suggest that, as for HIV-1, differences in viral evolutionary rates explain variation in disease progression of HIV-2. A recent study showed that HIV-2 *env* evolutionary rates in fast disease progressors are approximately double that of slow progressors (9). Furthermore, evidence of positive selection on conserved positions in *env* was associated with slower disease progression. This study used the same population which we have analysed in this chapter with similar methods for disease progression stratification.

Polymorphisms in HIV-2 *gag* sequences have been implicated in multiple aspects of disease progression. The p24 region of *gag* differs between patients with low and high HIV-2 viral loads, a key determinant of immune activation and disease progression. Specifically, patients with low viral loads often have prolines in p24 positions 119, 159 and 178 (10). Prolines at position 159 and 178 are associated with enhanced proteasomal processing of a $CD8^+$ T cell epitope 165-DRFYKSLRA-173. This epitope is associated with greater Gag-specific T cell responses as determined by ELISpot and is a frequent target for T cell responses in patients with low viral loads (11). Therefore, it is possible that p24 sequences directly influence adaptive immune responses. Proline at position 119 in p24 alters the conformation of the capsid structure and is associated with increased sensitivity to TRIM5α (12,13). HIV-1 NL4-3 chimeras which contain HIV-2 *gag* sequences are more susceptible to restriction by human TRIM5α than the wildtype virus (14). HIV-2 p24 sequences derived from patients with advanced AIDS and long term non-progressors do not differ in their sensitivity to TRIM5α – though they are all more sensitive than HIV-1 (15).

There are no reported studies which have determined if intrahost evolution of HIV-2 *gag* is associated with disease progression. The HIV capsid plays a critical role in the replication

cycle and several indispensable functions including nucleotide supply to the replicating viruses post-entry, nuclear import, cyclophilin binding and host sensing via the cyclic-GMP-AMP synthase-Stimulator of Interferon Genes (CGAS-STING) pathway have been mapped to specific amino acid motifs (16–19). It is unclear whether variation in these motifs explain differences in HIV-2 disease outcomes.

Our hypothesis is that intrahost evolution of HIV-2 p24 differs between relative fast and slow disease progressors.

6.1.1 Specific aims

1. Describe amino acid sequence variation in the HIV-2 quasispecies in 12 patients and map how these sequence variants change over time

2. Measure site-specific selection pressure in intrahost evolution

3. Compare evolutionary rates in p24 between fast and slow disease progressors

6.2 Methods

Full details for materials and methods are discussed in chapter two.

6.2.1 Study participants

Study participants were drawn from the Guinea-Bissau Police cohort. Twelve HIV-2 infected patients were selected based on availability of plasma samples and meeting the following inclusion criteria:

i) HIV-2 monoinfected

ii) Antiretroviral therapy naïve at the time of plasma sample collection

iii) Longitudinal CD4% data available to classify patient as relative fast or slow disease progressors.

6.2.2 Stratification of patients into progressor groups

Briefly – patients longitudinal $CD4^+$ T cell percentage (CD4%) data were analysed in per-patient linear regression models. The CD4% at the midpoint in follow-up and the CD4% change rate (slope of the regression line) were extracted from these models. Patients were ranked from highest midpoint CD4% to the lowest, those above the mean were classified as slow progressors and those below the mean as fast progressors. Separately, patients were ranked from highest positive change rate to the lowest negative change rate and classified again. The midpoint CD4% and CD4% change rate were then transformed into proportional values, added together and averaged for each patient. This gave a combined coefficient for each patient which was then used to rank patients and stratify them into progressor groups for further analysis.

6.2.3 RNA extraction and PCR amplification

Briefly – total RNA was extracted from plasma samples which has been stored at -80˚C using the Qiagen RNeasy Lipid Tissue Mini Kit with minor modifications to the manufacturer's instructions. Following RNA extraction, a nested PCR was performed with 5µL of extracted RNA. The first step involved a one-step reverse transcription PCR (RT-PCR) using the SuperScript IV One-Step RT-PCR System with Platinum Taq DNA Polymerase. Immediately following the RT-PCR a second nested reaction using an inner set of HIV-2 *gag* primers was performed using the Dream Taq PCR kit (**table 6.1**).

6.2.4 Cloning and sequencing

PCR products were cloned using a TOPO-TA cloning kit (Invitrogen); 23 clones were routinely picked and amplified by colony-PCR using the Advantage 2 PCR kit. The colony-PCR used the same primers as the nested PCR reaction described above and gave an identical length amplicon as the insert. The colony PCR products were then purified using an in-house EXO-SAP protocol and sequenced with Macrogen Europe (Amsterdam, Netherlands) by sanger sequencing using the inner PCR primers.

Table 6.1: Primers used in the HIV-2 *gag* PCR

Primers	**Oligo sequence**	**Binding site** [1]	**Amplicon size** [2]
Outer primers			
HIV2CA_IF4	AAACATRTTGTGTGGGCAGCGA	1197 - 1218	1193
HIV2CA_IR2	CTGTCTWTCTGGGCARTTTGCC	2369 - 2390	
Inner primers			
HIV2CA_IF2	TAGTACAGAGACATCTAGYGGCAG	1411 - 1434	846
HIV2CA_IR3	TGYTGGGCTGCTGCRAATGGG	2237 - 2257	

[1] Binding site in reference to HIV-2 BEN from 5' to 3'. [2] Amplicon length shown here includes the primer sequences. Forward primers are shown 5' to 3' and reverse primers are shown 3' to 5'. Degenerate primers were selected to promote amplification on sequences with variable bases.

6.2.5 Sequence analysis

Raw sequence data was analysed in Geneious 2019.2 and mapped to the HIV-2 BEN.M30502 reference sequence and trimmed for quality (20). Contigs were constructed for each patient from single forward and reverse reads. Poor quality reads and contigs with HQ scores of less than 80% were not included in further downstream analysis. This score indicates that 80% of bases are high quality and that the likelihood of a false positive base reading is 1:10 000. Contigs were assembled into patient-specific alignments and mapped to HIV-2 BEN.M30502 to ensure that all patient sequences mapped to the same coding region of *gag*. Recombination between sequences grouped in an alignment can severely distort downstream phylogenies due to their non-tree like evolution (discussed in chapter one). We therefore performed a recombination screen on each patient alignment in RDP4 and removed all sequences showing evidence of recombination (21).

6.2.6 Bayesian phylogenetic analysis

All Bayesian phylogenetic parameters were specified in BEAUti v1.10.4 and run in BEAST v1.10.4 which uses Markov Chain Monte Carlo (MCMC) algorithms for phylogenetic inference with parallel processing by BEAGLE (22,23). Model output logs were analysed in Tracer v1.7 (24) and were considered to have converged by visual inspection of the posterior distribution chains as well as an effective sample size (ESS) of between 100 and 200 after a 10% sample burn-in per parameter. All Models were run in duplicate to determine if estimates reproducibly converged. Bayesian phylogenetic models were used to **(i)** perform viral subtyping **(ii)** reconstruct the most recent common ancestor (MRCA) for the patients' combined sequence alignment, **(iii)** estimate site-specific dN/dS ratios in p24 for each patient **(iv)** estimate relaxed molecular clock estimates for individual patients

and **(v)** estimate strict patient-specific molecular clock rates via a hierarchical phylogenetic model (HPM).

i) Viral subtyping was done by sampling HIV-2 *gag* sequences from the clonal sequences. *Gag* sequences for HIV-2 groups A – G as well as SIVsmm and HIV-1 were downloaded from LANL and an alignment of the patient HIV-2 sequences with these reference sequences was created (**table S6.1**). Viral subtyping was performed using a Bayesian phylogenetic approach. A strict molecular clock model, HKY substitution model and constant population size were specified. The Markov chain was run for $2x10^6$ iterations.

ii) Model parameters for the MRCA reconstruction are described in the appendix (**table S6.2**). The MRCA was used as a reference sequence for variant identification and mapping over time (discussed below).

iii) Renaissance counting procedures were performed using the following model parameters – a strict molecular clock model, GTR nucleotide substitution model, no site heterogeneity, a 1,2,3 codon partition, constant population size and $4x10^7$ MCMC chains.

iv) To determine combined, synonymous and nonsynonymous substitution rates derived from the patient-specific relaxed molecular clock models we used a set of software packages reported by Lemey *et al* (2). Briefly, for each patient nucleotide alignment a relaxed molecular clock model was created in BEUti and run in BEAST. Model parameters included an HKY nucleotide substitution model, gamma site heterogeneity, a constant population size and $2x10^8$ MCMC chains. For each patient 10 000 trees were simulated as a posterior sample distribution. From these 10 000 trees 200 were selected randomly after a 10% burn-in and separated into nucleotide substitution unit denoted and a time unit denoted trees. The substitution trees were then separated further using a bespoke HyPhy script into expected synonymous and nonsynonymous substitution trees. Next, these

substitution trees were analysed separately in conjunction with their respective time unit trees to generate estimates of the combined, synonymous and nonsynonymous substitution rates. Lastly, the substitution trees were used to estimate divergence as a function of time from the first timepoint to the last timepoint in the alignment. This allowed us to plot divergence over time.

v) A significant limitation of the relaxed molecular clock estimates described above are interpatient variability regarding number of sequences, number of timepoints and total follow-up time. The HPM allows for feedback across patient average estimates to improve patient specific estimates, thereby making use of a larger data set (all sequences linked in the HPM) to inform smaller partitions in the data (patient specific estimates). Another main difference in the HPM estimates versus those above are that HPMs assume a strict molecular clock per partition which does not vary for each taxon whereas the previous estimates assumed an uncorrelated relaxed molecular clock per patient. We partitioned the HPM by the combined coefficient, pooling estimates for fast and slow progressors into two groups. Further model parameters are shown in the appendix (**table S6**)

6.2.7 Variant identification and fingerprinting

We translated and aligned all 371 nucleotide sequences to the MRCA to determine where single amino acid polymorphisms (SAAPs) in the p24 region of *gag* occurred. We categorised SAAPs as majority variants if they occurred in more than 50% of sequences; minority SAAPs prevalence was 0.3 – 50%. Polymorphisms found in a single sequence were designated as private SAAPs and not included in further analysis. Private single nucleotide polymorphisms identified in viral quasispecies using cloning and sanger sequencing are often not found using next-generation sequencing platforms and probably represent a combination of PCR and sequencing errors (25).

To test whether the frequency of variants significantly differed between progressor groups we divided the sequence alignment by progressor status (fast versus slow progressor). We then created new alignments of 1000 sequences per progressor group, randomly selected via a bootstrapping procedure with replacement. We then compared the amino acid frequencies by progressor group using the Viral Epidemiology Signature Pattern Analysis (VESPA) tool from LANL (26). VESPA identifies signature patterns which differ between alignments and reports frequencies of the specific amino acids in each alignment.

6.2.8 Statistical analysis

Uni and bivariate statistical analysis was done in R studio, for non-parametric pair-wise testing Mann Whitney-U (MW) tests were used and for parametric data Student's T Test (27,28). Comparison of differences in proportions was done with the Fisher's Exact Test (FET), reported as an odds ratio (OR). Correlation statistics were calculated using the Spearman correlation coefficient. Linear regression models were analysed in R. Statistical significance was determined at a P value of < 0.05. The Holm-Bonferroni method was used to correct for multiple comparisons. The fingerprinting matrix and bar plots were produced in Excel (Microsoft Office), pirate plots using the YARRR package in R and all other figures with the ggpubr package in R (29,30).

6.3 Results

For this analysis we included 12 HIV-2 infected patients from the Guinea-Bissau Police cohort. All patients were male with a median age at enrolment of 29.50 years. The median midpoint CD4% was 27.54% and fast progressors had a significantly lower CD4% than slow progressors (18.89% vs 32.25%. MW $P = 0.048$). The median CD4% change rate was -0.05 per year; fast progressors CD4% decreased by 0.94% per year and slow progressors increased by 1.04% per year (MW $P = 0.005$) (**table 6.2**). The median follow-up time, calculated as the time from first to last sample per patient, was 7.5 years. The median time from HIV-2 detection was 9.3 years (**table S6.3**).

Table 6.2: Summary clinical data for included patients

Patient	**Age** [1]	**Sex**	**Progressor** [2]	**Sequences** [3]	**Midpoint CD4%**	**CD4% change rate**
DL1751	28	male	fast	52	31.10	-1.75
DL2086	30	male	fast	30	25.75	-1.60
DL2224	36	male	slow	7	32.91	0.70
DL2233	25	male	slow	16	24.83	1.09
DL2948	19	male	slow	23	29.33	1.04
DL2986	39	male	slow	30	32.80	-0.05
DL3105	28	male	slow	30	32.25	1.45
DL3242	29	male	fast	26	31.11	-0.94
DL3346	38	male	fast	13	18.89	0.06
DL3440	39	male	fast	43	16.33	-0.05
DL3461	32	male	fast	32	15.00	-0.37
DL3641	27	male	fast	68	18.00	-1.89

[1] Age at enrolment into the Police cohort. [2] Progressor status was determined using the combined coefficient described in experimental chapter two. [3] Total number of clonal sequences per patient

The final *gag* sequence spanned a 735 nucleotides long single open reading frame which mapped to nucleotide 1460 – 2194 on HIV-2 BEN.M30502. This sequence included the first 687 nucleotides, from 5' to 3', of the p24 region of *gag* which translated to 229 amino acids. Twenty-five plasma samples yielded PCR products from the 12 patients; nine

patients had multiple PCR products (each from a different timepoint) and three had only a single sample successfully amplified (**table S6.4**). From the 575 sequenced clones we removed poor quality sequences and those with stop codons which left 403 clonal sequences from 12 patients. We excluded 32 sequences which showed evidence of recombination from further analysis leaving us with 371 sequences (median: 30 per patient) for the final analysis. Patient derived HIV-2 *gag* sequences clustered with HIV-2 group A in the subtyping analysis (**figure S6.1**). Patient sequences formed monophyletic clusters with high posterior support values, we were therefore satisfied that there was no contamination between patient samples, and that labelling errors were unlikely to have occurred. There was no clear clustering pattern by progressor status in a maximum clade credibility (MCC) tree (**figure 6.1**).

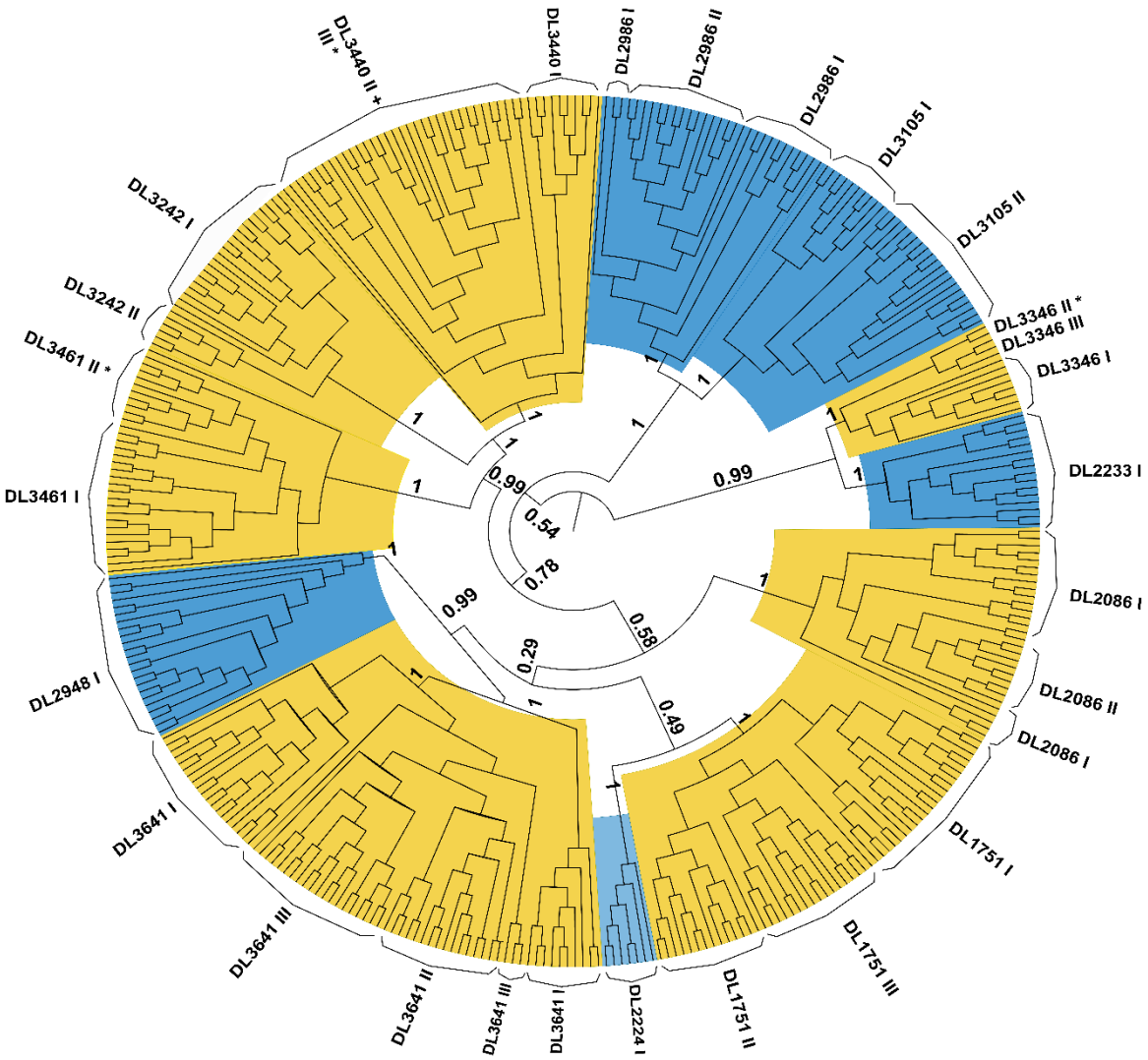


Figure 6.1: cladogram of all HIV-2 *gag* sequences included in phylogenetic analysis. This MCC cladogram represents the average of the posterior tree samples generated in the MRCA reconstruction. There is clear monophyletic clustering of patient-specific sequences with high posterior support values of 1.0. Fast disease progressors are highlighted in yellow and slow progressors in blue. An asterisk indicates that patient taxa were mixed by timepoint.

6.3.1 Variant identification and fingerprinting in HIV-2 p24

All 371 patient sequences were used to reconstruct the MRCA at the root of the MCC tree shown in figure 1. Four models were created and models 1 and 2 converged on similar estimates (**table S6.2**). The MRCA from model 2 aligned well with HIV-2 BEN.M30502, differing at 10 amino acid positions, and was used for variant identification.

We identified 61 positions where patient sequences differed from the MRCA and a greater number of sites varied within the fast progressor subset versus the slow progressor set (29 versus 15, excluding private variants, **figure 6.2**). Sixty-eight unique variants were identified and 28 of these were private variants. Of the 40 remaining variants one was a major variant and 39 were minor variants in the context of the larger sequence alignment. Important structural/functional positions which have previously been linked to either the capsid nucleotide pore, nuclear import of the capsid and NONO binding were completely conserved across all sequences (16–19). Two positions, I85 and E96, in the cyclophilin binding loop showed variation from the MRCA in fast progressors, though they were conserved in slow progressors. Furthermore, the specific motif in this loop which has been linked to TRIM5α binding and restriction, positions 87 to 91, were completely conserved across all patients (31). The frequency of amino acid variants in HLA class I epitopes differed between progressor groups though these differences were small and not statistically significant (**figure 6.2**). Three positions varied significantly in their amino acid signature between fast and slow disease progressors. At position 6, 12 and 119 the most common motif in slow progressors was AVP whereas in fast progressors it was GIA (**table 6.3**).

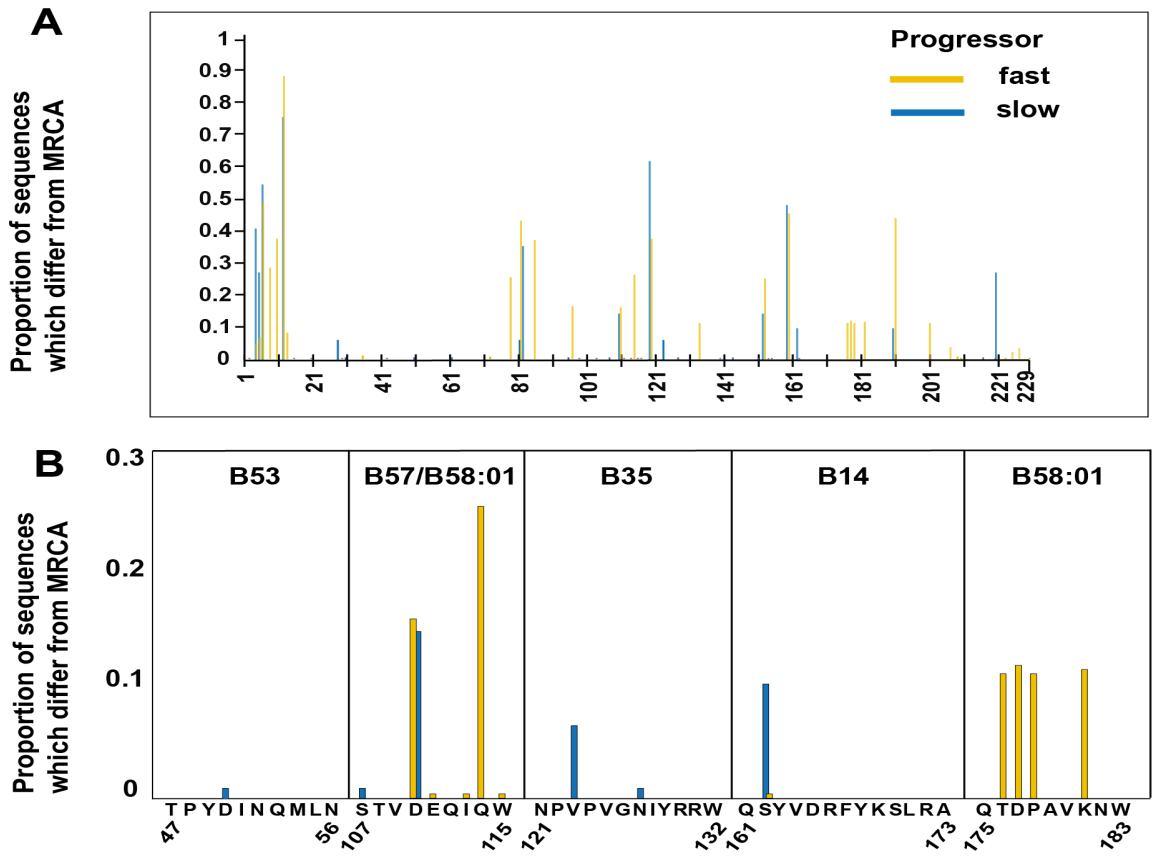


Figure 6.2: Sequence diversity in HIV-2 p24 by progressor status. A) The bar plot shows the proportion of sequences from fast and slow progressors which differs from the MRCA (y-axis) in their respective alignments at each position in HIV-2 p24 (x-axis). **B)** HLA class I epitopes are shown in each panel with the proportion of sequences for fast and slow progressors which differs from the MRCA. X-axis positions are relative to HIV-2 BEN.M30502 p24. HLA types which select for these epitopes are shown at the top of the panels.

Table 6.3: VESPA output summary for progressor groups

Position	Residue	Slow [1]	Fast [1]	OR	95% CI	*P*
6	G	444	506	1.28	1.07 - 1.53	0.006
	A	556	494			
12	V	772	199	2295.90	4.24×10^{2} - 4.50×10^{15}	$< 2.2 \times 10^{-16}$
	I	0	626			
119	A	363	619	4.15	3.41 - 5.08	$< 2.2 \times 10^{-16}$
	P	637	261			

[1] Progressor groups were assigned by using the combined coefficient. Position is for the amino acid on p24. The OR for position 12 was calculated using a FET with Haldane-Anscombe correction. Numbers of sequences in the slow and fast columns are taken from the bootstrapped sequence alignments.

Many of the variants were stable in follow-up and there was little evidence of specific variants being consistently selected out (**figure 6.3**). This suggests that negative selection is likely to have predominated as the main evolutionary force within patients.

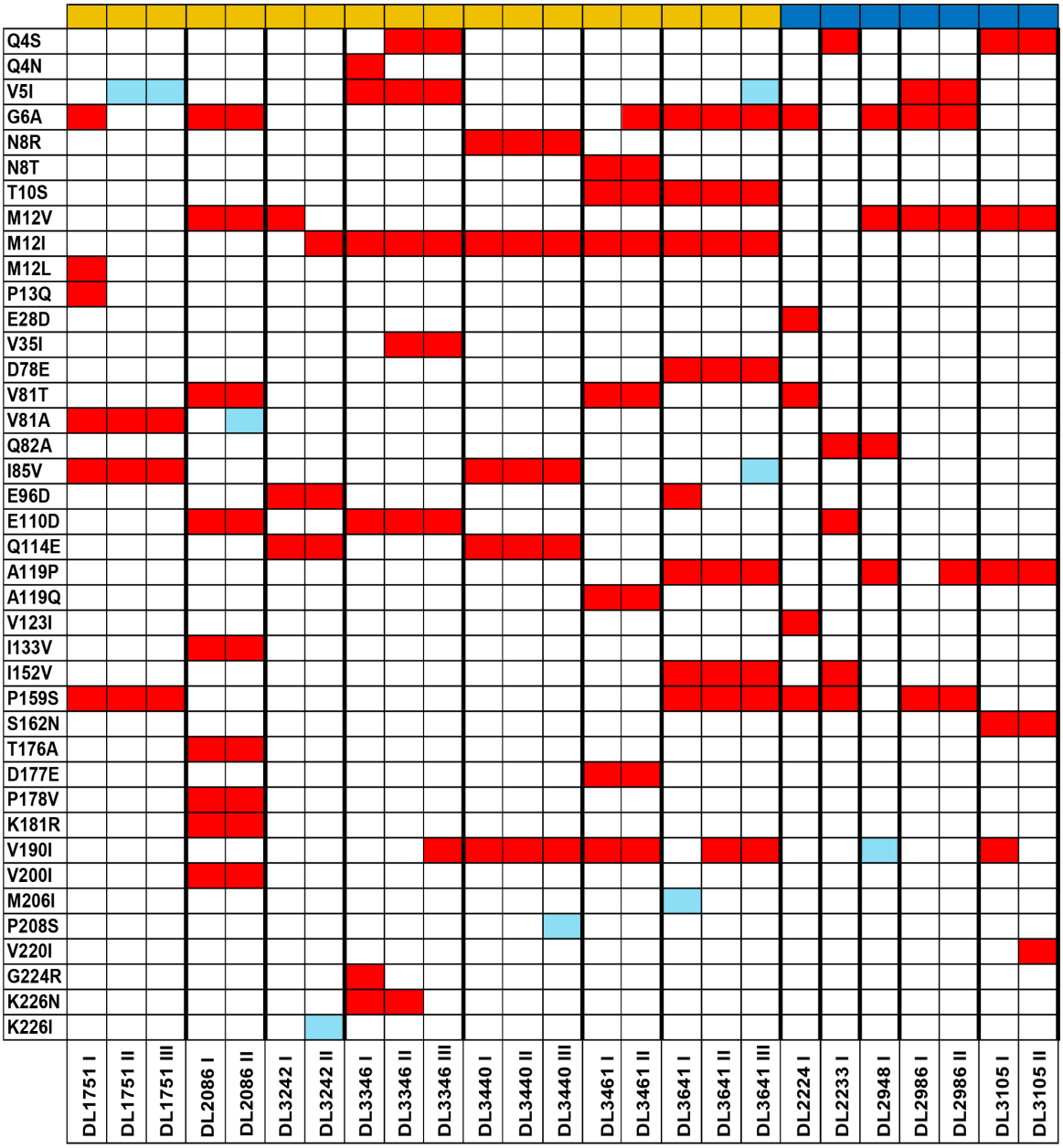


Figure 6.3: Fingerprinting of identified major variants. The y-axis here shows the major variants with the MRCA amino acid at a position used as the reference. This fingerprinting matrix shows where variants were minority or majority variants within a patient-specific timepoint. The bottom x-axis shows patient and timepoint labels. A variant found in less than 50% (minority) of sequences within a patient-specific timepoint is coloured light blue, if greater than 50% (majority) it is coloured red. Progressor labels on the top x-axis coloured yellow = fast progressors, blue = slow progressors

6.3.2 Estimation of site-specific selection pressures in the p24 region of *gag*

To quantify selection pressures in *gag* we used a previously described Renaissance counting procedure to estimate site-specific dN/dS ratios for each patient (32). To estimate selection pressure the dN/dS ratio is calculated by dividing the nonsynonymous substitution rate by the synonymous substitution rate (a ratio of less than 1.0 will indicate negative selection; a ratio equal to 1.0 neutral selection and a ratio greater than 1.0 positive selection at a site). Renaissance counting models did not converge for two of the 12 patients, these patients had a single timepoint for analysis with limited nucleotide diversity across clonal sequences. Negative selection predominated for all patients, with only one fast progressor showing a signature of positive selection at one site (**figure 6.4**).

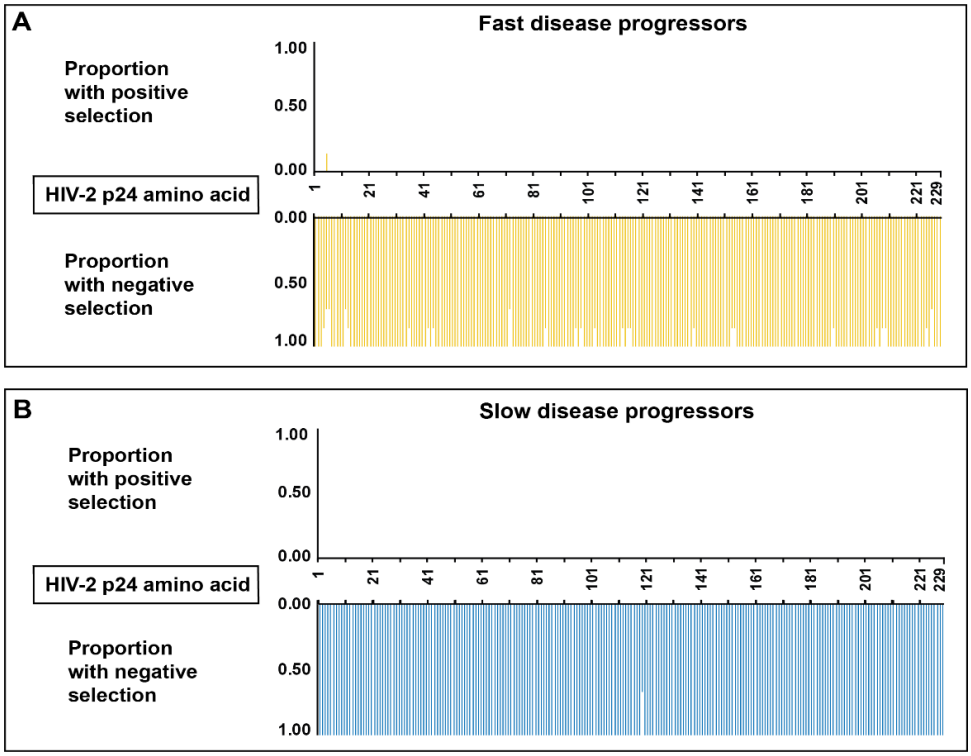


Figure 6.4: Per-site selection pressures in p24 by progressor groups. Bar plots which indicate the proportion (y-axis) of patients who have evidence of positive or negative selection pressure at a specific amino acid position in HIV-2 p24 (x-axis). Fast progressors are show in **panel A** and slow progressors in **panel B**.

6.3.3 Comparison of *gag* evolutionary rates by disease progressor status using patient-specific relaxed molecular clock models

Though there was no significant difference in site-specific selection pressures between fast and slow disease progressors we hypothesised that global substitution rates in *gag* may differ between them. The results for the relaxed clock estimates are reported for nine patients who had sequences from multiple timepoints available.

The median substitution rate in *gag* for all patients was 4.11×10^{-3} substitutions/site/year (s/s/y). Fast progressors had a significantly higher substitution rate than slow progressors (5.11×10^{-3} vs 1.97×10^{-3} s/s/y, MW $P < 2.2 \times 10^{-16}$, **figure 6.5**). The median synonymous substitution rate for all patients was 3.63×10^{-3} s/s/y and was also significantly higher in fast than slow progressors (4.47×10^{-3} vs 1.61×10^{-3} s/s/y, MW $P < 2.2 \times 10^{-16}$, **figure 6.5**). The median nonsynonymous substitution rate for all patients was 4.14×10^{-4} s/s/y and was significantly higher in fast than slow progressors (5.92×10^{-4} vs 2.66×10^{-4} s/s/y, MW $P < 2.2 \times 10^{-16}$, **figure 6.5**). The overall dN/dS ratio for *gag* in this sample was 0.13 and the difference between fast and slow progressors was small but highly significant (0.13 vs 0.17, MW $P < 2.2 \times 10^{-16}$). The results from the relaxed clock models agreed with the site-specific selection pressure estimates in that negative selection was clearly the main mode of evolution in both progressor groups.

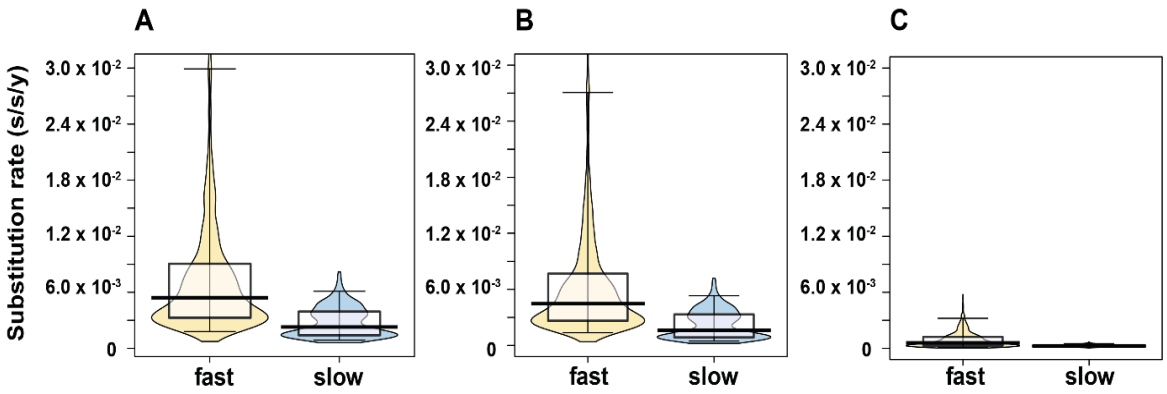


Figure 6.5: Joint, synonymous and nonsynonymous substitution rates for progressor groups. Pirate plots showing mean molecular-clock rates (y-axis) by progressor groups (x-axis). **A)** Joint molecular clock rates **B)** Synonymous substitution rates **C)** Nonsynonymous substitution rates by progressor groups stratified by the combined coefficient.

Synonymous and nonsynonymous divergence in *gag* increased linearly with time in both fast and slow progressors (**figure 6.6**). Linear regression models where fitted to each patient's divergence over time and the global r^2 for synonymous and nonsynonymous divergence was 0.94 and 0.82 respectively. Relaxed molecular clock estimates are summarised in the appendix (**table S6.5**).

Two patients had divergence data before and after the onset of AIDS (**figure S6.2**). Linear regression coefficients approximated the divergence per year for these two patients and were similar before and after AIDS. Patient DL3242 regression coefficients pre-AIDS and post-AIDS was 7.56×10^{-5} and 7.79×10^{-5}. Similarly, for patient DL2986 the coefficient for both regression models were 2.49×10^{-4}. Three patients had progressed to AIDS before p24 sequencing and four patients did not progress to AIDS during follow-up. Nonsynonymous substitutions rates were significantly higher in patients who had progressed to AIDS prior to the sequencing data (8.00×10^{-4} vs 4.84×10^{-4} s/s/y, MW $P = 8.9 \times 10^{-12}$).

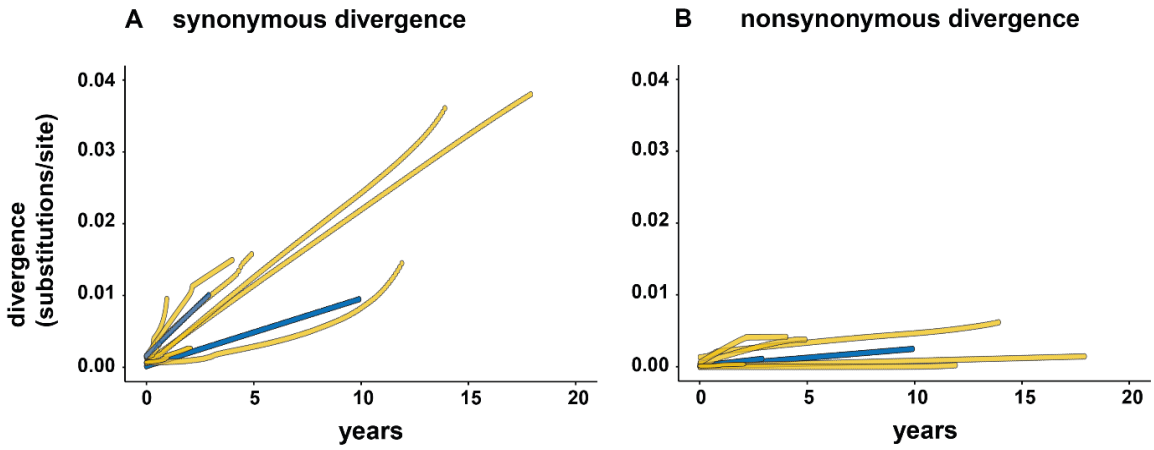


Figure 6.6: Synonymous and nonsynonymous divergence over time by progressor groups. The accumulated divergence (y-axis) from the first analysed sample for each patient indicated by time = 0 on the x-axis. Fast progressors are coloured in yellow, slow progressors in blue. **A)** Synonymous divergence over time **B)** Non-synonymous divergence over time.

6.3.4 Correlation of CD4% kinetic data with *gag* evolutionary rates estimated by Hierarchical Phylogenetic Models

We assessed a series of HPMs to generate estimates of evolutionary rates in *gag* for each patient with multiple timepoints and then averaged these estimates across progressor groups. The HPM modelling parameters and additional results are described in the appendix (**table S6.6 and S6.7**). Finally, we determined whether the HPM evolutionary rates correlated with the CD4% data which was used to stratify patients into progressor groups

The median molecular clock rate for all patients was 3.09×10^{-3} substitutions/site/year and was significantly higher in fast versus slow disease progressors (3.28×10^{-3} vs 1.7×10^{-3} s/s/y, MW $P < 2.2 \times 10^{-16}$). Median molecular clock rates correlated negatively with both CD4% change rate/year ($\rho = -0.52$, $P = 0.16$) as well as the midpoint CD4% ($\rho = -0.78$, $P = 0.02$, **figure 6.7**). These correlations were moderate to strongly negative, however, only the correlation between the molecular clock rate and midpoint CD4% was significant. To test

whether our results were biased by the model parameters we selected we performed a sensitivity analysis with narrower scales for the HPM hyperpriors (33). The median molecular clock estimates were similar across hyperprior scale values (**table S6.8**). The median clock rate across all patients for a scale of 10, 100 and 1000 was $3.02x10^{-3}$, $3.01x10^{-3}$ and $3.09x10^{-3}$ s/s/y respectively.

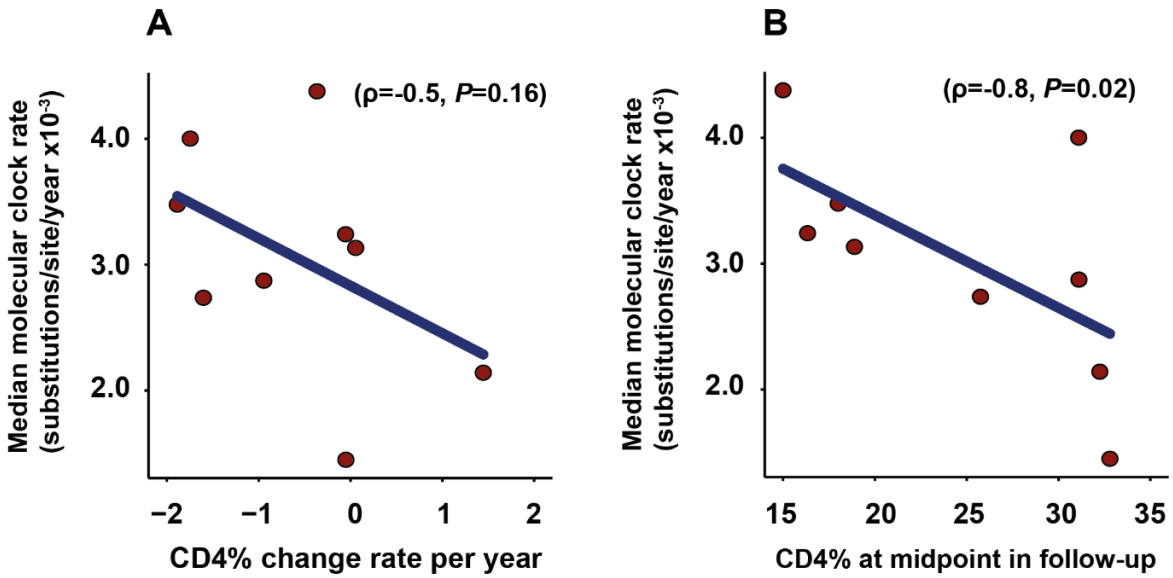


Figure 6.7: Correlation of median molecular clock rates derived from HPM with CD4% decline rates and midpoint CD4%. A) Scatterplot correlating a patient-specific estimated median molecular clock (y-axis) with their CD4% decline rate (x-axis). **B)** Scatterplot correlating a patient-specific estimated median molecular clock (y-axis) with their midpoint CD4% level in follow-up (x-axis). Results for spearman correlations are shown in the top right corners of each panel.

6.4 Discussion

We successfully performed the first study of intrahost evolution of HIV-2 *gag* and have built on previous results from this cohort (9). We have confirmed that viral evolutionary rates are higher in fast disease progressors and that this association is stronger for the CD4% level than the decline rate. This contrasts with findings for HIV-1, but agree with previous studies of *env* evolutionary rates in HIV-2 (2,9). The CD4% decline rate in HIV-2 infection is significantly slower than for HIV-1 infection, and therefore the effect of evolutionary rates on this parameter might require a larger sample size to detect (34).

Sequence variation from the MRCA was greater in fast progressors. It is possible that this was a result of including more sequences from fast progressors in the analysis, however, we think this is unlikely given that nonsynonymous divergence was significantly higher for fast progressors. With a sustained higher nonsynonymous substitution rate, we would expect a greater accumulation of amino acid diversity in fast progressors compared to slow progressors. Sequence diversity correlates positively with disease progression in HIV-1 and our results suggest a similar finding for HIV-2 (35).

Both between and within hosts, most of the important structural/functional amino acid motifs are conserved. At three sites on p24 the most common amino acid differed between fast and slow progressors. Position 12 flanks a histidine at position 11 which forms part of the nucleotide channel on the capsid (16). The M12I substitution characterised fast progressor sequences while slow progressors most common residue was valine. At position 119 slow progressors often had a proline and fast progressors an alanine. P119 has been associated with lower viral loads and enhanced sensitivity to TRIM5α (10,12,13). However, the latter finding has been contested (14). An important caveat is that the studies which

have mapped specific capsid structures to amino acid motifs were primarily based on HIV-1. However, we aligned HIV-2 BEN and HIV-1 HXB2 and found that these positions were conserved between the two viruses. This is in keeping with the conserved nature of retroviral capsids (15,17,36).

Fingerprinting of major variants showed that the majority were stable over time. This showed that negative selection was dominant in HIV-2 p24 evolution. There are three possible explanations for this: Firstly, the capsid is limited in its adaptive capacity – though this does not explain why we observed significant differences in diversity between patients. Secondly, the median time from HIV-2 detection to plasma sample collection was 9.29 years. In *gag* the primary driver of adaptive mutations will probably be CTL responses (37,38). Cytotoxic lymphocyte responses drive HIV adaptation in the early stages after infection, and therefore these changes were probably not captured by our sampling strategy (39,40). A further limitation is that we do not have *HLA* class I genotypes for these patients, making it difficult to determine if there was prior immune escape in CTL epitope regions.

Nonsynonymous substitution rates are significantly higher before the onset of AIDS in HIV-1 infection (2). The primary reason explaining this shift in nonsynonymous rates is collapse of the immune response in advanced immunosuppression. However, studies which have described this phenomenon were based on HIV-1 *env* where an important driver of immune pressure is neutralising antibodies. Patients with AIDS had a significantly higher nonsynonymous substitution rate than those at an earlier disease stage. In addition, nonsynonymous substitution rates increased linearly before and after the onset of AIDS at a constant rate in two patients. This drop in nonsynonymous substitution rates around the time of AIDS onset may not occur in HIV-2 p24.

In HIV-1 p24 synonymous and nonsynonymous substitution rates are comparable ($1.58\pm1.31x10^{-3}$ s/s/y and $2.48\pm4.77x10^{-3}$ s/s/y respectively) (33). In our analysis, HIV-2 nonsynonymous substitution rates are an order of magnitude lower than synonymous rates. It is possible that the sequence space which HIV-2 is able to explore during adaptation to the host is limited. Furthermore, our analysis supports findings from a study of interhost evolution of HIV-2 p24, which showed little evidence of positive selection shaping HIV-2 evolution (41). This is surprising as strong Gag-specific CTL responses are common in HIV-2 long term non-progressors (42). The significantly higher synonymous substitution rates suggest that viral RC and immune activation are responsible for the elevated evolutionary rates in fast progressors (43).

An important aspect of our study is that it used the same methods to determine disease progressor status and evolutionary rates as Palm *et al*. This allows us to accurately compare HIV-2 *env* and *gag* evolution. The HPM estimates of molecular clock rates are far higher for *env* than *gag* in this cohort. The global evolutionary rate in *env* was $23.50x10^{-3}$ s/s/y (95% HPD: 20.30 – 26.60) which is 7.5-fold greater than the average for *gag* at $3.09x10^{-3}$ s/s/y (95% HPD: 1.77 – 4.40).

HPM estimates had narrower 95% highest posterior density than the individual relaxed clock models which is an expected effect of using HPM (44). We had good agreement in results between relaxed clock models and the HPM. In addition, hyperprior scales did not significantly affect molecular clock estimates which strongly suggests that our estimates reflected the data and not model parameters.

Significant limitations to our study include that we were unable to sample an equal number of fast and slow progressors. In addition, all our included patients were men who are more likely to have rapid disease progression (shown in chapter four). Despite this our results

largely agreed with those of Palm *et al* – particularly that evolutionary rates correlate with disease progression in HIV-2 infection.

It would be particularly interesting to determine HLA class I types for these patients which would allow us to perform further analysis around CTL escape variants. CTL pressure and the compensatory mutations in HIV-1 *gag* can significantly affect viral RC and also increase sensitivity to TRIM5α restriction (4,45). The functional implications of many of the major variants we have described are unclear, and further *in vitro* experiments would be valuable in determining whether any of the variants or haplotypes of the variants affect viral RC. Lastly, immune activation and viral load have a strong influence on disease progression in HIV-2 infection. Understanding how *gag* affects these parameters may inform progress toward therapeutic or preventative vaccines for HIV-2.

6.5 Section summary

1. The HIV-2 p24 region of *gag* evolved under negative selection
2. Fast progressors accounted for a greater number of unique variants than slow progressors
3. Evolutionary rates were significantly higher in fast progressors
4. Evolutionary rates correlated negatively with midpoint CD4%

6.6 References for chapter 6

1. Theys K, Libin P, Pineda-Peña A-C, Nowé A, Vandamme A-M, Abecasis AB. The impact of HIV-1 within-host evolution on transmission dynamics. Curr Opin Virol. 2018 Feb 1;28:92–101.

2. Lemey P, Kosakovsky Pond SL, Drummond AJ, Pybus OG, Shapiro B, Barroso H, et al. Synonymous substitution rates predict HIV disease progression as a result of underlying replication dynamics. PLoS Comput Biol. 2007 Feb 16;3(2):e29.

3. Claiborne DT, Prince JL, Scully E, Macharia G, Micci L, Lawson B, et al. Replicative fitness of transmitted HIV-1 drives acute immune activation, proviral load in memory CD4+ T cells, and disease progression. Proc Natl Acad Sci U S A. 2015 Mar 24;112(12):E1480-9.

4. Prince JL, Claiborne DT, Carlson JM, Schaefer M, Yu T, Lahki S, et al. Role of transmitted Gag CTL polymorphisms in defining replicative capacity and early HIV-1 pathogenesis. PLoS Pathog. 2012;8(11):e1003041.

5. Selhorst P, Combrinck C, Ndabambi N, Ismail SD, Abrahams M-R, Lacerda M, et al. Replication Capacity of Viruses from Acute Infection Drives HIV-1 Disease Progression. Kirchhoff F, editor. J Virol. 2017 Apr 15;91(8):e01806-16.

6. Sousa AE, Carneiro J, Meier-Schellersheim M, Grossman Z, Victorino RMM. CD4 T cell depletion is linked directly to immune activation in the pathogenesis of HIV-1 and HIV-2 but only indirectly to the viral load. J Immunol. 2002 Sep 15;169(6):3400–6.

7. Skar H, Borrego P, Wallstrom TC, Mild M, Marcelino JM, Barroso H, et al. HIV-2 genetic evolution in patients with advanced disease is faster than that in matched HIV-1 patients. J Virol. 2010;84(14):7412–5.

8. MacNeil A, Sankalé J, Meloni ST, Sarr AD, Mboup S, Kanki P. Long-Term Intrapatient Viral Evolution during HIV-2 Infection. J Infect Dis. 2007;195(5):726–33.

9. Palm AA, Lemey P, Jansson M, Månsson F, Kvist A, Szojka Z, et al. Low Postseroconversion $CD4^+$ T-cell Level Is Associated with Faster Disease Progression and Higher Viral Evolutionary Rate in HIV-2 Infection. Paraskevis D, editor. MBio. 2019 Jan 8;10(1).

10. Onyango CO, Leligdowicz A, Yokoyama M, Sato H, Song H, Nakayama EE, et al. HIV-2 capsids distinguish high and low virus load patients in a West African community cohort. Vaccine. 2010 May;28:B60–7.

11. Jallow S, Leligdowicz A, Kramer HB, Onyango C, Cotten M, Wright C, et al. The presence of prolines in the flanking region of an immunodominant HIV-2 gag epitope influences the quality and quantity of the epitope generated. Eur J Immunol. 2015;45(8):2232–42.

12. Song H, Nakayama EE, Yokoyama M, Sato H, Levy JA, Shioda T. A Single Amino Acid of the Human Immunodeficiency Virus Type 2 Capsid Affects Its Replication in the Presence of Cynomolgus Monkey and Human TRIM5 s. J Virol. 2007 Jul 1;81(13):7280–5.

13. Miyamoto T, Yokoyama M, Kono K, Shioda T, Sato H, Nakayama EE. A Single Amino Acid of Human Immunodeficiency Virus Type 2 Capsid Protein Affects Conformation of Two External Loops and Viral Sensitivity to TRIM5α. Lee Y-M, editor. PLoS One. 2011 Jul 28;6(7):e22779.

14. Takeuchi JS, Perche B, Migraine J, Mercier-Delarue S, Ponscarme D, Simon F, et al. High level of susceptibility to human TRIM5α conferred by HIV-2 capsid sequences. Retrovirology. 2013;10(1):50.

15. Mamede JI, Damond F, Bernardo A de, Matheron S, Descamps D, Battini J-L, et al. Cyclophilins and nucleoporins are required for infection mediated by capsids from circulating HIV-2 primary isolates. Sci Rep. 2017 Mar 27;7:45214.

16. Jacques DA, McEwan WA, Hilditch L, Price AJ, Towers GJ, James LC. HIV-1 uses dynamic capsid pores to import nucleotides and fuel encapsidated DNA synthesis. Nature. 2016 Aug 10;536(7616):349–53.

17. Lahaye X, Gentili M, Silvin A, Conrad C, Picard L, Jouve M, et al. NONO Detects the Nuclear HIV Capsid to Promote cGAS-Mediated Innate Immune Activation. Cell. 2018;

18. Matreyek KA, Engelman A. The requirement for nucleoporin NUP153 during human immunodeficiency virus type 1 infection is determined by the viral capsid. J Virol. 2011 Aug 1;85(15):7818–27.

19. Schaller T, Ocwieja KE, Rasaiyaah J, Price AJ, Brady TL, Roth SL, et al. HIV-1 Capsid-Cyclophilin Interactions Determine Nuclear Import Pathway, Integration Targeting and Replication Efficiency. Aiken C, editor. PLoS Pathog. 2011 Dec 8;7(12):e1002439.

20. Kearse M, Moir R, Wilson A, Stones-Havas S, Cheung M, Sturrock S, et al. Geneious Basic: An integrated and extendable desktop software platform for the organization and analysis of sequence data. Bioinformatics. 2012;

21. Martin DP, Murrell B, Golden M, Khoosal A, Muhire B. RDP4: Detection and analysis of recombination patterns in virus genomes. Virus Evol. 2015 Mar 1;1(1).

22. Drummond AJ, Suchard MA, Xie D, Rambaut A. Bayesian Phylogenetics with BEAUti and the BEAST 1.7. Mol Biol Evol. 2012 Aug;29(8):1969–73.

23. Ayres DL, Darling A, Zwickl DJ, Beerli P, Holder MT, Lewis PO, et al. BEAGLE: An Application Programming Interface and High-Performance Computing Library for Statistical Phylogenetics. Syst Biol. 2012 Jan 1;61(1):170–3.

24. Rambaut A, Drummond AJ, Xie D, Baele G, Suchard MA. Posterior Summarization in Bayesian Phylogenetics Using Tracer 1.7. Susko E, editor. Syst Biol. 2018 Sep 1;67(5):901–4.

25. Iyer S, Casey E, Bouzek H, Kim M, Deng W, Larsen BB, et al. Comparison of major and minor viral SNPs identified through single template sequencing and pyrosequencing in acute HIV-1 infection. PLoS One. 2015 Aug 28;10(8).

26. Korber B, Myers G. Signature pattern analysis: a method for assessing viral sequence relatedness. AIDS Res Hum Retroviruses. 1992 Sep;8(9):1549–60.

27. R Core Team. R Language Definition V. 3.1.1. Https://WwwR-ProjectOrg/. 2014;

28. RStudio team, RStudio. RStudio: Integrated development environment for R. J Wildl Manage. 2015;

29. Kampstra P. Beanplot: A Boxplot Alternative for Visual Comparison of Distributions. J Stat Softw. 2008 Nov 10;28(Code Snippet 1):1–9.

30. Kassambara A. ggpubr: "ggplot2" Based Publication Ready Plots. R Package. 2018.

31. Ylinen LMJ, Keckesova Z, Wilson SJ, Ranasinghe S, Towers GJ. Differential Restriction of Human Immunodeficiency Virus Type 2 and Simian Immunodeficiency Virus SIVmac by TRIM5 Alleles. J Virol. 2005 Sep 15;79(18):11580–7.

32. Lemey P, Minin VN, Bielejec F, Kosakovsky Pond SL, Suchard MA. A counting renaissance: combining stochastic mapping and empirical Bayes to quickly detect amino acid sites under positive selection. Bioinformatics. 2012 Dec 15;28(24):3248–56.

33. Raghwani J, Redd AD, Longosz AF, Wu CH, Serwadda D, Martens C, et al. Evolution of HIV-1 within untreated individuals and at the population scale in Uganda. PLoS Pathog. 2018 Jul 1;14(7).

34. Esbjörnsson J, Månsson F, Kvist A, da Silva ZJ, Andersson S, Fenyö EM, et al. Long-term follow-up of HIV-2-related AIDS and mortality in Guinea-Bissau: a prospective open cohort study. lancet HIV. 2018 Nov 1;

35. Troyer RM, Collins KR, Abraha A, Fraundorf E, Moore DM, Krizan RW, et al. Changes in Human Immunodeficiency Virus Type 1 Fitness and Genetic Diversity during Disease Progression. J Virol. 2005 Jul 15;79(14):9006–18.

36. Rihn SJ, Wilson SJ, Loman NJ, Alim M, Bakker SE, Bhella D, et al. Extreme genetic fragility of the HIV-1 capsid. PLoS Pathog. 2013;9(6):e1003461.

37. de Silva TI, van Tienen C, Onyango C, Jabang A, Vincent T, Loeff MFS van der, et al. Population dynamics of HIV-2 in rural West Africa: comparison with HIV-1 and ongoing transmission at the heart of the epidemic. AIDS. 2013 Jan 2;27(1):125–34.

38. Leligdowicz A, Yindom L-M, Onyango C, Sarge-Njie R, Alabi A, Cotten M, et al. Robust Gag-specific T cell responses characterize viremia control in HIV-2 infection. J Clin Invest. 2007 Oct 1;117(10):3067–74.

39. Ganusov V V, Goonetilleke N, Liu MKP, Ferrari G, Shaw GM, McMichael AJ, et al. Fitness costs and diversity of the cytotoxic T lymphocyte (CTL) response determine the rate of CTL escape during acute and chronic phases of HIV infection. J Virol. 2011 Oct;85(20):10518–28.

40. Leviyang S, Ganusov V V. Broad CTL Response in Early HIV Infection Drives Multiple Concurrent CTL Escapes. PLoS Comput Biol. 2015 Oct;11(10):e1004492.

41. de Silva TI, Leligdowicz A, Carlson J, Garcia-Knight M, Onyango C, Miller N, et al. HLA-associated polymorphisms in the HIV-2 capsid highlight key differences between HIV-1 and HIV-2 immune adaptation. AIDS. 2018 Mar 27;32(6):709–14.

42. de Silva TI, Peng Y, Leligdowicz A, Zaidi I, Li L, Griffin H, et al. Correlates of T-cell-mediated viral control and phenotype of CD8+ T cells in HIV-2, a naturally contained human retroviral infection. Blood. 2013 May 23;121(21):4330–9.

43. Grossman Z, Meier-Schellersheim M, Paul WE, Picker LJ. Pathogenesis of HIV infection: what the virus spares is as important as what it destroys. Nat Med. 2006 Mar 6;12(3):289–95.

44. Suchard MA, Kitchen CMR, Sinsheimer JS, Weiss RE. Hierarchical Phylogenetic Models for Analyzing Multipartite Sequence Data. Syst Biol. 2003;52(5):649–64.

45. Merindol N, El-Far M, Sylla M, Masroori N, Dufour C, Li J, et al. HIV-1 capsids from B27/B57+ elite controllers escape Mx2 but are targeted by TRIM5α, leading to the induction of an antiviral state. Emerman M, editor. PLOS Pathog. 2018 Nov 12;14(11):e1007398.

Appendix 6

Supplementary details for most recent common ancestor reconstruction

Table S6.1: Accession numbers for sequences used in HIV-2 subtyping analysis:

HIV/SIV	Group	Accession number
HIV-2	A	A.GW.2010.TD062.MH681609
HIV-2	A	A.IN.1995.CRIK_147.DQ307022
HIV-2	A	A.CI.1988.UC2.U38293
HIV-2	A	A.PT.-.ALI.AF082339
HIV-2	A	A.DE.-.BEN.M30502
HIV-2	A	A.DE.-.PEI2_KR_KRCG.U22047
HIV-2	A	A.NL.2001.RH2.7.MF595863
HIV-2	B	B.CI.-.IC763088.AF073842
HIV-2	B	B.CI.-.IC3983.AF073841
HIV-2	B	B.CI.-.IC2380.AF073840
HIV-2	B	B.CI.-.IC763129.AF073843
HIV-2	B	B.JP.2001.IMCJ_KR020_1_.AB100245
HIV-2	B	B.CI.-.20_57.AB485671
HIV-2	B	B.CI.-.20_56.AB485670
HIV-2	C	L33077
HIV-2	D	L33083
HIV-2	E	L33087
HIV-2	F	F.SL.1993.93SL2F.U75441
HIV-2	F	KP890355
HIV-2	G	AF208027
HIV-1	A	02A1.GH.1997.97GH-AG2.AB052867
HIV-1	A1	A1.UG.1992.UG029.AB098332
HIV-1	A1	A1.UG.1992.UG029.AB098333
HIV-1	AE	01_AE.TH.1995.95TNIH022.AB032740
HIV-1	AE	01_AE.TH.1995.95TNIH047.AB032741
HIV-1	AE	01_AE.JP.1993.93JP_NH1.AB052995
HIV-1	AE	01_AE.JP.1993.NH25_93JPNH25T_93JP_NH2_5T.AB070352
HIV-1	AG	02_AG.GH.1997.97GH-AG1.AB049811
HIV-1	B	B.FR.1983.LAI-J19.A07867
HIV-1	B	B.JP.-.PATIENT_IMS1.AB074049
HIV-1	B	B.JP.-.PATIENT_IMS1.AB074050
HIV-1	C	C.IN.1993.93IN101.AB023804
SIVsmm	SIVsmm	SMM.US.-.RM174.V2.tf.KU182923
SIVsmm	SIVsmm	SMM.US.-.FTq.tf.KU182920
SIVsmm	SIVsmm	SMM.US.2011.SIVsmE660-FL8.JQ864086
SIVsmm	SIVsmm	SMM.US.-.SME543.U72748

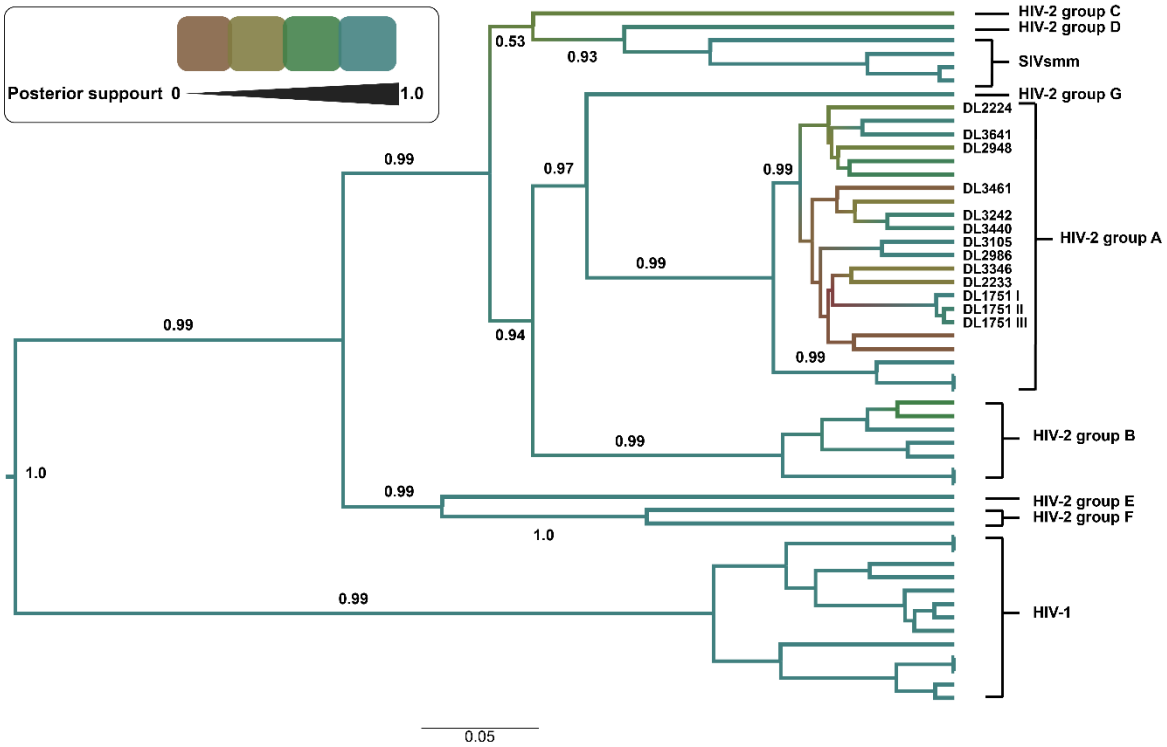


Figure S6.1: Phylogenetic tree of subtyping analysis for HIV-2 *gag* sequences. The MCC shows posterior support values for all internal branches with the colour scale shown in the top left corner. High support values of 0.99 indicate that the cloned *gag* sequences were HIV-2 group A. Scale bar shows genetic distance from the tree root.

Table S6.2: Most recent common ancestor reconstruction phylogenetic model parameter summaries

model	**clock**	**substitution model**	**heterogeneity**	**codon partition**	**population**	**MCMC chains**	**C**
MRCA1	strict	GTR	gamma	none	constant	$2.5x10^8$	yes
MRCA2*	**strict**	**GTR**	**none**	**none**	**constant**	$2.5x10^8$	**yes**
MRCA3	relaxed	GTR	none	none	constant	$5.0x10^8$	no
MRCA4	strict	HKY	gamma	1,2, 3	constant	$2.5x10^8$	yes

*The MRCA from model 2 (highlighted in bold) was used for variant calling and fingerprinting. C = converged

Supplementary details for patient epidemiological data and sample details

Table S6.3: summary epidemiological data for included patients

Patient	HIV-2 detected [1]	Last date [2]	ART date	Total follow-up time [3]	AIDS date [4]
DL1751	13/02/1990	24/02/2011	24/02/2011	21.04	31/07/2002
DL2086	29/07/1995	26/09/2013		18.18	
DL2224	19/02/1991	17/09/2013		22.59	08/08/2000
DL2233	19/04/1991	16/01/2008		16.76	03/12/1993
DL2948	15/08/2001	13/03/2006		4.58	13/03/2006
DL2986	27/04/1993	30/01/2010		16.77	10/06/2004
DL3105	26/06/1999	07/10/2008		9.29	26/06/2004
DL3242	18/08/1993	25/09/2013		20.12	13/11/2007
DL3346	10/11/1993	18/12/2009	18/12/2009	16.12	
DL3440	10/07/2004	01/07/2013	01/07/2013	8.98	
DL3461	04/01/2001	07/06/2010	07/06/2010	9.43	09/12/2005
DL3641	12/06/2000	23/09/2013		13.29	23/09/2013

[1] HIV-2 detected refers to patients who have an estimated date of infection based on serosurvey data as well as those who were HIV-2 positive at inclusion into the study. [2] Last date for which data was included. [3] Time from HIV-2 detection to last date. [4] AIDS was defined according to the WHO staging criteria as either i) $CD4^+$ count < 200cells/μL ii) CD4% < 14% iii) WHO stage IV illness diagnosed.

Table S6.4: summary sample data for included patients

Patient	Plasma sample	HIV-2 detected [1]	Timepoint	Follow-up time [2]	Years post HIV-2 detection
DL1751	I	13/02/1990	03/07/2004	0.00	14.39
	II	13/02/1990	28/11/2006	2.41	16.80
	III	13/02/1990	07/10/2009	5.27	19.66
DL2086 [3]	I	21/07/1995	01/01/2004	0.00	8.45
	II	21/07/1995	26/09/2013	9.74	18.20
DL2224	I	19/02/1991	19/05/2008	0.00	17.26
DL2233	I	19/04/1991	14/11/2006	0.00	15.58
DL2948	I	15/08/2001	01/07/2004	0.00	2.88
DL2986	I	27/04/1993	24/09/1997	0.00	4.41
	II	27/04/1993	25/10/2008	11.09	15.51
DL3105	I	26/06/1999	25/06/2004	0.00	5.00
	II	26/06/1999	06/10/2008	4.28	9.29
DL3242	I	18/08/1993	07/11/1994	0.00	1.22
	II	18/08/1993	25/09/2013	18.90	20.12
DL3346	I	10/11/1993	04/04/1996	0.00	2.40
	II	10/11/1993	20/10/2009	13.55	15.95
	III	10/11/1993	11/05/2010	14.11	16.51
DL3440	I	10/07/2004	23/10/2007	0.00	3.29
	II	10/07/2004	30/03/2009	1.44	4.72
	III	10/07/2004	23/10/2009	2.00	5.29
DL3461	I	04/01/2001	29/09/2009	0.00	8.74
	II	04/01/2001	10/05/2010	0.61	9.35
DL3641	I	12/06/2000	08/12/2004	0.00	4.49
	II	12/06/2000	16/09/2008	3.78	8.27
	III	12/06/2000	10/03/2010	5.25	9.75

[1] HIV-2 detected refers to patients who have an estimated date of infection based on serosurvey data as well as those who were HIV-2 positive at inclusion into the study. [2] Follow-up time measured in years from the first sample date listed in the timepoint column. [3] Sample I for patient DL2086 did not have a date label. We therefore used the midpoint in follow-up time for this patient and used an uncertainty correction of ± 9 years in all models which required a date for the sample.

Supplementary details for HIV-2 capsid molecular clock estimates

All relaxed clock models were run using the same parameters as described in the methods section. Only patients with multiple timepoints were included

Table S6.5: Relaxed molecular clock estimates

Patient	Median clock rate	95% HPD [1]	Progressor [2]
DL1751	$7.49x10^{-3}$	$3.84x10^{-3}$, $1.00x10^{-2}$	Fast
DL2086	$4.59x10^{-3}$	$7.67x10^{-4}$, $1.00x10^{-2}$	Fast
DL2224	$1.16x10^{-3}$	$4.76x10^{-4}$, $1.95x10^{-3}$	Slow
DL2233	$3.64x10^{-3}$	$1.49x10^{-3}$, $6.07x10^{-3}$	Slow
DL2948	$2.64x10^{-3}$	$1.02x10^{-3}$, $3.96x10^{-3}$	Fast
DL2986	$3.97x10^{-3}$	$1.87x10^{-3}$, $6.88x10^{-3}$	Fast
DL3105	$3.32x10^{-3}$	$1.03x10^{-3}$, $9.67x10^{-3}$	Fast
DL3242	$1.80x10^{-2}$	$2.31x10^{-3}$, $5.00x10^{-2}$	Fast
DL3346	$1.20x10^{-2}$	$4.01x10^{-3}$, $2.00x10^{-2}$	Fast

[1] Patient relaxed clock model estimates for molecular clock rates with their 95% highest posterior density are shown. [2] Progressor status as determined by the combined coefficient.

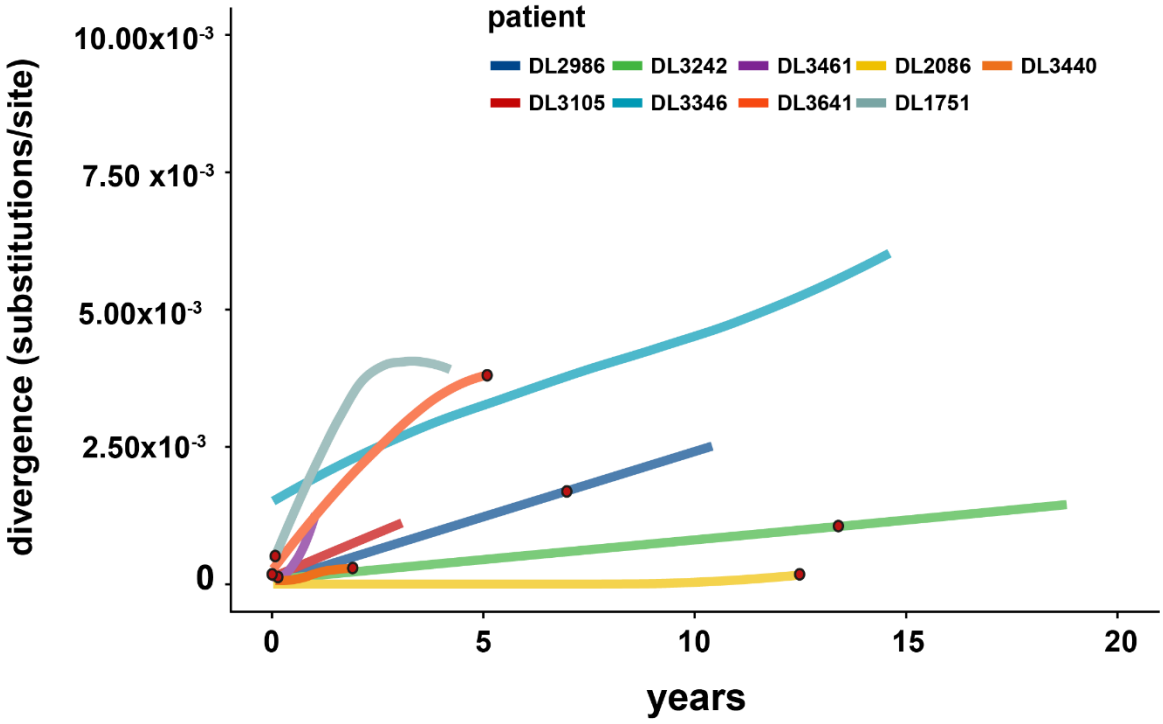


Figure S6.2: Nonsynonymous divergence over time. Patient divergence models are shown above, divergence on the y-axis and years from the first to last sample for each patient on the x-axis. Red dots placed on each divergence curve indicate when that patient was diagnosed with AIDS relative to this timeline. Dots placed at the beginning of DL1751, DL3461 and DL3105 indicate that AIDS had been diagnosed before our sequence data starts. Dots at the end of the curve indicate AIDS occurred after our sequence data. DL3346 did not progress to AIDS during follow-up.

Supplementary details for HIV-2 Hierarchical Phylogenetic Model strict clock estimates

Here we have summarised the model parameters used to run HPM. HPM7 was finally used and the strict clock estimates for each patient are shown in table S7. Only patients with multiple timepoints were included.

Table S6.6: Hierarchical Phylogenetic Model parameters used to estimate molecular clock rates

model	substitution model	heterogeneity	codon partition	population	MCMC chains	**C**
HPM1	GTR	none	none	constant	$5x10^8$	no
HPM2	GTR	none	none	constant	$5x10^8$	no
HPM3	GTR	none	none	constant	$5x10^8$	no
HPM4	GTR	none	none	constant	$5x10^8$	yes
HPM5	HKY	none	none	constant	$5x10^8$	no
HPM6	JC	none	none	constant	$5x10^8$	no
HPM7 [1]	**HKY**	**gamma**	**1,2 and 3**	**constant**	$6x10^8$	**yes**
HPM8	HKY	gamma	1,2 and 3	exponential	$6x10^8$	no
HPM9	GTR	gamma	1,2 and 3	constant	$6x10^8$	no

[1] Model 7 (highlighted in bold) was used for HPM strict clock estimates. C = converged

Table S6.7: Strict molecular clock estimates from the Hierarchical Phylogenetic Model

Patient	**Median clock rate**	**95% HPD** [1]	**Progressor** [2]
DL1751	$4.00x10^{-3}$	$2.62x10^{-3}$, $6.40x10^{-3}$	Fast
DL2086	$2.74x10^{-3}$	$1.76x10^{-3}$, $3.70x10^{-3}$	Fast
DL2224	$1.45x10^{-3}$	$5.83x10^{-4}$, $2.33x10^{-3}$	Slow
DL2233	$2.14x10^{-3}$	$1.02x10^{-3}$, $4.40x10^{-3}$	Slow
DL2948	$2.87x10^{-3}$	$1.77x10^{-3}$, $3.92x10^{-3}$	Fast
DL2986	$3.13x10^{-3}$	$2.18x10^{-3}$, $4.24x10^{-3}$	Fast
DL3105	$3.24x10^{-3}$	$1.16x10^{-3}$, $5.27x10^{-3}$	Fast
DL3242	$4.38x10^{-3}$	$2.45x10^{-3}$, $1.00x10^{-2}$	Fast
DL3346	$3.48x10^{-3}$	$2.42x10^{-3}$, $5.11x10^{-3}$	Fast

[1] Patient HPM estimates for molecular clock rates with their 95% highest posterior density are shown. [2] Progressor status as determined by the combined coefficient.

Hyperprior scales were adjusted for each HPM. Our first model had used an unspecified hyperprior scale of 1000. We narrowed the scale to 10 and 100 for separate models. The narrower scale specifies a narrowed range of parameter values which are tested in the MCMC algorithm.

Table S6.8: Results of Hierarchical Phylogenetic Model sensitivity analysis

		DL3346	DL3440	DL3461	DL3641	DL2086	DL1751	DL2986	DL3105	DL3242
HPM10	Median clock rate	$3.04x10^{-3}$	$3.03x10^{-3}$	$6.26x10^{-3}$	$3.67x10^{-3}$	$2.42x10^{-3}$	$4.19x10^{-3}$	$1.17x10^{-3}$	$2.86x10^{-3}$	$2.52x10^{-3}$
	95% HPD	$1.97x10^{-3}$, $4.13x10^{-3}$	$6.85x10^{-4,}$ $5.589x10^{-3}$	$2.77x10^{-3}$, $1.22x10^{-2}$	$2.28x10^{-3}$, $5.49x10^{-3}$	$1.6x10^{-3}$, $3.4x10^{-3}$	$2.49x10^{-3}$, $6.39x10^{-3}$	$4.80x10^{-4}$, $2.06x10^{-3}$	$1.17x10^{-3}$, $5.06x10^{-3}$	$1.59x10^{-3}$, $3.82x10^{-3}$
HPM100	Median clock rate	$3.08x10^{-3}$	$3.05x10^{-3}$	$4.77x10^{-3}$	$3.54x10^{-3}$	$2.58x10^{-3}$	$3.91x10^{-3}$	$1.30x10^{-3}$	$2.51x10^{-3}$	$2.70x10^{-3}$
	95% HPD	$2.14x10^{-3}$, $4.20x10^{-3}$	$7.93x10^{-4}$, $5.38x10^{-3}$	$2.21x10^{-3}$, $1.07x10^{-2}$	$2.32x10^{-3}$, $5.3x10^{-3}$	$1.68x10^{-3}$, $3.56x10^{-3}$	$2.55x10^{-3}$, $6.18x10^{-3}$	$4.97x10^{-4}$, $2.12x10^{-3}$	$9.40x10^{-4}$, $4.83x10^{-3}$	$1.62x10^{-3}$, $3.69x10^{-3}$
HPM1000	Median clock rate	$3.13x10^{-3}$	$3.24x10^{-3}$	$4.38x10^{-3}$	$3.48x10^{-3}$	$2.74x10^{-3}$	$4.00x10^{-3}$	$1.45x10^{-3}$	$2.14x10^{-3}$	$2.87x10^{-3}$
	95% HPD	$2.18x10^{-3}$, $4.24x10^{-3}$	$1.16x10^{-3}$, $5.27x10^{-3}$	$2.45x10^{-3}$, $1.00x10^{-2}$	$2.42x10^{-3}$, $5.11x10^{-3}$	$1.76x10^{-3}$, $3.70x10^{-3}$	$2.62x10^{-3}$, $6.40x10^{-3}$	$5.83x10^{-4}$, $2.33x10^{-3}$	$1.02x10^{-3}$, $4.40x10^{-3}$	$1.77x10^{-3}$, $3.92x10^{-3}$

[1] Patient HPM estimates for molecular clock rates with their 95% highest posterior density are shown. Estimates are shown for three HPM hyperprior scales: 10, 100 and 1000. Increasing scales allow for wider ranges of prior probabilities

Chapter 7

Conclusions and future perspectives

The HIV pandemic is well into its third decade and medical science has been unable to produce a vaccine or cure despite Herculean efforts. The global incidence of HIV-1 is falling and more people are on ART than ever before (1). However, the number of people with HIV who require lifelong therapy and monitoring is increasing. This is placing a large strain on health care systems in poor countries. HIV-related illness is still the leading cause of death in Sub-Saharan Africa and is driving an ongoing TB pandemic. It seems unlikely that we will be able to treat our way out of this pandemic. The early promise of the HPTN 052 trial has been followed up with mixed results for reducing the incidence of HIV by scaling up treatment programmes. The United Nations 90-90-90 targets are highly unlikely to be met in 2020 and HIV transmission is probably going to continue for the foreseeable future (2–4).

In contrast to HIV-1, the HIV-2 epidemic is likely to reach extinction around 2068 with minimal intervention by humans. The decreasing HIV-2 prevalence is a cause for celebration in West Africa, however, this also means that the window for studying this virus is closing (5). HIV-2 infection offers important avenues for further study as it may represent a natural model for a functional cure (6). HIV elite-controllers offer an example of what a functional cure might look like. These patients have relapse-free remissions of HIV disease, in the absence of treatment, which is sustained long term, and often associated with low to

absent plasma viral loads (6). The prevalence of spontaneous HIV-1 elite-controllers is low, whereas this is a more common phenotype in HIV-2 infected patients.

Most adult HIV-1 elite controllers will have a narrow range of "protective" *HLA* class I genotypes or *CCR5* deletions. However, as discussed in the previous chapters disease progression in HIV-1 infected children and in HIV-2 infected adults have not been robustly linked to these pathways (7–9). These data show that there are probably many different mechanisms which can influence HIV disease progression.

To partially understand these mechanisms, we completed four research projects which focused on three categories:

7.1 Single nucleotide polymorphism association studies of *TRIM5* and *TRIM22* with immunological, viral and clinical outcomes of HIV infection

We found that these genes are highly polymorphic within sub Saharan-African populations, with no significant linkage between loci within them. Despite considerable diversity in *TRIM5* and *TRIM22,* SNPs in these genes did not contribute substantially to disease outcomes in either perinatal HIV-1 infection or adult HIV-2 infection. In HIV-1 infected children, we reported that substitutions in the coiled-coil of *TRIM22* may associate with modest effects on $CD4^+$ counts when corrected for age and CMV viraemia. The latter two covariates had significant effects on $CD4^+$ counts. Median pVL did not vary between long-term non-progressors and the rest of the sample. In HIV-2 infected adults a pair of SNPs in *TRIM22* associated with longitudinal $CD4^+$ T cell percentage (CD4%) data, and again these

effects were modest. *TRIM* genotype did not vary significantly between fast and slow disease progressors.

There are several conclusions we can draw from these association studies. The first is *TRIM5* and *TRIM22* do not show evidence of linkage in two divergent African populations despite their proximity on chromosome 11. Different pathogens may have driven the divergent evolution of *TRIM5* and *TRIM22*. Secondly, genotype of *TRIM5* and *TRIM22* did not associate with strong effects on the range of tested outcomes. HIV evades and uses a stunning number of cellular systems to maintain successful viral replication. This means that there is likely to be a lot of redundancy built into these systems (10). HIV disease progression is a complex trait to perform genetic association studies on, we would therefore expect effect sizes of individual SNPs to be small. This means that, either, our positive results represent false positives which will "regress to the null hypothesis" in larger samples (11); or *TRIM22* has significant effects on HIV-1 and -2 disease outcomes in a diverse range of populations. Given that *TRIM22* targets viral transcription indirectly, and that both HIV-1 and HIV-2 respond to the relevant Sp1 driven pathways this is a possibility.

There is clearly a need for further host genetic association studies with HIV disease outcomes. Modern humans originated in Africa within the past few hundred thousand years and migrated out of Africa in small groups over the past 80 000 years (12). These migrations caused genetic bottlenecks which has resulted in non-Africans having similar haplogroups and linkage patterns. In addition, Africans have higher genetic diversity, more variability in genetic structure and less linkage between loci than non-Africans (13). The genome wide association studies (GWAS) which have reported significant effects of *HLA* and *CCR5* genotype on HIV disease outcomes routinely have samples of a few thousand individuals (10,14–16). HIV-infected people of European descent have made up most of these GWAS

recruits. This correlates with the larger GWAS field where most studies, and the majority of individuals recruited, are people of European ancestry (12). Future GWAS or whole gene sequencing studies should aim to include many African patients with adequate epidemiological data to use in association studies. Cohorts of HIV-1 infected children and adolescents and HIV-2 infected adults are often small, numbering at most a few hundred individuals. It would be helpful to establish collaborations between diverse sites and pool data as is commonly done for meta-analyses. This is particularly relevant for HIV-2 infected cohorts in Europe where a small number of individuals receive detailed follow-up in a high-resource setting.

7.2 Reporting a HIV-2 disease progressor stratification system based on $CD4^+$ T cell kinetic data over long-term follow-up

Studies which have stratified HIV-1 infected patients into disease progressor groups have focused on the elite controllers and long term non-progressors. This is often a small fraction of HIV-1 infected individuals, which not only limits the generalisability of these studies but also decreases statistical power for further analyses (6,17). In the Guinea-Bissau Police cohort, we designed a stratification system which was able to include all patients with longitudinal CD4% data. We stratified patients into two, relatively balanced groups – fast and slow disease progressors. This stratification created clear groups with major differences in immunological and clinical outcomes which was robust to estimated time since infection. Seroconverter estimates for time to AIDS did not vary significantly from seroprevalent patients. We unfortunately could not assess the effect of this stratification on mortality rates due to a high censoring rate. However, based on existing literature it seems probable that fast progressors had a higher mortality rate than slow progressors.

Our analysis was biased by including a greater proportion of males, and it would be valuable to repeat the stratification procedure and time to event analyses on a larger sample which included more females. We have access to the data from the Caió community cohort and intend to do this in the near future. Should our analysis' results be replicable it would provide a basis for extending this system to other HIV-2 cohorts. This may be useful in identifying factors which are associated with disease progression in larger pooled analyses. Continuous covariates can be correlated directly with the midpoint CD4%, decline rate or combined coefficient. In addition, if we can identify patients relatively soon in infection who are unlikely to progress to AIDS within their lifetime it may provide a justification for randomising these slower progressing patients to receive ART or placebo in controlled trials to assess their treatment benefit. However, if the current RCTs of ART in HIV-2 infected individuals show a significant benefit of ART to all HIV-2 infected patients it would not be ethical to randomise patients as described.

7.3 Comparing intrahost evolution of HIV-2 *gag* between relative fast and slow disease progressors.

We have generated the first sequence dataset of intrahost evolution of HIV-2 *gag.* We identified three positions in p24 which differed in their amino acid based on progressor status. The A119P substitution has been previously associated with lower HIV-2 viral loads, and in our analysis was more common amongst sequences from slow disease progressors. In addition, this substitution may increase the sensitivity of the capsid to TRIM5α. The V12I substitution has not been previously described in relation to HIV-2 disease outcome. This position flanks a conserved histidine at position eleven which forms part of the capsid nucleotide pore and may have functional relevance as a result.

We further characterised sites that were strictly conserved within and between hosts. The p24 region of *gag* evolved under negative selection, and this did not differ between patients based on disease stage. Viral evolutionary rates were significantly higher in fast progressors compared to slow progressors. Furthermore, evolutionary rates correlated negatively with the midpoint CD4%, but not the decline rate. A logical extension to this project would be to determine if the substitutions which were more common amongst fast progressors improved viral replicative capacity (RC). If novel T cell epitopes can be identified which include such substitutions it may progress the task of developing a therapeutic vaccine.

The recently reported capsid specific inhibitor antiretroviral agent GS-CA1 potently inhibits a range of HIV-1 strains (18). GS-CA1 binds to conserved residues on the capsid and inhibits nuclear import. It is robust to escape mutations, and when escape variants do emerge their fitness is decreased relative to the wild-type virus. GS-CA1 has activity against HIV-2 and can be administered subcutaneously with long acting formulations. An analogue, GS-6207, is currently in phase one trials. Our data increase the number of within host p24 sequences for further analysis with regards to this new class of ART.

The methods which we used to sequence the quasispecies are likely to be superseded by next generation sequencing platforms. There has been exceptional progress in using this technology to generate data for phylogenetic analysis. Whole gene sequencing for HIV-2 has been reported and expanding this method will allow researchers to determine whether linkage is significantly influencing viral evolution. In addition, it would determine whether viral RC is a complex trait linked to multiple sites on the viral genome or strongly influenced by single residues (19).

Approximate word count: 37 000

7.4 References for chapter 7

1. AIDSinfo | UNAIDS [Internet]. [cited 2019 Feb 1]. Available from: http://aidsinfo.unaids.org/
2. Levi J, Raymond A, Pozniak A, Vernazza P, Kohler P, Hill A. Can the UNAIDS 90-90-90 target be achieved? A systematic analysis of national HIV treatment cascades. BMJ Glob Heal. 2016;1(2).
3. Huerga H, Van Cutsem G, Ben Farhat J, Puren A, Bouhenia M, Wiesner L, et al. Progress towards the UNAIDS 90-90-90 goals by age and gender in a rural area of KwaZulu-Natal, South Africa: A household-based community cross-sectional survey. BMC Public Health. 2018 Mar 2;18(1).
4. Edet A, Akinsola H, Bessong PO. Virologic and immunologic responses of patients on highly active antiretroviral therapy in a rural community health centre in Limpopo, South Africa: A retrospective study. South Afr J HIV Med. 2019;20(1).
5. Olesen JS, Jespersen S, da Silva ZJ, Rodrigues A, Erikstrup C, Aaby P, et al. HIV-2 continues to decrease, whereas HIV-1 is stabilizing in Guinea-Bissau. AIDS. 2018 Jun 1;32(9):1193–8.
6. Esbjörnsson J, Jansson M, Jespersen S, Månsson F, Hønge BL, Lindman J, et al. HIV-2 as a model to identify a functional HIV cure. AIDS Res Ther. 2019;16(1):1–12.
7. Tang J, Wilson CM, Schaen M, Myracle A, Douglas SD, Kaslow RA. CCR2 and CCR5 genotypes in HIV type 1-infected adolescents: Limited contributions to variability in plasma HIV type 1 RNA concentration in the absence of antiretroviral therapy. AIDS Res Hum Retroviruses. 2002;18(6):403–12.
8. Bansal A, Yue L, Conway J, Yusim K, Tang J, Kappes J, et al. Immunological control of chronic HIV-1 infection: HLA-mediated immune function and viral evolution in adolescents. AIDS. 2007 Nov 30;21(18):2387–97.
9. Yindom L-M, Leligdowicz A, Martin MP, Gao X, Qi Y, Zaman SM a, et al. Influence of HLA class I and HLA-KIR compound genotypes on HIV-2 infection and markers of disease progression in a Manjako community in West Africa. J Virol. 2010;84(16):8202–8.
10. Naranbhai V, Carrington M. Host genetic variation and HIV disease: from mapping to mechanism. Immunogenetics. 2017 Aug 10;69(8–9):489–98.
11. Spiegelhalter DJ. The art of statistics : learning from data. 1st ed. Penguin, editor. London: Random House; 426 p.
12. Sirugo G, Williams SM, Tishkoff SA. The Missing Diversity in Human Genetic Studies. Cell. 2019;177(1):26–31.
13. Campbell MC, Tishkoff SA. African genetic diversity: Implications for Human Demographic History, Modern Human Origins, and Complex Disease Mapping. 2008;

14. Shrestha S, Irvin MR, Taylor KD, Wiener HW, Pajewski NM, Haritunians T, et al. A genome-wide association study of carotid atherosclerosis in HIV-infected men. AIDS. 2010 Feb 20;24(4):583–92.

15. McLaren PJ, Coulonges C, Ripke S, van den Berg L, Buchbinder S, Carrington M, et al. Association study of common genetic variants and HIV-1 acquisition in 6,300 infected cases and 7,200 controls. PLoS Pathog. 2013;9(7):e1003515.

16. McLaren PJ, Coulonges C, Bartha I, Lenz TL, Deutsch AJ, Bashirova A, et al. Polymorphisms of large effect explain the majority of the host genetic contribution to variation of HIV-1 virus load. Proc Natl Acad Sci U S A. 2015 Nov 24;112(47):14658–63.

17. Gurdasani D, Iles L, Dillon DG, Young EH, Olson AD, Naranbhai V, et al. A systematic review of definitions of extreme phenotypes of HIV control and progression. AIDS. 2014 Jan 14;28(2):149.

18. Yant SR, Mulato A, Hansen D, Tse WC, Niedziela-Majka A, Zhang JR, et al. A highly potent long-acting small-molecule HIV-1 capsid inhibitor with efficacy in a humanized mouse model. Nat Med. 2019;25(9):1377–84.

19. James KL, de Silva T, Brown K, Whittle H, Taylor S, McVean G, et al. Low-bias RNA sequencing of the HIV-2 genome from blood plasma. Kirchhoff F, editor. J Virol. 2018 Oct 17;93(1).