

https://dspace.org/

Co-evolution of antibody lineages and HIV epitopes

Mabvakure, Batsirai Macdonald

2019
https://hdl.handle.net/10539/29746

Mabvakure, Batsirai Macdonald, Co-evolution of antibody lineages and HIV epitopes, University of the Witwatersrand, Johannesburg, <http://hdl.handle.net/10539/29746>

Downloaded from a DSpace repository. https://dspace.org/

Co-evolution of antibody lineages and HIV epitopes

Batsirai Macdonald Mabvakure

A thesis submitted in fulfilment of the requirements for the degree of
Doctor of philosophy (PhD)
University of the Witwatersrand, Johannesburg

Supervisors: Prof Penny Moore, Prof Lynn Morris and Dr Cathrine Scheepers

March, 2019

Declaration

I, Batsirai Macdonald Mabvakure (Student number: 892141), declare that this thesis is my own, unaided work. It is being submitted for the Degree of Doctor of Philosophy at the University of the Witwatersrand, Johannesburg. It has not been submitted before for any degree or examination at any other University.

Batsirai B. Mabvakure 18 March 2019

Signed: 

Publications

The following publications arose from the work presented in this thesis.

- **Mabvakure B.M**, Scheepers C, Garrett N, Abdool Karim S, Williamson C, Morris L, and Moore P.L. Positive selection at key residues in the HIV Envelope distinguishes broad and strain-specific plasma neutralizing antibodies. Journal of Virology. 2018 Dec 19. pii: JVI.01685-18. doi: 10.1128/JVI.01685-18

- **Mabvakure B.M**, Lambson B.E, Ramdayal K, Masson L, Kitchin D, Ali M, Abdool Karim S, Williamson C, Passmore J, Martin D, Scheepers C, Moore P.L, Harkins G.W and Morris L. Evidence for both intermittent and persistent compartmentalization of HIV-1 in the female genital tract. Journal of Virology. 2019 Mar 6. pii: JVI.00311-19. doi: 10.1128/JVI.00311-19

- **Mabvakure B.M**, Rott R, Dobrowsky L, van Heusden P, Morris L, Scheepers C, and Moore P.L. Application of a low-cost bioinformatics cluster for the analysis of large antibody sequence data. (Manuscript in preparation)

Co-authorships

- Van Eeden C, Wibmer C.K, Scheepers C, Richardson S, Nonyane M, Lambson B, Mkhize N, Vijayakumar B, Sheng Z, Stanfield-Oakley S, Bhiman J, Bekker V, Hermanus T, **Mabvakure B.M**, Ismail A, Moody A, Wiehe K, Garrett N, Karim SA, Dirr H, Fernandes M, Sayed Y, Shapiro L, Ferrari G, Haynes B, Moore P.L and Morris L. V2-directed vaccine-like antibodies from HIV-1 infection identify a novel K169-binding light chain motif with broad ADCC activity. Cell Reports. 2018 Dec 11;25(11):3123-3135.e6. doi: 10.1016/j.celrep.2018.11.058

- Richardson SI, Chung AW, Natarajan H, **Mabvakure B.M**, Mkhize N.N, Garrett N, Abdool Karim S.S, Moore P.L, Ackerman M.E, Alter G, Morris L. HIV-specific Fc effector function early in infection predicts the development of broadly neutralizing antibodies. PLoS Pathog.2018; 14(4): e1006987. https://doi.org/10.1371/journal.ppat.1006987

Conference outputs from this thesis

- **Oral**: **Mabvakure B.M**, Scheepers C, Garrett N, Abdool Karim S, Williamson C, Morris L and Moore P.L. Targeted viral selection is associated with the development of HIV broadly neutralizing antibodies. HIV Research for Prevention (HIVR4P), Spain, 25 October 2018

- **Oral**: **Mabvakure B.M**, Lambson B.E, Ramdayal K, Masson L, Kitchin D, Ali M, Abdool Karim S, Williamson C, Passmore J, Martin D, Scheepers C, Moore P.L, Harkins G.W and Morris L. Evidence for both intermittent and persistent compartmentalization of HIV-1 in the female genital tract. University of the Witwatersrand Faculty of Health Sciences Research Day and Postgraduate Expo, Johannesburg, South Africa. 6 September 2018

- **Oral**: **Mabvakure B.M**, Lambson B.E, Ramdayal K, Masson L, Kitchin D, Ali M, Abdool Karim S, Williamson C, Passmore J, Walzl G, Martin D, Scheepers C, Moore P.L, Harkins G.W and Morris L. Evolutionary pathways of HIV-1 subtype C viruses in the female genital tract. HIV Dynamics and Evolution Conference, Leavenworth, Washington, USA. 22 April 2018

- **Poster**: **Mabvakure B.M**, Scheepers C, Nonyane M, Lambson B, Madzorera S, Kitchin D, Bhiman J, Abdool Karim S, Williamson C, Morris L and Moore P.L. Broadly neutralizing antibodies to the HIV-1 V1V2 region are associated with distinct viral evolutionary pathways. University of the Witwatersrand Molecular Biosciences Research Thrust Postgraduate Research Day, 30 November 2017

- **Oral**: **Mabvakure B.M**, Scheepers C, Nonyane M, Lambson B, Madzorera S, Kitchin D, Bhiman J, Wibmer K, Abdool Karim S, Williamson C, Morris L and Moore P.L. Diversity analyses of HIV-1 envelope glycoproteins in HIV-infected individuals with and without broadly neutralizing antibodies. University of the Witwatersrand Faculty of Health Sciences Research Day and Postgraduate Expo, 1 September 2016

- **Poster**: **Mabvakure B.M**, Lambson B, Ramdayal K, Masson L, Moore P, Abdool Karim S, Scheepers C, Williamson C, Martin D, Passmore J, Harkins G and

Morris L. Analysis of HIV-1 subtype C envelope sequence diversity in the female genital tract from early infection reveals no evidence for compartmentalization. 21st International AIDS Conference (AIDS 2016). Inkosi Albert Luthuli International Convention Centre, Durban, South Africa, 18-22 July 2016

- **Poster**: **Mabvakure B.M**, Scheepers C, Nonyane M, Lambson B, Madzorera S, Kitchin D, Bhiman J, Wibmer K, Abdool Karim S, Williamson C, Morris L and Moore P.L. Diversity analyses of HIV-1 envelope glycoproteins in HIV-infected individuals with and without broadly neutralizing antibodies. 21st International Bioinformatics Workshop on Virus Evolution and Molecular Epidemiology (VEME), Korea University College of Medicine, Seoul, Republic of Korea, 14-19 August 2016

Abstract

The HIV envelope glycoprotein (Env) is enormously variable, making it very difficult to represent this diversity in a vaccine. Exposure of the immune system to the continuously diversifying viral quasispecies leads to an “arms race” that contributes to the evolution of the antibody response, and virus escape. Another consequence of this within-host HIV diversity is that viral populations may differ between compartments, with phenotypic effects that may impact the sensitivity of viruses to vaccines/microbicides. The aim of the study is to investigate the role of HIV envelope diversity in driving development of strain-specific or broadly neutralizing (bNAb) responses among hosts, and compartmentalization within hosts, using low-cost computational infrastructure.

We first compared viral diversity among hosts by studying eight HIV-1 infected participants who developed bNAbs and six donors who did not develop bNAbs over three years of infection. Defining the viral evolution that contributes to breadth is a major focus of the field so as to elicit bNAbs through vaccination. Comparison of Envs from individuals with bNAbs and strain-specific antibodies revealed that gross evolutionary measures such as Env diversity, evolutionary rates and number of viral infections are not sufficient to distinguish the two groups. However, bNAb donors had more positive selection within targeted sites, with some residues shared between bNAb donors. The positive selection at these common sites increased with the onset of neutralization breadth. The association of shared, positively selected sites with the emergence of breadth highlights the importance of diversity at these positions in bNAb development.

We next investigated the viral properties in different compartments within the same individual, as this is important for understanding the potential efficacy of HIV-directed antibodies and microbicides. We compared longitudinal viral Envs between the genital tract and blood plasma of four women. We found evidence of compartmentalization in two of the four individuals. In all four individuals there was restricted viral migration from the genital tract to the systemic blood circulation suggesting a mucosal sieve effect. However, this restricted viral migration was more pronounced in the two individuals who had evidence of compartmentalization. We identified several sites in the Env that differed between viruses in the two compartments, and that arose through

positive selection and/or indels. This resulted in charge distribution differences which may differentially impact the sensitivity of these viruses to antibody responses. Overall, these findings show that HIV viruses in the female genital tract may evolve differently from those in plasma, and develop properties that may potentially impact the efficacy of vaccines.

These studies relied on high numbers of Env sequences isolated at multiple timepoints, making these analyses computationally intensive. We thus designed and implemented a low-cost bioinformatics cluster for analysing large sequencing datasets (including next generation sequencing). Cluster design included meeting memory requirements, parallelization and long computational times. In addition to supporting viral studies described above, we used the cluster to trace an antibody lineage isolated from an HIV-infected donor, CAP228. The antibody lineage, CAP228-16H/19F, exhibited genetic characteristics similar to the monoclonal antibody, CH58, isolated from the RV144 human vaccine trial. Our analyses identified the unmutated common ancestor (UCA) that initiated the CAP228-16H/19F lineage, and showed that this was derived from a novel allele that encodes an amino acid DDxD motif that was responsible for binding to the V1V2 epitope on Env. This is in contrast to previous studies which suggested that binding was limited to an unusual ED motif that occurs rarely in the immunoglobulin repertoire. The identification of the DDxD motif thus demonstrated that there is a greater potential for eliciting these antibody responses through vaccination as the alleles encoding this motif are common in the population. Overall, this approach to high-performance computing is important as an alternative for data processing challenges faced in low-income communities that are unable to afford advanced computational infrastructure.

In conclusion, firstly, these analyses show the importance of diversity at specific amino acid residues in driving bNAb development, which may have valuable implications for vaccination design. Secondly, the study addresses the important subject of whether HIV compartmentalization occurs or not between the female genital tract and plasma, with findings that may impact on vaccine design. Lastly, we demonstrate the development and application of cheaper data processing infrastructure, an approach which may help researchers from low-income settings to overcome challenges in performing computationally intensive analyses.

Acknowledgements

Special gratitude to my supervisors Professor Penny Moore, Professor Lynn Morris and Dr Cathrine Scheepers for the opportunity to work in the lab and the amazing supervision and guidance. I would like to thank Dr Bronwen Lambson for the great insights, and my office mates (David Sacks, Rebecca van Dorsten, Aga Szydlik, Thandeka Moyo, Simone Richardson, Boitumelo Motsoeneng, Vincent Hlatshwayo and Dale Kitchin) for the insightful and fun discussions, and encouragement greatly needed throughout the course of the study. I am grateful to Raymond Rott, Lesley Drobowsky, Mushal Ali (Bioinformatics Sequencing Core NICD) and Peter van Heusden (South African National Bioinformatics Institute) for teaching and supporting me with the computational aspects of the study. Many thanks to Carina Kriel for all the help with the administration stuff, and making sure that I had all the resources I needed to be a happy student. I would like to thank donors at the Center for the AIDS Programme of Research in South Africa (CAPRISA) for contributing the precious samples used in the study, the CAPRISA staff for managing the cohort and collecting the samples, and past and present lab members who were involved in the generation of Single Genome Env Amplicons and performing neutralization assays. Special thanks to my wife Petronella Mabvakure for her continuous support and encouragements. I acknowledge funding from the Department of Science and Technology through the National Research Foundation (NRF) South African Research Chairs Initiative (SARChI), the Poliomyelitis Research Foundation (PRF grant number: 16/66) and the South African Medical Research Council (MRC) through the SHIP program. Last but not least, I am grateful to God for ordering my steps and directing my path of life.

Table of Contents

Declaration ..... *i*

Publications ..... *ii*

Conference outputs from this thesis ..... *iii*

Abstract ..... *v*

Acknowledgements ..... *vii*

List of Figures ..... *xii*

List of Tables ..... *xiv*

Ethics ..... *xv*

List of abbreviations ..... *xvi*

1 Introduction *1*

1.1 First recognition of HIV and spread of AIDS **1**

1.2 Phylogenetic studies of the origin of HIV **2**

1.3 Transmission of HIV **4**

1.4 Structure of HIV **5**

1.4.1 Genome organization of HIV-1 5

1.4.2 The HIV Env glycoprotein 7

1.5 HIV infection and viral diversification **9**

1.5.1 The HIV replication cycle 9

1.5.2 Intra-host evolution of HIV-1 11

1.5.3 HIV compartmentalization 12

1.6 Measures to treat and prevent HIV infection **14**

1.6.1 HIV treatment 14

1.6.1 HIV prevention 15

1.7 Host response to HIV **17**

1.7.1 Structure, diversity and multiplicity of antibodies 17

1.7.2 Development of binding and neutralizing responses against HIV Env 20

1.7.3 The development of bNAbs during infection 20

1.7.4 Broadly neutralizing antibody (bNAb) targets 21

1.8 Co-evolution of antibody lineages and HIV epitopes **24**

1.9 Computational challenges of viral and antibody evolutionary analyses **26**

1.10 Objectives and aims **27**

1.10.1 Overall project objective 27
1.10.2 Specific research aims 27

2 Positive selection at key residues in the HIV Envelope distinguishes broad and strain-specific plasma neutralizing antibodies 28

2.1 Abstract 29

2.2 Introduction 30

2.3 Materials and Methods 32

2.3.1 HIV infected participants and Ethics 32
2.3.2 Single Genome Envelope Amplification and Sequencing 32
2.3.3 Sequence Alignments and Data Partitioning 33
2.3.4 Estimating Diversity, Nucleotide Substitution Rates and Time of Infection 33
2.3.5 Identifying Sites Under Positive Selection Pressure 34

2.4 Results 35

2.4.1 Env diversity and evolutionary rates do not distinguish broad and strain-specific donors with equivalent viral loads 35
2.4.2 Comparison of Env evolution within antibody-targeted regions between bNAb and non-bNAb donors 38
2.4.3 Participants with bNAbs show high levels of positive selection within targeted epitopes 40
2.4.4 Common sites in the V2, C3 and V3 regions are targeted by bNAbs 42
2.4.5 Increased positive selection is associated with neutralization breadth 45

2.5 Discussion 48

3 Evidence for both intermittent and persistent compartmentalization of HIV-1 in the female genital tract 52

3.1 Abstract 53

3.2 Introduction 54

3.3 Materials and methods 56

3.3.1 Participants 56
3.3.2 Collection of CVL samples 56
3.3.3 Isolation of HIV RNA from CVL supernatant and plasma 57
3.3.4 Single genome amplification (SGA) and gp160 sequencing 57
3.3.5 Measurement of cytokines 58
3.3.6 Sequence alignments 59
3.3.7 Phylogenetic analyses 59
3.3.8 Bayesian analysis of viral compartmentalization 59
3.3.9 Analysis of HIV migration between CVL and plasma compartments 60
3.3.10 Identifying amino acid differences in viruses from different compartments 60

3.4 Results 61

3.4.1 HIV-1 gp160 amplification from the female genital tract over time 61
3.4.2 Association between CVL cytokine levels and envelope PCR during acute infection 62
3.4.3 Compartment-specific clusters in plasma and CVL samples 64
3.4.4 Statistical analyses of compartmentalization 65
3.4.5 HIV migration between genital tract and plasma compartments 67
3.4.6 Comparison of evolutionary differences between sequences in plasma and CVL 68
3.4.7 Comparison of CVL and plasma amino acid sequences in participants with evidence of compartmentalization 70
3.5 Discussion 72
4 Design and application of a low-cost bioinformatics cluster for resource-limited settings 80
4.1 Abstract 81
4.2 Introduction 82
4.3 Methods 84
4.3.1 Development of a multi-node cluster 84
4.3.2 Cluster configuration 85
4.3.3 Bioinformatics cluster file system 85
4.3.4 Package manager and installation of bioinformatics packages 88
4.4 Results 90
4.4.1 Cluster performance with computationally intensive programs 90
4.4.2 Cluster performance with large datasets 91
4.4.3 Easy integration of bioinformatics programs on the cluster 93
4.4.4 Case study: Antibody analysis of the CAP228-16H/19F antibody lineage using SONAR 94
4.5 Discussion 103
5 General Discussion and Conclusions 105
5.1 General Discussion 105
5.2 General Conclusion 108
6 References 110
7 Appendices 151
7.1 Appendix A: Mabvakure et al, Journal of Virology, 2018 (Proofs) 152
7.2 Appendix B: Van Eeden et al, Cell Reports, 2018 168
7.3 Appendix C: Ethics Certificate 182
7.4 Appendix D: Supervisor's letter 183

7.5 Appendix E: Support letter - Publications included in the thesis 184

7.6 Appendix F: Certificate of submission - student 186

7.7 Appendix G: Certificate of submission - supervisors..... 188

7.8 Appendix H: Turn-it-in Report 190

List of Figures

Figure 1.1: Estimated numbers of people living with HIV in 2016. 2
Figure 1.2: Origin of HIV. 3
Figure 1.3: Global distribution of HIV-1 viral subtypes. 4
Figure 1.4: Genome organization and structure of HIV. 7
Figure 1.5: Schematic of the *env* gene and structure of the HIV Env glycoprotein ... 8
Figure 1.6: Overview of HIV replication cycle 10
Figure 1.7: High evolutionary rates of HIV. 11
Figure 1.8: Limited migration of viruses between different anatomical tissues may result in HIV compartmentalization.. 13
Figure 1.9: The generation of antibody diversity through gene recombination. 18
Figure 1.10: B cell response to antigenic stimulation. 19
Figure 1.11: Antibody-HIV virus co-evolution. 21
Figure 1.12: BNAb targets, potency and neutralization breadth. 22
Figure 2.1: Viral diversity and evolutionary rates are not the sole determinants of bNAb development. 37
Figure 2.2: Comparison of diversity and nucleotide substitution rates in epitope regions targeted by participants with/without bNAbs. 39
Figure 2.3: BNAb participants have more sites under positive selection than non-bNAb individuals. 41
Figure 2.4: Comparison of positively selected sites in the V2 region of participants with and without bNAbs 43
Figure 2.5: Comparison of positively selected sites in the C3 region of participants with and without breadth 44
Figure 2.6: Increased positive selection at the onset of neutralization breadth 46
Figure 3.1: Association between cytokine levels, STIs and gp160 amplification in CVL at acute infection 63
Figure 3.2: Maximum likelihood trees of CVL and plasma Env sequences from four donors constructed in PhyML. 65
Figure 3.3: Statistical analysis of compartmentalization using BaTS 66
Figure 3.4: Viral migration events between CVL and plasma compartments 68
Figure 3.5: Longitudinal Env sequence diversity analyses in plasma and CVL 69
Figure 3.6: Identification of HIV-1 Env glycoprotein sites that distinguish CVL and plasma sequences in individuals with evidence of compartmentalization 71

Figure 4.1: Bioinformatics cluster architecture 87
Figure 4.2: Access to bioinformatics programs on the cluster..... 89
Figure 4.3: Cluster performance with large datasets of antibody repertoire data..... 92
Figure 4.4: NGS analysis flowchart. 94
Figure 4.5: Visualization of the CAP228 NGS light chain sequence data. 99
Figure 4.6: Visualization of the CAP228 NGS heavy chain sequence data 100
Figure 4.7: Identifying clonally related sequences following SONAR analysis..... 102

List of Tables

Table 2.1: Description of HIV-infected participants, plasma neutralization breadth, set-point viral load and antibody targets. 35
Table 3.1: List of 18 CAPRISA individuals who donated early infection CVL samples used in this study. 62
Table 4.1: Comparison of the cluster performance with that of a high specification ordinary machine 91
Table 4.2: Table showing the number of sequences and processes for the CAP228 longitudinal dataset before running SONAR. 98

Ethics

Both the parent study involving the CAPRISA cohorts and this particular study were granted ethics clearance under the following certificate numbers: M080470 for the parent study and M160340 for this study. The ethics certificate for this study can be found in Appendix C.

List of abbreviations

AAV	Adeno-associated viruses
ADCC	Antibody-dependent cell-mediated cytotoxicity
AI	Association Index
AIC	Akaike Information Criterion
AIDS	Acquired Immunodeficiency Syndrome
API	Application Programming Interface
ARVs	Antiretrovirals
BaTS	Bayesian Tip-association Significance testing
BEAGLE	Broad-platform Evolutionary Analysis General Likelihood Evaluator
BEAST	Bayesian Evolutionary Analysis by Sampling Trees
BF	Bayes factor
BIC	Bayesian Information Criteria
BLAST	Basic Local Alignment Search Tool
bNAb	Broadly neutralizing antibody
BV	Bacterial Vaginosis
CA	Capsid
CAPRISA	Centre for the AIDS Programme of Research in South Africa
CCL	Chemokine ligand
CCR5	CC chemokine receptor 5
CD4bs	Cluster of differentiation 4 binding site
CDR	Complementarity-determining region
CH	Constant heavy chain
CPU	Central Processing Unit
CRFs	Circulating Recombinant Forms
CTL	Cytotoxic T lymphocytes
CVL	Cervico-vaginal lavage
CXCR4	C-X-C Chemokine Receptor Type 4
D gene	Diversity gene
DAMBE	Data analysis for molecular biology and evolution
DNA	Deoxyribonucleic Acid
DST	Department of Science and Technology
ED	Glutamic-Aspartic amino acid pair

EIA	Enzyme immunoassay
Env	Envelope
ESS	Effective sample sizes
Fab	Antigen-binding fragment
Fc	Crystallizable fragment
FR	Framework region
FUBAR	Fast Unbiased Bayesian Approximation
G	Gigabytes
G-CSF	Granulocyte Colony Stimulating Factor
Gag	Group-specific antigen
GARD	Genetic algorithm for recombination detection
GM-CSF	Granulocyte Macrophage Colony-Stimulating Factor
gp120	Glycoprotein120
gp160	Glycoprotein160
gp41	Glycoprotein41
GPUs	Graphics Processing Units
H chain	Heavy chain
HIV	Human Immunodeficiency Virus
HLA	Human Leukocyte Antigen
HPC	High-performance computing
HR	Heptad repeat
HSV-2	Herpes simplex virus 2
HVTN	HIV Vaccine Trials Network
HyPhy	Hypothesis testing using Phylogenies
IFN	Interferon
Ig	Immunoglobulin
IL	Interleukin
IMGT	International immunogenetics information system
IN	Integrase
J gene	Joining gene
L chain	Light chain
LAN	Local Area Network
LANL	Los Alamos National Laboratory
LTO	Linear Tape-Open
M	Major

mAb	Monoclonal antibody
MAC	*Mycobacterium avium* complex
MC	Maximum exclusive single-state clade
MCC	Maximum Clade Credibility
MCMC	Markov chain Monte Carlo
MCP	Monocyte Chemotactic Protein
MEGA6	Molecular Evolutionary Genetics Analysis Version 6
MEME	Mixed Effects Model of Evolution
MIP	Macrophage Inflammatory Protein
MPER	Membrane proximal external region
MRC	Medical Research Council
MUSCLE	Multiple Sequence Comparison by Log-Expectation
N	Non-Major, non-Outlier
NCBI	National Center for Biotechnology Information
Nef	Negative Regulatory Factor
Neg	Negative
NFS	Network file system
NGS	Next Generation Sequencing
NIAID	National Institute of Allergy and Infectious Diseases
NIH	National Institutes of Health
O	Outlier
OPC	*Oropharyngeal candidiasis*
PBMCs	Peripheral blood mononuclear cells
PCP	*Pneumocystis jirovecii*
PCR	polymerase chain reaction
PEAR	Paired-end read merger
PEP	Post-Exposure Prophylaxis
Pol	Polymerase
Pos	Positive
PrEP	Pre-Exposure Prophylaxis
PRF	Poliomyelitis Research Foundation
PS	Fitch Parsimony Score
RANTES	Regulated upon Activation, Normal T cell Expressed, and Secreted
RNA	Ribonucleic Acid
RPM	Redhat Package Manager

RT	Reverse Transcriptase
SANBI	South African National Bioinformatics Institute
SARCHi	South African Research Chairs Initiative
sCD40L	Soluble CD40 Ligand
SGA	Single genome amplification
SHM	Somatic hypermutation
SIV	Simian Immunodeficiency Virus
SLURM	Simple Linux Utility for Resource Management
SONAR	Software for ontogenic analysis of antibody repertoires
STIs	Sexually transmitted infections
T	Terabytes
T/F	Transmitted/founder
TAP	Treatment as prevention
TdT	Terminal Deoxynucleotidyl Transferase
tMRCA	Time of the most recent common ancestor
TNF	Tumor Necrosis Factor
UCA	Unmutated common ancestor
UNAIDS	The Joint United Nations Programme on HIV and AIDS
UPS	Uninterruptible Power Supply
URFs	Unique Recombinant Forms
USA	United States of America
V gene	Variable gene
Vif	Virion Infectivity Factor
VL	Viral loads
Vpr	Viral Protein R
Vpu	Viral Protein U
VRC	Vaccine Research Center
WHO	World Health Organisation

1 Introduction

1.1 First recognition of HIV and spread of AIDS

Acquired Immunodeficiency Syndrome (AIDS) is a disease caused by a lentivirus called the Human Immunodeficiency Virus (HIV). AIDS presents as a set of symptoms such as fever, fatigue, swollen lymph nodes, and opportunistic infections that arise later during infection as a result of a weakened immune system (1–3). The opportunistic infections associated with HIV include *cryptococcal meningitis*, HIV-related encephalopathy, Cytomegalovirus disease (particularly CMV retinitis), *Mycobacterium avium* complex (MAC), *Oropharyngeal candidiasis* (OPC) or Thrush, *Pneumocystis jirovecii* (PCP) pneumonia, and some cancers that include Kaposi's Sarcoma, Lymphoma and invasive cancers of the anus or cervix (1, 4, 5). HIV infection has also been associated with dementia accompanied by motor dysfunction, speech problems and behavioural change (6, 7).

AIDS was first recognized in 1981 in the USA, and HIV was later identified as the causative agent in 1983 through repeated isolation of the virus from infected patients (2, 3). HIV was later also isolated retrospectively from historic samples collected from patients in Kinshasa in the Democratic Republic of Congo as far back as 1959 (8). HIV spread was rapid in the 1980s, and AIDS quickly became a global epidemic. A total of 38,401 cases of AIDS had been reported in 85 countries in Africa, the Americas, Asia, Europe, and Oceania by the end of 1986 (9). The number of cases went up to 71,751 the following year and continued to grow to 142,000 in 1989 (in 145 countries) and to 307,000 by the end of 1999. This figure has continued to rise over the past 3 decades and as of 2018, 70 million people have been infected with HIV and about 35 million people have died since the start of the epidemic (9).

HIV is most prevalent in Africa, followed by Asia and the Pacific, Western and Central Europe and North America, Latin America and the Caribbean, East Europe and Central Asia, and Middle East and North Africa (Figure 1.1) (10). Sub-Saharan Africa has around 66% of the total HIV infections in the world, with women accounting for more than half of these (11). As a single country, South Africa has the largest number of

people living with HIV-1, with more than 7 million infected (12, 13). Most of these countries that are particularly hard hit by the epidemic have high rates of poverty that promote unsafe sexual practices, limited access to preventative strategies as governments cannot afford to provide subsidies, and high rates of human rights abuse (14–18). Socio-economic factors are therefore pivotal in the spread of HIV.

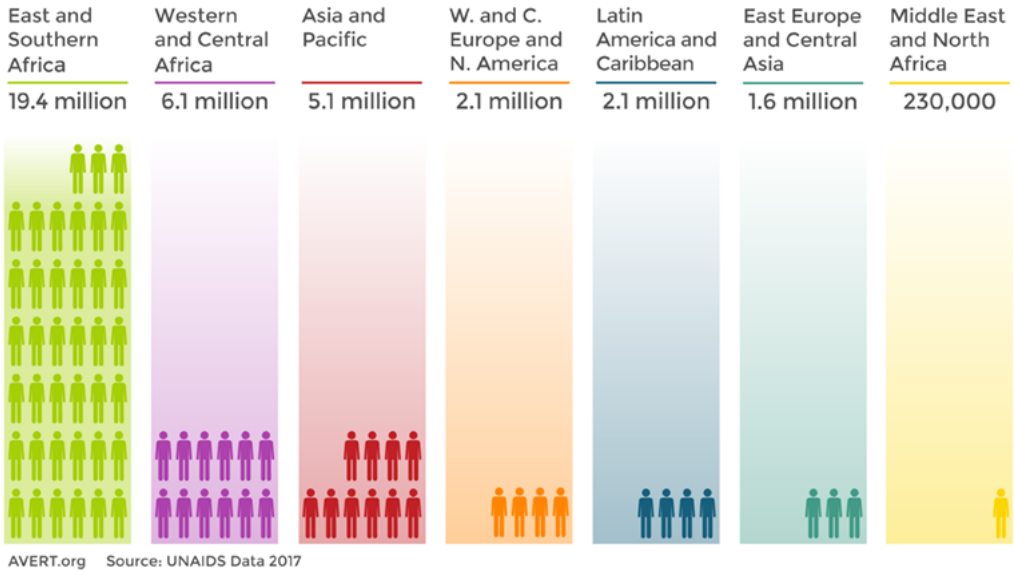


Figure 1.1: Estimated numbers of people living with HIV in 2016. The most affected regions are East and Southern Africa (19,4 million), followed by Western and Central Africa (6,1 million), Asia Pacific (5,1 million), Western and Central Europe and North America (2,1 million), Latin America and the Caribbean (2,1 million), Eastern Europe and Central Asia (1,6 million) and lastly Middle East and North Africa (230,000). Source: Avert, 2017

1.2 Phylogenetic studies of the origin of HIV

Phylogenetic studies have identified two closely related types of HIV, namely HIV-1 and HIV-2. HIV-2 is most prevalent in West Africa and was first identified in a Senegalese woman in 1985. However, HIV-2 is less virulent and progresses more slowly to AIDS than the more widespread HIV-1 (19–21). Sequence analysis shows high levels of similarity between HIV-1 and Simian Immunodeficiency Virus (SIV) from West and Central African chimpanzees, and between HIV-2 and SIV from West African Sooty Mangabeys (Figure 1.2), suggesting that these were the original sources of the zoonotic transmissions of the ancestors of these viruses (22, 23). These studies concluded that HIV-1 and HIV-2 were transmitted to humans through multiple zoonotic

events from chimpanzees and sooty mangabeys respectively (Figure 1.2). Bayesian statistical analyses have estimated the first transmission of HIV-1 to humans to have occurred in Kinshasa around 1920, and many studies suggest that this is where the global epidemic started (24, 25).

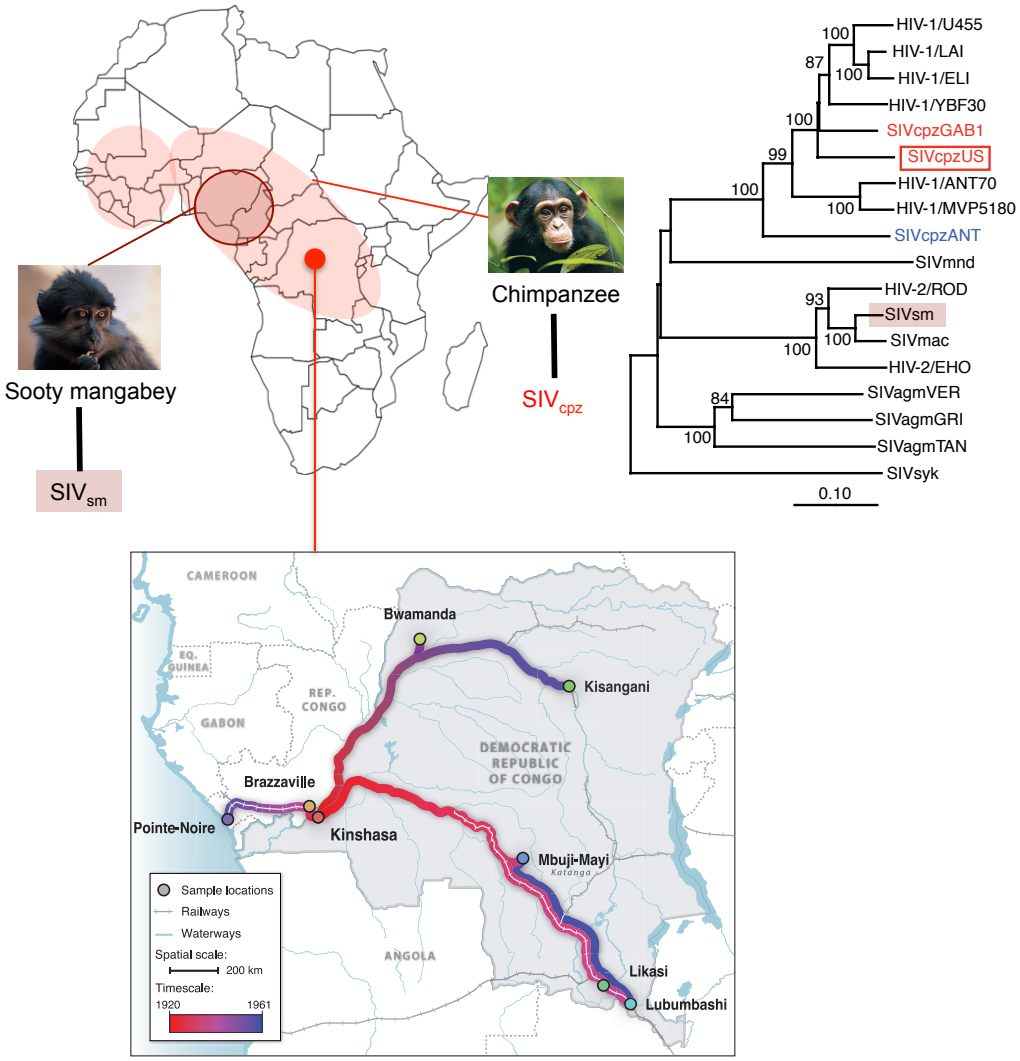


Figure 1.2: Origin of HIV. The top left map shows where the SIV hosts (Sooty Mangabeys and Chimpanzees) are located. Phylogenetic studies show that HIV originated from SIV from these sooty mangabeys and chimpanzee species from West and Central Africa, respectively (shown in the phylogenetic tree at the top right). The bottom map shows that the HIV pandemic originated in Kinshasa, Central Africa before it became a global epidemic (Gao et al. 1999).

HIV-1 is further categorized into three groups, referred to as M (major), N (non-M, non-O) and O (outliers), with the M group being further sub-divided into subtypes A – K. HIV-1 M, N, O and P each reflect separate introductions into the human population

and group O and P is most likely derived from gorillas (3, 26, 27). Subtype C is responsible for more than 48% of global infections, and is dominant in Africa and in India, whereas subtype B viruses are responsible for most infections in Europe and the Americas (Figure 1.3) (22). Central Africa has the most diverse HIV subtypes globally, likely a consequence of the emergence of the epidemic in this region. In addition to subtypes, circulating recombinant forms (CRFs) and unique recombinant forms (URFs) are increasingly being recognised.

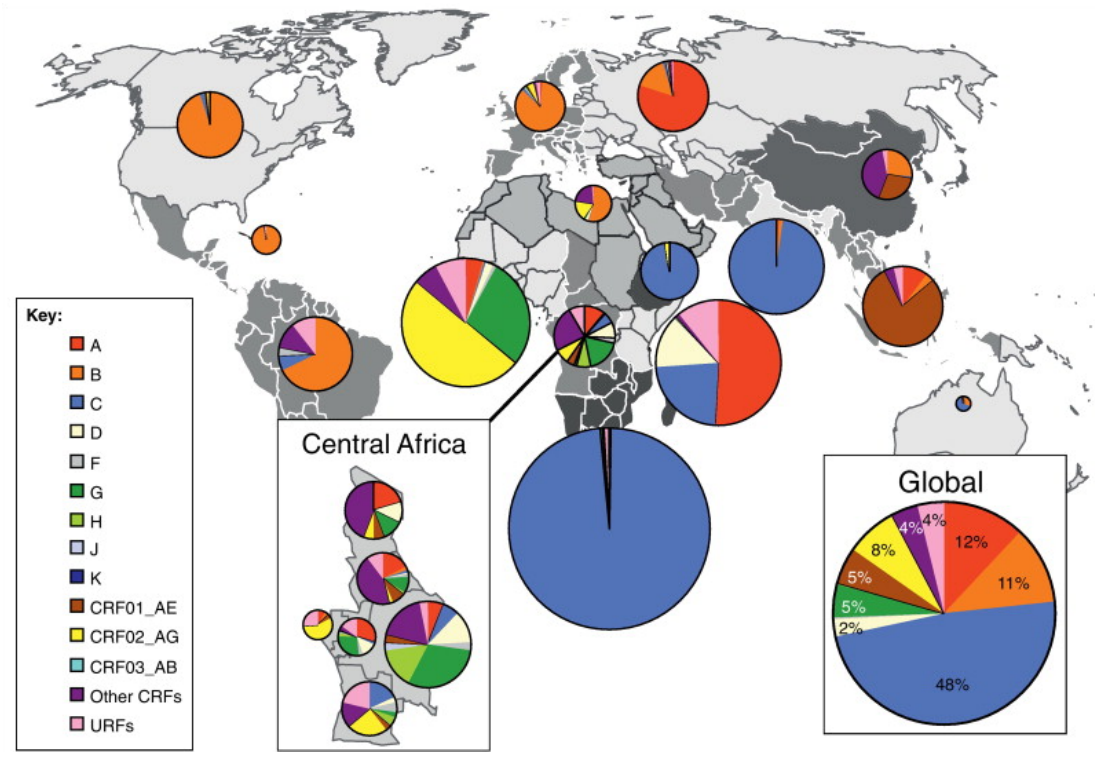


Figure 1.3: Global distribution of HIV-1 viral subtypes. Subtype C is dominant in sub-Saharan Africa whereas subtype B is more dominant in Europe and America (Hemelaar 2013). CRF; circulating recombinant form. URF; unique recombinant form.

1.3 Transmission of HIV

The global spread of HIV amongst humans has been largely as a result of sexual transmission, with the global epidemic now driven by heterosexual transmission (28). Transmission through intravenous drug use and from mother to child during birth and

breast-feeding also occurs, but these are not the major modes of transmission globally (28). In heterosexual intercourse, women are at a greater risk of acquiring HIV. The reasons for this increased risk are multifactorial, and include physical damage to the vaginal lining during intercourse, the larger exposed surface area of the vagina, and an enrichment of cells in the vaginal tissue lining that favour HIV acquisition. Together with socioeconomic issues, these factors account for the disproportionately higher infection rates amongst women compared to men (29–32). Sexually transmitted Infections (STIs) that cause genital inflammation have also been reported to elevate HIV shedding in the genital tract thereby increasing the chance of infection (33–36). The risk of HIV transmission is also increased when the infecting partner has high viral loads, and studies suggest that there is a 2.5-fold increase in transmission for every 10-fold increase in viral load (37, 38).

Many studies have sought to understand whether specific viral attributes are associated with infection. A key finding of such studies has been that approximately 80% of sexually transmitted infections are initiated by a single viral variant, an effect referred to as a genetic bottleneck (39–46). This bottleneck is less stringent in cases such as intravenous drug use and in the context of high levels of inflammation, resulting in more frequent transmission of multiple strains at the same time (multi-variant infection) (47, 48). The genetic bottleneck suggests that viruses that initiate HIV infection, which are referred to as transmitted/founder (T/F) viruses, have phenotypic features that increase their chances of productive infection. Indeed, several studies have indicated that T/F viruses show enhanced cell-free infectivity, improved interactions with dendritic cells and increased IFN (interferon)-α resistance (49, 50).

1.4 Structure of HIV

1.4.1 Genome organization of HIV-1

The 9,7kb long HIV genome consists of two single-stranded, non-covalently linked, unspliced, positive-sense RNA molecules (Figure 1.4). These RNA molecules have genes that code for the various viral proteins that are required for viral replication (which is described in further detail below). The *gag* gene produces the Gag precursor

protein that is further processed into the Matrix, Capsid, Nucleocapsid, p1, p2 and p6 proteins. The *pol* gene encodes the Viral Protease, Reverse Transcriptase and Integrase enzymes. The *tat* and *rev* genes code for viral regulatory factors that are important for gene expression, and the *vif*, *vpr, vpu*, and *nef* genes code for accessory or auxiliary proteins that are not essential for replication (51). Vif (Viral Infectivity Factor) promotes infectivity as viral particles are defective in its absence. Vpr (viral protein R) plays a role in nuclear import of pre-integration complexes, cell growth arrest, transactivation of cellular genes and induction of cellular differentiation. Vpu (Viral protein U) is responsible for CD4 degradation in the endoplasmic reticulum and the enhancement of virion release from the plasma membrane of HIV-1 infected cells. Vpu is present in HIV-1 but absent in HIV-2. The Nef protein has a number of functions that include down-regulation of CD4 and Major Histocompatibility Complex (MHC) class 1 molecules. The *env* gene is responsible for coding the HIV envelope glycoprotein which is the focus of vaccine development, and of this thesis, because it is the major target of the humoral immune response (described in more detail below).

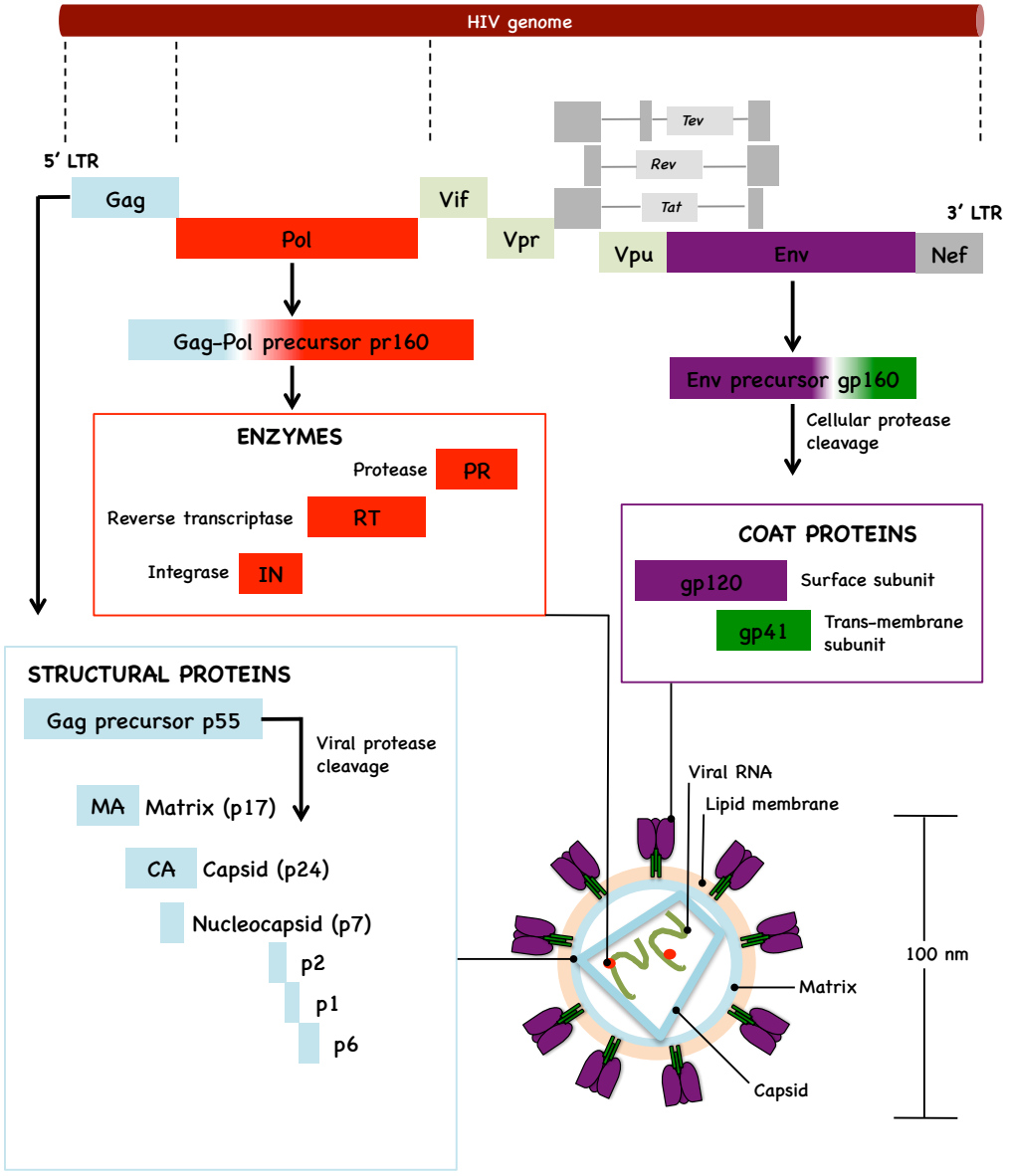


Figure 1.4: Genome organization and structure of HIV. The genome is 9,7 kilobases long and encodes enzymes, structural (matrix, capsid and nucleocapsid) and envelope coat proteins (surface and trans-membrane subunits) necessary for viral replication. The genome is enclosed by structural proteins and surrounded by the lipid membrane (Ben Haij et al. 2011).

1.4.2 The HIV Env glycoprotein

HIV particles are surrounded by a spherical lipid bi-layer membrane, which surrounds the matrix and the capsid that protects the viral genome and proteins such as Reverse Transcriptase and Integrase. Embedded in the lipid bilayer are trimeric envelope glycoprotein spikes (shown in purple and green in Figure 1.4). The HIV envelope glycoprotein is the target of antibody responses and consists of three molecules each

of gp120 and gp41 proteins (Figure 1.5A and B). The gp120 molecules play an important role in receptor binding, co-receptor tropism, virus sensitivity to neutralization by antibodies, virus pathogenesis, and disease progression. The gp120 can be partitioned into five hypervariable regions (V1-V5), which are interspersed with five conserved regions (C1-C5) (Figure 1.5A) (52). The gp41 protein consists of an N terminal region that contains a hydrophobic fusion domain followed by N and C terminal heptad repeat (HR) regions that have important roles in fusion (53–55). The gp41 includes the tryptophan-rich membrane proximal external region (MPER) of HIV gp41 (53–55).

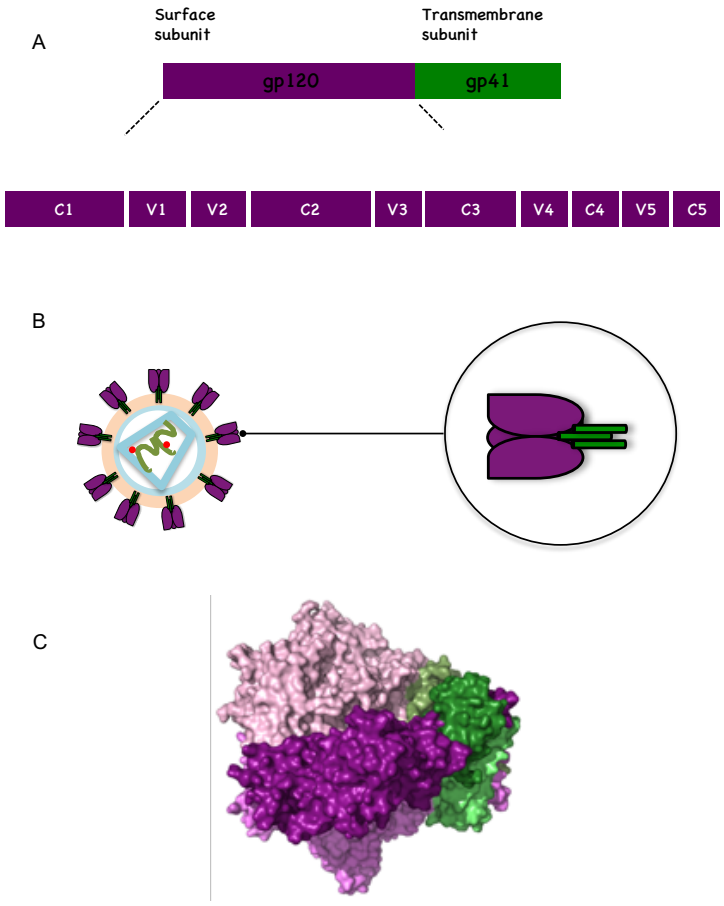


Figure 1.5: Schematic of the *env* gene and structure of the HIV Env glycoprotein within the viral particle. (A) The gp120 and gp41 genes are shown in purple and green respectively. The five variable regions of the gp120 molecule (V1-5) are interspersed with five constant regions (C1-5). (B) Schematic of the HIV virus particle and the Env glycoprotein, with the gp120 and gp41 subunits shown in purple and green respectively. (C) Model of the HIV Env trimer, with the shades of purple (gp120) and green (gp41) representing the three monomers that form a single trimeric structure.

The Env glycoprotein exists as a heterotrimeric complex that consists of three gp120 molecules that interact non-covalently with three transmembrane gp41 molecules (56)

(Figure 1.5C). The gp120 subunits have a compact globular configuration and are assembled into a trimer through intra- and inter-protomer interactions at the apex that involve the V1/V2 and V3 loop structures (57). The constant regions fold into a core structure, whereas the variable regions, with the exception of V5, form loops that are bracketed by disulfide bridges (58). The V2 loop is localized above the CD4 binding site (CD4bs), and the V3 loop is beneath the V1V2 hairpin with the tip pointing toward the trimer axis and interacting with the V2 base from the adjacent protomer (57). Crystallography studies show that the envelope glycoprotein structure folds into three major domains namely the inner domain, the outer domain and the bridging sheet (58, 59). The inner domain consists of a two-helix, two-strand bundle at the termini-distal end and a five-stranded β-sandwich at its termini-proximal end (58, 59). The stacked double barrel of the outer domain consists of a proximal mixed directional β-sheet, and shares one contiguous hydrophobic core and barrels located at the inner-outer domain interface (58, 59). The β-sheet links the outer and inner domains to form a third domain, the bridging sheet (58, 59). The gp41 is located at the base of the trimer (57).

1.5 HIV infection and viral diversification

1.5.1 The HIV replication cycle

During a productive infection by HIV, the Env protein adheres to the host cell by first binding to the primary cell receptor, CD4. Most of the residues involved in CD4 binding are located in the C3, C4 and V5 regions of the Env. This is followed by binding to a cellular chemokine co-receptor CCR5 or CXCR4, and then fusion of the virus and the host cell membrane follows afterwards (Figure 1.6) (60). Fusion requires domains in the gp41 that are critical for catalyzing the membrane fusion reaction between the viral and host lipid bilayers during virus entry (51).

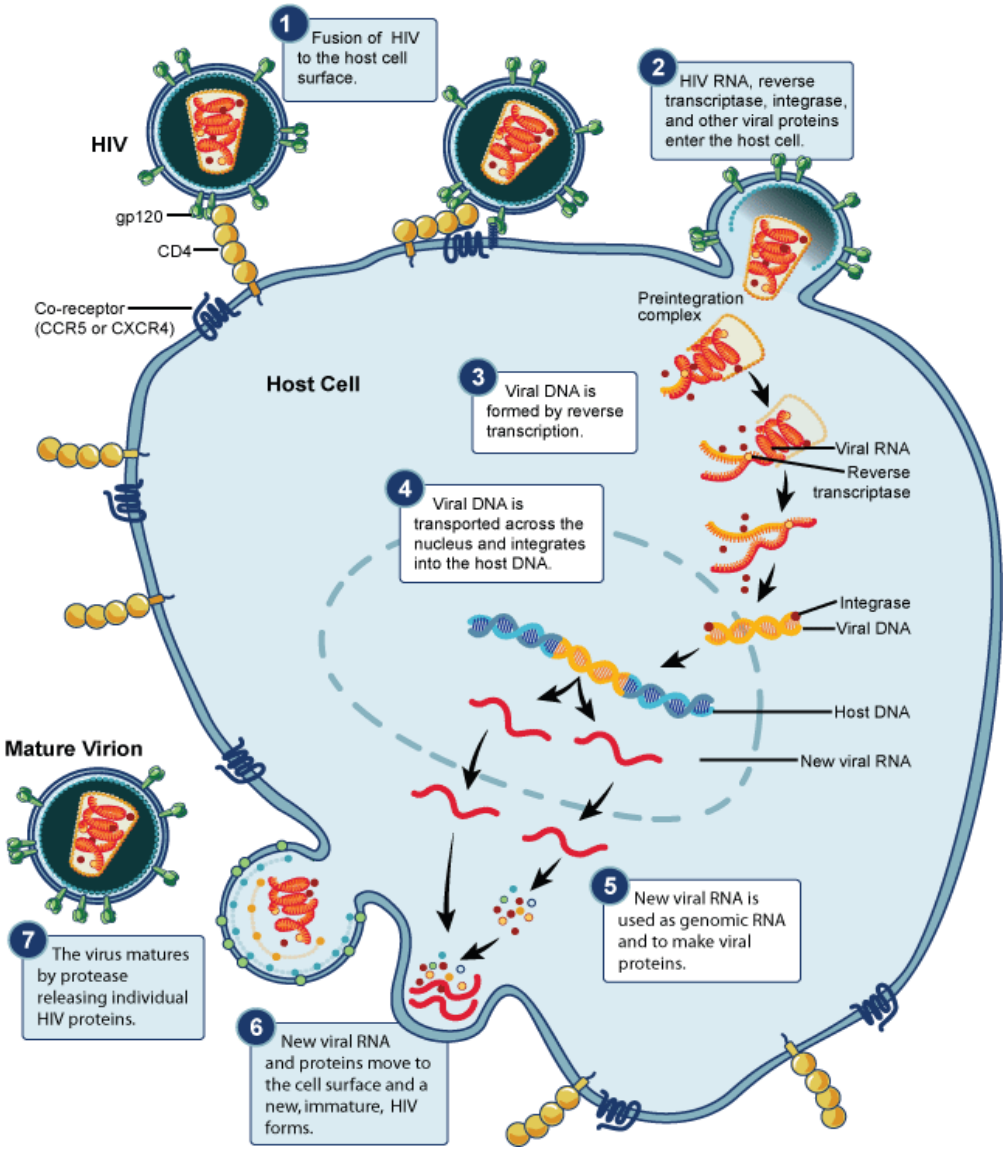


Figure 1.6: Overview of HIV replication cycle. HIV-1 attaches itself to the host cell, fuses into the cell and releases its genetic material into the cytoplasm. This is followed by reverse transcription and integration into the host genome and transcription to produce new particles that bud off as mature HIV-1 virions (Freed 2001).

Following fusion, the viral core is released into the host cell cytoplasm. The capsid is shed (in a process referred to as uncoating), and the viral RNA and proteins are converted to the pre-integration complex and the reverse transcription complex (61). The Reverse Transcriptase enzyme creates a copy of viral DNA from the RNA, which is imported into the nucleus to integrate with the host DNA using the Integrase enzyme (51, 62). The integrated provirus then starts directing viral replication by encoding viral RNA, structural, regulatory and accessory proteins. These nucleic and protein molecules move to the host cell membrane where they assemble and bud off to form mature HIV virions (Figure 1.6).

1.5.2 Intra-host evolution of HIV-1

HIV-1 is characterised by extremely high levels of diversity, particularly within the Env glycoprotein, which is the sole target of neutralizing antibody responses, described in more detail below (57, 63, 64). As an adaptive means to evade these responses during infection, the HIV Env has higher evolutionary rates compared to other genes, and may achieve an intra-host diversity of 25-35% (65, 66). These extraordinarily high levels of variability are a general phenomenon of many RNA viruses, which have mutation rates ranked second highest compared to all other microorganisms (Figure 1.7A and B).

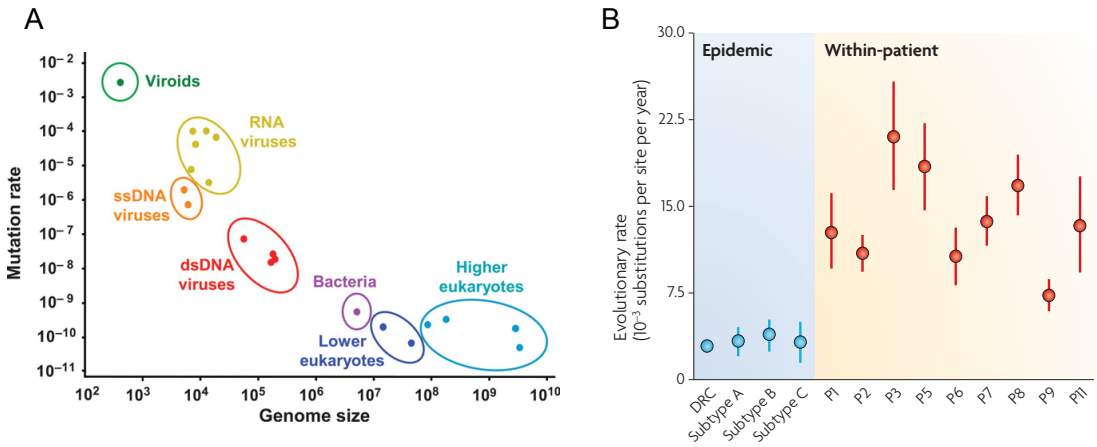


Figure 1.7: High evolutionary rates of HIV. (A) RNA viruses have very high evolutionary rates compared to DNA viruses, bacteria and eukaryotes. (B) Inter-host (epidemic) evolutionary rates of HIV-1 are lower compared to within patients, denoted by "P". (Source: Pybus and Rambaut, 2009; Gago et al, 2009)

There are several contributors to the high levels of variation. The virus Reverse Transcriptase introduces errors during viral replication as it has no proof reading mechanism (67). Another source of variation during replication is from the host RNA Polymerase II that also lacks proof reading function (67, 68). These two enzymes introduce errors at a combined rate of approximately 2×10^{-5} per nucleotide per replication cycle (67, 69, 70). Another major contributor to viral diversity is recombination. HIV-1 has very high recombination rates between the two RNA molecules that are packaged into the same virion (71–76). During replication, the

Reverse Transcriptase jumps from one RNA strand to another (template switching) resulting in the creation of a novel recombinant DNA genome containing sequences derived from both parental RNA (72). The effect of recombination on viral diversification is enhanced by infection of a cell with multiple variants, providing a great leap in diversity, specifically in Env. Studies have shown that individuals with single variant infections had lower mean *env* diversity (0.073% (0 - 0.566%)) compared to those with multiple infections (0.593% (0.026 - 5.98%)) (77). The importance of high HIV envelope diversity is in increasing the chances of viral escape from antibody responses and antiretroviral therapies such as entry inhibitors. Another consequence of HIV diversification is the creation of viral populations that differ between anatomical tissues. This is discussed in greater detail below.

1.5.3 HIV compartmentalization

During intra-host evolution of HIV, viruses in different anatomical tissues may start evolving independently, a phenomenon known as compartmentalization (Figure 1.8). This occurs as a consequence of anatomical barriers, tissue-specific immune selective pressures and/or distinct target cell populations in different tissues. HIV compartmentalization results in the creation of distinct viral variants that can enable rapid adaptation to a variety of selection pressures thereby contributing to viral fitness and disease progression (78).

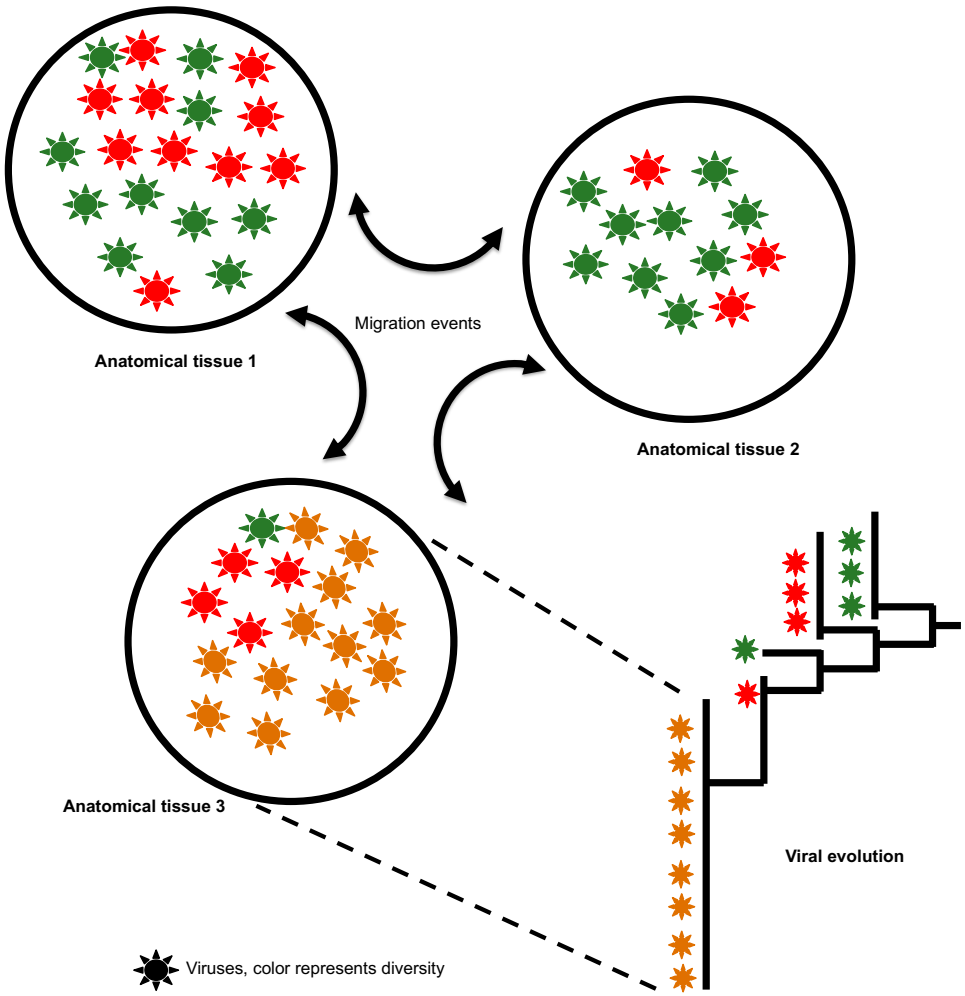


Figure 1.8: Limited migration of viruses between different anatomical tissues may result in HIV compartmentalization. The circles represent different compartments, and the stars represent the viral diversity in these tissues. The number of stars represents abundance. The phylogenetic tree shows the relatedness of the sequences and how compartmentalized viruses cluster together.

HIV compartmentalization has been reported in a number of tissues including the brain, breast milk, semen, cerebral spinal fluid, kidneys, peripheral blood mononuclear cells (PBMCs), liver, lungs, saliva, bone marrow, placenta and epidermal Langerhans cells (78–86). The strongest evidence of HIV compartmentalization has been detected in the brain and cerebrospinal fluid (87, 88). Drug resistant strains in the brain and other health associated complications such as dementia have been reported as a result of tissue-specific evolution of HIV in the brain (85, 89).

There is also convincing evidence of tissue-specific compartmentalization in the male genital tract (90–94). However studies in the female genital tract have been

inconclusive (95–98). Most studies in the female genital tract were cross-sectional analyses as opposed to longitudinal studies. Where longitudinal analyses were performed they involved limited time-points (95, 96). Earlier work also revealed that some of the studies might have been biased because of the inclusion of monotypic and low diversity sequences (95, 96). Furthermore, few of these studies used complete gp160 envelope sequences, and the tree-based phylogenetic methods employed for the analyses are limited by sequence availability (95–108). Since women are especially vulnerable to HIV, it is important to understand the viral populations replicating in the female genital tract in order to develop effective microbicides and vaccines which would prevent acquisition. This was also a focus of this thesis, forming the basis of Chapter 3.

1.6 Measures to treat and prevent HIV infection

1.6.1 HIV treatment

HIV treatments have been designed that suppress HIV viral load by affecting different stages of the HIV replication process as described above, targeting vulnerable conserved sites on the HIV structure including the envelope. These antiretroviral drugs (ARVs) have been proven to be effective in reducing the viral levels to undetectable levels when taken consistently (109–113). The United Nations 90-90-90 goals policy has thus been implemented to ensure that all HIV-infected individuals receive ARVs from time of infection regardless of CD4 cell count (114). The United Nations 90-90-90 goals aim to achieve 90% of all people living with HIV knowing their status, 90% of those people receiving sustained antiretroviral therapy and viral suppression in 90% of people on ARVs by 2020 (114).

Several milestones have been achieved using antiretroviral treatment, including reduced transmissions, and improved life expectancy (115–117). These antiretroviral drugs, however, have to be taken daily for life which may be accompanied by side-effects that are sometimes debilitating (116). In addition, some marginalized communities in low-income countries have limited access to this medication, and access to ARVs in some communities is associated with stigmatization (116). Some of

these challenges have led to non-adherence to treatment, resulting in the development and spread of drug resistant viral strains that are not responsive to treatment (116, 118). Surveillance data indicates that a new HIV epidemic of drug resistant strains is looming, which may be very difficult to control with the current medications (118, 119). In addition, despite South Africa having the largest ARV program in the world, statistics show that the number of new infections in this and other regions is still very high (12). It is therefore crucial to develop strategies that prevent HIV infection.

1.6.1 HIV prevention

HIV prevention strategies are likely the most feasible means to eradicate HIV as these stop further spread of the virus. Since HIV is commonly transmitted via sexual intercourse, prevention measures that target this route include targeting risky behavioural practices by teaching safe sex practices such as faithfulness to one's partner, knowing one's HIV status, use of condoms and encouraging sex abstinence in teenagers (120). Intravenous drug users are also at a very high risk of acquiring HIV through sharing infected needles (121, 122), and as such they have been classified as key population group for prevention measures.

Medical interventions to prevent HIV infection include medical male circumcision, Pre-exposure Prophylaxis (PrEP) in HIV uninfected people, and treatment as prevention (TAP) in HIV positive individuals. Studies have shown that male circumcision reduces the chances of HIV infection in men, and many programs have been put in place to promote medical male circumcision (123–127). PrEP aims to prevent HIV acquisition in HIV negative individuals who are especially vulnerable to HIV infection, such as partners of infected individuals (128). However, unlike in high-income countries, access to these drugs is sometimes limited to select countries, such as South Africa, and only to high-risk groups such as sex workers (129). These programs require government funding which is unaffordable for most countries in the long-term (130–132). Lastly, studies have shown that HIV infected people who use ARVs consistently for treatment and achieve viral suppression are unable to transmit the virus, providing more opportunities for HIV prevention through treatment (133, 134). Although these breakthroughs have undoubtedly reshaped the HIV prevention field, an effective vaccine remains elusive.

1.6.1.1 Development of an HIV vaccine by active immunization

A major focus of prevention research is the development of a vaccine, which has been shown to be the most effective medical intervention for many infectious diseases (135). Vaccines have a proven track record of eliminating such epidemics, and have the potential to elicit immunity against HIV in the uninfected population. A successful vaccination strategy achieves herd immunity in the population, preventing further infections until the epidemic eventually dies out from the population. HIV vaccine development has been complicated by the rapid evolutionary dynamics and variability of HIV, described above (136). Despite these challenges, there have been great advances in vaccine development against HIV-1. The RV144 Phase III human clinical trial of a vaccine consisting of ALVAC-HIV (vCP1521) with a boost of the AIDSVAX gp120 vaccine (an envelope protein segment from HIV subtypes B and E) showed 30% efficacy (137), and provided proof of concept that it is possible to elicit antibodies against the envelope glycoprotein via vaccination. This opened up many opportunities for research to improve the vaccination design, and many follow-up trials are underway (138). However this vaccine failed to elicit neutralizing antibodies, which are the protective mechanism for many licenced vaccines. Several parallel vaccine approaches therefore focus on understanding HIV Env evolution and the development of HIV-specific neutralizing antibodies during infection, as a template for vaccine design (139). These studies and their implications are discussed in more detail below in section 1.7.

1.6.1.2 Passive vaccination approaches to prevent HIV infection

Though a licenced vaccine able to elicit neutralizing antibodies has not yet been developed, several antibodies have been isolated from HIV infected individuals that are very effective in neutralizing the majority of HIV strains. These have the potential to be used in passive immunizations (140–143). These monoclonal antibodies are referred to as broadly neutralizing antibodies (bNAbs), and described in more detail in section 1.7.3. One such monoclonal antibody, called VRC01, is currently being tested

in a phase III trial by the HIV Vaccine Trials Network (HVTN) to assess its efficacy in preventing infection, and to define the protective dose.

Passive administration of is however complicated by high costs. There is significant effort being invested in improving the half-life of these antibodies, to reduce the frequency and volumes of transfusions. Antibody engineering is an active field, including making bi- or tri-specific antibodies, and smaller versions of antibodies, such as single chains, to improve tissue access. Delivery of these antibodies to the human circulatory system is also very important, and several studies aim to find more efficient ways than the traditional intravenous and subcutaneous administration. These methods include the use vaginal gels (144), adeno-associated viruses (AAV) (145), and engineered *lactobacillus* bacteria (146). Although these show promise, the development of a traditional vaccine which would provide life-long immunity still remains preferable, and relies on a deep understanding of the host immune response to HIV infection.

1.7 Host response to HIV-1

Following HIV-1 infection, the human immune system is activated to produce both T and B cell responses that are directed towards the HIV-1 proteins. The T and B cell responses occur early in infection, and target the HIV-1 structural, regulatory, accessory and Env proteins (147, 148). The B cell responses include neutralizing antibodies that are directed solely towards the HIV-1 envelope glycoprotein. The section below will briefly describe the structure of the antibodies and their development. This is important to enable better understanding how the antibodies function, and to inform the elicitation of such antibodies via vaccination.

1.7.1 Structure, diversity and multiplicity of antibodies

HIV-1 directed antibodies, like all other antibodies, consist of two heavy and two light chains that form the variable and constant regions (Figure 1.9). The human body is well adapted to counter different types of pathogens it encounters, including the large

diversity of HIV-1. This adaptation enables the human body to produce a large antibody repertoire that has been estimated to be greater than 10^{11} molecules per individual (149–152). This vast repertoire is encoded by a small group of variable (*V*), joining (*J*) and diversity (*D*) genes that determine the variable regions (Figure 1.9). There are 51 *V*, 27 *D*, 6 *J* heavy chain segments and this large number of gene segments available for encoding antibody chains makes a substantial contribution to antibody diversity (153). In addition, the joining of these segments, in a process known as combinatorial diversification, greatly increases antibody diversity, as any of the 51 *V* segments in the human heavy-chain pool can be joined to any of the 6 J segments and any of the 27 D segments to encode at least 8,262 (51 × 6 × 27) different heavy-chain V regions (Figure 1.9) (150, 153). Further, the joining of the different genes occurs through the random addition of up to six nucleotides by the enzyme terminal deoxynucleotidyl transferase (TdT) which also increases the diversity. The gene combinations result in different antibody characteristics that can potentially engage an antigen. Antibody light chains are classified as lambda (λ**)** or kappa (κ) and use only the *V* and *J* genes. (Figure 1.9). In addition to the diversity encoded by the variable regions, there are five classes of human antibodies namely IgA, IgD, IgE, IgG and IgM based on the heavy chain they contain and they are important for effector function (154).

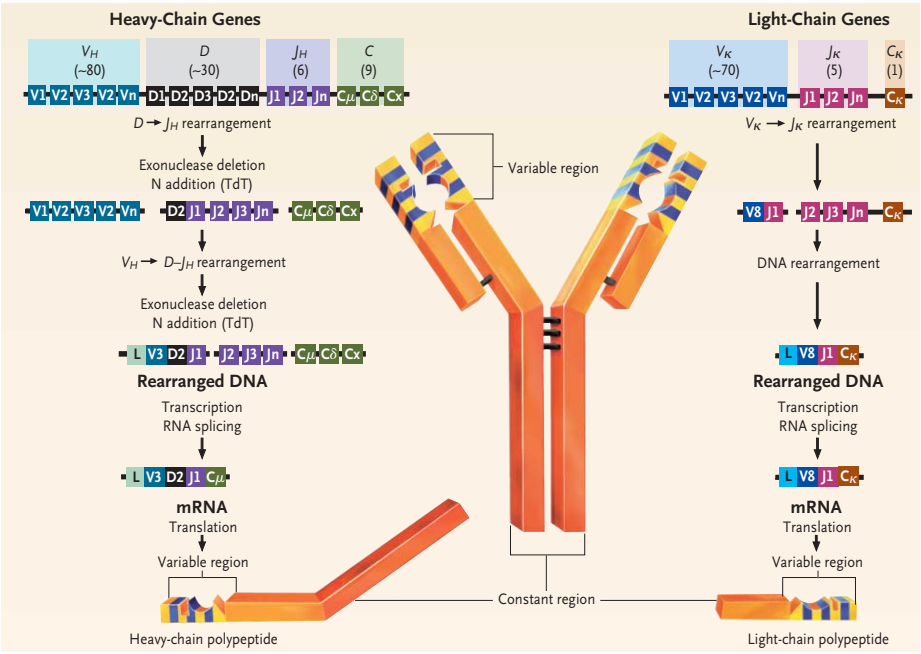


Figure 1.9: The generation of antibody diversity through gene recombination. Different D - J and D - V gene segments join and an endonuclease excises the remaining segments. The rearranged D - J and D - V genes are joined to form the V – D – J coding unit. The κ or λ light chains are formed by random rearrangement of Vκ and Jκ (or Vλ and Jλ) genes and joining to a light-chain gene of the constant region (Schwartz 2003).

A few weeks following HIV-1 and other infections, B cells are stimulated to proliferate into antibody-secreting cells (Figure 1.10). Activated B cells migrate to the bone marrow where they produce HIV-1-specific antibodies that work by preventing, reducing and clearing infection (155). They also undergo clonal expansion to make more B cells that will produce similar antibodies (Figure 1.10). All of the antibodies that are derived from a specific B cell constitute an antibody lineage, and this first B cell that initiates a lineage is termed an unmutated common ancestor (UCA). Their antigen specificity/affinity subsequently increases through somatic hypermutations which create gradual differences in related antibodies. Immunoglobulin V-region genes accumulate mutations at a rate of about one base pair change per 10^3 base pairs per cell division (156). This process is referred to as affinity maturation and is key for the generation of efficient antibody responses.

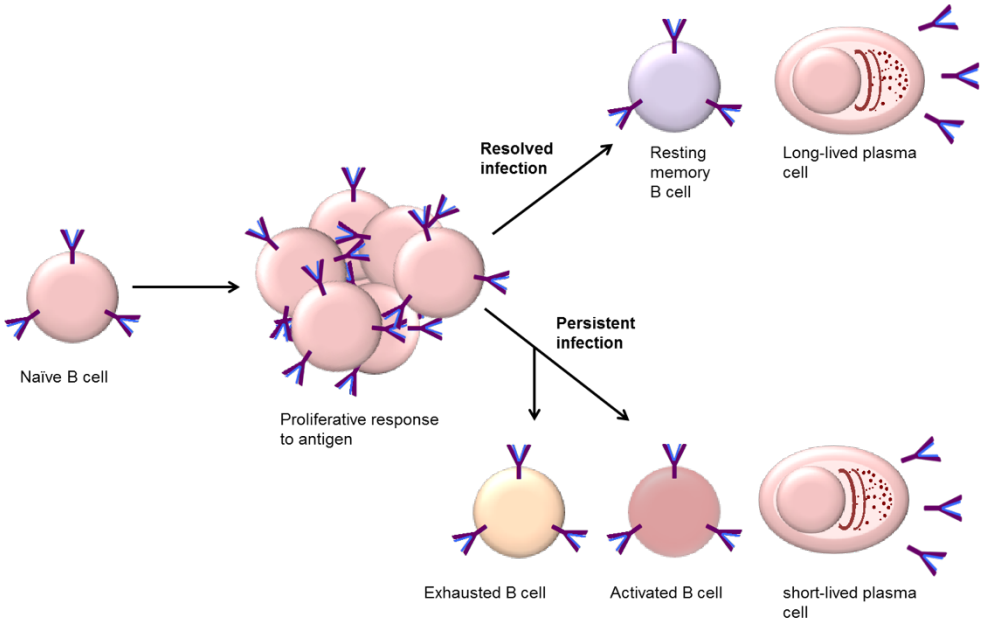


Figure 1.10: B cell response to antigenic stimulation. After antigenic stimulation, naïve B cells starting proliferating and produce activated B cells that produce antibodies. If an infection is resolved, B cells become memory cells that can be activated to produce antibodies if the infection recurs. However, during a persistent infection, B cells may become exhausted (Moir and Fauci 2009).

1.7.2 Development of binding and neutralizing responses against HIV-1 Env

Antibodies against HIV-1 are first detected as immune complexes, followed by free antibodies directed at the gp41 subunit, and then antibodies against the gp120 (157, 158). Although the first detectable responses are capable of binding HIV-1, they are not able to neutralize the virus (prevent entry of the viral particle into host cells) or apply selective pressure resulting in virus evolution (157). The initial strain-specific neutralizing responses are detectable weeks to months after infection (159–162). HIV-1 neutralizing antibodies frequently target epitopes in the C3V4 domain, the base of the V3 loop, V1V2 regions, and the CD4bs (163, 164). Although strain-specific neutralizing antibodies are produced in all HIV-1-infected individuals, viruses easily evade these strain-specific antibodies and thus are of limited value for vaccination design. However, in some people, bNAbs develop and these are described in more detail below.

1.7.3 The development of bNAbs during infection

BNAbs develop naturally during infection in approximately 15-30% of HIV-1 infected individuals but generally only after years of infection (165–168). The development of bNAbs has been associated with duration of HIV-1 infection, high viral loads and viral diversity suggesting that extended exposure to viral antigens and virus antibody co-evolution is required, which described in more detail in section 1.8 (161, 165). Unlike most antibodies, bNAbs tend to have unusual characteristics such as long CDRs, skewed germline usage, high levels of somatic hypermutation associated with extensive affinity maturation and polyreactivity (169–172). Because of the rarity of bNAb development in natural infection, studies of antibody-virus co-evolution are essential to understand the events that lead to their production. Such studies have shown that some strain-specific antibodies continue to evolve to become bNAbs that have the ability to neutralize the majority of HIV-1 strains (173) (Figure 1.11).

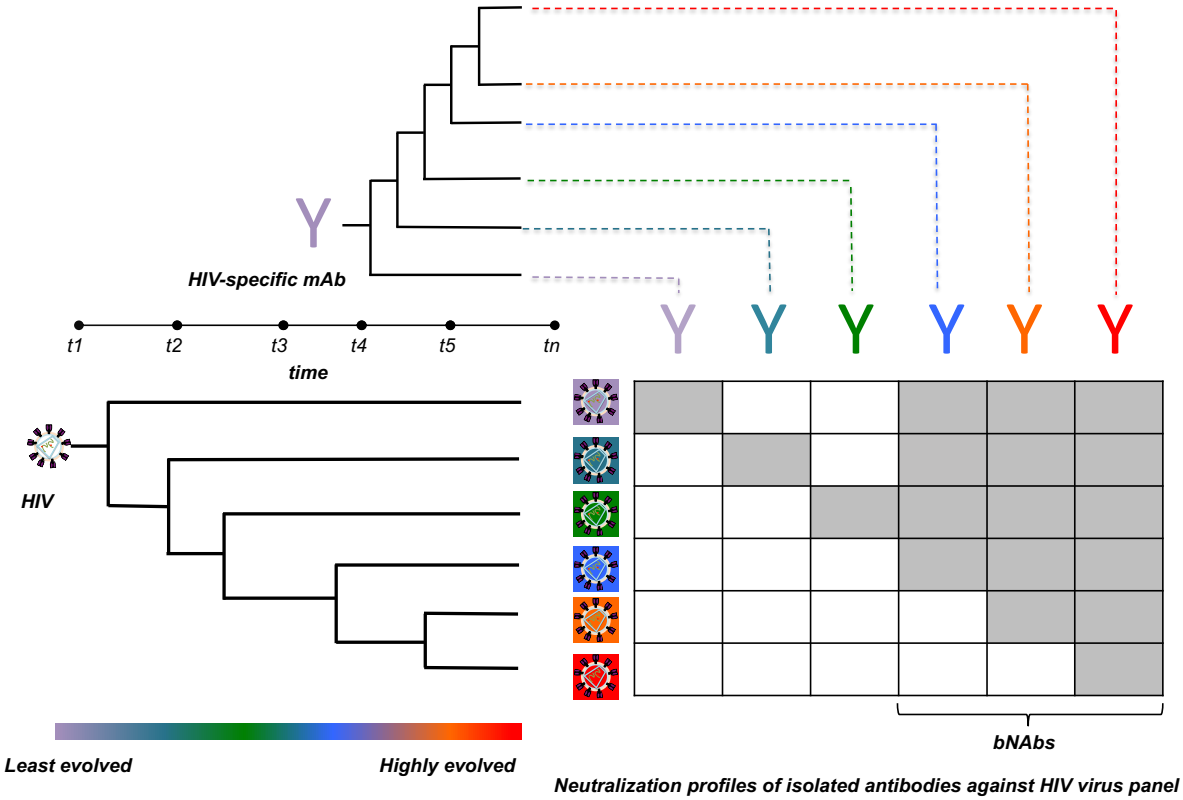


Figure 1.11: Antibody-HIV-1 virus co-evolution. The schematic phylogenetic trees show evolution of viruses and antibodies respectively. The table on the right shows the neutralization of viruses (shown on the y-axis) by the antibodies (shown on the x-axis). Continuous antigenic stimulation causes a specific antibody lineage to mature through somatic hypermutation. Some antibodies mature to become bNAbs that have the ability to neutralize the majority of HIV-1 strains.

1.7.4 Broadly neutralizing antibody (bNAb) targets

The isolation of monoclonal antibodies from HIV-1 infected donors has enabled us to identify bNAb targets. To date, five bNAb target sites have been identified on the HIV-1 envelope trimer (Figure 1.12). These include the V1V2 site, the N332 supersite, the CD4bs, the gp120-gp41 interface (including the fusion peptide) and the membrane proximal external region (MPER) (174), each of which is further described below.

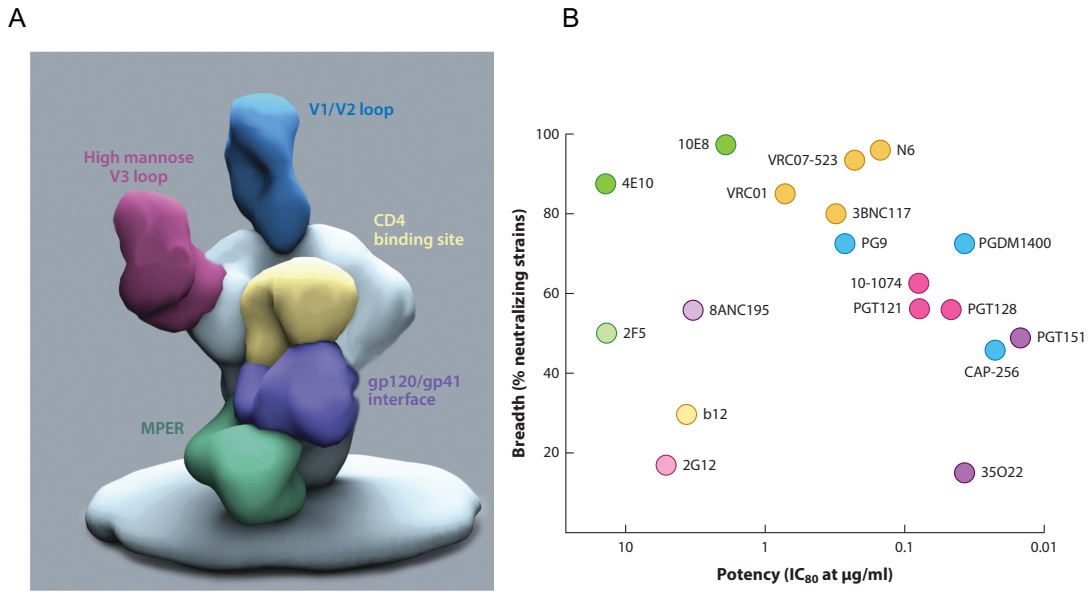


Figure 1.12: BNAb targets, potency and neutralization breadth. (A) Model of the HIV-1 trimer showing the different bNAb targets, (B) Potency and neutralization breadth of the different isolated broadly neutralizing antibodies. Source: Bonsignori et al., 2017; Gama and Koup, 2018.

1.7.4.1 V1V2 site

The V1V2 region is highly variable in terms of sequence, length, and the extent of glycosylation (175, 176), which poses a challenge for vaccine design. Nonetheless, many HIV-1 infected individuals develop antibody responses against the V1V2 site, and as such bNAbs against this site are important in vaccine design (165, 177–179). The antibodies are highly potent (shown in blue, Figure 1.12) and emerge relatively early in infection compared to other bNAbs (180–182). These bNAbs, exemplified by PG9/PG16, have an extraordinarily long CDRH3 loops, (≥ 24 amino acids), which helps them to penetrate the dense glycan shield on the Env spike trimer and access the protein surface below (183–185). Most isolated antibodies targeting these regions have shown dependence on N-linked glycans at sites 156 and 160, as well as other amino acids in the B and C strand, such as residues 166 and 169 in the V2 region (184, 186–190).

1.7.4.2 C3V4 site

Another common target of bNAbs is the C3V4. BNAbs targeting this region are dependent on the N332 and N301 glycans, and are commonly referred to as V3-glycan supersite, high mannose V3 loop or N332-supersite targeting bNAbs (191). The bNAbs can approach this region from variable angles by making use of different high-mannose glycans at several sites (192–198). These bNAbs also recognize discontinuous amino acids in the C3, V3 and V4 regions, including a 4 amino acid linear motif (GDIR) in the HIV-1 V3 loop (192, 197–201). C3V4 bNAbs include PGT121, PGT128 and PGT135 (shown in violet in Figure 1.12) (174, 194, 195, 199). C3V4 bNAbs are of interest for vaccine development because, like V1V2 responses, they are relatively common during infection. Furthermore, unlike other bNAb classes, they may not require extensive somatic hypermutation to develop neutralization breadth (173, 198).

1.7.4.3 CD4 binding site

The CD4bs is highly conserved, as it is required for virus infectivity. CD4bs bNAbs have been isolated from multiple HIV-1 infected individuals, and they are among the most broad and potent antibodies (shown in orange in Figure 1.12) (173). Examples of CD4bs antibodies include VRC01, VRC07, CH31, 8ANC131, CH235, CH103, HJ16 and CH98 (170, 188, 202–209). Some isolated CD4bs bNAbs, referred to as VRC01 class, share similar heavy *V* and *J* gene usage suggesting commonality of the antibody evolutionary pathway in multiple individuals (209, 210). However, such CD4bs bNAbs are likely to be difficult to elicit through vaccination as they have extremely high levels of somatic hypermutation, suggesting complex or prolonged maturation pathways (170, 204, 208, 211).

1.7.4.4 gp120-gp41 interface

Antibodies against the gp120-gp41 interface target conserved peptide and glycan segments in this region that exist on the pre-fusion conformation of the Env trimer (174, 212). Interface bNAbs are very diverse as the epitopes are not completely overlapping (213). For example, PGT151 is highly trimer-specific, exclusively binding to the closed

Env conformation and potently neutralizing the virus through interactions with complex tri- and tetra-antennary glycans at positions 611 and 637 within gp41, and protein residues K490, T499, R500, R503 in gp120 (214, 215). BNAbs such as 35022 on the other hand bind the gp120-gp41 interface at a site closer to the viral membrane than PGT151 (174, 212). Another interface bNAb, 8ANC19540 binds to glycans at 276, 234, and 637, with a footprint falling between that of PGT151 and 35022. This bNAb can bind to both closed and open conformations of the Env (216–218). Other examples of bNAbs targeting this region include 8ANC195, 3BC315/3BC176, N123-VRC34.01, ACS202 and CAP248-2B (shown in purple in Figure 1.12) (212, 215–217, 219).

1.7.4.5 MPER site

The fifth bNAb site, the MPER region, comprises the last 23 amino acids, from residues 660 to 683, of the extracellular domain of gp41 just before the transmembrane domain (220). This region is highly conserved and critical for cell fusion and infectivity (221, 222). Examples of MPER antibodies include 4E10, 10E8 and 2F5 (223–225). The bNAbs against the MPER exhibit the broadest neutralization capability (up to 98% of circulating HIV-1 subtypes), however they often have lower potencies compared to other HIV-1 neutralizing antibodies (shown in green in Figure 1.12) (199, 226). MPER-specific bNAbs may however, be difficult to elicit by vaccination for a number of reasons that include tolerance control and immunoregulation, sequestration of the epitope in the lipid membrane, and exposure of the epitope only transiently during viral entry (227, 228).

1.8 Co-evolution of antibody lineages and HIV-1 epitopes

The discovery of bNAbs in HIV-1 infected participants has brought about renewed enthusiasm in the vaccine field as they are effective at neutralizing the majority of HIV-1 strains and have shown to be protective in non-human primates (210, 229). However, bNAbs have not yet been elicited through vaccination, and this is a major focus of global research efforts. Co-evolutionary studies have the potential to elucidate intra-molecular (between amino acids in an epitope/paratope) and inter-molecular (between

paratope and epitope) interactions, and explain co-evolution in terms of co-adaptation, where changes in one protein are followed by changes in another protein (230).

A number of co-evolutionary studies have been performed in infected subjects to understand how bNAbs develop (182, 186, 231, 232). These studies have shed light on the arms race between the antibody and the virus which provides important information for vaccine strategies aimed at eliciting bNAbs. The studies performed thus far have investigated the developmental pathways of bNAbs targeting the V1V2, CD4bs, MPER and the C3V4 site (182, 186, 198, 201, 231–233).

Comparison of the evolutionary rates of three broad pathways showed that high mutation rates in both virus and antibody were associated with rapid bNAb development (232). Studies that have used viral and antibody data from multiple time-points during infection identified mutations that contributed to breadth, and the viral variants that engaged the UCA of bNAb lineages (186). The mAbs that matured to become broad co-evolved with the envelope glycoproteins to tolerate emerging viral escape mutations (186). Apart from showing how bNAb lineages evolve with viral variants, these studies have shed light on how some antibodies within bNAb lineages fail to tolerate viral envelope mutations, hindering their further evolution (186). Other antibodies within specific lineages became highly mutated but not acquiring breadth in the process suggesting they did not acquire the necessary mutations for neutralization breadth (186).

While the developmental pathways of some broad lineages have been elucidated, very little is known about the development of antibody lineages that fail to become broadly neutralizing, or of the viral variants that shape their evolution. This thesis therefore compared viral evolution in individuals with broad and non-broad responses to identify the unique virological features associated with bNAb development (234), with implications for vaccine design (Chapter 2).

1.9 Computational challenges of viral and antibody evolutionary analyses

Studying virus-antibody co-evolution involves sequencing pathogen and host samples from multiple time-points in an individual during infection. In antibody sequencing, successful identification of the relevant antibody lineages is dependent on producing as many antibody reads as possible during sequencing, to capture the diversity of the repertoire. Although sequencing technologies have dramatically improved, the amount of data generated requires high-performance computational capacity for analysis. Ordinary computers cannot manage the immense amount of data generated during the analyses, and the memory requirements to perform the computations, and/or require many hours of computation.

Similarly, understanding HIV-1 phylodynamics, like all other viruses, requires bioinformatics tools that have been developed to study the evolutionary, ecological and epidemiological processes that shape viral phylogenies (235). A better understanding of intra-host HIV-1 phylodynamics will assist in the development of a vaccine (236, 237). This will also be particularly important to other viral pathogens such as dengue and influenza. Bayesian phylodynamic tools are increasingly becoming useful to reconstruct the evolutionary histories of viruses such as HIV-1, to identify virus population demographics and to quantify dispersal patterns between host species. The Bayesian phylodynamic models are based on the Bayes' theorem that derives the posterior probability as a consequence of a prior probability and a likelihood function derived from a probability model for the data to be observed. These tools are therefore very robust, as they also apply Markov chain Monte Carlo (MCMC) methods for sampling stochastically from a probability distribution. Phylodynamic modelling can be performed by computational programs such as the Bayesian Evolutionary Analysis by Sampling Trees (BEAST), which is a statistical, graphical user program for parameter estimation and hypothesis testing of evolutionary models from molecular sequence data (238).

However, BEAST is computationally intensive and parallelises calculations for large state models by making use of many-core algorithms for statistical phylogenetics under arbitrary molecular evolutionary models on Graphics Processing Units (GPUs) (239). It uses BEAGLE (Broad-platform Evolutionary Analysis General Likelihood Evaluator)

that provides both an API (Application Programming Interface) and implementations in Java, C and for CUDA-based GPUs (239). Such phylodynamics programs run efficiently on clusters and servers designed to run for longer hours, that have lots of memory and storage space for the generated data. In this thesis, the requirement for many of these computational tools led us to explore the development of a low-cost computational cluster to enhance our computational capacity (Chapter 4 of this thesis)

1.10 Objectives and aims

1.10.1 Overall project objective

The aim of this study was to investigate HIV-1 viral envelope glycoprotein evolutionary dynamics, both among hosts with differing immune responses, and within hosts, between different tissues, using low-cost computational infrastructure. We utilised longitudinal samples from individuals enrolled in Centre for AIDS Programme of Research in South Africa (CAPRISA), to address three specific aims, as follows:

1.10.2 Specific research aims

1. To compare the evolution of the HIV-1 envelope glycoproteins between individuals with strain-specific and bNAb responses, to assess whether the unique viral evolutionary pathways in bNAb donors may inform vaccine design.
2. To investigate compartmentalization of HIV-1 between plasma and cervico-vaginal lavages using envelope glycoprotein sequences, to assess whether this might impact sensitivity to vaccines/microbicides.
3. To develop a low-cost bioinformatics cluster to perform computationally intensive viral and antibody evolutionary analyses.

2 Positive selection at key residues in the HIV Envelope distinguishes broad and strain-specific plasma neutralizing antibodies

Batsirai M. Mabvakure[a,b], Cathrine Scheepers[a,b], Nigel Garrett[c], Salim Abdool Karim[c,d], Carolyn Williamson[c,e], Lynn Morris[a,b,c] and Penny L. Moore[a,b,c]

[a]Center for HIV and STIs, National Institute for Communicable Diseases, National Health Laboratory Service, Johannesburg.
[b]School of Pathology, Faculty of Health Sciences, University of the Witwatersrand, Johannesburg.
[c]Centre for the AIDS Programme of Research in South Africa (CAPRISA), Kwa-Zulu Natal,
[d]Department of Epidemiology, Columbia University, New York, New York, United States of America.
[e]Division of Medical Virology, Institute of Infectious Disease and Molecular Medicine, University of Cape Town

Published article: Mabvakure et al. Journal of Virology. 2018 Dec 19. pii: JVI.01685-18. doi: 10.1128/JVI.01685-18

2.1 Abstract

The development of HIV broadly neutralizing antibodies (bNAbs) has previously been shown to be associated with viral evolution and high levels of genetic diversity in the HIV envelope (Env) glycoprotein. However, few studies have examined Env evolution in those who fail to develop breadth, to assess whether bNAbs result from distinct evolutionary pathways. We compared Env evolution in eight HIV-1 infected participants who developed bNAbs to six donors with similar viral loads who did not develop bNAbs over three years of infection. We focused on Env V1V2 and C3V4 as these are major targets for both strain-specific neutralizing antibodies (nAbs) and bNAbs. Overall evolutionary rates (ranging from 9.92×10^{-3} to 4.1×10^{-2} substitutions/site/year) and viral diversity (from 1.1% to 6.5%) across Env, and within targeted epitopes, did not distinguish bNAb donors from non-bNAb donors. However, bNAb participants had more positively selected residues within epitopes than those without bNAbs, and several of these were common among bNAb donors. Comparison of the kinetics of strain-specific nAbs and bNAbs indicated that selection pressure at these residues increased with the onset of breadth. These data suggest that highly targeted viral evolution rather than overall envelope diversity, is associated with neutralization breadth. The association of shared, positively selected sites with the onset of breadth highlights the importance of diversity at specific positions in these epitopes for bNAb development, with implications for the development of sequential and cocktail immunization strategies.

2.2 Introduction

A preventative HIV vaccine is likely to require broadly neutralizing antibodies (bNAbs) that have the ability to neutralize the majority of HIV strains (173, 240). Animal studies have shown that bNAbs are able to protect from virus infection (241), however, no vaccine has thus far been able to elicit bNAbs. In contrast, longitudinal studies have shown that bNAbs develop in approximately 10-30% of chronically HIV infected patients (165, 167, 168). These bNAbs target five sites on the HIV trimer, namely the V1V2 region, the CD4 binding site, the membrane proximal external region (MPER), the V3 glycan supersite and the gp41-gp120 interface, which includes the fusion peptide (174). Although most donors develop some degree of cross-reactivity over many years of infection (242), some donors have neutralizing antibody responses that remain strain-specific, neutralizing only the infecting strain, despite equivalent duration of infection and high viral loads (172). The reason strain-specific responses fail to mature into bNAbs in all HIV-infected individuals is an important unanswered question in the field.

Large cohort studies have shown that neutralization breadth is associated with several host factors including ethnicity, HLA genotype, CD4 T cell loss, and circulating T follicular helper cells (165, 167, 168, 243–246). However, viral factors such as subtype, high viral loads, virus diversity and duration of HIV-1 infection have been shown to be strongly associated with breadth, highlighting the importance of high levels of antigenic stimulation in driving the maturation of bNAbs (167, 168, 247, 248).

Studies of the developmental pathways of bNAbs have shown that these develop from strain-specific precursors that mature to acquire breadth through a prolonged co-evolutionary pathway, with emerging viral mutations driving antibodies to tolerate epitope diversity (173, 181, 182, 186, 201, 206, 231, 232, 249, 250). This has led to a vaccine framework that aims to incorporate variation within immunogens, an approach that has recently shown promise in small-animal immunization studies (251–256). However, similarly detailed studies of viral evolution in donors who maintain strain-specific responses, especially in the context of high levels of antigenic stimulation, have not been performed. Comparative analyses of virus evolutionary pathways in

these individuals and those who develop bNAbs may provide useful insights into the viral characteristics and mechanisms that drive antibody maturation towards breadth.

Robust strain-specific neutralizing responses develop in almost all infected individuals, regardless of whether they later develop breadth, and exert significant selection pressure on the HIV envelope glycoprotein (Env) (257, 258). This results in increased Env diversity and divergence from the transmitted/founder (T/F) virus and selection of neutralization resistant strains. Escape is associated with increased variable loop length and glycosylation that may shield underlying epitopes making it difficult to neutralize the virus (159, 160, 164, 259–261). Importantly, many of the strain-specific neutralizing responses target the same overall regions of the HIV envelope as bNAbs, such as the V1V2 and C3V4 regions (162, 181), and drive viral diversification within these regions (161, 164, 181).

To assess why these strain-specific responses fail to acquire breadth, we have used longitudinal single genome-derived full Env sequencing data and matched antibody kinetics and mapping data from 14 participants in the Centre for the AIDS Programme of Research in South Africa (CAPRISA) 002 cohort, six of whom developed strain-specific antibodies and eight who developed bNAbs. Interestingly, we did not observe significant differences in nucleotide substitution rates in participants with or without bNAbs, and observed high viral diversity across Env and within epitopes, in both groups. However, we identified several sites under positive selection that were common to donors with bNAbs. The tight clustering of these sites on the Env trimer suggest that the precise targeting of viral epitopes by early bNAb precursors, and subsequent viral escape pathways, determines whether these antibodies acquire breadth. Defining these sites highlights developmental differences in strain-specific and bNAbs which may inform future vaccine design strategies.

2.3 Materials and Methods

2.3.1 HIV infected participants and Ethics

Longitudinal plasma samples were collected from 14 women chronically infected with HIV-1 subtype C who were enrolled into the CAPRISA 002 and 004 cohorts (262, 263). Of these, 8 participants (CAP8, CAP177, CAP206, CAP357, CAP248, CAP255, CAP256 and CAP257) developed bNAbs (defined by their ability to neutralize more than 40% of a multi-subtype panel at 3 years) and 6 donors (CAP45, CAP88, CAP200, CAP225, CAP228 and CAP229) who did not develop bNAbs (with plasma breadth less than 10%) despite approximately 3 years of HIV infection (165) (Madzhivandila et al. unpublished).

Evidence of HIV infection in participants was determined using two rapid HIV-1 antibody tests (Determine, Abbott 89 Laboratories; Tokyo and Capillus, Trinity Biotech, Jamestown, NY). False positives and negatives were eliminated by using PCR for the negative samples (Ampliscreen v1.5, Roche Diagnostics, Rotkreuz, Switzerland) and HIV enzyme immunoassay (EIA, BEP2000, Dade Behring, Marburg, Germany) for positive samples. Viral loads were estimated using the Cobas Amplicor HIV-1 monitor test version 1.5 (Roche Diagnostics, Rotkreuz, Switzerland). The time of infection was estimated as either the midpoint between the last antibody negative and first antibody positive visits, or 14 days prior to an RNA-positive, antibody negative sample.

Written informed consent to conduct research on stored specimens was obtained from all women at enrolment, and ethics clearance for the use of plasma samples was obtained from the Human Research Ethics Committee (Medical) from the University of Witwatersrand (MM040202), the University of Cape Town (025/2004) and CAPRISA at the University of KwaZulu Natal (E013/04).

2.3.2 Single Genome Envelope Amplification and Sequencing

HIV-1 RNA was isolated from plasma samples and full-length envelope genes were amplified from single genomes at multiple time points by nested PCR, and sequenced

using Sanger sequencing as described (264). Accession numbers for *env* sequences are EF203968-82, FJ443149-58, FJ443223-39, FJ443275-366, FJ443390-93, FJ443492, FJ443499, FJ443501-04, FJ443508, FJ443519, FJ444632, GQ999980, GK999985, GK999989, HQ625597-99, HQ625604-05, KC863376-78, KC863380-92, KC863446-56, KC863458-75, KC863542-53, KC863555-68, KC863571, KF241776, KF996579, KF996581, KF996584, KF996586-88, KF996590, KF996591, KF996593-96, KF996598, KF996600-01, KF996604-30, KF996632-99, KF996700-16, KT698223-27, KU198436, MF572810-29 and MK205449-6213.

2.3.3 Sequence Alignments and Data Partitioning

Nucleotide sequence alignments were made in MUSCLE (265), and codon aligned and manually edited in Bioedit and Se-Al (266, 267). The longitudinal envelope (*env)* glycoprotein sequences were analysed as a whole, and also subsequently partitioned into variable and constant regions for subsequent analyses using Se-Al software (267). The partitions were defined according to the HXB2 reference sequence as follows: V1 (131-156), V2 (157-196), V3 (296-330), C3 (329-383) and V4 (385-418). In some participants, the V2 and V4 regions contained several INDELS that were sometimes difficult to accurately determine their HXB2 numbering despite the fact that they had been confidently aligned.

2.3.4 Estimating Diversity, Nucleotide Substitution Rates and Time of Infection

Overall Env, C3V4 and V1V2 diversity was calculated using the Poisson correction distance model implemented in MEGA6 (268, 269). The model assumes equality of substitution rates among sites and equal amino acid frequencies while correcting for multiple substitutions at the same site (268, 269). The presence of temporal signal was estimated in TempEst using maximum likelihood trees constructed with PhyML (270, 271). Phylogenetic analyses were performed using BEAST v1.8.4 (272). Best-fit nucleotide substitution models were selected based on the Bayesian Information Criteria (BIC) and Akaike Information Criterion, corrected (AICc) values estimated

using MEGA 6 software (268). The best-fit demographic and clock models were selected with marginal likelihood estimations using the generalized path sampling and stepping stone methods with 50 path steps (273, 274). The Monte Carlo Markov chains were run until convergence (Effective Sample Sizes above 200), and the chain lengths were between 50-500 million steps. The log files generated were analysed in Tracer (275), and the Maximum Clade Credibility (MCC) tree annotated in TreeAnnotator and viewed in FigTree (276).

2.3.5 Identifying Sites Under Positive Selection Pressure

We performed natural selection analyses using Fast Unbiased Bayesian Approximation (FUBAR) that identifies codons under positive selection (277, 278). Codons that had more non-synonymous compared to synonymous changes, and supported by a posterior probability greater than 0.9 were reported to be evolving under positive selection. The pipeline for selection analysis involved first analyzing recombination in sequences using the genetic algorithm for recombination detection (GARD) to avoid phylogenetic bias when running FUBAR (277, 278). We employed two strategies in the selection analyses, which we termed the snapshot and cumulative analyses, respectively. The snapshot analysis involved analyzing sequence data from two consecutive time points, whereas the cumulative analysis involved integrating overall sequences in the dataset at a particular time point.

2.4 Results

2.4.1 Env diversity and evolutionary rates do not distinguish broad and strain-specific donors with equivalent viral loads

We first assessed viral diversity in 6 non-bNAb donors (with <10% breadth) and 8 bNAb donors (with >40% breadth) over three years of infection (Table 2.1). Since viral diversity is closely linked to viral load, we selected where possible non-bNAb participants with similarly high viral loads to the bNAb donors. Two participants CAP45 and CAP228, who had consistently low viral loads were also included as we have extensively characterized their neutralizing responses (Figure 2.1A). We generated between 29 and 171 single genome derived *env* sequences per donor from multiple time points spanning 3 years (Table 2.1). Maximum likelihood analysis of transmitted/founder viruses from bNAb and non-bNAb donors confirmed that these were phylogenetically unlinked (data not shown).

Table 2.1: Description of HIV-infected participants, plasma neutralization breadth, set-point viral load (at 12 months post-infection) and antibody targets.

Participant	% neutralization breadth at 3 years[#]	Number of *env* sequences	Number of timepoints	Set-point viral load (RNA copies/ml)	Antibody targets (bNAbs shown in bold)	References
CAP257	82	63	6	8260	**V1V2, CD4 binding site**	Wibmer et al., 2013
CAP256	77	171	13	178000	**V1V2**	Moore et al., 2011,Doria-Rose., 2014
CAP248	59	120	18	64600	**gp120-gp41 interface**	Wibmer et al., 2017
CAP177	52	144	7	18000	**C3V4**, V1V2	Moore et al., 2009, Moore et al., 2012
CAP255	47	148	13	18200	**C3V4**	Lacerda et al., 2013
CAP206	47	130	9	315000	**MPER**, C3V4	Morris et al., 2011, Unpublished
CAP357*	45	76	7	11596	not mapped	Unpublished
CAP8	42	70	5	39300	not mapped	Lacerda et al., 2013
CAP88	8	83	8	38700	V1V2, C3V4	Moore et al., 2009
CAP228	7	100	6	1520	V1V2, C3V4	Unpublished
CAP229	7	35	3	24600	V1V2	Unpublished
CAP225	5	29	3	21500	not mapped	Unpublished
CAP200	2	53	4	108000	C3V4	Unpublished
CAP45	0	41	6	556	not mapped	Moore et al., 2009

*Madzhivandila et al, unpublished; #Gray et al., 2011

Since dual infection results in increased genetic diversity, we used BEAST to compare the estimated time of infection (tMRCA), based on the sequence diversity, to the actual time of infection for each donor (279, 280). For dual infection, sequence analysis predicts a greater estimated duration of infection compared to the known time of infection; indicated by orange shading for CAP256 (a donor known to be superinfected) but not for CAP88 (Figure 2.1B). Twelve out of fourteen datasets had sufficiently strong temporal signals (as determined by the positive slope (rate) and correlation coefficient) for phylogenetic molecular clock analysis. The two donors that were excluded from this analysis were CAP225 and CAP229. BEAST analysis indicated single infections in 6 donors: CAP88, CAP206, CAP248, CAP255, CAP357 and CAP257 (Figure 2.1C, bNAb participants are ranked by breadth), consistent with the fact that most HIV-1 infections are initiated by a single variant (41, 281). In contrast, we saw evidence of dual infection (indicated by asterisks in Figure 2.1C) for 5 donors: CAP200, CAP228, CAP8, CAP177 and CAP256. Dual infections were observed both in individuals with bNAbs (CAP8, CAP177 and CAP256), and without bNAbs (CAP200, CAP228).

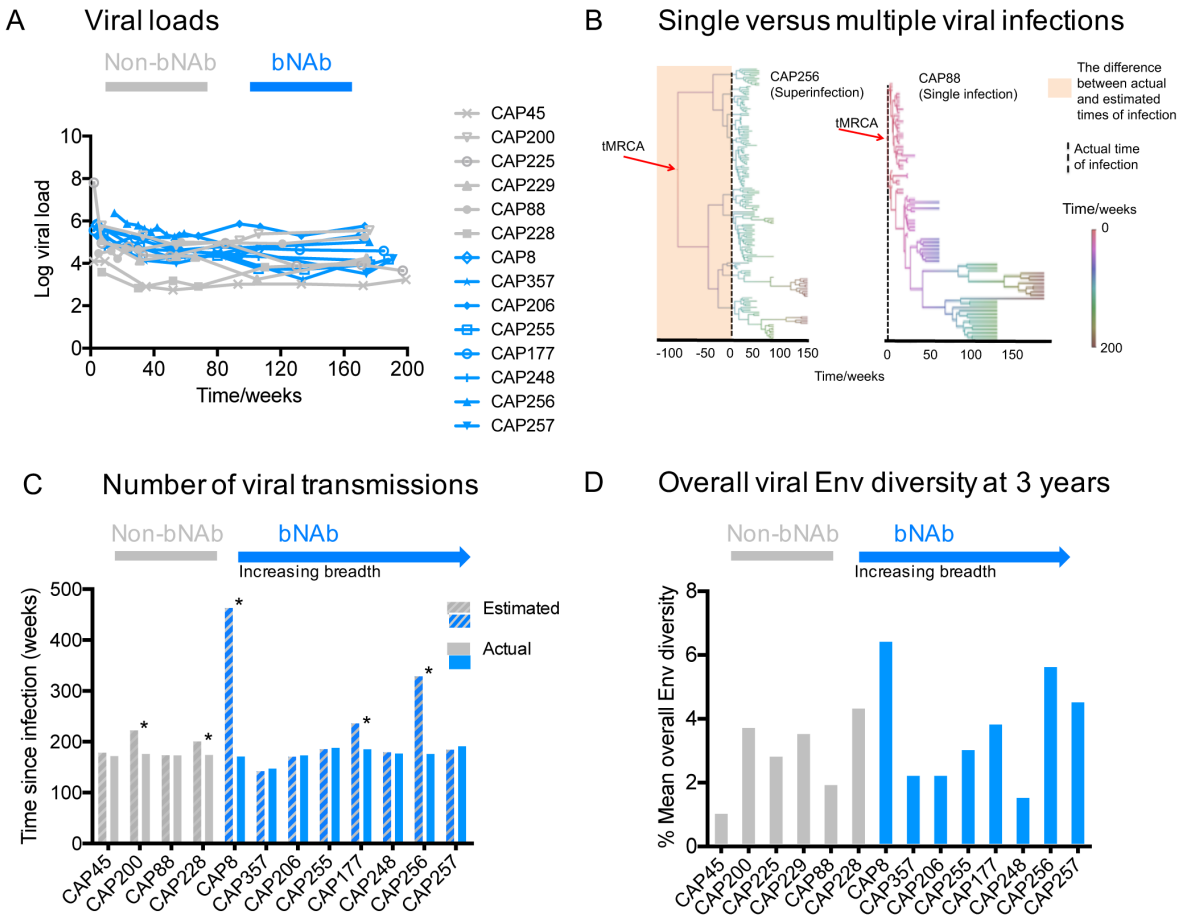


Figure 2.1: Viral diversity and evolutionary rates are not the sole determinants of bNAb development. (A) Longitudinal viral loads of all participants, with bNAb donors in blue and strain-specific donors in grey (B) Maximum clade credibility trees of CAP88 and CAP256 constructed in BEAST showing the estimated time of infection (tMRCA) compared to actual time of infection. Similar analyses were performed for all donors. (C) Comparison of estimated and actual time of infection in 12 participants. The asterisk (*) denotes possible dual infection. (D) Comparison of overall amino acid envelope diversity at three years in participants with/without bNAbs.

We next compared the levels of overall Env glycoprotein diversity in participants with and without bNAbs at 3 years post-infection (Figure 2.1D). We observed the highest overall diversity in CAP228, CAP8, CAP256 and CAP257 with a mean overall diversity >4% amino acid differences per site at 3 years post infection (Figure 2.1D). With the exception of CAP257, this is likely a result of their infection with multiple viruses or variants (Figure 2.1C). Among these donors, CAP228, who failed to develop breadth, had relatively high diversity despite low viral loads (Figure 2.1A). However, overall, there was no statistically significant difference in Env diversity between participants with or without bNAbs ($p=0.7758$, two-tailed Mann-Whitney test). Similarly, evolutionary rates across the envelope glycoprotein, based on nucleotide substitution rates estimated in BEAST, showed no significant differences between these groups

(data not shown).

2.4.2 Comparison of Env evolution within antibody-targeted regions between bNAb and non- bNAb donors

We have previously mapped the targets of both strain-specific and bNAb responses in 8 participants to the V1V2 and/or C3V4 regions (Table 2.1, Figure 2.2A). This enabled us to compare the Env evolutionary pathways between bNAb and non-bNAb donors specifically within these targeted regions. The V1V2 and C3V4 regions were defined according to the HXB2 reference sequence as amino acid positions 131-196 and 329-418 respectively. We first compared two donors with V1V2-directed bNAbs (CAP256 and CAP257) with four donors who had strain-specific V1V2 responses (CAP88, CAP228, CAP229, CAP177) (165, 181, 189, 250) (Table 2.1). Although CAP177 developed bNAbs to the C3V4 region, this participant also had a strain-specific V1V2 response (164) and is included as a strain-specific donor for this epitope. The two participants with V1V2-directed bNAbs (CAP256 and CAP257) had higher levels of V1V2 amino acid diversity (20.7% and 14.5%, respectively) compared to non-bNAb participants (ranging from 6.9 to 11.9% amino acid differences per site, though this difference failed to reach significance) (Figure 2.2B). Similarly, the two bNAb participants showed slightly higher V1V2 evolutionary rates than non-bNAb donors, however these differences were also not significant, with estimates within the same high posterior density intervals (Figure 2.2C).

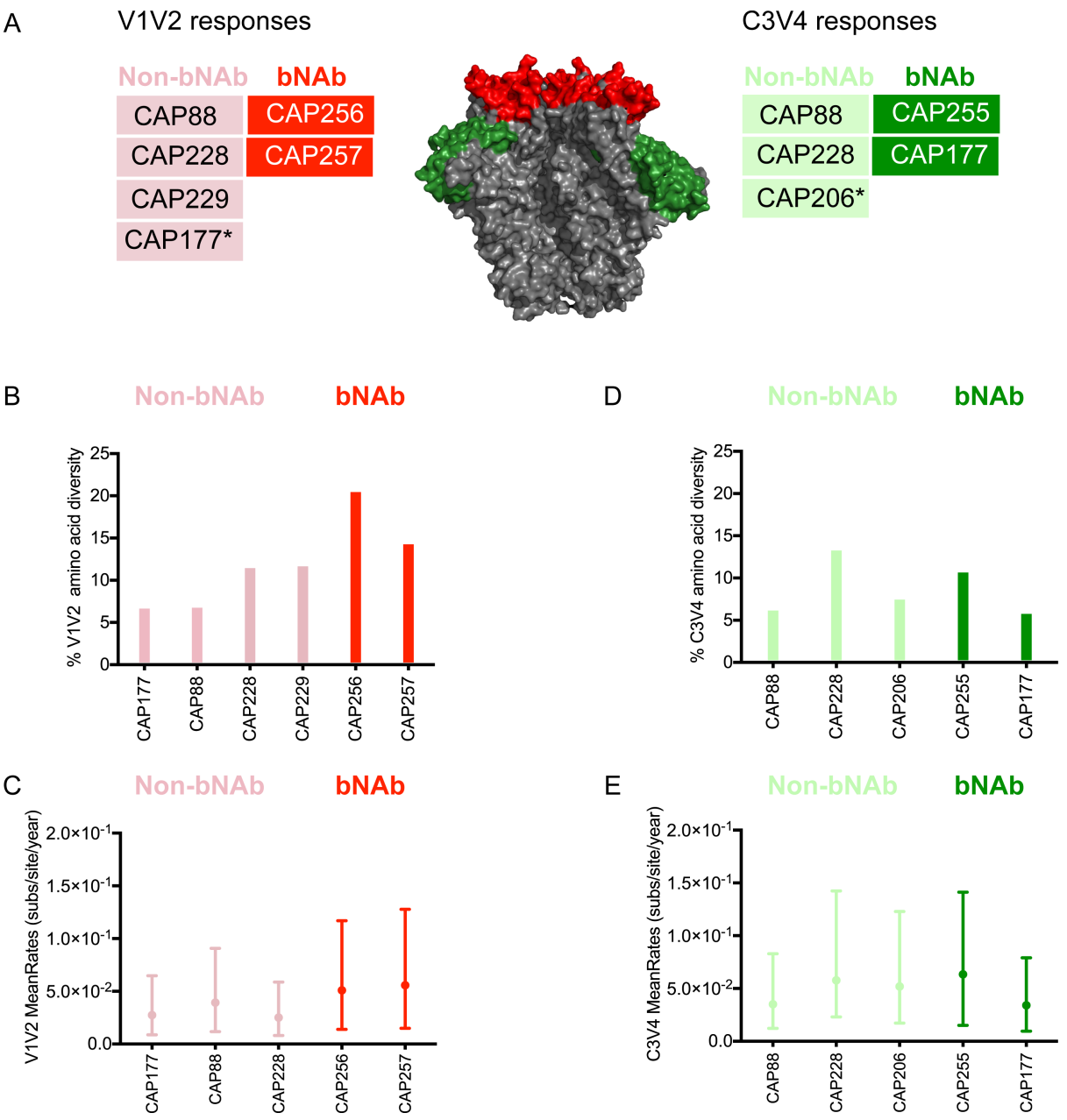


Figure 2.2: Comparison of diversity and nucleotide substitution rates in epitope regions targeted by participants with/without bNAbs. (A) Schematic of an Env trimer showing the antibody targets defined in previous studies. Participants with V1V2 responses are shown in red and those with C3V4 responses are shown in green. The bold and faded colors represent participants with and without bNAbs respectively. The asterisk (*) denotes participants who have bNAb responses to other regions, but also developed strain-specific antibodies to V1V2 or C3V4. Epitope diversity (average number of amino acid differences per site) and Nucleotide substitution rates within V1V2 (B and C) and C3V4 (D and E) in participants with/without bNAbs. Broad participants arranged in order of increasing neutralization breadth.

We similarly compared two donors who developed N332-specific bNAbs (CAP177 and CAP255, dark green in Figure 2.2A) with three donors shown to have strain-specific C3V4 responses (CAP88, CAP206, and CAP228, light green) (Figure 2.2A). CAP206 developed bNAbs that targeted the MPER (165) but in this analysis we

focused on the strain-specific C3V4 response. BNAb donors showed no difference in overall C3V4 diversity and nucleotide substitution rates compared to the strain-specific C3V4 responders (Figure 2.2D and 2E). Overall, analysis of viral evolution within antibody targeted regions did not show evidence of unique viral evolutionary pathways for donors with N332-directed bNAbs.

2.4.3 Participants with bNAbs show high levels of positive selection within targeted epitopes

We compared levels of positive selection within epitope regions in bNAb and non-bNAb donors, by estimating the ratio of non-synonymous to synonymous substitutions per site (282). We first compared two approaches for assessing selection: a cumulative and a snapshot selection analyses. In the cumulative analysis, all sequences at preceding time points were combined, whereas in the snapshot analysis, we analyzed sequence data only from two consecutive time points. The cumulative analysis identified the same positively selected sites as the snapshot analysis as well as additional sites, and unlike the snapshot analysis retained all sequences for analysis (data not shown). All subsequent analyses were therefore done using the cumulative analysis.

As expected, the number of positively selected sites increased over time in all participants regardless of whether they developed breadth or not (Figure 2.3 A and D). However, participants with V1V2-directed bNAbs exhibited a rapid increase in the number of positively selected sites (peaking at 26-27 amino acid residues) in V1V2 within the first year, compared to those without bNAbs (6-11 amino acid residues), and this increased level of positive selection persisted over 3 years of infection (Figure 2.3A-C). Similarly, participants with C3V4-directed bNAb responses had an increased number of positively selected sites (peaking at 11-12 amino acid residues at 1 year) in C3V4 during the first year of infection compared to strain-specific donors (3-4 amino acid residues at 1 year) (Figure 2.3D-F**),** though this difference was less pronounced than that observed in V1V2, and was not sustained in later years. These differences between bNAb/non-bNAb donors were driven largely by the C3 region (Figure 2.3D-F), consistent with the fact that bNAbs to this region are known to be focused on the

N332 glycan in C3, whereas strain-specific antibodies also target the more sequence variable V4 region (162, 283).

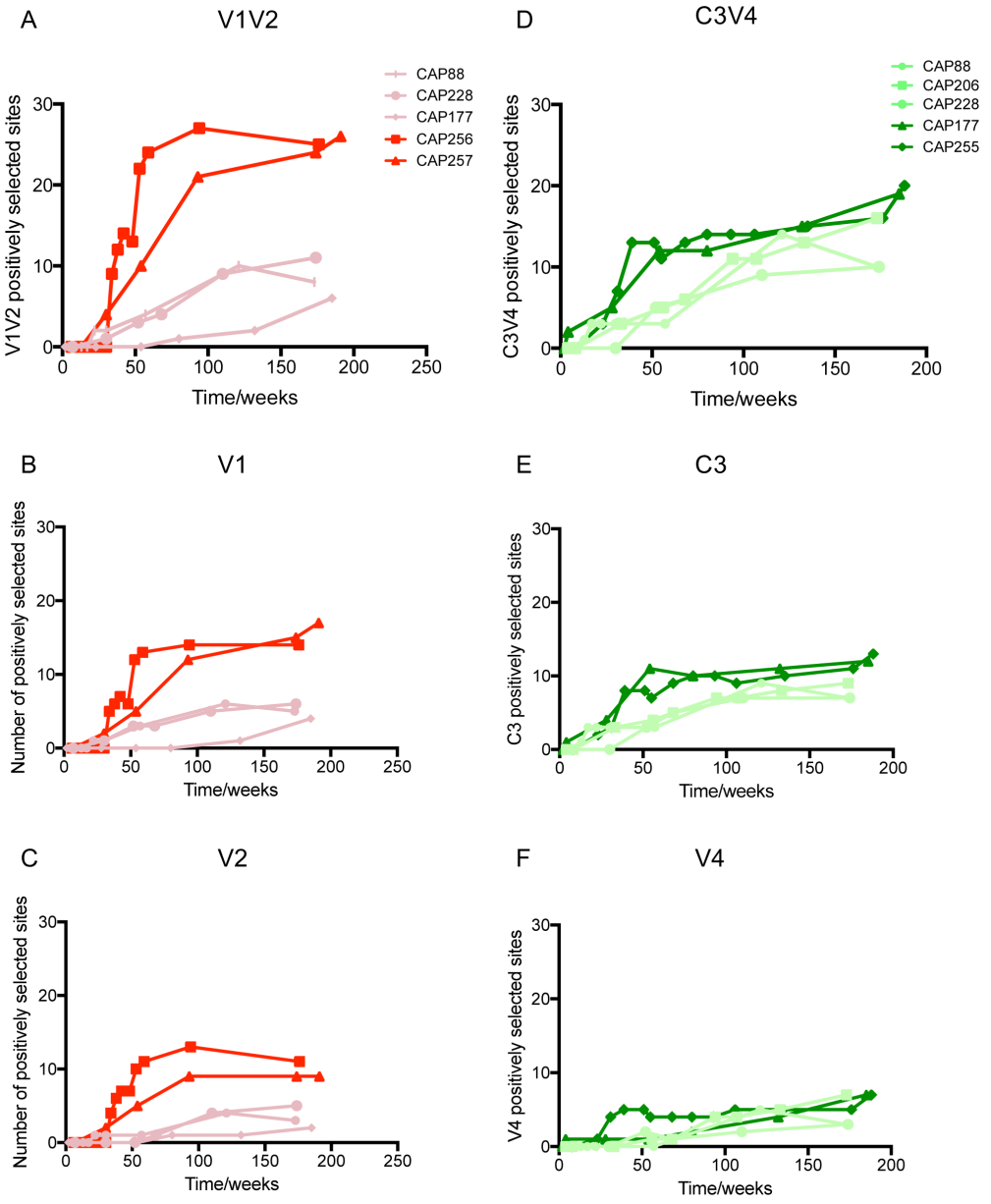


Figure 2.3: BNAb participants have more sites under positive selection than non-bNAb individuals. Positive selection was higher in the V1V2 region of participants that developed bNAbs (dark red) compared to those with strain-specific neutralizing antibodies (dull red) (A-C). In participants that had C3V4 responses, positive selection was higher in the C3V4 region of bNAb participants (dark green) compared to those with strain-specific neutralizing antibodies (dull green) only in the first year of infection (D-F).

2.4.4 Common sites in the V2, C3 and V3 regions are targeted by bNAbs

We next compared the specific sites under positive selection between participants with and without bNAbs. We focused on the V2 and C3 regions only, as the V1 and V4 regions were highly variable and prone to indels, making accurate identification of specific sites difficult. Using FUBAR, we identified positively selected sites in the V2 and C3 regions in each of the donors, and compared their positions between bNAb and non-bNAb donors.

In the V2 region, we identified 28 positively selected sites across all donors with V1V2 responses. bNAb donors CAP256 and CAP257 had 13 and 11 positively selected sites in V2 respectively (Figure 2.4A). Of these, six sites (160 and 162 which together form the highly conserved N160 glycosylation sequon, 166, 169, 181, 185 and 187) were common in both V2 bNAb participants (CAP256 and CAP257) (highlighted in red in Figure 2.4A and B). Four of these sites were also targeted by at least one of the non-bNAb donors indicating that strain-specific V1V2 antibodies overlapped somewhat with bNAbs. Many of these sites (160/162, 166, 169 and 181) have previously been reported to mediate escape from V1V2-directed bNAbs (186, 188, 210, 284), and all are located proximal to one another at the apex of the envelope trimer (Figure 2.4B). However, overall positive selection at these sites, particularly at sites 166 and 181, was less common in donors who developed strain-specific V1V2 responses (Figure 2.4C).

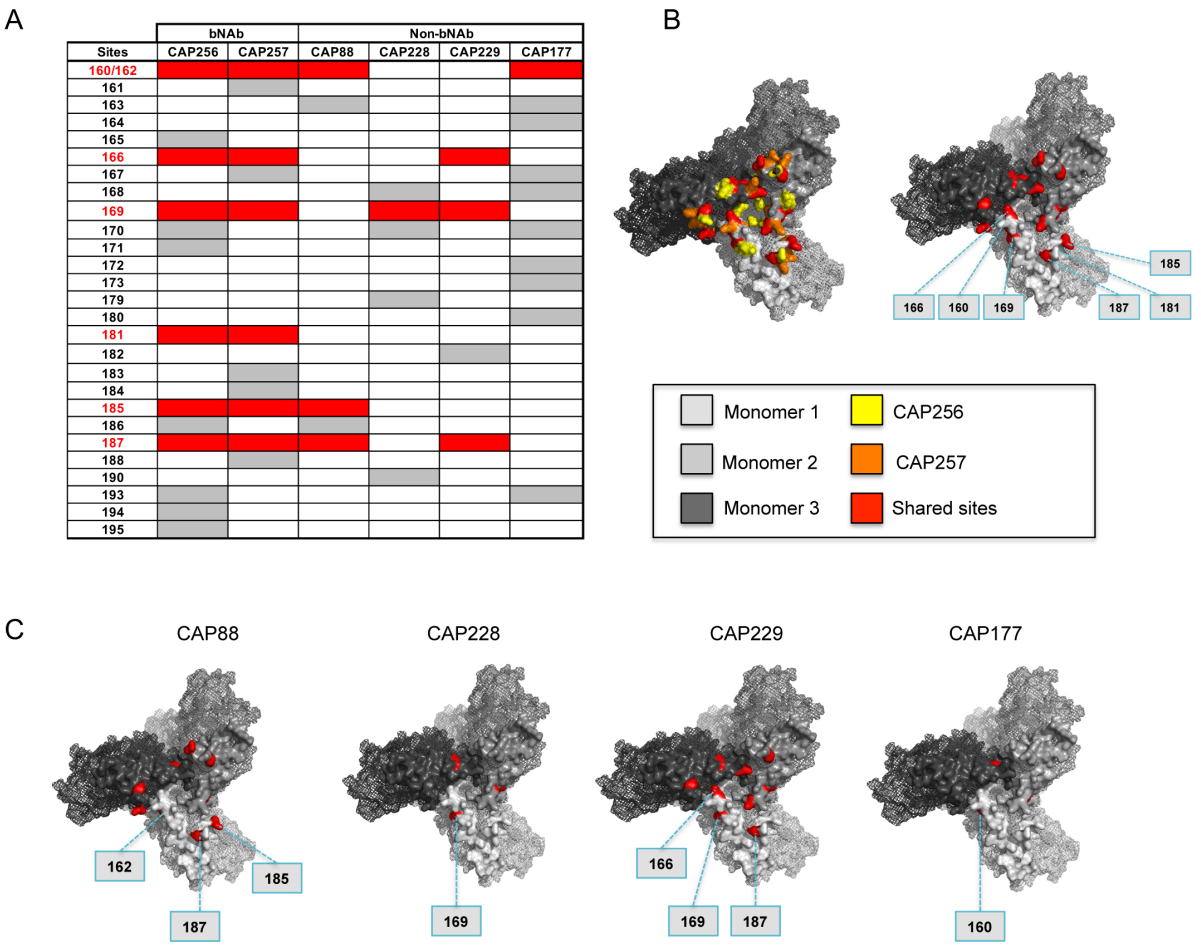


Figure 2.4: Comparison of positively selected sites in the V2 region of participants with and without bNAbs. (A) Table showing all sites under positive selection in the V2 region of participants with V1V2 responses. Positively selected sites common to both bNAb donors are highlighted in red. (B) Trimer model showing all sites under positive selection in participants with V2 responses (left) and common sites under positive selection in CAP256 and CAP257 (right). (C) bNAb sites also under positive selection in non-bNAb participants.

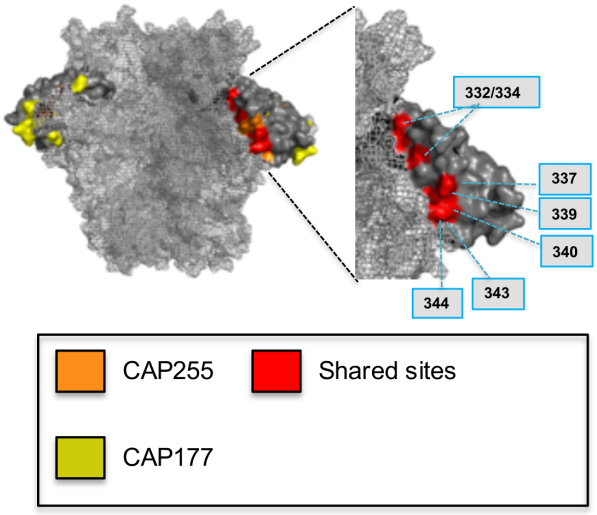
Similarly, 22 positively selected sites were identified within the C3 region of the 5 participants with C3V4 responses (Figure 2.5A**)**. Of these, seven sites (332/334 which form part of mutually exclusive glycosylation sequons, 337, 339, 340, 341, 343 and 344) were common in both participants who developed bNAbs targeting the N332 supersite (highlighted in red in Figure 2.5A and B). Unlike the two bNAb participants, donors with strain-specific responses to the C3V4 region did not show evidence of positive selection at residue 332 or 334. In addition to the glycan, residues 341 and 344 were not under positive selection in any of the three strain-specific donors (CAP88, CAP206 and CAP228). Sites 337, 339, 340 and 343, under positive selection in both bNAb donors were also sometimes under positive selection in participants who failed to develop bNAbs (Figure 2.5C). Of note, we identified other common sites (351, 354,

355, 356 and 357) under selection in both CAP177 and CAP206 that clustered together within the C3 region on the envelope trimer, (highlighted in yellow in Figure 2.5B) suggesting another distinct epitope within this region.

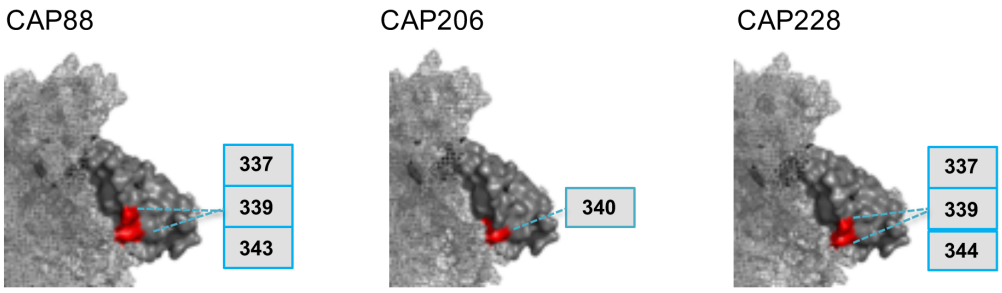
A

	bNAb		Non-bNAb		
Sites	CAP177	CAP255	CAP88	CAP206	CAP228
332/334					
335					
336					
337					
339					
340					
341					
343					
344					
347					
350					
351					
354					
355					
356					
357					
359					
361					
364					
375					
379					

B



C



D

Sites	bNAb		Non-bNAb		
	CAP177	CAP255	CAP88	CAP206	CAP228
325	D325N 54-132wks	D325N 55wks, 176-188wks	-	-	D325N 52-68wks
330	Y330H 54-185wks	H330Y 68-80wks		-	H330Y 52-68wks

Figure 2.5: Comparison of positively selected sites in the C3 region of participants with and without breadth. (A) Table showing all sites under positive selection in the C3 region of participants with C3V4 responses. Positively selected sites common to both bNAb donors are highlighted in red. (B) Trimer model showing all sites under positive selection in participants with C3V4 responses (left) and common sites under positive selection in bNAb donors CAP256 and CAP257 (right). (C) bNAb sites also under positive selection in non-bNAb participants. (D) Analysis of sites that toggle in the V3 GDIRQAH motif of participants with C3V4-directed responses.

BNAbs to the N332 supersite have also been shown to depend on the ^{324}GDIRQAH^{330} motif at the base of the V3 loop (285). We therefore analysed selection at these sites. Although none reached statistical significance for positive selection, we observed toggling at sites 325 and 330 in both bNAb donors (CAP177 and CAP255) but also in strain-specific donor CAP228. No such toggling was observed in CAP206 and CAP88 **(**Figure 2.5D). Sites 325 and 330 have been reported to impact neutralization sensitivity when these mutations are accompanied by changes at other sites such as 301, 324, 326 and 327 (285). The similarity of toggling of V3 residues in CAP177, CAP255 and CAP228 might suggest a similar epitope for both broad antibodies and some strain-specific antibodies, raising the question of how they differ in their mode of recognition.

2.4.5 Increased positive selection is associated with neutralization breadth

We next assessed whether positive selection at key sites increased with the acquisition of breadth. For this analysis we used previously published data defining the kinetics of bNAbs (grey shading, Figure 2.6) and strain-specific responses to the same epitope (indicated by the dotted line, Figure 2.6) (164, 165, 181, 182, 186, 250, 259). Selected residues are shown compared to the T/F virus, except for CAP256 where the comparison is with the superinfecting virus previously shown to elicit bNAbs in this donor (182). For V1V2 bNAb donors CAP256 and CAP257, we observed positive selection at sites 169/185 and 166/169/181, respectively, prior to the development of breadth, likely driven by the earlier strain-specific responses (Figure 2.6A). In CAP256, the positively charged K169 in the superinfecting virus was replaced with a 169Q, which is rare among subtype C viruses, but present in the primary infecting virus, and confers partial escape from V2 directed antibodies (182). Similarly, in CAP257, a charge-changing minority K169E mutation emerged prior to the development of breadth. In both donors, positive selection expanded to additional sites including the N160 glycan with the onset of neutralization breadth. In CAP257, additional, mutually exclusive glycans, at positions at 185 and 187 were selected for after bNAbs development.

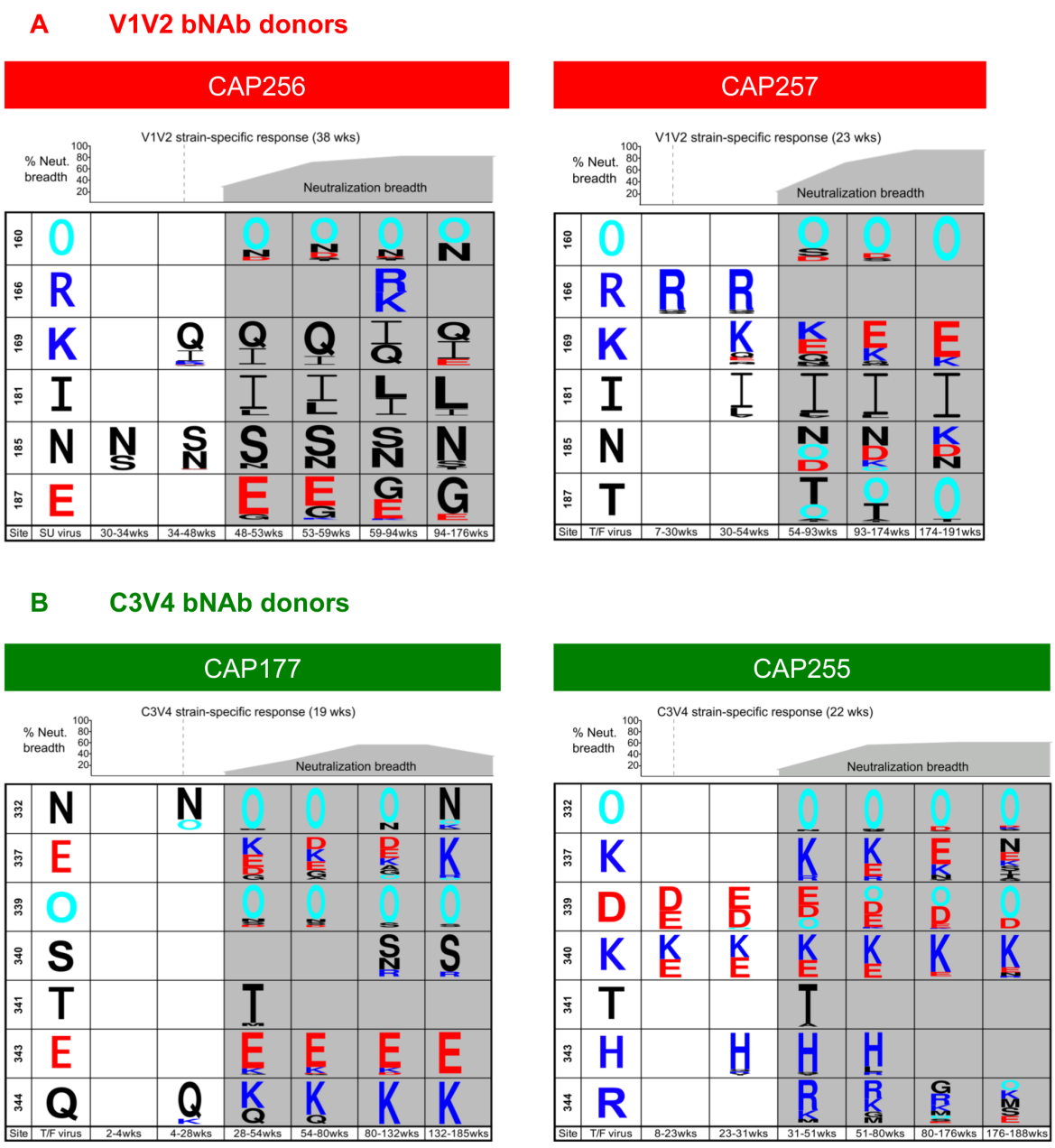


Figure 2.6: Increased positive selection at the onset of neutralization breadth. Breadth is shown in grey, glycans are shown in cyan, positively charged amino acids in blue, negatively charged amino acids in red, and non-charged in black. (A) Participants with V1V2-directed bNAbs showed increased numbers of positively selected bNAb sites (160, 166, 169, 181, 185 and 187) as breadth developed. (B) Participants with C3V4 bNAbs accumulated pressure at sites 332, 337, 339, 340, 341, 343 and 344 with breadth.

In C3V4 donors CAP177 and CAP255, strain-specific responses drove positive selection at two to three sites (N332 and 344 for CAP177 and sites N339 and 340 for CAP255) (Figure 2.6B). However the emergence of bNAbs was associated with a significant increase in levels of positive selection. For CAP177, we have previously shown that the T/F does not contain a N332 glycan, and that early strain-specific nAbs

drive selection of this glycan, which is subsequently lost as bNAbs develop (250). For CAP255, the N332 glycan is under positive selection, but is nonetheless largely maintained in most late viral variants. Overall, we observed a significant increase in the number of key sites exhibiting selection pressure with the onset of breadth (Figure 2.6), both within the V1V2 and C3V4 bNAb donors.

2.5 Discussion

The study of virus-antibody co-evolution in infected donors who develop bNAbs has provided crucial insights into how viral variation drives antibody lineages towards breadth. However, most studies have focused on bNAb donors making the assumption that specific Env evolutionary patterns are unique to this group. The existence of infected donors who develop strain-specific responses raises the question of whether viral evolutionary pathways in these donors differ from the well-described bNAb donors. The purpose of this study was therefore to compare the evolutionary pathways of HIV envelope glycoproteins in bNAb and non-bNAb donors to identify whether there are distinct viral features that are associated with breadth. Moreover, we used previously generated data describing the targets of both strain-specific antibodies and bNAbs to compare viral evolution specifically within epitope regions. We showed that neither overall viral diversity, nor local diversity in targeted regions is sufficient for breadth. However, we observed higher numbers of positively selected sites in broad neutralizers, and identified several positively selected sites in the V2 and C3 regions that were common to bNAb donors, and limited or absent in participants without neutralization breadth. Selection pressure at these sites also increased with onset of neutralization breadth, highlighting the role of targeted viral selection in the development of breadth.

One of the consistent factors associated with the development of bNAbs in cohort studies is the presence of high levels of viral diversity (167, 168, 172, 247, 248). However, in this study, Env diversity at 3 years of infection showed no significant differences between bNAb and non-bNAb donors. While this may appear to contradict previous cohort studies, it is important to note that most non-bNAb donors in this study were selected to have similarly viral load to the bNAb donors. As viral load and diversity are linked, this analysis indicates that in donors with similar viral loads, overall diversity does not distinguish bNAb donors from non-bNAb donors, and suggests that additional factors influence breadth. The lack of difference in overall Env diversity between the two groups is consistent with several studies showing that bNAbs often represent only a small proportion of the overall neutralizing responses in an infected donor (286) (Kitchin et al. unpublished). Although most of the strain-specific neutralizing antibodies that comprise the overall neutralizing response fail to acquire breadth (as is evident

from the low proportion of bNAb donors in infected cohorts), they nonetheless drive substantial diversity through viral escape. Indeed, there appears to be no clear association between overall autologous neutralization titers and breadth (165, 287, 288).

Superinfection has been reported to be important in the development of breadth although the effect is small (182, 289, 290). Our study included two donors where the tMRCA and actual time of infection differed significantly, suggesting (CAP8) or confirming (CAP256) superinfection. Both donors developed bNAbs, and CAP256 has been intensively studied over many years (182, 186, 189, 291). However, in CAP256 we have recently demonstrated that superinfection played no significant role in driving breadth (288). Furthermore, within the CAPRISA cohorts, we see no association between superinfection and breadth (288).

Many studies defining viral escape from bNAbs have confirmed that despite the conserved nature of these epitopes, HIV can utilize multiple pathways to escape (181, 190, 250). For this reason, bNAbs fail to confer any clinical benefit to those donors in whom they develop (165). In donors who develop breadth, extensive toggling with epitopes has been shown to precede the development of breadth (186, 231). Our observation of higher numbers of positively selected sites in bNAb donors, particularly within the first year of infection, supports the notion that this early but targeted viral toggling is required for the development of breadth, and differentiates bNAb donors from those who remain strain-specific.

The underlying reason for a higher number of positively selected sites in bNAb donors is of interest for HIV vaccine design. This is unlikely to be a consequence of higher titers, as many of the strain-specific responses we have previously mapped have equivalently high titers to bNAb responses (164, 257). The targeting of intrinsically more conserved epitopes may result in fitness costs for some escape mutations, and therefore higher levels of compensatory mutations at proximal sites (292). Alternatively, it is possible that antibodies with the potential to mature towards breadth bind the virus using a wider ”footprint” i.e, they have more contact sites, and therefore drive early escape mutations at more sites. The finding that mutations at many of these

positively selected sites, when introduced into heterologous viruses, confer neutralization escape from bNAbs may support the latter possibility (186, 210, 231, 285, 293, 294).

The identification of sites common to bNAb donors, and their clustered location on the trimeric Env may emphasise the importance of accurate early targetting by bNAbs. Studies of the ontogeny of V1V2-directed bNAb donors have confirmed the need for precursors encoding long third complementarity determining regions to penetrate the glycans shielding the V2 C-strand (where many of the positively selected sites we identify here are located). The tight clustering of positively selected sites for C3V4 bNAb donors in this study may suggest that initial targetting is important here too, though mature members of V3-directed bNAb lineages show much more promiscuity in their angles of approach as they mature to accommodate sequence variation and glycosylation (185, 192–194, 197, 295, 296).

We selected 14 participants for the in-depth epitope study because of the availability of both sequence and mapping data. Furthermore, both V1V2 and N332-directed bNAbs are among the most common bNAb specificities elicited during infection (165, 168, 177, 178, 287), and may require relatively less somatic hypermutation to acquire breadth (182, 198, 201). This makes these bNAb specificities attractive vaccine targets. However, this study could be improved by performing similar analyses on additional participants with/without bNAbs, including other less commonly elicited bNAb specificities. Since the number of participants in this study was relatively small, larger cohorts and inclusion of donors infected by other viral subtypes would be valuable in future. We also utilized single genome Sanger sequencing data with limited depth, whereas sequences generated through next generation sequencing would give a more accurate representation of viral diversity. Nonetheless, this study has highlighted key differences between the evolutionary pathways of HIV *env* glycoproteins in participants with/without bNAbs. Furthermore, it has clarified that amino acid diversity at key positions within the V1V2 and C3V4 epitopes is more important than overall *env* diversity. The association of shared, positively selected sites with the onset of breadth highlights the importance of diversity at these positions in bNAb development. Designing immunogens against HIV that include targeted diversity within these common sites may thus be critical in the maturation of V1V2 and C3V4-

directed bNAbs.

3 Evidence for both intermittent and persistent compartmentalization of HIV-1 in the female genital tract

Batsirai M. Mabvakure[a,b*], Bronwen E. Lambson[a,b*], Kavisha Ramdayal[c], Lindi Masson[d], Dale Kitchin[a], Mushal Allam[e], Salim Abdool Karim[f], Carolyn Williamson[d,f,g], Jo-Ann Passmore[d,f,g], Darren P. Martin[d], Cathrine Scheepers[a,b], Penny L. Moore[a,b,f], Gordon W. Harkins[c] and Lynn Morris[a,b,f]

[a]Centre for HIV and STIs, National Institute for Communicable Diseases of the National Health Laboratory Service, Johannesburg, South Africa

[b]Faculty of Health Sciences, University of the Witwatersrand, Johannesburg, South Africa

[c]South African MRC Bioinformatics Unit, South African National Bioinformatics Institute, University of the Western Cape, South Africa

[d]Institute of Infectious Disease and Molecular Medicine, University of Cape Town, Cape Town, South Africa

[e]Sequencing Core Facility, National Institute for Communicable Diseases of the National Health Laboratory Service, Johannesburg, South Africa

[f]Centre for the AIDS Programme of Research in South Africa (CAPRISA), KwaZulu-Natal, South Africa

[g]National Health Laboratory Service, Johannesburg, South Africa

* Joint first authors – BM performed all bioinformatics analyses and wrote the paper. BEL generated viral Env sequences

Article published: Mabvakure et al. 2019. Journal of Virology. Mar 6. pii: JVI.00311-19. doi: 10.1128/JVI.00311-19

3.1 Abstract

HIV-1 has been shown to evolve independently in different anatomical compartments but studies in the female genital tract have been inconclusive. Here we compared the evolution of the HIV-1 envelope (Env) glycoprotein in cervico-vaginal lavages (CVL) to matched plasma samples over 2-3 years of infection for evidence of compartmentalization. HIV-1 gp160 amplification from CVL was achieved for only four of 18 acutely infected women and this was associated with the presence of pro-inflammatory cytokines and/or measurable viremia in the CVL. Maximum likelihood trees and divergence analyses showed that all four individuals had compartment-specific clusters by monophyly at all or some timepoints. Two participants (CAP177 and CAP217) had CVL diversity patterns that differed from those in plasma and showed restricted viral flow from the CVL. However, time-stamped maximum clade credibility trees showed that only CAP177 had differences in evolutionary patterns with nucleotide substitution rates higher in plasma than CVL sequences. Three statistical tests of compartmentalization revealed evidence of persistent compartment-specific evolution in CAP177, while in CAP217 this was intermittent. Lastly, we identified several sites in Env that distinguished viruses in these two compartments. For CAP177, amino acid differences arose largely through positive selection while indels were more common in CAP217 and in both cases, these contributed to substantial charge changes located across the Env. Our data indicates that, in some women, distinct HIV-1 populations with different Env genetic features can arise in the genital tract, which could potentially impact sensitivity to vaginal microbicides and vaccine-elicited antibodies.

3.2 Introduction

Globally, there were almost two million new HIV-1 infections during 2017, with the majority occurring among women in sub-Saharan Africa via vaginal intercourse (297). HIV-1 subtype C is the predominant circulating genetic subtype in southern Africa where the epidemic is mostly concentrated, particularly in women (297). It is not well understood whether viral populations in the female genital tract are genetically distinct from those in the blood compartment and if so, how these populations emerge and evolve over time. Such studies on the properties of HIV viral populations in the female genital tract are important for the design of effective microbicides and vaccines that have the potential to reduce vaginal HIV transmissions in this region (40).

Compartmentalization is well described in HIV infection, with these analyses generally using envelope (Env) glycoprotein sequences which shows the greatest diversity (78, 298). A number of factors favour compartmentalization such as barriers between tissues, differential host cell availability and tissue-specific immune selection pressures. While differences in viral replication rates in different compartments are likely the main driver of compartmentalization, genetic drift also likely contributes (78, 299). As a result, compartmentalized viral populations often possess distinct tissue-specific phenotypic characteristics when compared to plasma viruses, including differences in cell tropism, the extent of glycosylation, and drug resistance (78). These independent viral populations may produce variants with advantageous phenotypes for viral proliferation when compared to parental strains (300). Compartmentalization is particularly well described in the cerebrospinal fluid and the central nervous system, where it has been reported to affect responses to treatment and is associated with various brain complications (80, 83–86, 88). There is also convincing evidence that the semen constitutes a distinct compartment where locally produced HIV lineages evolve under different selection pressures to those in the blood (91, 93, 301–303). Relative to blood derived viral populations, semen-derived populations show less diversity, decreased levels of positive selection, decreased CXCR4 coreceptor utilization and altered glycosylation patterns (92). These differences may be at least in part attributable to an enrichment of cytokines and chemokines in the seminal tracts of HIV-infected men that promote T cell activation and viral replication (94).

Studies of HIV compartmentalization in the female genital tract have shown evidence of tissue-specific differences in some, but not all of the HIV-infected women (97, 103–106, 304–309). In certain individuals, compartmentalization was shown to be associated with distinct viral features including coreceptor usage, number of N-linked glycosylation sites, neutralization resistance and viral diversity (97, 104, 304, 308, 310). The temporal scope of most of these studies has largely been limited by the use of cross-sectional sampling. Moreover, investigations that used more advanced computational and statistical analyses failed to find evidence of compartmentalization, and suggested such studies can be biased by monotypic and low diversity sequences (95, 96). A more recent study showed that there was higher HIV-1 Env sequence diversity in the vaginal tract early in infection than there was in the blood (311). This suggested the possibility of either restrictions in movement of viruses from the genital tract to the systemic blood compartment, or differences in HIV evolution between these compartments (311). However, this study utilized only the C2-V3-C3 region of the Env and might have missed other tissue-specific signatures elsewhere in Env. The only detailed study of HIV compartmentalization in the female genital tract that made use of complete gp160 sequences analysed only a small number of sequences and did not include longitudinal samples from the same individuals (106).

In this study, we amplified longitudinal complete HIV-1 subtype C gp160 envelopes sequences from the cervico-vaginal lavage (CVL) supernatant and plasma samples, and analysed these for evidence of compartmentalization using statistical methods. We present evidence for the existence of separate populations that are maintained over time in the genital tract and plasma compartments of some study participants. These participants showed an asymmetric pattern of viral migration between the genital tract and plasma with fewer movements to the plasma compartment, suggesting a possible mucosal sieve. We conclude that compartmentalization may occur in some individuals yielding viruses in the genital tract that are genetically distinct from those in the plasma.

3.3 Materials and methods

3.3.1 Participants

This study made use of participants from the CAPRISA 002 Acute Infection study, a cohort of 245 uninfected high-risk women, which was established in KwaZulu-Natal in 2004 (262). The parent study was reviewed and approved by the research ethics committees of the University of KwaZulu-Natal (E013/04), the University of Cape Town (025/2004), and the University of the Witwatersrand (MM040202). All participants provided written informed consent for sample storage for research purposes. Ethics approval for this sub-study was obtained from the University of the Witwatersrand (M160340).

Stored plasma and CVL samples from 18 women (ages 18–59 years) collected during acute HIV infection were selected for this study. The time of infection was defined as the mid-point between the last HIV-1 antibody negative test and the first HIV-1 antibody positive test or estimated to be 14 days prior to a positive RNA polymerase chain reaction (PCR) assay when the HIV-1 Enzyme Immunoassay (EIA) was negative. Plasma and CVL samples were collected at multiple matched timepoints from acute to chronic HIV infection (2–190 weeks post-infection).

3.3.2 Collection of CVL samples

The cervix was bathed with 10ml of phosphate buffered saline before aspirating the fluid from the posterior fornix. This was repeated with the same fluid 2 to 3 times. The collected fluid was centrifuged at 400g for 10 minutes to fractionate the cellular component from the supernatant. The supernatant was aliquoted and stored at -70^{0}C. CVL samples were not collected during menstruation and supernatants with macroscopic blood contamination were not used. The presence of microscopic blood was detected with Roche Cobas combur dipsticks which detect 5 to 10 erythrocytes/µl (1+ score) to 250 erythrocytes/µl (4+ score). Viral load per ml was measured in the 1 ml CVL supernatants with Nuclisens Easyq HIV1 version 1.2 which has a limit of detection of 50 copies per ml.

3.3.3 Isolation of HIV RNA from CVL supernatant and plasma

Two different extraction procedures were tested to isolate viral RNA from CVL, either the VitalVirus followed by QIAamp Viral RNA MiniKit (VitalVirus+QIAamp) or QIAamp Viral RNA MiniKit on its own. The µMACS™ VitalVirus HIV Isolation Kit (Miltenyi Biotec, Bergisch Gladbach, Germany) captures virus particles with magnetic MicroBeads that bind to host derived CD44 in the viral envelope. Viral RNA was then isolated from the captured concentrated, intact virions with the QIAamp® Viral RNA Mini Kit (Qiagen, Hilden, Germany).

Using the QIAamp Viral RNA MiniKit, RNA was isolated from 0.2ml - 1.75ml CVL supernatant or from 140µl plasma. When sample volumes exceeded 0.6ml, the supernatant was centrifuged at 23,000g for 1 hour at 4^0C and the supernatant was discarded except for 140µl in the base of the tube. In order to ensure that all PCR products had been amplified from cDNA and not DNA, during RNA purification an on-column DNase digestion step (RNase-Free DNase Set, Qiagen) was added to the QIAamp procedure. The QIAamp extraction procedure on its own yielded more amplicons from CVL than VitalVirus+QIAamp and so the former was used for all subsequent amplifications.

3.3.4 Single genome amplification (SGA) and gp160 sequencing

Viral RNA was reverse transcribed to cDNA using SuperScript™ III Reverse Transcriptase (Invitrogen, Carlsbad, CA) and the *env* primer OFM19 (264). The complete *env* was amplified in a nested PCR reaction. In order to ensure that *env* genes were amplified from single cDNA copies, cDNA prepared from plasma was diluted before PCR until fewer than 30% of reactions were positive. In most cases, it was not necessary to dilute the CVL cDNA as few reactions were positive. Amplicons were directly sequenced using the ABI PRISM Big Dye Terminator Cycle Sequencing Ready Reaction kit (Applied Biosystems, Foster City, CA) and resolved on an ABI 3100 automated genetic analyzer. The full length *env* sequences were assembled and

edited using Sequencher v.4.5 software (Gene Codes, Ann Arbor, MI). Amplicons with double peaks or interrupted reading frames were not included in the sequence analysis.

3.3.5 Measurement of cytokines

The concentrations of 20 cytokines were measured in CVL samples from acute infection using Luminex Multiplex Flow Cytometric assays. After thawing, CVL samples were pre-filtered by centrifugation using 0.2 µm cellulose acetate filters (Sigma, USA). Concentrations of eotaxin/CCL11, fractalkine/CX_3CL1, granulocyte colony stimulating factor (G-CSF), interleukin (IL)-1α, IL-8/CXCL8, IL-12p40, IL-15, monocyte chemotactic protein (MCP)-1/CCL2, macrophage inflammatory protein (MIP)-1α/CCL3, MIP-1β/CCL4, RANTES/CCL5, soluble CD40 ligand (sCD40L), were measured using Human Cytokine LINCOplex kits (LINCO Research, MO, USA) according to the manufacturer's protocol. The sensitivity of these kits ranged between 0.3 and 18.3 pg/ml for each of the 12 cytokines measured. Concentrations of IL-1β, IL-2, IL-6, IL-7, IL-10, IL-12p70, granulocyte macrophage (GM)-CSF, and TNF-α were measured in CVL using High Sensitivity Human Cytokine LINCOplex kits. The sensitivity ranged between 0.01 and 0.48 pg/ml for each of the 8 cytokines measured with high sensitivity. Data was collected using the Bio-Plex™ Suspension Array Reader (Bio-Rad Laboratories Inc®) and a 5 PL regression formula was used to calculate sample concentrations from the standard curves. Data was analysed using BIO-plex manager software (version 4). Cytokine concentrations that were below the lower limit of detection of the assay were reported as the mid-point between the lowest concentration measured for each cytokine and zero.

Multivariate statistical analyses were performed using STATA™ (StataCorp, Texas, USA) and a hierarchical clustering analysis using R. Confirmatory factor analysis was used to group cytokines according to biological functions and to generate factor scores for each participant. Factor scores are linear combinations of the concentrations of each cytokine in a factor, weighted according to their factor loadings. Cytokines were grouped as follows: Pro-inflammatory (IL-1α, IL-1β, IL-6, IL-12p40, IL-12p70, TNFα), hematopoietic (G-CSF, GM-CSF), chemokines (IL-8, fractalkine, eotaxin, MCP-1, MIP-

1α, MIP-1β, RANTES), anti-inflammatory (IL-10) and adaptive immune mediators (IL-2, IL-7, IL-15, sCD40L). Factor scores were compared using Mann-Whitney U test and P-values <0.05 were considered statistically significant.

3.3.6 Sequence alignments

Env nucleotide sequences from each participant were initially aligned with ClustalO (312) using HXB2 as the reference sequence. Sequence alignments were then codon aligned using the Se-Al v2.0a11 (http://tree.bio.ed.ac.uk/software/seal/) sequence alignment editor (267) to avoid the inclusion of DNA encoding frame shifts and mistranslated sequences that would affect downstream analyses.

3.3.7 Phylogenetic analyses

Pairwise genetic distances between sequences were calculated using MEGA v6 (268). Maximum likelihood trees were constructed using PHYML (271) implemented in Recombination Detection Program (313). Nucleotide substitution saturation was analysed in DAMBE7 (314). Maximum clade credibility (MCC) trees were produced for CVL and plasma sequences in Bayesian Evolutionary Analyses by Sampling Trees (BEAST) v 1.8.4 (238, 272, 315). The phylogenetic trees were viewed in Figtree (276).

3.3.8 Bayesian analysis of viral compartmentalization

Posterior distributions containing 1000 trees for both plasma and CVL sequences were constructed in BEAST for all timepoints, and used as input for the Bayesian Tip-association Significance testing (BaTS) analysis (316). Timepoints that only had sequences from one compartment type, and identical sequences or low diversity sequences that could potentially inflate the appearance of structure, were excluded in subsequent BaTS analyses. The BaTS analysis performs statistical analyses for evidence of lineage movements between locations (the female genital tract and blood plasma) indicative of compartmentalization using the Association Index (AI), Fitch

Parsimony Score (PS) and the Maximum exclusive single-state clade (MC) tests (316). Evidence of compartmentalization was inferred when significant results for at least two of the three statistical methods were obtained.

3.3.9 Analysis of HIV migration between CVL and plasma compartments

Phylogeography analyses of viral movements between CVL and plasma were performed in BEAST v1.10 (315). The direction of viral flow was determined by applying a discrete diffusion asymmetric model with a Bayesian stochastic search variable selection (BSSVS) and Bayes factor tests to identify the most statistically supported viral movements between the two compartments (317, 318). The average number of migration events between the genital tract and plasma compartments was estimated by reconstructing the state change counts using Markov jumps (319). Bayes Factors support of >5 was considered as significant statistical evidence for viral movement between compartments (317).

3.3.10 Identifying amino acid differences in viruses from different compartments

Amino acid sites in CVL-derived sequences that were different to those in plasma-derived sequences were identified using the Highlighter tool (http://www.hiv.lanl.gov/). The consensuses of plasma-derived sequences were used to identify amino acid sites that were different to those found in the CVL-derived sequences. Only codons in the non-overlapping regions of the Env were analysed. The frequencies of any differences per site and the percentage differences were then calculated. Sites that were evolving under positive selection were detected using the Fast Unbiased Bayesian Approximation (FUBAR) and Mixed Effects Model of Evolution (MEME) methods implemented in HyPhy (277, 278, 320).

3.4 Results

3.4.1 HIV-1 gp160 amplification from the female genital tract over time

CVL samples were obtained from 18 women enrolled in the CAPRISA acute infection study in Durban, South Africa within 2-15 weeks of HIV-1 infection (Table 3.1A). Women were not menstruating at the time of sample collection and there was no evidence of macroscopic blood contamination in CVL samples. Most samples were found to have low numbers of erythrocytes typical of normal mucosal vascularization (321). All participants had detectable plasma viral loads, but only two (CAP177 and CAP217) had detectable CVL viral loads (Table 3.1A). Full length gp160 Env sequences were successfully amplified from the CVL samples from these two participants plus two others (CAP261 and CAP270) who had CVL viral loads of less than 50 copies/ml.

Longitudinal CVL samples from 2.5–3.5 years following acute infection were used to isolate additional sequences from all four participants. CVL viral loads were detectable in at least one timepoint in all cases (Table 3.1B). Amplicons were obtained from the CVL supernatants at all seven timepoints tested for CAP270, while fewer timepoints yielded amplicons for CAP177, CAP217 and CAP261 (5/10, 4/9 and 3/6 timepoints respectively). The total number of sequences from CVL varied from 1 to 19 per timepoint and totalled 42–62 over time for the different participants (Table 3.1B). Similar numbers of HIV-1 gp160 Env sequences from plasma were obtained for all participants at matching timepoints.

Table 3.1: (A) List of 18 CAPRISA individuals who donated early infection CVL samples used in this study. Clinical and demographic features and gp160 amplification results are shown. (B) Longitudinal CVL samples from four CAPRISA donors where gp160 was successful.

A

Participant	Age	Feibig stage	Weeks pi	CD4	Plasma VL	CVL VL	Dipstick*	CVL gp160 PCR
CAP261	18	VI	9	311	821 000	<50	Hb3+/4+	+
CAP177	37	III	4	435	698 000	640	Hb-	+
CAP270	24	V	8	492	76 800	<50	Hb2+/3+	+
CAP217	20	VI	9	375	75 600	1 580	Hb2+/3+	+
CAP137	23	VI	7	494	360 000	<50	Hb -	-
CAP266	21	VI	7	447	238 000	<50	Hb3+	-
CAP276	25	nd	15	365	67 100	nd	Hb3+	-
CAP129	59	VI	4	616	55 900	<50	Hb1+	-
CAP200	32	V/VI	6	560	52 500	<50	Hb4+	-
CAP40	31	nd	3	197	51 600	nd	Hb1+	-
CAP136	25	V	2	581	41 600	<50	Hb1+	-
CAP271	25	VI	6	606	33 100	<50	Hb3+	-
CAP275	22	VI	8	535	26 600	nd	Hb-	-
CAP222	19	V	6	454	15 800	<50	Hb2+	-
CAP262	24	V	7	366	6 560	<50	Hb>4+	-
CAP220	25	V	3	354	2 070	nd	Hb-	-
CAP265	19	VI	12	892	2 000	<50	Hb1+/2+	-
CAP221	25	V/VI	5	800	547	<50	Hb4+	-

B

Participant	Weeks pi	CD4	Plasma VL	CVL VL	Dipstick*	Plasma gp160 PCR	CVL gp160 PCR
CAP177	2	nd	359 000	nd	nd	18	nd
	4	435	698 000	640	Hb-	12	18
	28	444	152 000	nd	Hb-	11	0
	54	370	42 100	630	Hb-	11	19
	80	347	59 200	nd	Hb-	8	5
	132	352	43 800	nd	Hb-	4	1
	186	233	38 800	nd	Hb3+/4+	18	19
						82	**62**
CAP217	2	nd	3 260 000	nd	nd	18	nd
	9	375	75 600	1 580	Hb2+/3+	20	20
	60	638	18 400	<50	Hb4+	6	5
	110	346	1 940	nd	Hb1+	6	5
	190	339	30 000	nd	Hb2+	19	19
						69	**49**
CAP261	9	311	821 000	<50	Hb3+/4+	19	19
	59	471	65 400	2 000	Hb1+	11	13
	137	438	27 500	nd	Hb-	10	10
						40	**42**
CAP270	8	492	76 800	<50	Hb2+/3+	10	15
	15	470	103 200	nd	Hb2+	nd	1
	21	358	687 000	260	Hb1+	8	8
	33	315	310 000	6 500	nd	14	3
	58	308	738 000	3 100	Hb3+	12	11
	85	161	3 710 000	nd	Hb1+	7	5
	129	189	1 290 000	nd	Hb1+	10	12
						61	**55**

nd = not done; pi = post infection; Hb = hemoglobin

*The presence of microscopic blood contamination was detected with Roche Cobas combur dipsticks which detect 5 to 10 erythrocytes/µl (1+ score) to 250 erythrocytes/µl (4+ score).

3.4.2 Association between CVL cytokine levels and envelope PCR during acute infection

We investigated whether there was correlation between Env amplification from CVL and local immune activation in the genital tract, which we determined by measuring cytokine levels (322, 323). CVL samples from acute infection were available from 14

of the 18 participants for this analysis. Participants were grouped according to the concentrations of 20 cytokines measured in their CVL using unsupervised hierarchical clustering. Three of the four participants with amplifiable gp160 clustered together on the heatmap, suggesting an association with higher concentrations of cytokines (Figure 3.1A). Statistical analysis revealed that the pro-inflammatory cytokine cluster was significantly associated with CVL gp160 amplification (Mann-Whitney U test, p=0.024) (Figure 3.1B).

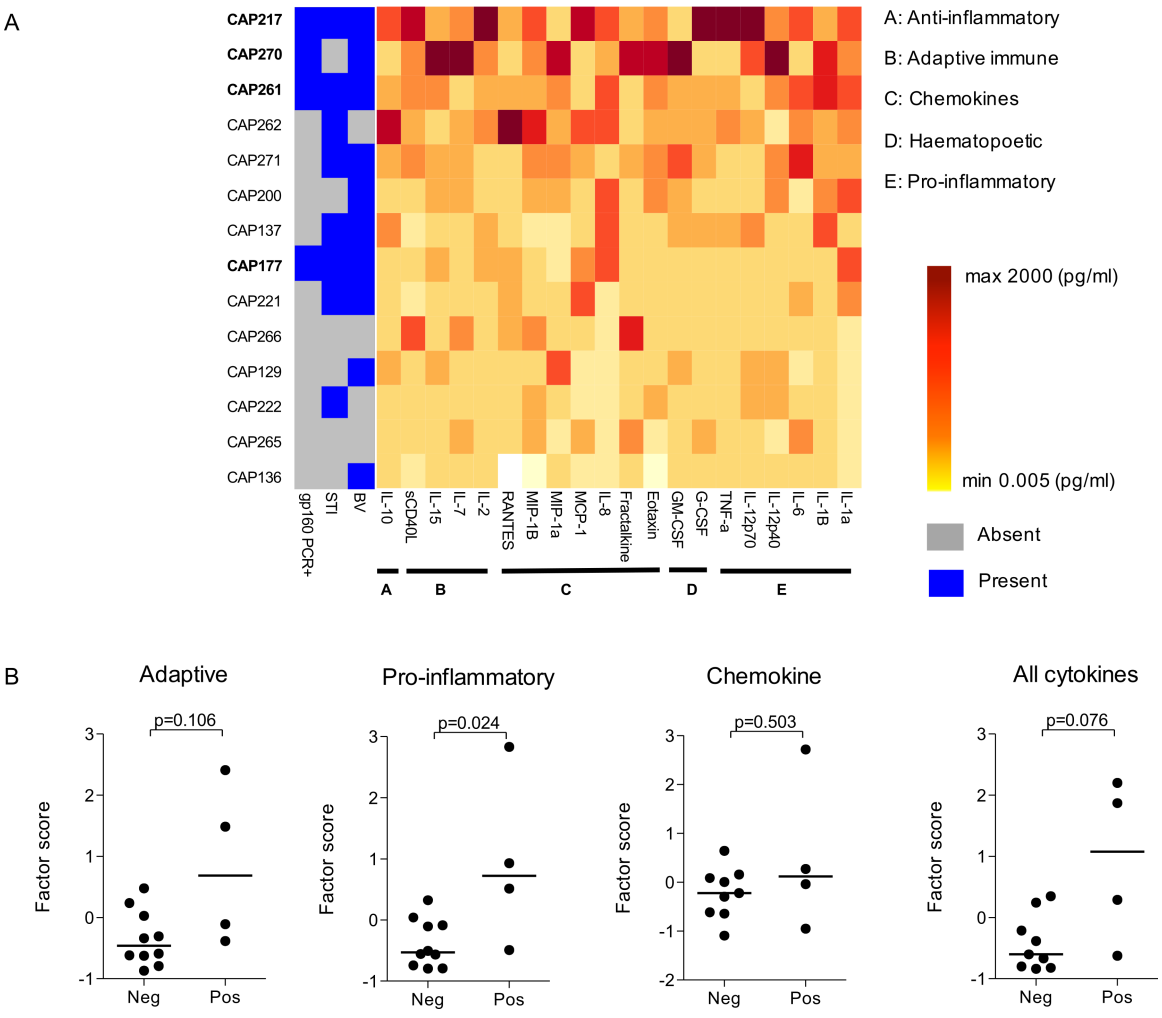


Figure 3.1: Association between cytokine levels, STIs and gp160 amplification in CVL at acute infection. (A) Heat map showing CVL cytokine levels, HIV gp160 PCR results, sexually transmitted infections (STIs) and bacterial vaginosis (BV). Participants with amplifiable CVL gp160 are shown in bold. (B) Confirmatory factor analysis was used to group cytokines according to biological functions and generate factors scores for each cytokine group for each participant. Mann-Whitney U test was used to evaluate differences in factor scores between women with amplifiable gp160 (Pos) and those without (Neg). Lines indicate medians. P-values <0.05 were considered statistically significant.

Sexually transmitted diseases (STIs) and bacterial vaginosis (BV) frequently result in genital tract inflammation, and also increase HIV replication rates in the female genital tract via pro-inflammatory signalling pathways (324–326). We investigated if the ability to amplify gp160 in the genital tract was related to the presence of STIs and BV. Eight of the 14 participants had evidence of an active STI (Figure 3.1A), including the three participants from whom CVL gp160 was amplified. This included CAP177 (*Neisseria gonorrhoeae* positive), CAP261 (HSV-2 PCR positive) and CAP217 (*Mycoplasma genitalium* positive) who all had infections known to be associated with increased HIV shedding (327–330). However, the prevalence of STIs was not associated with gp160 amplification (Fisher's Exact, p=0.5804). Although all four participants with amplifiable gp160 had BV, 6/10 participants from whom gp160 was not amplifiable also had BV and the difference was not statistically significant (Fisher's Exact, p=0.2507; Supplementary Table 3.1). It was however found that the difference in CVL VL between those with amplifiable gp160 and those without approached significance (p=0.0659).

We next performed logistic regression to evaluate whether the relationship between inflammation and amplification was independent of STIs, BV and CVL VL. It was found that STIs, BV and CVL VL were not significantly associated with pro-inflammatory cytokine factor scores (p=0.296, p=0.212 and p=0.141, respectively). Furthermore, despite the small sample size, pro-inflammatory factor scores remained significantly associated with amplification after adjusting for STIs and CVL VL (p=0.0357 and p=0.0303, respectively) and approached significance after adjusting for BV (p=0.0507). Therefore, although pro-inflammatory cytokine production may be influenced by STIs, BV and CVL VL, this analysis suggests that inflammation is associated with viral amplification regardless of the cause.

3.4.3 Compartment-specific clusters in plasma and CVL samples

To determine if sequences in the CVL differed from those in plasma, we constructed maximum likelihood trees to identify monophyletic clades comprised of sequences sharing common phenotypic characteristics and ancestry. For all four participants the plasma and CVL sequences were largely intermingled on the trees, although there was some evidence of sequences grouping together by timepoint and compartment (Figure 3.2). This suggested the possibility of compartmentalization in these individuals. We

therefore further investigated if there were evolutionary differences between the sequences from the plasma and CVL compartments of these individuals.

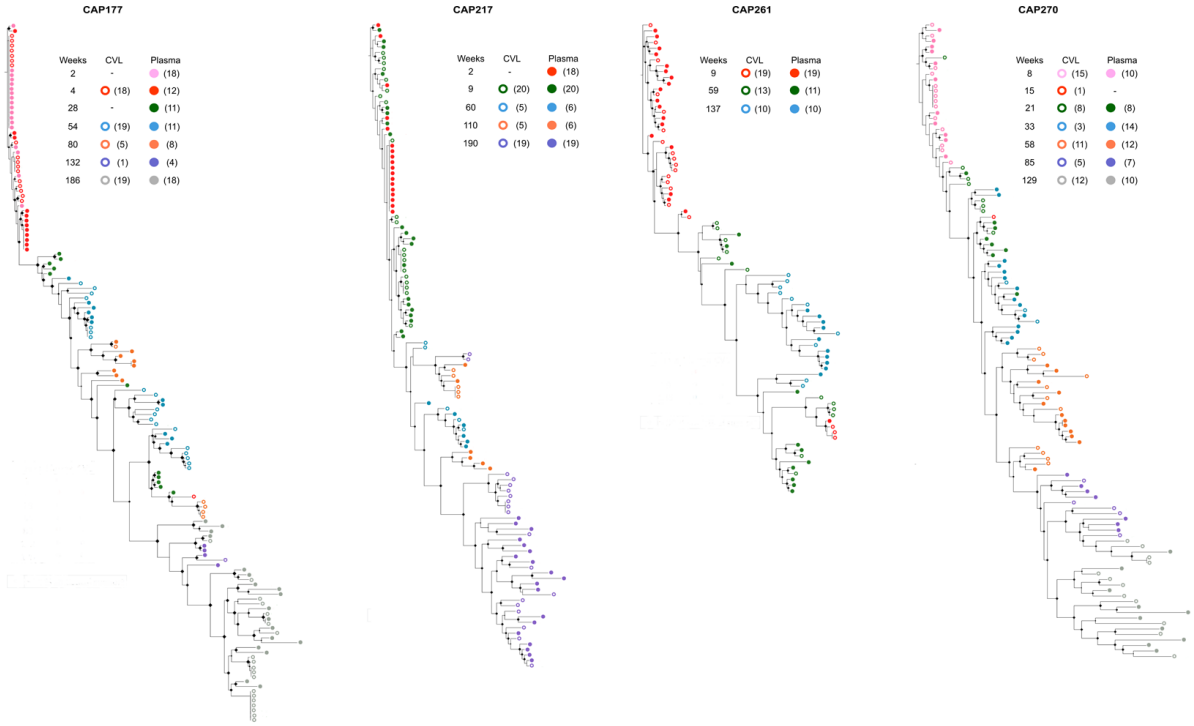


Figure 3.2: Maximum likelihood trees of CVL and plasma Env sequences from four donors constructed in PhyML. Plasma and CVL sequences are indicated by filled and open circles respectively, and the circle color represents the different timepoints. Node support greater than 0.348 is represented by black diamonds. The number of sequences included is shown in brackets.

3.4.4 Statistical analyses of compartmentalization

We next performed statistical tests for evidence of compartmentalization using BaTS (316). We first analysed all sequences from all timepoints and then repeated the analysis with identical sequences removed to rule out the possibility of inaccurate inferences (Supplementary Table 3.1). There was strong evidence of compartmentalization in CAP177 and CAP217 regardless of whether identical sequences were included or not (Figure 3.3A). Monophyletic clade size statistic for CVL was significantly higher than for plasma for both CAP177 and CAP217, providing more support for genital tract compartmentalization. This was consistent with the hypothesis that viruses move less frequently from the genital tract to the plasma compartment than they do in the opposite direction. There was no

compartmentalization in CAP261, with or without identical sequences, while for CAP270 the evidence was lost when identical sequences were removed, suggesting the possibility that bursts of local replication could account for the observed compartmentation in this participant (Figure 3.3A).

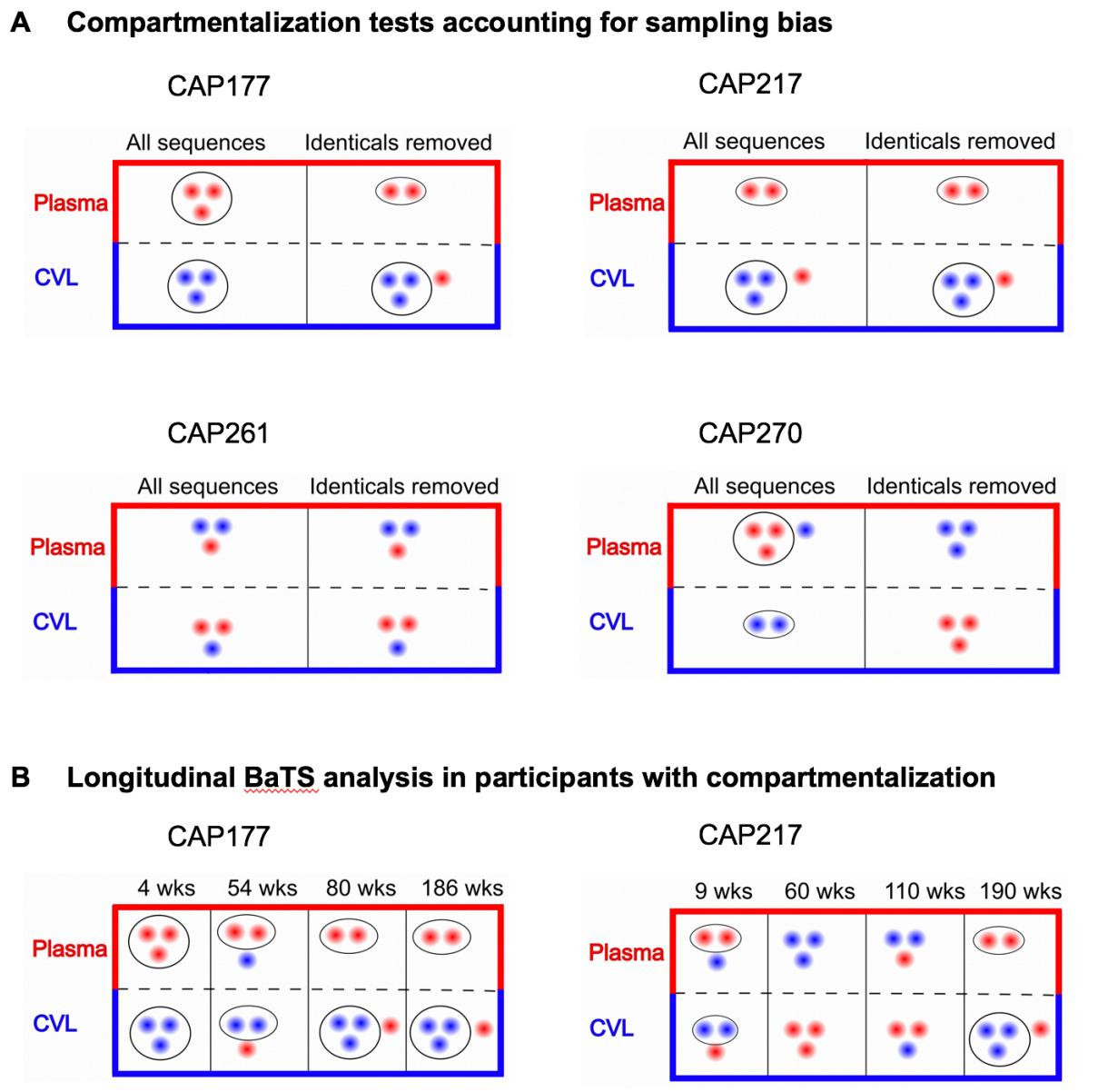


Figure 3.3: Statistical analysis of compartmentalization using BaTS. (A) Analysis of all sequences for each of the 4 participants with/without identical sequences removed. (B) Longitudinal analysis of sequences from 2 participants with evidence of compartmentalization. Each dot represents an independent statistical test for compartmentalization. Circles indicate cases where at least 2 tests were significant ($p<0.05$) for that compartment. Blue and red represent CVL and plasma respectively. Complete compartmentalization is shown by three circled dots in all the two compartments. The timepoints and numbers of sequences used in this analysis is shown in Supplementary Tables 1 and 2.

Having confirmed evidence of compartmentalization in CAP177 and CAP217, we next traced whether this changed during the course of infection by analysing the data at 4 timepoints with the most number of sequences (Supplementary Table 3.2). We identified compartment-specific clusters at all timepoints in CAP177, and at the first and last timepoints in CAP217 (9 and 190 weeks) (Figure 3.3B). We also observed significantly higher monophyletic clade size statistic for CVL than plasma, providing more support for genital tract compartmentalization than for plasma. In summary, the Bayesian statistical analyses suggested compartmentalization in two of the four participants. In CAP177 the evidence of compartmentalization was persistent, whereas in CAP217 it was intermittent.

3.4.5 HIV migration between genital tract and plasma compartments

We next investigated whether the compartmentalization in CAP177 and CAP217 was associated with restricted movement of viruses between the two compartments. To do this, we estimated the numbers of distinct migration events during the histories of the sampled viruses using Markov jumps in BEAST (315, 319). The sequence phylogenies from all four individuals yielded evidence of more viral migration events from the plasma to the genital tract compartment (Markov counts from 16 to 23) than from the genital tract to the plasma compartment (Figure 3.4). This was particularly evident for the CAP177 and CAP217 sequence phylogenies, both of which displayed evidence of only two instances of viruses moving from the genital tract to plasma (i.e. two Markov counts) during the evolutionary histories of all the sampled viruses. In contrast, the CAP261 and CAP270 phylogenies displayed evidence of eight and nine genital tract to plasma migration events, respectively. There was good statistical support for occurrence of at least some movements from the genital tract to plasma as indicated by the Bayes factors (BFs). Specifically, in all the participants other than CAP270, the BFs for CVL to plasma movements was >5 indicating that there was greater than 5 times support for the occurrence of these movements than for their absence.

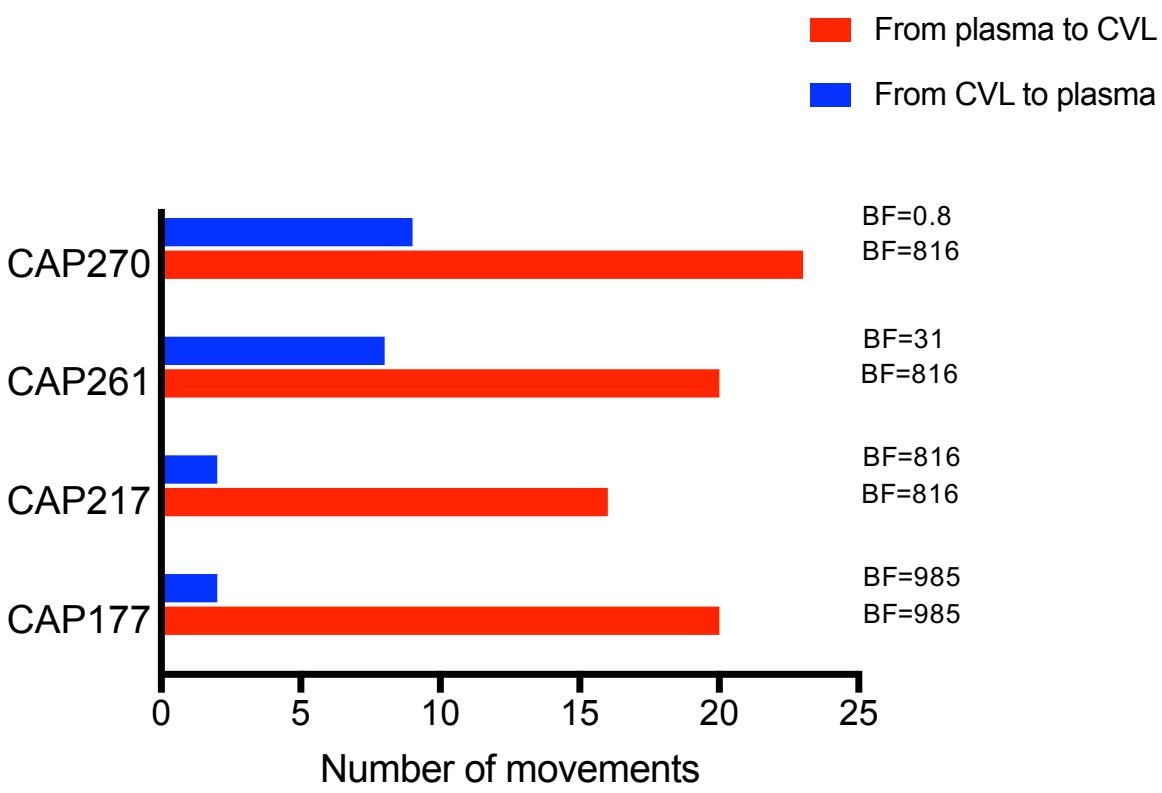


Figure 3.4: Viral migration events between CVL and plasma compartments. Migration events estimated as Markov counts in BEAST and the Bayes factor (BF) support for the movements between anatomical compartments are shown. Bayes Factor values > 5 indicate statistical support for the movements.

3.4.6 Comparison of evolutionary differences between sequences in plasma and CVL

As compartmentalization may arise due to differential viral population diversification in tissues, we estimated the average pairwise genetic distances between the sequences sampled at each of the timepoints in the separate compartments. For CAP177, CAP217 and CAP261 patterns of CVL-derived virus diversity differed from those seen in plasma-derived viruses (Figure 3.5). However, only CAP261 displayed higher viral sequence diversity in CVL compared to plasma at all timepoints, a finding that has been noted in another study (311).

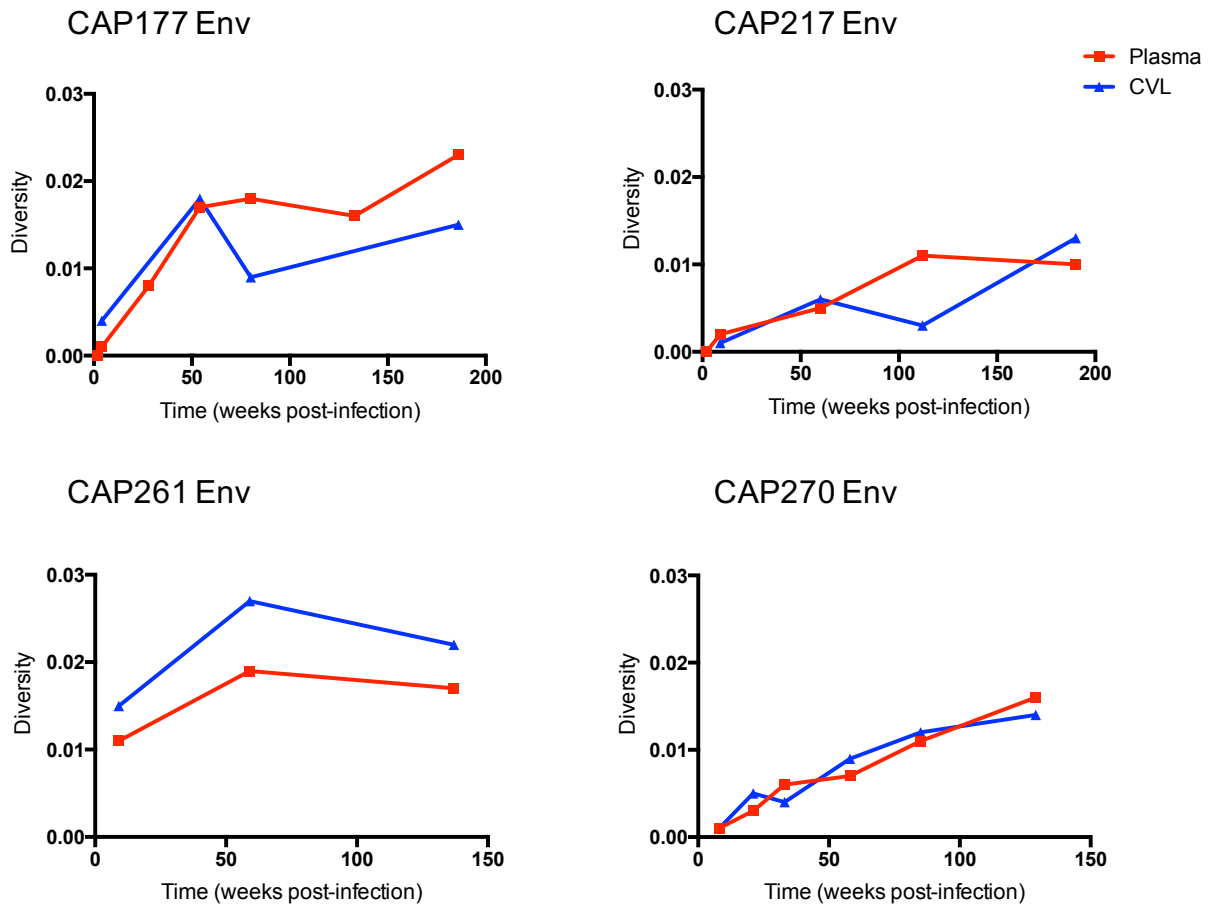


Figure 3.5: Longitudinal Env sequence diversity analyses in plasma and CVL (shown in red and blue respectively) of four individuals. The average nucleotide pairwise distances were estimated in MEGA6.

We next constructed maximum clade credibility trees in BEAST for plasma and CVL sequences separately for each participant to assess whether the observed diversity patterns were as a result of differences in evolutionary rates and ancestry of the viral lineages in the separate anatomical compartments (238). We used all sequences obtained at all available timepoints for these analyses. In three of the four participants no differences in either estimated nucleotide substitution rates, or estimated times to most recent common ancestors (tMRCAs) were noted between the CVL and plasma sequences. For CAP177, however, the estimated tMRCA of the plasma-derived sequences (185 weeks, 95% High Posterior Density (HPD) 184-186 weeks) was more recent than that of the CVL derived sequences (267 weeks, 95% HPD 200-345 weeks). In this individual viral evolutionary rates were also inferred to be higher in plasma (4.5×10^{-4} substitutions/site/week , 95% HPD 3.5×10^{-4} - 5.5×10^{-4} subs/site/week) than in CVL (2.2×10^{-4} substitutions/site/week , 95% HPD 1.5×10^{-4} - 2.9×10^{-4}

subs/site/week). In addition, for both CAP177 and CAP261 the estimated tMRCA for all sequences was greater than the actual time of infection, supporting the hypothesis that these individuals were initially infected by multiple variants. We confirmed in DAMBE (data analysis in molecular biology and evolution) version 7 that our rate estimates were not biased by nucleotide substitution saturation at all codon positions.

3.4.7 Comparison of CVL and plasma amino acid sequences in participants with evidence of compartmentalization

We next compared amino acids in the Env sequences at all sites in the non-overlapping regions to identify compartment-specific differences between the plasma and CVL derived viruses. Using the highlighter tool on the HIV LANL website, we identified a total of 43 and 31 amino acid residues in CAP177 and CAP217 Env sequences that were either more or less common in CVL compared to plasma viruses, by a margin of 40-100% (Figure 3.6A). We performed the comparisons using plasma and CVL sequences at timepoints that showed evidence of compartmentalization (Supplementary Table 3.3). The relative frequencies of synonymous and nonsynonymous substitutions within the codons encoding these amino acids estimated using FUBAR and MEME suggested that these amino acid differences were potentially a consequence of positive selection in CAP177, and both positive selection and insertions/deletions in CAP217 (Figure 3.6A). These amino acid differences resulted in charge differences at similar positions in the Envs of the plasma and CVL derived sequences suggesting that different amino acids are preferred at these sites in the different compartments. In CAP217, insertions were more common in the V1V2 regions of CVL-derived sequences than they were in plasma-derived sequences (Supplementary Table 3.3). These amino acid differences also increased with time in both participants (Figure 3.6). When we compared the distribution of these sites on the Env of both participants, they were mostly in the V4 and C4 regions in CAP177 Env sequences, and in the V1V2 regions in CAP217 Env sequences (Figure 3.6B and C). In summary, these results suggest that the compartmentalization observed in CAP177 and CAP217 was potentially associated with distinct, and possibly compartment-specific selection acting at codon sites and/or the differential accumulation of insertions and deletions in CVL compared to plasma.

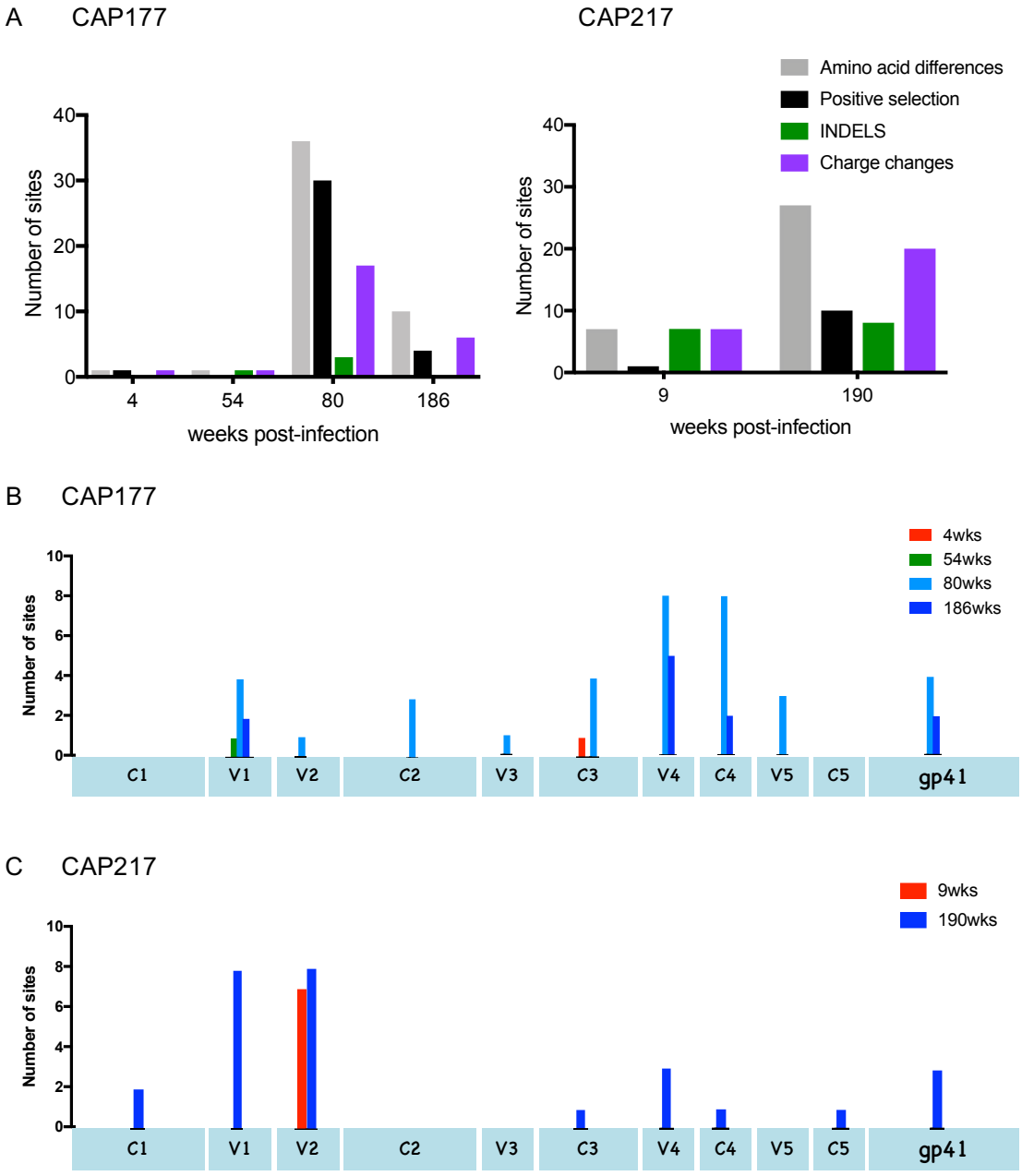


Figure 3.6: Identification of HIV-1 Env glycoprotein sites that distinguish CVL and plasma sequences in individuals with evidence of compartmentalization. (A) Number of sites that are different between plasma and CVL sequences shown in grey (p-value $\leqslant 0.05$), positive selection on those sites estimated in FUBAR and MEME in black, INDELS on the sites shown in green, and the amino acid differences that resulted in charge changes are shown in purple. (B and C) Distribution of the amino acid differences on the glycoprotein Envs over time, the bars are colored by timepoint.

3.5 Discussion

Our study provides evidence of compartmentalization of HIV-1 subtype C in the genital tract in two of the four women studied. Using multiple analytical approaches, we showed that compartmentalization could be persistent or intermittent. Furthermore, we identified Env features in viruses from the genital tract that differentiated them from those found in plasma. In all four cases there was evidence of restricted migrations of viruses from the genital tract to plasma, which was more pronounced in individuals with compartmentalization. These data suggest that the female genital tract and the plasma can in some cases, act as separate compartments, each with a virus population that evolves with a degree of independence from the other compartment.

Studies focusing on the detection and characterization of viral populations in the female genital tract have proven difficult due to challenges in collecting genital tract samples and the generally low viral loads in this compartment (311, 331–333). We made use of CVL samples which were collected at multiple timepoints from women in the CAPRISA cohort enrolled during acute HIV-1 infection and followed through to chronic infection (322). Although sampling by means of CVL yields a good representation of a large genital tract area, the high degree of genital fluid dilution likely contributed to our failure to amplify Env from most of the CVL samples. Other studies that have had greater success amplifying HIV sequences from the female genital tract have made use of swabs or endocervical cytobrushes (332, 334–337). However, these methods capture only the viral population at the specific site in the cervix from which samples were taken and cannot be used to study cytokine levels. The use of CVL samples allowed us to discover that HIV amplification was significantly associated with levels of pro-inflammatory cytokines in the female genital tract. Furthermore, all four women from which viral sequences in CVL samples were amplifiable at the acute infection stage had clinical or laboratory evidence of an STI and/or detectable CVL viral load. Local immune activation associated with STIs often results in an influx of leukocytes, elevated cytokine concentrations and increased HIV shedding that may have favoured viral replication (337, 338).

We used a combination of genetic distance-based methods, Bayesian methods and natural selection analyses of sequences sampled from plasma and the genital tract to detect and characterize compartmentalization of viral populations at these sites.

Previous studies have used approaches such as genetic diversity and phylogenetic analyses to infer compartmentalization (105, 106, 306, 309), but as shown here, it is difficult to achieve robust conclusions based on the results from a single analytical approach. Until recently, previous compartmentalization studies did not account for low diversity and monotypic sequences which could occur because of resampling or local viral replication (95, 96). We accounted for these potentially confounding effects repeating analyses with identical sequences removed and ensuring that almost equal numbers of SGA-derived gp160 sequences were analyzed from CVL and plasma at matched timepoints. The number of sequences impacted the analyses in one participant (CAP270): time points with the most sequences (19-20 sequences at 9 and 190 weeks) showed strong statistical support compared to the time-points with the lowest number of sequences that lacked statistical power (6 in plasma and 5 in CVL at 60 and 110 weeks). This was however, not the case for CAP177 where the results were consistent at all timepoints regardless the number of sequences. We confirmed evidence of compartmentalization with a combination of approaches that examined different aspects of viral evolution and identified individual-specific residues in Env that differentiated CVL-derived viruses from those in the plasma.

Evidence of HIV compartmentalization between the female genital tract and plasma was strongly supported in two of the four participants. In CAP177, compartmentalization was persistent with evidence for restricted viral movements from the genital tract to the plasma found at all four timepoints up to 186 weeks post infection. In CAP217 compartmentalization was only detected at the first timepoint and at 190 weeks post infection and not at the other two timepoints. Intermittent compartmentalization of HIV populations between the plasma and female genital tract has not been observed previously, but has been reported between the male genital tract and the blood (91). CAP177 and CAP217 were the only two individuals who had detectable levels of virus in CVL at acute infection, suggesting that local viral replication and diversification may have favoured subsequent compartmentalization in the genital tract HIV population from that in the blood.

Two of the study participants, CAP261 and CAP270, displayed no convincing evidence of compartmentalization in that both CVL and plasma populations displayed similar patterns of diversity, indistinguishable evolutionary rates and tMRCA estimates. There were also similar numbers of migration events detected in both directions between the

genital tract and plasma. Other studies have similarly reported a lack of compartmentalization between the female genital tract and blood (103, 106, 304). By studying longitudinal sampling we have further shown that compartmentalization can occur sporadically in some individuals.

Regardless of whether there was strong evidence of compartmentalization, all four participants consistently showed higher numbers of virus migration events from the plasma to the genital tract compartment. Low levels of virus in the CVL may have compromised our ability to detect migration events from the compartment. However, it has been well described that diversity rather than viral population size is a key factor in detecting compartmentalization (92, 97, 339). In general, levels of diversity were similar in both compartments despite the higher viral levels in plasma. Our ability to quantitate migration events in both directions is therefore likely due to the fact that the levels of diversity were sufficiently high in both compartments.

The Env amino acid composition of the 2 individuals with evidence of compartmentalization showed distinctive differences, including varying charges. Such differences have been reported to significantly alter key virologic properties, including cellular tropism and transmission (340). We speculate that the T/F viruses in these individuals potentially had properties that favored the genital tract over the plasma compartment. Individuals with no compartmentalization may have had T/F viruses with amino acid properties that were equally well adapted to the plasma and genital tract compartments. Persistent compartmentalization was associated with distinct evolutionary rates and tMRCA differences between the plasma and CVL-derived sequences, whereas intermittent compartmentalization was associated with more subtle evolutionary differences. A recent study showing that early HIV-1 viruses from the endocervix had higher levels of genetic diversity than those from blood, is consistent with the independent evolution of viruses in the female genital tract and/or the presence of a mucosal sieve effect between these two compartments (311).

In conclusion, compartmentalization of HIV populations between the female genital tract and plasma likely does occur in some individuals despite frequent movements of viruses from the blood to the female genital tract. Our data indicates that in cases where compartmentalization occurs, distinct genetic features of viral Env populations

in the genital tract may impact the efficacy of microbicides and vaccines designed to provide mucosal immunity.

Supplementary Table 3.1: Compartmentalization tests using datasets with monotypic sequences included and excluded respectively. Tests performed using tree-based statistics AI (Association Index), PS (Fitch Parsimony Score) and MC (Maximum exclusive single-state clade) implemented in BaTS. Significant p-values that provide evidence of compartmentalization are shown in red text.

Data with monophyletic sequences included

Participant	n(Plasma)	n(CVL)	Statistic	Observed Mean	p-value
CAP177	82	62	AI	3,08	0,00
			PS	28,19	0,00
			MC (CVL)	7,56	0,01
			MC (Plasma)	9,98	0,02
CAP217	69	49	AI	3,35	0,00
			PS	26,64	0,00
			MC (CVL)	8,00	0,00
			MC (Plasma)	5,63	0,19
CAP261	40	41	AI	3,45	0,14
			PS	22,22	0,01
			MC (CVL)	4,00	0,33
			MC (Plasma)	4,03	0,31
CAP270	62	54	AI	4,68	0,01
			PS	32,12	0,00
			MC (CVL)	4,24	0,25
			MC (Plasma)	7,05	0,02

Data with monophyletic sequences excluded

Participant	n(Plasma)	n(CVL)	Statistic	Observed Mean	p-value
CAP177	46	38	AI	2,35	0,00
			PS	19,74	0,00
			MC (CVL)	5,37	0,03
			MC (Plasma)	3,83	0,39
CAP217	51	49	AI	3,09	0,00
			PS	25,58	0,00
			MC (CVL)	8,00	0,01
			MC (Plasma)	4,17	0,25
CAP261	22	23	AI	1,65	0,09
			PS	12,06	0,03
			MC (CVL)	3,02	0,54
			MC (Plasma)	3,08	0,45
CAP270	28	29	AI	2,94	0,43
			PS	17,83	0,22
			MC (CVL)	3,46	0,50
			MC (Plasma)	3,00	0,40

Supplementary Table 3.2: Longitudinal compartmentalization tests for CAP177, CAP217, CAP261 and CAP270 using BaTS. Tests performed using tree-based statistics AI (Association Index), PS (Fitch Parsimony Score) and MC (Maximum exclusive single-state clade). Significant p-values for evidence of compartmentalization are shown in red text. Timepoints with samples from only one of the two compartments and/or one sequence only were excluded from the analysis.

Participant	Timepoint	n(Plasma)	n(CVL)	Statistic	Observed Mean	p-value
CAP177	4 wks	12	18	AI	0,13	0,00
				PS	2,91	0,00
				MC (CVL)	8,45	0,00
				MC (Plasma)	9,13	0,00
	54 wks	11	19	AI	0,68	0,00
				PS	6,30	0,01
				MC (CVL)	4,62	0,09
				MC (Plasma)	3,00	0,07
	80 wks	8	5	AI	0,26	0,03
				PS	2,00	0,01
				MC (CVL)	4,00	0,01
				MC (Plasma)	3,90	0,13
	186 wks	18	19	AI	1,09	0,00
				PS	8,01	0,00
				MC (CVL)	7,01	0,00
				MC (Plasma)	3,85	0,43

Participant	Timepoint	n(Plasma)	n(CVL)	Statistic	Observed Mean	p-value
CAP217	9 wks	20	20	AI	1,50	0,00
				PS	11,28	0,01
				MC (CVL)	3,53	0,35
				MC (Plasma)	3,47	0,34
	60 wks	5	6	AI	0,42	0,24
				PS	3,84	0,59
				MC (CVL)	1,91	0,25
				MC (Plasma)	2,22	0,44
	110 wks	6	5	AI	0,34	0,08
				PS	2,35	0,02
				MC (CVL)	3,00	0,14
				MC (Plasma)	4,00	0,06
	190 wks	19	19	AI	1,23	0,01
				PS	7,88	0,00
				MC (CVL)	10,00	0,00
				MC (Plasma)	3,41	0,42

Supplementary Table 3.3: Plasma and CVL amino acid differences in (A) CAP177 and (B) CAP217 alignments (p-value=0.05 using Fisher's test). The sites that are shown in grey were significantly evolving under positive selection using FUBAR and MEME. The sites that have asterisks (*) are number positions on the alignment.

CAP177

A

Time/weeks	HXB2 site	Mutation	Region	Total CVL sequences	Total Plasma sequences	CVL sequences with specified mutation	Plasma sequences with specified mutation	% difference	Plasma charge	CVL charge
4	355	E	C3	18	12	18	1	91,67	N	-
54	141	deletion	V1	19	11	7	0	100,00	N	
80	142	I	V1	5	8	5	1	87,50	N	N
80	*156	H	V1	5	8	4	1	84,38	N	+
80	148	T	V1	5	8	4	0	100,00	-	N
80	151	N	V1	5	8	4	0	100,00	+	N
80	160	T	V2	5	8	4	0	100,00	N	N
80	281	I and S	C2	5	8	4	0	100,00	N	N
80	293	K	C2	5	8	4	0	100,00	-	+
80	295	E	C2	5	8	4	0	100,00	+	-
80	330	Y	V3	5	8	4	0	100,00	+	N
80	337	E	C3	5	8	4	0	100,00	-	-
80	354	S	C3	5	8	4	0	100,00	N	N
80	355	R	C3	5	8	4	0	100,00	N	+
80	383	L	C3	5	8	4	0	100,00	N	N
80	389	E	V4	5	8	4	0	100,00	+	-
80	392	N	V4	5	8	4	0	100,00	N	N
80	406	N (NNGGN motif in CVL)	V4	5	8	5	2	75,00	N	N
80	407	N (NNGGN motif in CVL)	V4	5	8	4	0	100,00	N	N
80	408	G (NNGGN motif in CVL)	V4	5	8	4	1	84,38	N	N
80	409	G (NNGGN motif in CVL)	V4	5	8	5	2	75,00	-	N
80	410	N (NNGGN motif in CVL)	V4	5	8	4	0	100,00	-	N
80	416	I	V4	5	8	4	0	100,00	N	N
80	429	G	C4	5	8	4	0	100,00	-	N
80	448	D	C4	5	8	4	0	100,00	N	-
80	460	S	C4	5	8	4	0	100,00	N	N
80	461	deletion	C4	5	8	4	0	100,00	N	
80	462	deletion	C4	5	8	4	0	100,00	N	
80	463	T	C4	5	8	4	0	100,00	N	N
80	464	N	C4	5	8	4	0	100,00	N	N
80	deletion (*475)	T	C4	5	8	4	0	100,00	N	N
80	*476	I (insertion)	V5	5	8	4	0	100,00		N
80	*478	deletion	V5	5	8	4	0	100,00	N	
80	467	I	V5	5	8	4	0	100,00	N	N
80		H	gp41	5	8	4	1	84,38	-	+
80	644	R	gp41	5	8	4	1	84,38	+	+
80	648	D	gp41	5	8	4	1	84,38	-	-
80	833	T	gp41	5	8	4	0	100,00	N	N
186	133	D	V1	19	18	1	6	84,21	N	-
186	144	D	V1	19	18	5	0	100,00	N	-
186	400	L (insertion)	V4	19	18	13	3	75,64		N
186	402	T (insertion)	V4	19	18	13	4	67,52		N
186	403	N (insertion)	V4	19	18	13	4	67,52		N
186	407	I and S	V4	19	18	12	2	82,41	N	N
186	460	H	C4	19	18	5	0	100,00	N	+
186	461	K (insertion)	C4	19	18	5	0	100,00		+
186	561	V	gp41	19	18	0	7	100,00	N	N
186	644	R	gp41	19	18	0	7	100,00	+	+

Supplementary Table 3.3 (continued):

CAP217

B

Time/weeks	HXB2 site	Mutation	Region	Total CVL sequences	Total Plasma sequences	CVL sequences with specified mutation	Plasma sequences with specified mutation	% difference	Plasma charge	CVL charge
9	*206	K (Insertion)	V2	20	20	10	17	41,18		+
9	*207	E (Insertion)	V2	20	20	10	17	41,18		-
9	*208	T (Insertion)	V2	20	20	10	17	41,18		N
9	*209	S (Insertion)	V2	20	20	10	17	41,18		N
9	*210	N (Insertion)	V2	20	20	10	17	41,18		N
9	*211	S (Insertion)	V2	20	20	10	17	41,18		N
9	*212	S (Insertion)	V2	20	20	10	17	41,18		N
190	87	G	C1	19	19	10	1	90,00	-	N
190	130	D	C1	19	19	10	0	100,00	N	-
190	*136	deletion	V1	19	19	10	0	100,00	N	
190	*137	deletion	V1	19	19	10	0	100,00	N	
190	*144	deletion	V1	19	19	10	1	90,00	N	
190	*145	deletion	V1	19	19	10	1	90,00	N	
190	*146	deletion	V1	19	19	10	1	90,00	N	
190	*147	deletion	V1	19	19	10	1	90,00	N	
190	*154	deletion	V1	19	19	10	1	90,00	N	
190	*155	deletion	V1	19	19	10	1	90,00	N	
190	171	Q	V2	19	19	17	7	58,82	+	N
190	181	L	V2	19	19	0	8	100,00	N	N
190	186	E	V2	19	19	9	0	100,00	+	-
190	*205	N (insertion)	V2	19	19	10	0	100,00		N
190	*206	K (insertion)	V2	19	19	10	0	100,00		+
190	*208	T (insertion)	V2	19	19	10	0	100,00		N
190	*209	S (insertion)	V2	19	19	9	0	100,00		N
190	*213	K	V2	19	19	10	0	100,00	-	+
190	345	I	C3	19	19	7	0	100,00	N	N
190	397	D	V4	19	19	15	7	53,33	N	-
190	399	T	V4	19	19	10	2	80,00	+	N
190	416	L	V4	19	19	10	0	100,00	N	N
190	429	G	C4	19	19	10	0	100,00	-	N
190	492	K	C5	19	19	13	5	61,54	+	+
190	818	T	gp41	19	19	10	0	100,00	N	N
190	851	L	gp41	19	19	10	0	100,00	N	N
190	853	I	gp41	19	19	9	17	47,06	N	N

4 Design and application of a low-cost bioinformatics cluster for resource-limited settings

Batsirai M. Mabvakure[a,b], Raymond Rott[c], Leslie Dobrowsky[c], Peter Van Heusden[d], Lynn Morris[a,b,e], Cathrine Scheepers[a, b] and Penny L. Moore[a,b,e]

[a]Center for HIV and STIs, National Institute for Communicable Diseases of the National Health Laboratory Service, Johannesburg.

[b]School of Pathology, Faculty of Health Sciences, University of the Witwatersrand, Johannesburg.

[c]Bridge-the-Gap, Johannesburg.

[d]South African National Bioinformatics Institute, University of the Western Cape.

[e]Centre for the AIDS Programme of Research in South Africa (CAPRISA), Kwa-Zulu Natal

4.1 Abstract

Next Generation Sequencing (NGS) technologies have revolutionized biological research by generating genomic data that were once unaffordable by traditional first-generation sequencing technologies. These sequencing methodologies provide an opportunity for in-depth analyses of host and pathogen genomes as they are able to sequence millions of templates at a time. However, these large datasets can only be efficiently explored using bioinformatics analyses requiring huge data storage and computational resources adapted for high-performance processing. High-performance computing allows for efficient handling of large data and tasks that may require multi-threading and prolonged computational times, which is not feasible with ordinary computers. However, high-performance computing resources are costly, and therefore not always readily available in low-income settings. We report the establishment of an affordable high-performance computing bioinformatics cluster consisting of 3 nodes, constructed using ordinary desktop computers and open-source software including Linux-Fedora, SLURM Workload Manager and a Conda package manager. This cluster out-performed high specification desktop computers to analyse large sets of antibody data and perform complex viral phylodynamics analyses. This has demonstrated that it is possible to construct high-performance computing capacity capable of analyzing large NGS data from relatively low-cost hardware and entirely free (open-source) software, even in resource limited settings. Such a cluster design has broad utility beyond bioinformatics to other studies that require high performance computing.

4.2 Introduction

Next Generation Sequencing (NGS) technologies have revolutionized the sequencing landscape through their ability to generate millions of sequences at a time, and, provided new insights into pathogen and host evolution (341–344). Furthermore, the development of cheaper NGS systems means that this is no longer only the domain of well-resourced centers of excellence and is now becoming a standard tool in many molecular labs (345). For example, the Oxford Nanopore Minion costs just under US$1,000, and each flow cell can generate 10–20 Gb of DNA sequence data (346). Although NGS technologies are getting increasingly portable and cheaper, data management is lagging behind (347). The amount of data generated using NGS makes it challenging to analyze using desktop computers, as this comes with significant memory and storage requirements.

Some bioinformatics programs required for such analyses are computationally intensive, requiring high processing power and long computational times either because of the large datasets being analyzed or the complex calculations and simulations performed (323, 348). Clusters and servers have been tailored to perform such analyses, however these are very expensive to establish. This means that researchers from low-income settings (where pathogen NGS data would be particularly useful to confront high burdens of disease) might not be able to afford to buy or rent the necessary computational resources. In settings where existing computational resources are available but scarce, internet is often a limiting factor. Transfer of large datasets between high-performance computing centers and the research centers is often very expensive, or the internet connections are too slow to be effective. In addition, some of the research data (for example patient data) may contain personal and sensitive information that require high levels of protection, making it risky to transfer into public domains. Also, many institutions have firewalls that prevent users from accessing outside servers and clusters to analyze their data, forcing researchers to perform analyses using ordinary desktop computers.

Our laboratory is focused on studies of viral-host evolution in the context of HIV infection and vaccination. Understanding the interplay between the host and virus is a major basis of HIV vaccine design (349). We use bioinformatics analyses of Sanger and NGS to analyze viral and antibody gene evolution. The human antibody repertoire

is enormous, at greater than 10^{11} molecules per individual (149–152). NGS technologies are therefore ideally suited for antibody repertoire analyses as they are able to sequence millions of antibody reads at a time. Similarly, the HIV envelope protein, which is the target of antibody responses during HIV infection, varies by as much as 30% between infected individuals (65, 66). However, the amount of data generated through these types of studies has resulted in our laboratory encountering the computational challenges described above.

To enhance our analyses, we have developed a low-cost bioinformatics cluster that is easy to build, capable of analyzing large NGS datasets and performing phylodynamic analyses, and less reliant on internet data usage. This system, which is amenable to analysis of virtually any large dataset, will enhance the ability of under-resourced researchers to independently interrogate large datasets to address locally relevant scientific problems.

4.3 Methods

4.3.1 Development of a multi-node cluster

4.3.1.1 Computer hardware

A 3-node (20 thread) cluster, consisting of a master node and two subsidiary nodes was developed using ordinary personal computer workstations (Figure 4.1, Supplementary Table 4.1). The master node (Bio-linux) is an Intel(R) Xeon(R) CPU E3-1220 v3 @ 3.10GHz, with 4 CPU cores. It has the x86_64 architecture with 32G memory (RAM). Nodes 01 and 02 are Intel(R) Core(TM) i7-3770 CPU @ 3.40GHz machines, 4 cores per socket, 2 threads per core and a total of 8 threads. They have the x86_64 architecture and real memory of 4 Gigabytes (G). We increased the memory of each to the maximum capacity so that they could handle larger datasets and more tasks. The master node, Node 01 and the Bio-linuxnode 2 were all upgraded to 28G of memory. We also increased the space on the hard drive from 500G to 1 Terabytes (T) and installed 11T and 50T storage drives for raw and processed data respectively (Figure 4.1, Supplementary Table 4.1). Files are also backed up to a file server that is further backed up to Linear Tape-Open (LTO) tapes.

4.3.1.2 Operating system and packages

Fedora release 23 was installed on all the nodes. The standard installation of linux was used, which does not have a graphical user interface. Fedora uses RPM (Redhat Package Manager) and the packages were installed from their repositories using either “yum install” or “dnf install”. The packages and all file components were extracted during installation and stored in the correct locations on the system (the default location being /var/lib/rpm). Cluster users were created using the Fedora dashboard.

4.3.1.3 Networking

Users login to the Master node, which is connected to the external Internet (Figure 4.1). All the other nodes are connected via a Local Area Network (LAN) to the master node. External access to the cluster is by ssh on a non-standard port to reduce the risk of port scanning by automated bots. Users login by ssh on their terminal using a username and password supplied by the system administrator.

4.3.1.4 Power

The electricity from the main electrical supply passes through a generator and Uninterruptible Power Supply (UPS) before passing through a secondary local UPS, to avoid disruption of analyses due to power failure.

4.3.2 Cluster configuration

4.3.2.1 SLURM

Simple Linux Utility for Resource Management (SLURM) was installed to manage the cluster resources. SLURM allocates exclusive and/or non-exclusive access to resources (computer nodes) to users for a defined duration of time by providing a framework for starting, executing, and monitoring jobs (normally a parallel job) on the set of allocated nodes and arbitrates contention for resources by managing a queue of pending jobs. This means users can all work simultaneously on the cluster. SLURM was first installed on the Master node, and the process was repeated on the additional nodes. The number of nodes can be increased in future depending on the laboratory computational requirements.

4.3.3 Bioinformatics cluster file system

Data and scientific software packages are shared between nodes in the cluster using Network file system (NFS). The /opt/conda2 directory (for scientific software) and /home directory (for data being analyzed) are exported from the master node and mounted on the same paths on the worker nodes. This allows scripts submitted to the SLURM scheduler to run on a consistent environment on all nodes on the cluster. We installed a program called Ansible to synchronize user information on all the nodes.

The cluster has two storage facilities, 11T for raw and a 50T for processed data (Figure 4.1A). The storage for raw data is accessible to the institutional sequencing core (where the Miseq instruments are housed) to upload data. Cluster users access the cluster from the home directory and are able to analyse the raw data already uploaded

in the storage using various bioinformatics tools installed in the opt directory. Cluster users have access to the green and blue areas (Figure 4.1B). The red area shows an example of system files only accessible to the cluster administrator.

The home directory where users login in is on the Master node, and by default all the jobs running write their output on the Master node disk space which only has 1T capacity. Once the diskspace became full the Master node would not be able to allocate other nodes to run jobs. To resolve this, we decongested the Master node by mounting a 11T volume to smb/cifs/share to increase the storage space on the share folder. Symbolic links (symlinks) were then created of all the user home folders to the share directory to prevent the master node from becoming congested (Figure 4.1B).

Bioinformatics cluster architecture

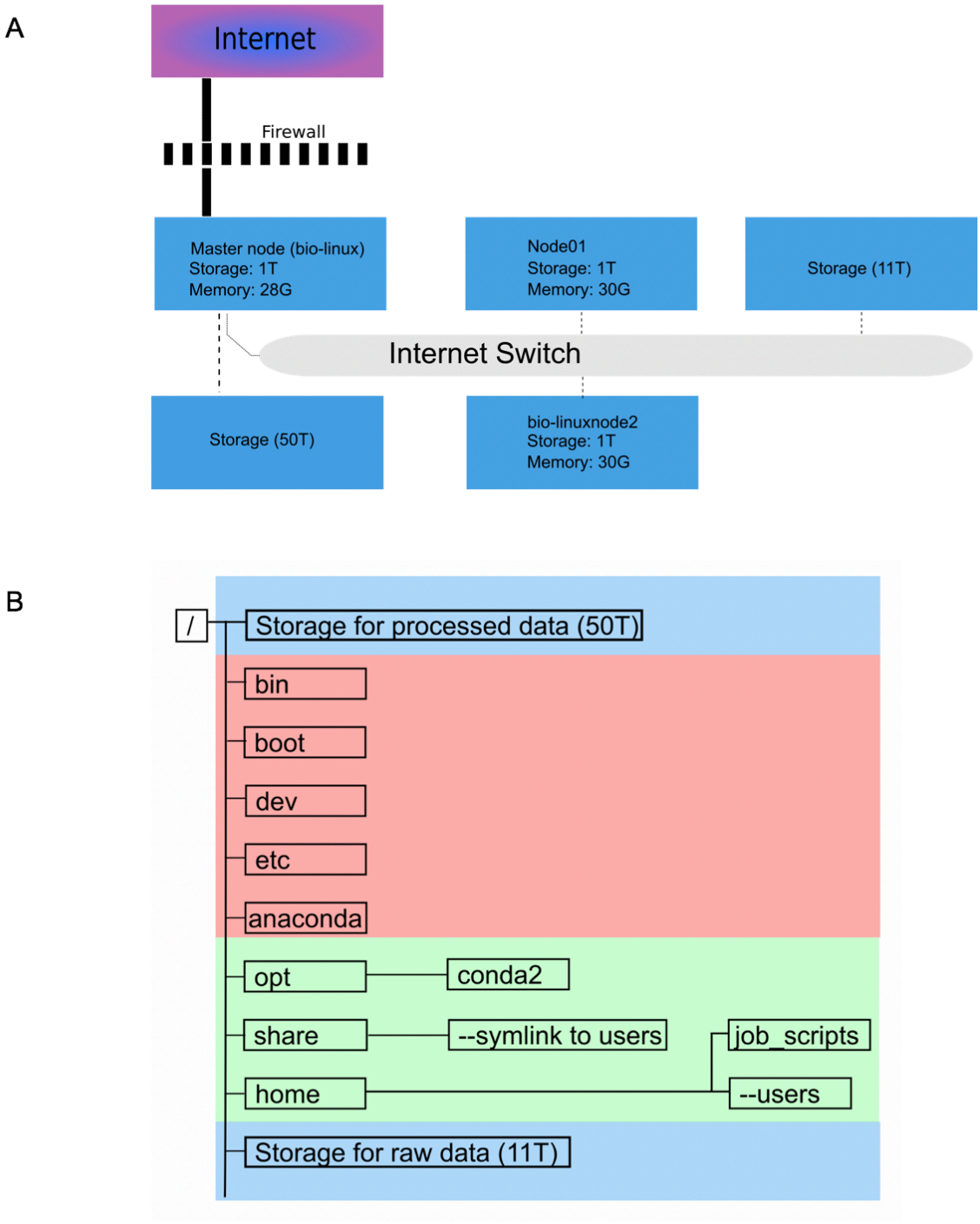


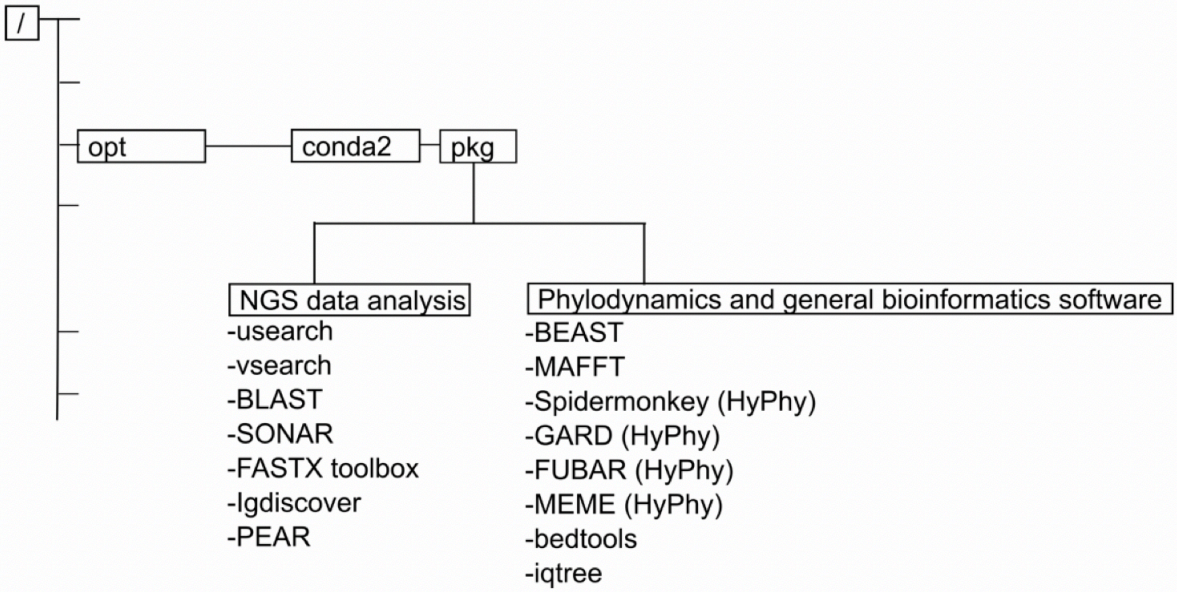
Figure 4.1: Bioinformatics cluster architecture. (A) Schematic description of the cluster architecture, storage and memory. (B) Bioinformatics cluster file system. The storage area is shown in blue, the user area for data analysis in green and the restricted system files in red.

4.3.4 Package manager and installation of bioinformatics packages

We installed several bioinformatics packages relevant to our studies of viral and antibody sequences. These are described below, however this could be replaced with tools relevant to local laboratory needs. We installed conda (https://conda.io/docs/intro.html) as our package manager. All the installed packages were located in /opt/conda2/ (Figure 4.1B and 4.2A). All programs were installed on the master node, and executable on all the computer nodes. Conda is a package manager application that quickly installs, runs, and updates packages and their dependencies. The conda command is the primary interface for managing installations of various packages. It can query and search the package index and current installation, create new environments, and install and update packages into existing conda environments. Creation of different environments is done for programs that might have conflicting requirements in terms of the dependencies used or versions of certain binaries (pre-built executables). The bioconda channel (http://bioconda.github.io/) was enabled within conda to provide access to its collection of over 4,000 open-source bioinformatics packages.

We used open-source tools and established pipelines developed by collaborators to perform bioinformatics analyses on viral and antibody sequence data. Bash scripts were used to execute the various bioinformatics programs allowing us to specify a given node and memory requirements to run the program, an example of which is shown in Figure 4.2B. Job scripts are stored in a folder accessible by all users, standardizing all parameters for performing analyses and saving time when performing similar analyses between donors.

A Bioinformatics programs on the cluster



B Example of a bash script to run programs on the cluster

```
#!/bin/sh

#SBATCH –w, --nodelist=node01
#SBATCH –mem=28000

vsearch -derep_fulllength *.fasta -output uniques.fasta -sizeout -relabel Uniq
```

Figure 4.2: Access to bioinformatics programs on the cluster. (A) Bioinformatics programs on the cluster. (B) Example of a bash script to run programs on the cluster.

4.4 Results

The establishment of this low-cost bioinformatics cluster has enabled us to perform analyses of large datasets that are impossible to process using ordinary machines. In this chapter, we summarise the outcomes of analyses which have been enabled by this cluster, several of which form the basis of ongoing (Scheepers et al, in preparation; Hlatswayo et al, unpublished data) and published studies (234, 350, 351).

4.4.1 Cluster performance with computationally intensive programs

Bioinformatics programs perform many analyses ranging from identifying signatures in sequences, to performing complex calculations, simulations and predictions. Some of these programs are very computationally intensive, requiring large memory and longer computational hours. An example of such computations performed in our laboratory are virus phylodynamics analyses that provide insights into intra-host evolution of HIV-1 by estimating the rates of evolution, selection, diversity, divergence, elucidating spatio-temporal distributions of viruses and identifying number of viral infections. These phylodynamic analyses are described in detail in chapters two and three of this thesis (234, 351).

To achieve this, we installed a number of phylodynamics programs on the cluster, including BEAST, which is computationally intensive. BEAST uses Bayesian statistical methods that require long computational times and memory requirements. BEAST ran faster on the bioinformatics cluster compared to an ordinary machine (164 and 284 hours respectively) to complete the same task (Table 4.1). Furthermore, the bioinformatics cluster analyzed 18 datasets at a time compared to one on the ordinary machine. This further highlights the benefits of the cluster in rapidly performing analyses with computationally intensive programs.

Table 4.1: Comparison of the cluster performance with that of a high specification ordinary machine. The cluster runs jobs faster compared to the ordinary machine, and also runs multiple jobs in parallel therefore reducing the total time to complete multiple analyses.

Machine	Av. speed & range (hours / mil steps)	Total time to complete 400 million steps (hours)	Number of analyses at a time
Cluster	0.41 (0.39 – 0.43)	164	18
Ordinary machine 2.8 GHz Core i7, 16 G mem	0.78 (0.25 – 1.01)	284	1

4.4.2 Cluster performance with large datasets

Antibody repertoire analyses rely on large datasets to accurately capture their evolution. We have thus far analyzed more than 3 T of NGS antibody repertoire data (Figure 4.3A), which far exceeds the capability of ordinary desktop computers. This data was obtained from four participants (CAP88, CAP228, CAP206 and CAP248) at multiple time-points ranging between 7-281 weeks post-infection (Figure 4.3B). Each analysis (2–10 million light and heavy chain antibody reads) took 8–168 computational hours, and the memory usage varied between 4-20G. The cluster enables users to run several jobs simultaneously, thus allowing multiple time-points to be analyzed concurrently.

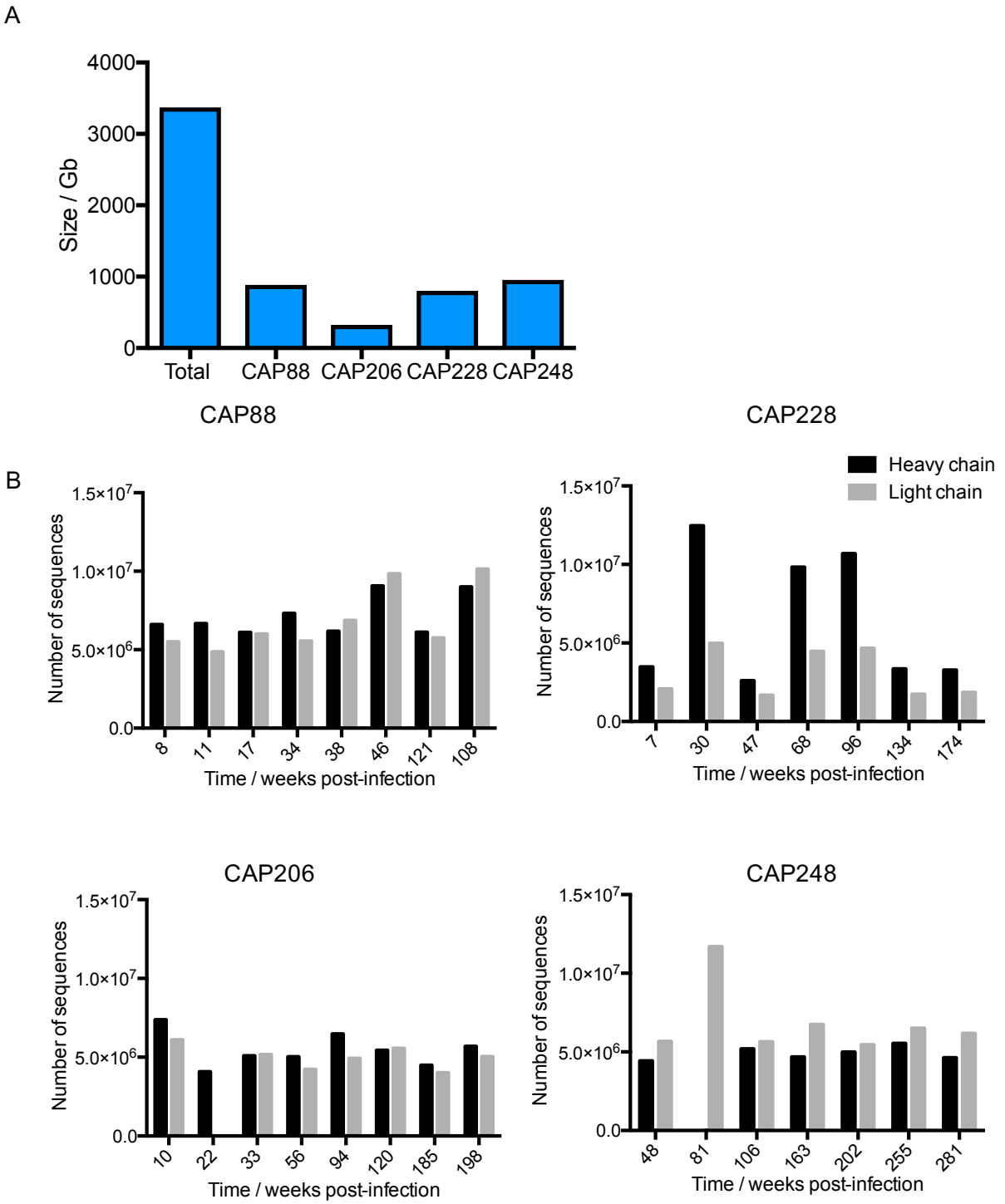


Figure 4.3: Cluster performance with large datasets of antibody repertoire data. (A) The total amount of data from the SONAR analyses, followed by the breakdown per participant for donors CAP88, CAP206, AP228 and CAP248 (still ongoing). (B) Number of antibody sequences analyzed using the bioinformatics cluster per donor and per time point. Antibody heavy and light chain antibody sequences data are shown in black and grey respectively.

4.4.3 Easy integration of bioinformatics programs on the cluster

Bioinformatics analyses of complex data often involves using different tools to perform specialized tasks as part of a pipeline or a workflow. An example is the antibody repertoire NGS data analysis that we use, which requires many steps, summarized in Figure 4.4, and for which all of the steps are performed on the cluster. The analysis steps involve the use of the SONAR pipeline which was developed by our collaborators at the University of Columbia (352). SONAR links several bioinformatics tools, including Clustal-Omega (265, 312), MUSCLE (265) and Basic Local Alignment Search Tool (BLAST+) (353), BEAST (272) and DNAML (354), to process the data and identify sequences related to an antibody sequence of interest (clonally related sequences) (Figure 4.4) (352).

The pipeline was installed as follows: dependencies were first installed using conda or as per the developer' instructions if not packaged within conda. SONAR was downloaded from https://github.com/scharch/SONAR and placed in /opt/conda2/pkgs/. After uncompressing the files, we installed the pipeline by following the command prompts after executing “setup.sh”. The command prompts allowed us to specify the paths to the installed dependencies. We then defined the python path using the command: “export PYTHONPATH=$PYTHONPATH: /opt/conda2/pkgs/sonar/”. Finally for all the cluster users to be able to use the bioinformatics programs, we edited the *.bashrc* scripts to include the path to all the programs.

The cluster therefore allows easy integration and automation of these bioinformatics programs, enhancing the laboratory’s throughput, with a specific example of these analyses described in more detail below.

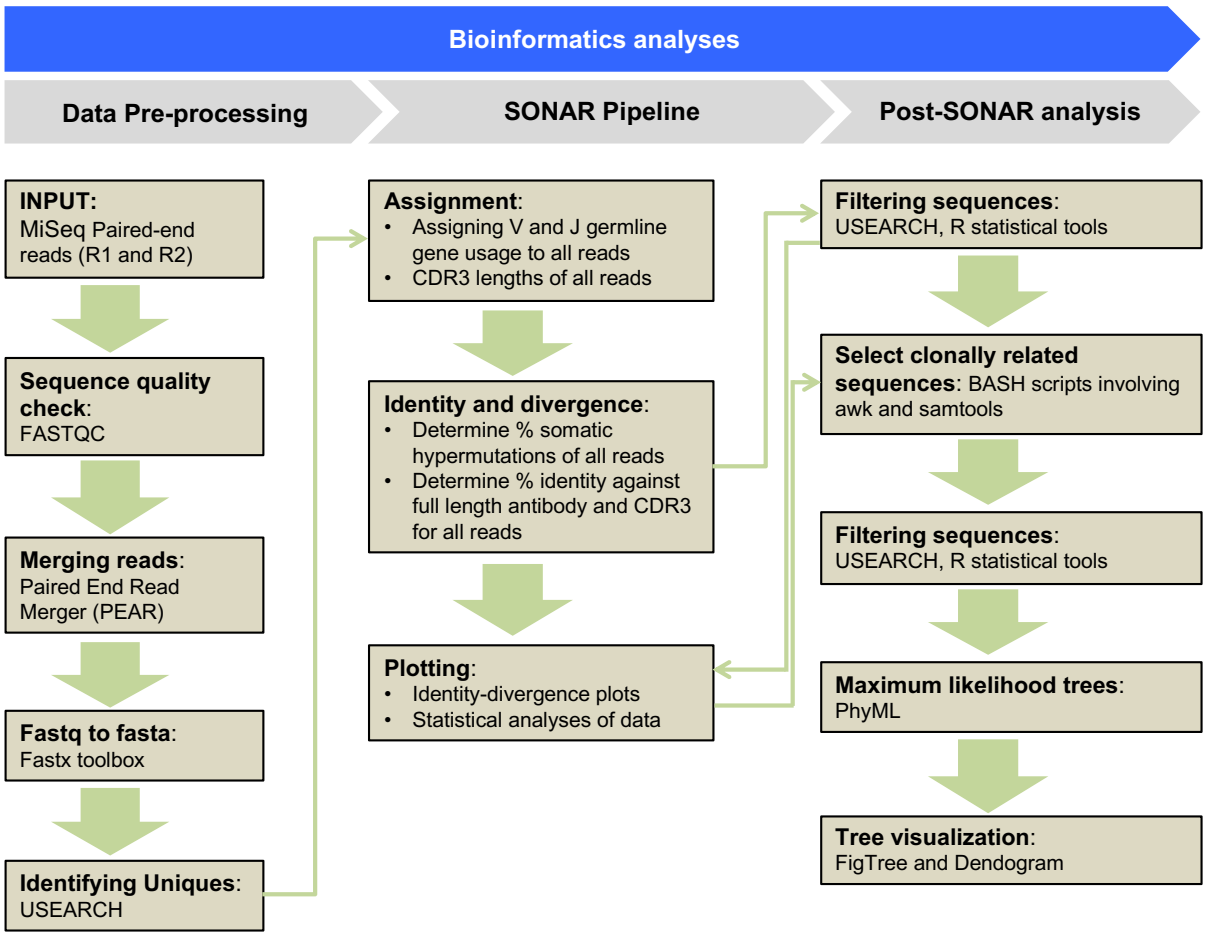


Figure 4.4: NGS analysis flowchart using the bioinformatics programs on the cluster. The antibody repertoire analysis involves first organising the data from the sequencing facility (data pre-processing), SONAR analysis and post SONAR analysis that involves selection of clonally related sequences. The arrows show the flow of data through the various stages, and in some instances data can go back and forth certain stages (shown by the thin arrows).

4.4.4 Case study: Antibody analysis of the CAP228-16H/19F antibody lineage using SONAR

Here we describe the application of the bioinformatics cluster to trace the antibody lineage which includes CAP228-16H and 19F (which are clonally related antibodies) from CAPRISA participant CAP228 (350). This study built on the RV144 phase III HIV human clinal trial in Thailand that achieved 30% efficacy in vaccinated individuals (137, 179). This the only protection ever to be achieved in HIV vaccination (137, 179). Protection was attributed to antibody responses directed against the V1V2 region, that showed antibody-dependent cell-mediated cytotoxicity (ADCC) activity. Two antibodies against the V1V2 were isolated from vaccinees and named CH58 and

CH59. The CH58 mAb used the IGHV5-51 heavy chain and IGLV6-57 light chain, and CH59 mAb used the IGHV3-9 and IGLV3-10 heavy and light chains respectively. Structural studies of these mAbs (CH58 and CH59) showed that binding was achieved by interactions with positively charged residues at positions 168, 169 and 171 in the V2 region of the envelope glycoprotein (188, 355). This binding was facilitated by a negatively charged glutamic-aspartic amino acid pair (ED motif) in the CDRL2 of the mAbs (350). However, this ED motif is only encoded by only two light chain germline genes (IGLV3-10 and IGLV6-57), which reduces the likelihood of these responses being elicited through vaccination. Our laboratory was therefore interested in exploring whether there are other modes of V2 binding by similar vaccine elicited antibodies that might have a greater probability of being present in germline genes.

To do this, single-cell sorting of antigen-specific B cells was performed using biotinylated cyclic V1V2 peptides comprising of gp120 residues 157 – 196 (350). The resulting mAbs, CAP228-16H/19F bound to the V1V2 epitope. Similarly to CH58, the CAP228 mAbs used the IGHV5-51 heavy chain gene, but an IGLV3-21 light chain gene which was different from that of CH58 (350). Comparison of the CAP228-16H/19F and CH58 mAbs showed striking similarities in that they both depend on the K169 residue, and both were able to mediate antibody-dependent cell-mediated cytotoxicity (ADCC) activity. However, unlike CH58 that used the ED motif, the CAP228-16H/19F mAbs used a DDxD motif which consisted of three aspartic acid residues creating a large anionic patch, and covering a larger surface area compared to CH58 (350). The CAP228-16H/19F mAbs exhibited broader ADCC activity through the interaction of the DDxD motif with a tyrosine residue at viral Env position 173, which is present in approximately ~65% of HIV-1 circulating strains(350).

We therefore traced the development of the CAP228-16H/19F lineage in infection, performing these analyses on the bioinformatics cluster using a number of programs and the SONAR pipeline described in the previous section. Firstly, we designed gene specific primers to the CAP228-16H/19F lineage that can be used to further identify lineage members including the unmutated common ancestor (UCA) that initiated the antibody lineage. Next generation sequencing (Illumina MiSeq NGS System) was then applied to generate related sequences at multiple timepoints using the gene specific primers. Data analysis was then performed in SONAR as described below.

4.4.4.1 Data Pre-processing

Prior to antibody lineage analysis, the data generated by MiSeq has to be pre-processed into the correct format as well as ensuring adequate quality (Figure 4.4). The pre-processing stage involves checking the data quality using FASTQC (356), merging reads using PEAR (357) and identifying unique sequences using USEARCH (358).

There are a number of quality controls during merging with PEAR that result in the reduction of the number of sequences. These quality controls include specifying a p-value of 0.001 (PHRED quality score of 33), a minimum overlap size of 10 and a minimum read length of 50 bases. The quality controls reduce false positive assembles and very short reads (357). After assembling the sequences, identical sequences are collapsed using USEARCH generating only unique sequences and their abundance information for further analyses. Collapsing the sequences helps to reduce the computational load in analyzing identical data. The input files for PEAR are in a fastq format whereas the input files for USEARCH and all the subsequent analyses are in a fasta format, we therefore convert the fastq files to fasta after merging using fastx toolbox (359).

4.4.4.2 Antibody Analysis using SONAR

After preparing the data for SONAR analysis, there are three bash scripts that have been prepared to run the different stages of the pipeline. The first bash script runs the first three annotation stages of SONAR namely BLAST V, BLAST J and finalizing appointments (1.1-blast_V.py, 1.2-blast_J.py and 1.3-finalize_assignments.py). The annotation stages run BLAST against a custom database of immunoglobulin genes (including those from IMGT) to characterize the *V(D)J* germline gene usage and the CDRs of the all the NGS transcripts and filter transcripts that do not match the length requirements (352). The last stage (finalizing assignments) summarizes all the information about the NGS transcripts characteristics and outputs it in a table (352).

The second bash script evokes the calculate identity divergence SONAR stage (2.1-calculate_id-div.pl) that first aligns the NGS reads from the previous stages to its assigned germline sequence and the known antibody sequence of interest. It then

calculates the percentage identity of each transcript to the antibody sequence of interest, and also the percentage divergence from the germline V gene (352). This data is also written an output file and will be used as input in the plotting stage.

The plotting stages are executed using the third bash script, and these stages (4.1-setup_plots.pl, 4.2-plot_histograms.R and 4.3-plot_identity_divergence.R) use information from the output tables to provide a graphical summary of the characteristics of the NGS transcripts (352). This includes the assigned *V* and *J* germline genes that all the NGS transcripts use, and the CDR lengths determined in the first stages of the analyses and identity divergence plots described in more detail in the data visualization section below.

4.4.4.3 Data visualization

Seven different time-points were sequenced covering 7-174 weeks post-infection with 1.6-6.2 million heavy chain sequence pairs and 841 533-2.3 million light chain sequence pairs generated for each time-point (Table 4.2) (350). Between 84-99% of these sequences were merged and further analyzed. The rest did not meet the criteria that included the minimum quality score, overlap size and read length as described above. The fastq data was converted to fasta and subsequently analysed in USEARCH. Approximately 88-97% of these data had a lot of sequences that only appeared once (singletons) which could either be a result of PCR or sequencing error or rare antibody sequences that were not sequenced at sufficient depth. More than 75% of all the light chain NGS reads used the IGLV3-21 gene of interest, and the IGLJ3 (Figure 4.5A and B). In addition, almost 30% of all the heavy chain NGS reads were successfully assigned used the IGHV5-51 germline gene, and over 10% used IGHJ3 (Figure 4.6A and B).

Table 4.2: Table showing the number of sequences and processes for the CAP228 longitudinal dataset before running SONAR. (A) Heavy chain data. (B) Light chain data (van Eeden et al. 2018).

A Heavy chains

	Sequencing		Merging				fastq to fasta	USEARCH		
Timepoint (wpi)	Total forward sequences	Total reverse sequences	Total Pairs	Sequences merged	Sequences not merged	% merged	Sequences converted	Uniques	Singletons	% Singletons
7	1,739,277	1,739,277	1,739,277	1,672,615	66,659	96.167	1,672,615	1,423,081	1,390,574	97.7
30	6,233,970	6,233,970	6,233,970	5,283,515	950,455	84.754	4,810,851	3,352,602	3,191,845	95.2
47	1,305,447	1,305,447	1,305,447	1,288,901	16,541	98.733	1,288,901	813,875	721,535	88.7
68	4,918,001	4,918,001	4,918,001	4,289,190	628,806	87.214	3,926,551	2,803,736	2,667,110	95.1
96	5,346,292	5,346,292	5,346,292	4,599,532	746,760	86.032	4,186,679	3,007,440	2,860,292	95.1
134	1,679,602	1,679,602	1,679,602	1,658,592	21,004	98.749	1,658,592	1,085,513	970,434	89.4
174	1,637,889	1,637,889	1,637,889	1,620,885	16,999	98.962	1,620,885	1,133,973	1,009,049	89.0

B Light chains

	Sequencing		Merging				fastq to fasta	USEARCH		
Timepoint (wpi)	Total forward sequences	Total reverse sequences	Total Pairs	Sequences merged	Sequences not merged	% merged	Sequences converted	Uniques	Singletons	% Singletons
7	1,040,776	1,040,776	1,031,646	1,033,740	7,025	99.324	1,031,646	542,278	488,593	90.1
30	2,492,586	2,492,586	2,492,586	2,458,050	34,536	98.614	2,298,777	1,354,022	1,250,465	92.4
47	841,533	841,533	841,533	833,304	8,229	99.022	831,653	447,517	412,214	92.1
68	2,237,276	2,237,276	2,237,276	2,207,671	29,604	98.677	2,058,079	1,281,287	1,193,995	93.2
96	2,334,525	2,334,525	2,334,525	2,305,316	29,209	98.749	2,154,627	1,230,325	1,141,038	92.7
134	868,960	868,960	868,960	863,117	5,843	99.328	861,414	489,319	450,900	92.1
174	927,751	927,751	927,751	918,542	9,208	99.007	916711	543,631	502,437	92.4

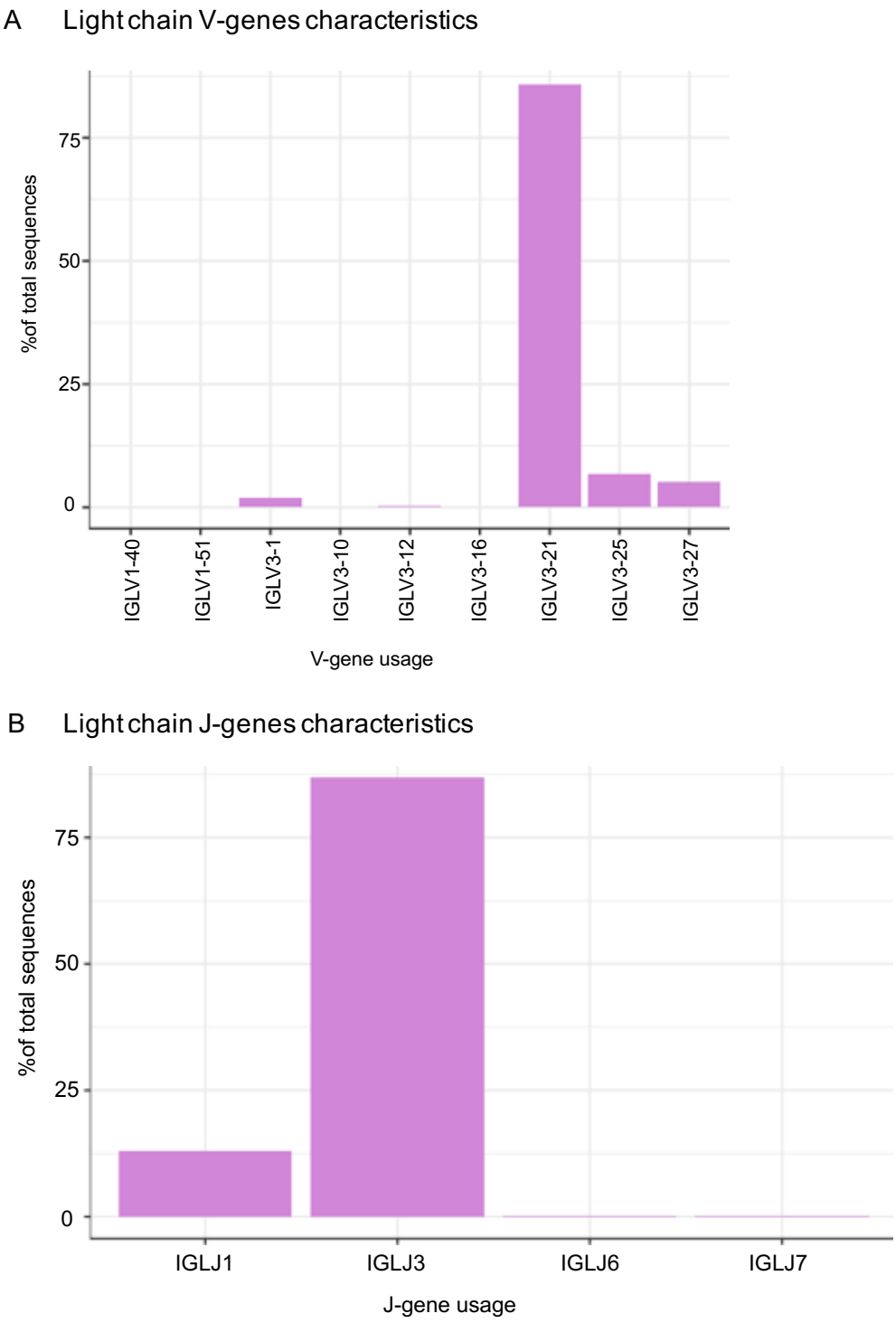


Figure 4.5: Visualization of the CAP228 NGS light chain sequence data, using the CAP228 30 weeks dataset as an example. The pipeline assigns V and J gene usage, and the CDR lengths. (A) Histogram showing the IGLV gene characteristics. (B) Histogram showing the IGLJ gene characteristics.

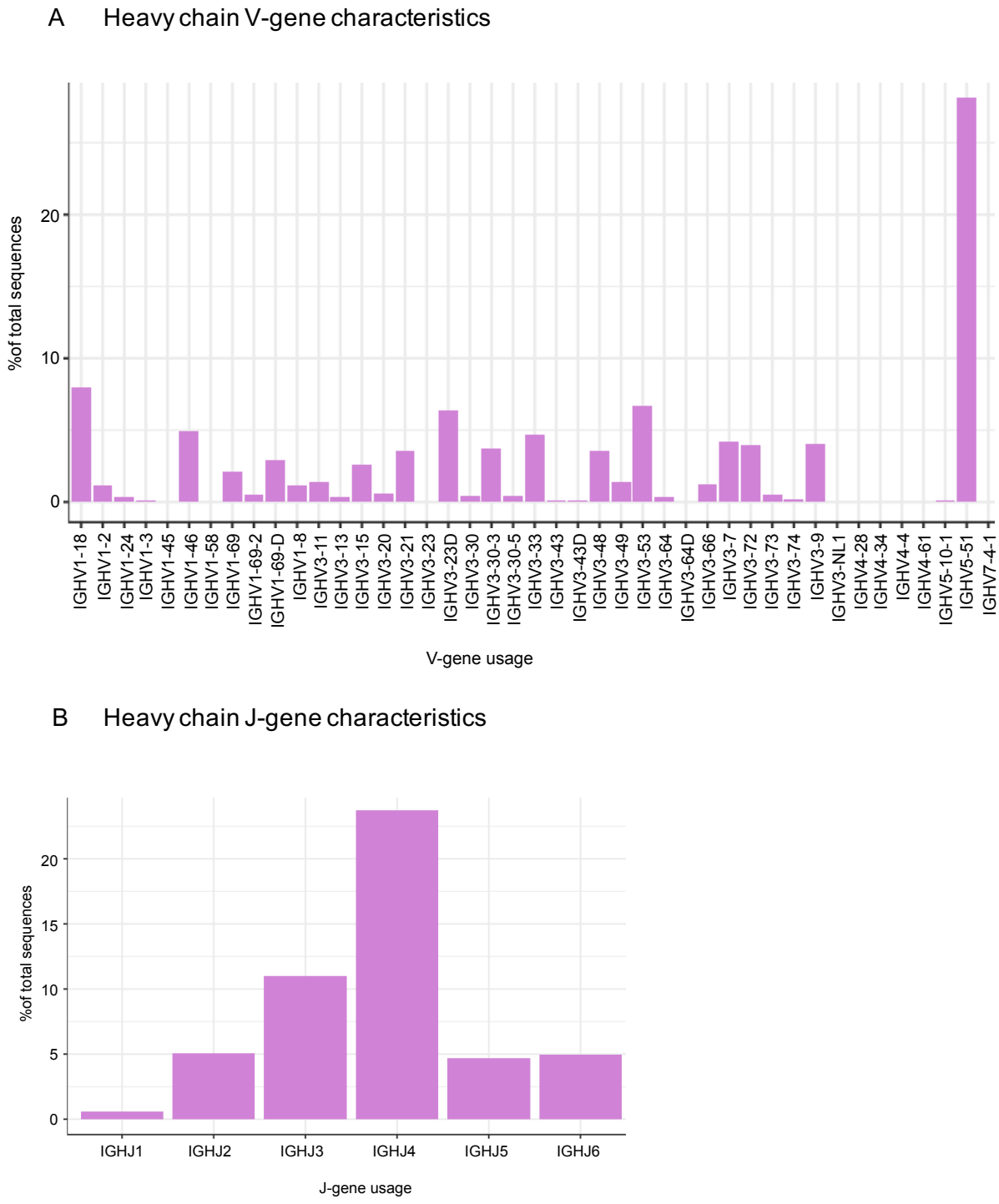


Figure 4.6: Visualization of the CAP228 NGS heavy chain sequence data, using the CAP228 30 weeks dataset as an example. The pipeline assigns V and J gene usage, and the CDR lengths. (A) Histogram showing the IGHV gene characteristics. (B) Histogram showing the IGHJ gene characteristics.

During SONAR analysis of the CAP228-16H/19F lineage, identity divergence plots were created to show the percentage identities of the sequences to the isolated mAbs and the percentage divergence of the sequences from the germline IGHV5-51 gene for the heavy chains and IGLV3-21 for the light chains (Figure 4.7A) (350). The

percentage divergence from the germline V-gene represents the somatic hypermutation (SHM) of the sequences. These plots highlight the presence of clonally related sequences to CAP228-16H throughout all time-points, by the presence of an “island” above 90% identity to CAP228-16H. These clonally related sequences were used to create phylogenetic trees of the lineage to trace their evolution and identify the UCA at 7 weeks (Figure 4.7A and B) (350). The UCA was closely related to IGLV3-21*01, however the latter encodes a YD motif rather than the DD motif that mediates binding of the CAP228 mAbs to their epitope. However, sequencing of the germline genes from CAP228 genomic DNA identified a novel IGLV3-21 allele (IGLV3-21*01mm) which encoded the CDRL2 DDxD motif (highlighted in blue in Figure 4.7C), suggesting that the DD encoding UCA was derived from this allele (350). Further, an analysis of the allelic frequencies estimated the prevalence of the DDxD-encoding IGLV3-21 alleles in the general population to be 60%, compared to the ED-encoding alleles that are less common. These data suggest there is a greater probability of eliciting these responses compared to those that are dependent on the ED motif in future vaccine studies.

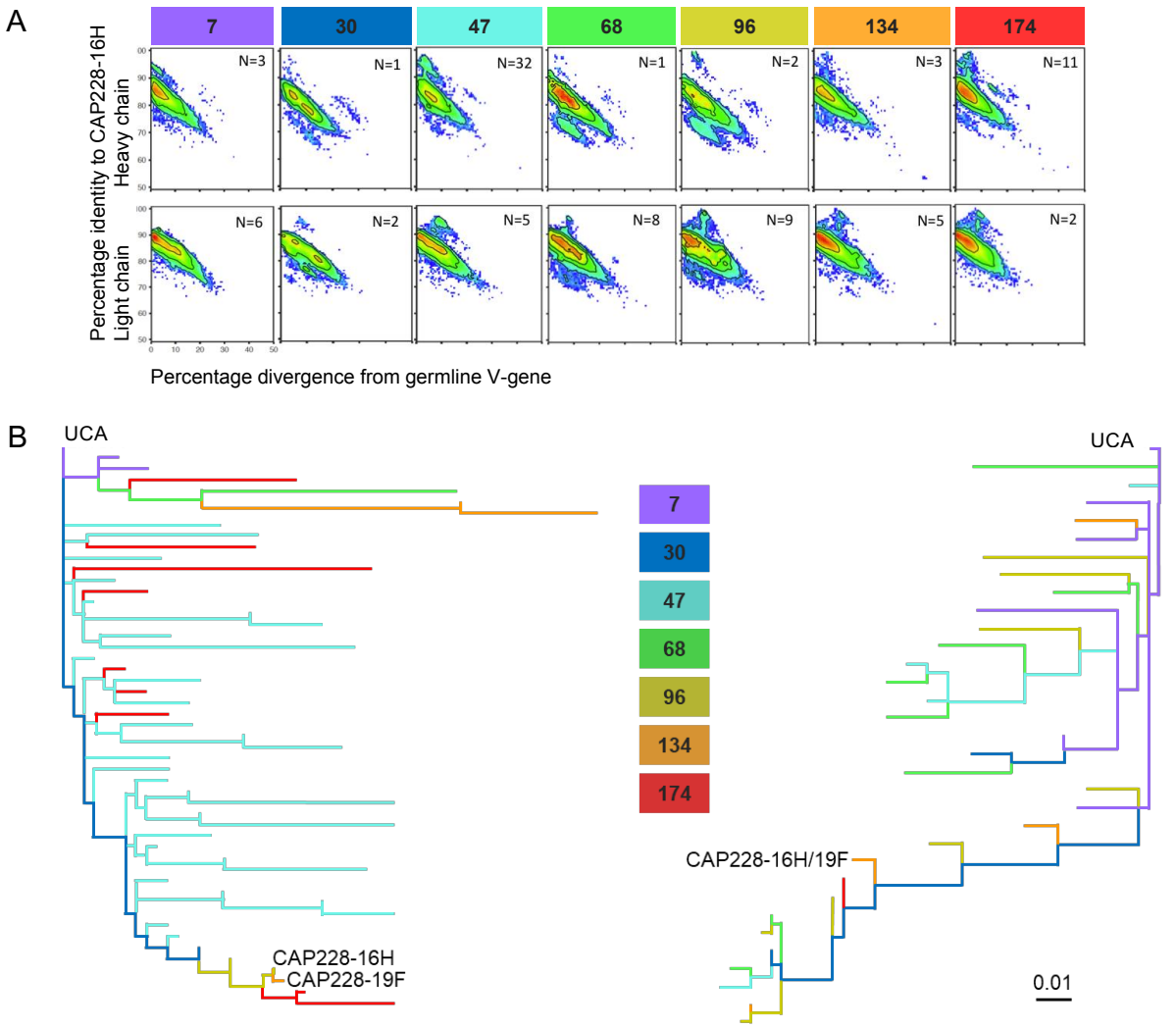


Figure 4.7: Identifying clonally related sequences following SONAR analysis. Analysis of the CAP228-16H/19F lineage has been used as an example. (A) Identity-divergence plots for heavy and light chains. (B) Maximum likelihood trees of the heavy and light clonally related sequences. The colours represent the different time-points the sequences were isolated. (C) Sequence alignments showing that CAP228-16H/19F lineage was initiated by a novel germline (IGLV3-21*01mm) that encodes the DDxD motif (van Eeden et al. 2018).

4.5 Discussion

Lack of infrastructure and limited resources are bottlenecks in research conducted in low to medium income settings. These regions are often also heavily burdened by infectious diseases such as HIV, malaria and tuberculosis, some of which are not priority research areas in well-resourced countries. Conducting disease surveillance and vaccine research to curb the spread of locally relevant pathogens is a public health priority. Our research laboratory is actively involved in HIV vaccine research, and this work involves generating and analysing huge amounts of sequencing data. We have successfully set-up a bioinformatics cluster using cheap and low specification CPUs and demonstrated its application in analyzing these large NGS datasets.

The cost of constructing this cluster was reduced by using old computers that were no longer used in the laboratory and entirely open-source software. Only the computer memory and storage space were upgraded at a small cost to accommodate more data. This approach is very feasible in resource limited laboratories that have access to old computers, or funding to purchase relatively inexpensive ones. We used open-source software to configure the cluster. There is a significant body of useful open-source software that also comes with good technical support. Most of this support comes from the community of users that interact on Google groups such as Gitter, GitHub, Biostars, Stack Overflow and other online platforms. We used Linux Fedora operating system, as it is open-source, stable and is sponsored and funded by Red Hat. Conda package manager is also open-source software that makes it easy and manageable to install bioinformatics packages. The SLURM scheduler is also open-source software with excellent community support.

There are a number of software packages that may be used for cluster configuration management such as Ansible, Chef and Puppet. Puppet has advantages when it comes to updating programs on the cluster, it automatically updates the other nodes, whereas Ansible has to be run to update any changes on the other nodes. We used Ansible because it is simple and easy to configure compared to the other two as our cluster is small, with less than 3 nodes.

A centralized cluster for running tasks makes it easy to standardize processes and install updates on programs run by users. In addition, it enables good management of

memory and storage resources as well as data security as users don't have to carry huge quantities of data around. Access to the cluster only requires use of the command line and an understanding of Unix systems. If a Windows operating system if preferred, installation of a program such as Putty provides a command line to execute Unix commands.

We applied the cluster to the analysis of the HIV envelope glycoproteins evolution within individuals (chapters 2 and 3), and to studying the development antibody responses (described in the previous section). Using this cluster, we have traced the antibody lineages of vaccine-like antibodies that developed during infection in CAP228 and identified a novel IGLV3-21 germline (IGLV3-21*01mm) that encodes the DDxD motif (350). The frequency of DDxD-motif alleles is more common compared to those that encode the previously characterized ED-motif. The isolated CAP228-16H/19F antibodies also showed ADCC activity which has been linked to protection in the RV144 trial (350, 360, 361). This case study demonstrated how usage of this cluster has helped overcome hurdles in data processing and analysis, and generated valuable insights into vaccine design.

We have demonstrated the building and implementation of a low-cost cluster for analyzing large NGS data and performing computationally intensive studies of intra-host evolution of HIV. The establishment of this low-cost cluster demonstrates another alternative on how researchers from low-income settings can solve global challenges using cheap resources. Such an approach of using low-cost technologies, and recycling/repurposing equipment to tackle complex scientific problems is highly relevant to Africa and has broader implications in advancing creativity, research and bringing about home-grown solutions to the challenges facing the continent.

5 General Discussion and Conclusions

5.1 General Discussion

Millions of people are still being infected with HIV decades since the first recognition of this virus (297). Understanding HIV diversity within individuals from acute to chronic infection, and its influence on the sensitivity and maturation of neutralizing antibody responses is very important for the development of vaccines and microbicides. However, these studies are sometimes complicated because of the large amount of data and the sophisticated analysis software that require specialized computational infrastructure. The purpose of this study was to investigate HIV envelope diversity, both among hosts with differing immune responses, and within hosts, between different tissues, using low-cost computational infrastructure.

Currently, no vaccine is able to elicit bNAbs that will prevent infection by global HIV strains. Several studies have implicated HIV Env diversity in the development of breadth (167, 168, 247, 248). However, Env evolution in individuals who fail to develop breadth despite mounting potent, strain-specific neutralizing responses has not been well-defined. Using longitudinal neutralization, epitope mapping and sequence data from 14 participants, we found that overall measures of viral diversity were similar in all donors (Chapter 2). This is unlike previous studies (167, 168, 362), probably because we specifically compared individuals with matched viral loads. However, the number of positively selected sites within Env epitopes was higher in bNAb participants than strain-specific donors. We further identified common sites that were positively selected as bNAbs developed. These data indicate that while viral diversity is required for breadth, this should be highly targeted to specific residues to shape the elicitation of bNAbs by vaccination.

Several current vaccine approaches aim to represent the Env diversity associated with bNAb development in regimens using immunogen cocktails, and through sequential boosts (139, 142, 363, 364). The identification of toggling sites in the Env that are unique to bNAb donors may thus have very important implications for vaccine design. These findings indicate that instead of using multiple Envs in a regimen, fewer immunogens will be required, but these will need to incorporate diversity at key sites

to engage the relevant antibody lineage(s), and drive them towards breadth. This approach will likely be more feasible compared to the current strategies as it will require fewer immunogens. Secondly, we identified key sites that are specific to bNAbs targeting two different sites of the Env (V1V2 and C3V4). This makes it possible to develop a vaccine capable of eliciting multiple bNAb responses by incorporating the "key diversity sites" for two or more epitopes into a single trimeric immunogen (139, 142, 365, 366). Current trimers have been able to elicit neutralizing antibodies, but thus far not bNAbs (366). This may be because such immunogens do not incorporate diversity targeted to specific residues, some of which have been identified in this study (139, 142, 365, 366). Overall, this study therefore leveraged a unique set of longitudinal sequence and antibody mapping data to inform the development of immunogens that seek to elicit bNAbs by vaccination.

Most HIV-1 infections in sub-Saharan Africa are acquired heterosexually through the genital mucosa (367). Understanding the properties of viruses replicating in the female genital tract, and whether these properties differ from those of more commonly studied viruses sampled from the blood, is therefore important. Using longitudinal CVL and plasma-derived sequences from four HIV-infected women we found fewer viral migrations from the genital tract to plasma than in the opposite direction, suggesting a mucosal sieve effect from the genital tract to the blood compartment (Chapter 3). Evidence for both persistent and intermittent compartmentation was detected in two of four individuals between the genital tract and plasma viruses during chronic infection. These data may explain previously conflicting findings regarding compartmentalization in the female genital tract (95, 96, 103, 105, 106, 304, 306, 309). Our study demonstrates that compartmentalization can either occur variably, probably depending on the characteristics of the T/F viruses, the strength of the mucosal sieve, and varying viral diversification pathways. In cases where compartmentalization occurred, comparison of CVL and plasma sequences indicated that distinct features of viral populations in the CVL may impact the efficacy of microbicides and vaccines designed to provide mucosal immunity. Phenotypic studies comparing the sensitivities of viruses from these compartments will provide further insight into these data. These findings show that it is not sufficient to sample HIV glycoproteins only from the plasma, as these do not necessarily represent the overall diversity in other compartments such as the genital tract. It is therefore important to also sample HIV from the mucosal tissues in

vaccine trials and epidemiological studies, so as to understand their diversity and properties, which in turn leads to development of better vaccines.

Phylodynamic analyses are complex in nature, often involving large amounts of data, and sophisticated analytical software. This can limit these kinds of studies in resource-limited settings such as Africa where the HIV epidemic, and the burden of many other pathogens, is greatest. We recycled old computational equipment, made improvements to memory and storage, and installed high-performance computing software (Chapter 4). This resulted in a low-cost supercomputer that was able to perform larger tasks with greater speed, as well as handling multiple jobs at the same time. In addition to enabling the studies of viral evolution detailed above, we also demonstrated the functionality of the cluster using a case study where we identified novel alleles in an HIV infected participant CAP228 that encoded antibodies against HIV-1 V1V2 peptides (350). Antibodies to the V1V2 site are important because the RV144 trial demonstrated that they can reduce risk of HIV infection (137, 179). Antibodies from CAP228 resemble these vaccine-elicited antibodies, but used a different light chain that encoded a more common K169-binding motif than those previously isolated. Importantly, this allele was discovered in an African population, where current vaccination trials are occurring, and using a similar immunogen as the RV144 trial (363). This means that there is a larger repertoire of B cells that may potentially be engaged by this vaccine to produce protective responses against HIV. In addition, this study provides more insights into African genes, a field that is yet to be explored in sufficient depth (368). Only a few studies have looked into African population genetics, although it is the most diverse (369). Establishment of this cluster creates more capacity to perform such studies, and provides an alternative strategy for other African researchers limited by computational power to conduct similar studies. This part of the thesis hopes to inspire communities that are resource-limited to tackle some of the local research challenges through application of simple and inexpensive approaches.

5.2 General Conclusion

Overall, these studies have contributed to understanding HIV Env diversity in different tissues within a host, as well as the influence of Env diversity on the development of varying antibody responses between different hosts. The study highlights the importance of diversity at specific amino acid residues, but not overall Env diversity, to potentially elicit bNAbs through vaccination. The study also confirms that HIV evolves variably within hosts, and in some donors results in compartmentalization of viruses in the female genital tract, which may impact sensitivity to vaccines. Further, these studies have demonstrated novel approaches to managing computational challenges faced by researchers in less resourced settings, to advance scientific research in low-income communities.

Supplementary Table 4.1: Estimated costs for establishing the bioinformatics cluster.

Item	Cost Rands)
3 * desktops	recycled from the lab (icore 7 workstations)
1 * computer screens	recycled from the lab
software	free (open-source)
Internet cables	recycled from the lab
Memory upgrades	8400
Uninterrupted Power Supply (UPS)	7000
Internet switch	3500
Cabinet with cooling fans	10000
Storage updates (50T)	72000

6 References

1. Sepkowitz KA. 2002. Opportunistic infections in patients with and patients without Acquired Immunodeficiency Syndrome. Clin Infect Dis 34:1098–1107.
2. Gallo RC, Montagnier L. 2003. The discovery of HIV as the cause of AIDS. N Engl J Med 349:2283–2285.
3. Sharp PM, Hahn BH. 2011. Origins of HIV and the AIDS pandemic. Cold Spring Harb Perspect Med 1:a006841.
4. Smith PD. 1988. Intestinal Infections in Patients with the Acquired Immunodeficiency Syndrome (AIDS). Ann Intern Med 108:328.
5. Horsburgh CR, Selik RM. 1989. The Epidemiology of Disseminated Nontuberculous Mycobacterial Infection in the Acquired Immunodeficiency Syndrome (AIDS). Am Rev Respir Dis 139:4–7.
6. Ho DD. 1989. The Acquired Immunodeficiency Syndrome (AIDS) Dementia Complex. Ann Intern Med 111:400.
7. Trillo-Pazos G, Diamanturos A, Rislove L, Menza T, Chao W, Belem P, Sadiq S, Morgello S, Sharer L, Volsky DJ. 2006. Detection of HIV-1 DNA in Microglia/ Macrophages, Astrocytes and Neurons Isolated from Brain Tissue with HIV-1 Encephalitis by Laser Capture Microdissection. Brain Pathol 13:144–154.
8. Zhu T, Korber BT, Nahmias AJ, Hooper E, Sharp PM, Ho DD. 1998. An African HIV-1 sequence from 1959 and implications for the origin of the epidemic. Nature 391:594–597.
9. Avert. 2018. History of HIV and AIDS overview. HIV Timeline.
10. Avert. 2017. Global HIV and AIDS statistics.
11. UNAIDS. 2015. AIDS by the numbers. AIDS by Numbers 2015.
12. UNAIDS. 2017. South Africa | UNAIDS.
13. Avert. 2018. HIV and AIDS in South Africa.
14. Parkhurst JO. 2010. Understanding the correlations between wealth, poverty and human immunodeficiency virus infection in African countries. Bull World Health Organ 88:519–526.
15. Shelton JD, Cassell MM, Adetunji J. 2005. Is poverty or wealth at the root of HIV? Lancet 366:1057–1058.
16. Quinn TC. 2005. HIV/AIDS in Women: An Expanding Epidemic. Science (80-) 308:1582–1583.
17. Hargreaves JR, Morison LA, Chege J, Rutenburg N, Kahindo M, Weiss HA,

Hayes R, Buve A. 2002. Socioeconomic status and risk of HIV infection in an urban population in Kenya. Trop Med Int Heal 7:793–802.

18. Hallman K. 2005. Gendered socioeconomic conditions and HIV risk behaviours among young people in South Africa. African J AIDS Res 4:37–50.
19. Barin F, Denis F, Allan JS, M'Boup S, Kanki P, Lee TH, Essex M. 1985. Serological evidence for virus related to simian T-lymphotropic retrovirus III in residence of West Africa. Lancet 326:1387–1389.
20. Clavel F, Guetard D, Brun-Vezinet F, Chamaret S, Rey M, Santos-Ferreira M, Laurent A, Dauguet C, Katlama C, Rouzioux C, Al. E. 1986. Isolation of a new human retrovirus from West African patients with AIDS. Science (80-) 233:343–346.
21. Marlink R, Kanki P, Thior I, Travers K, Eisen G, Siby T, Traore I, Hsieh C, Dia M, Gueye E, Et A. 1994. Reduced rate of disease development after HIV-2 infection as compared to HIV-1. Science (80-) 265:1587–1590.
22. Hemelaar J. 2013. Implications of HIV diversity for the HIV-1 pandemic. J Infect 66:391–400.
23. Gao F, Bailes E, Robertson DL, Chen Y, Rodenburg CM, Michael SF, Cummins LB, Arthur LO, Peeters M, Shaw GM, Sharp PM, Hahn BH. 1999. Origin of HIV-1 in the chimpanzee Pan troglodytes troglodytes. Nature 397:436–441.
24. Keele BF. 2006. Chimpanzee reservoirs of pandemic and nonpandemic HIV-1. Science (80-) 313:523–526.
25. Faria NR, Rambaut A, Suchard MA, Baele G, Bedford T, Ward MJ, Tatem AJ, Sousa JD, Arinaminpathy N, Pepin J, Posada D, Peeters M, Pybus OG, Lemey P. 2014. The early spread and epidemic ignition of HIV-1 in human populations. Science (80-) 346:56–61.
26. Hahn BH. 2000. AIDS as a zoonosis: Scientific and public health implications. Science (80-) 287:607–614.
27. Peeters M, D'Arc M, Delaporte E. 2014. Origin and diversity of human retroviruses. AIDS Rev 16:23–34.
28. Tebit D, Ndembi N, Weinberg A, E. Quinones-Mateu M. 2012. Mucosal Transmission of Human Immunodeficiency Virus. Curr HIV Res 10:3–8.
29. Hladik F, Hope TJ. 2009. HIV infection of the genital mucosa in women. Curr HIV/AIDS Rep 6:20–28.
30. Bomsel M, David V. 2002. Mucosal gatekeepers: Selecting HIV viruses for

early infection. Nat Med 8:114–116.
31. Lederman MM, Offord RE, Hartley O. 2006. Microbicides and other topical strategies to prevent vaginal transmission of HIV. Nat Rev Immunol 6:371–382.
32. Nicolosi A, Corrêa Leite ML, Musicco M, Arici C, Gavazzeni G, Lazzarin A. 1994. The efficiency of male-to-female and female-to-male sexual transmission of the human immunodeficiency virus: a study of 730 stable couples. Italian Study Group on HIV Heterosexual Transmission. Epidemiology 5:570–5.
33. Galvin SR, Cohen MS. 2004. The role of sexually transmitted diseases in HIV transmission. Nat Rev Microbiol 2:33–42.
34. Wall KM, Kilembe W, Vwalika B, Haddad LB, Hunter E, Lakhi S, Chavuma R, Htee Khu N, Brill I, Vwalika C, Mwananyanda L, Chomba E, Mulenga J, Tichacek A, Allen S. 2017. Risk of heterosexual HIV transmission attributable to sexually transmitted infections and non-specific genital inflammation in Zambian discordant couples, 1994–2012. Int J Epidemiol 46:1593–1606.
35. Barnabas SL, Dabee S, Passmore J-AS, Jaspan HB, Lewis DA, Jaumdally SZ, Gamieldien H, Masson L, Muller E, Maseko VD, Mkhize N, Mbulawa Z, Williamson A-L, Gray CM, Hope TJ, Chiodi F, Dietrich J, Gray G, Bekker L-G. 2018. Converging epidemics of sexually transmitted infections and bacterial vaginosis in southern African female adolescents at risk of HIV. Int J STD AIDS 29:531–539.
36. Masson L, Barnabas S, Deese J, Lennard K, Dabee S, Gamieldien H, Jaumdally SZ, Williamson A-L, Little F, Van Damme L, Ahmed K, Crucitti T, Abdellati S, Bekker L-G, Gray G, Dietrich J, Jaspan H, Passmore J-AS. 2018. Inflammatory cytokine biomarkers of asymptomatic sexually transmitted infections and vaginal dysbiosis: a multicentre validation study. Sex Transm Infect sextrans-2017-053506.
37. Fideli ÜS, Allen SA, Musonda R, Trask S, Hahn BH, Weiss H, Mulenga J, Kasolo F, Vermund SH, Aldrovandi GM. 2001. Virologic and immunologic determinants of heterosexual transmission of Human Immunodeficiency Virus Type 1 in Africa. AIDS Res Hum Retroviruses 17:901–910.
38. Quinn TC, Wawer MJ, Sewankambo N, Serwadda D, Li C, Wabwire-Mangen F, Meehan MO, Lutalo T, Gray RH. 2000. Viral load and heterosexual transmission of Human Immunodeficiency Virus Type 1. N Engl J Med 342:921–929.
39. Shaw GM, Hunter E. 2012. HIV Transmission. Cold Spring Harb Perspect Med

2:a006965–a006965.

40. Joseph SB, Swanstrom R, Kashuba ADM, Cohen MS. 2015. Bottlenecks in HIV-1 transmission: Insights from the study of founder viruses. Nat Rev Microbiol 13:414–425.
41. Keele BF, Giorgi EE, Salazar-Gonzalez JF, Decker JM, Pham KT, Salazar MG, Sun C, Grayson T, Wang S, Li H, Wei X, Jiang C, Kirchherr JL, Gao F, Anderson JA, Ping L-H, Swanstrom R, Tomaras GD, Blattner WA, Goepfert PA, Kilby JM, Saag MS, Delwart EL, Busch MP, Cohen MS, Montefiori DC, Haynes BF, Gaschen B, Athreya GS, Lee HY, Wood N, Seoighe C, Perelson AS, Bhattacharya T, Korber BT, Hahn BH, Shaw GM. 2008. Identification and characterization of transmitted and early founder virus envelopes in primary HIV-1 infection. Proc Natl Acad Sci 105:7552–7557.
42. Derdeyn CA, Decker JM, Bibollet-ruche F, Mokili JL, Muldoon M, Denham SA, Heil ML, Kasolo F, Musonda R, Hahn BH, Shaw GM, Korber BT, Allen S, Hunter E. 2004. Envelope-constrained neutralization-sensitive HIV-1 after heterosexual transmission. Science (80-) 303:2019–2022.
43. Wolfs TFW, Zwart G, Bakker M, Goudsmit J. 1992. HIV-1 genomic rna diversification following sexual and parenteral virus transmission. Virology.
44. Wolinsky SM, Wike CM, Korber BTM, Hutto C, Parks WP, Rosenblum LL, Kunstman KJ, Furtado MR, Muñoz JL. 1992. Selective transmission of human immunodeficiency virus type-1 variants from mothers to infants. Science (80-).
45. Zhang LQ, MacKenzie P, Cleland A, Holmes EC, Brown AJ, Simmonds P. 1993. Selection for specific sequences in the external envelope protein of human immunodeficiency virus type 1 upon primary infection. J Virol 67:3345–56.
46. Zhu T, Mo H, Wang N, Nam D, Cao Y, Koup R, Ho D. 1993. Genotypic and phenotypic characterization of HIV-1 patients with primary infection. Science (80-) 261:1179–1181.
47. Selhorst P, Masson L, Ismail SD, Samsunder N, Garrett N, Mansoor LE, Abdool Karim Q, Abdool Karim SS, Passmore J-AS, Williamson C. 2017. Cervicovaginal Inflammation Facilitates Acquisition of Less Infectious HIV Variants. Clin Infect Dis 64:79–82.
48. Bar KJ, Li H, Chamberland A, Tremblay C, Routy JP, Grayson T, Sun C, Wang S, Learn GH, Morgan CJ, Schumacher JE, Haynes BF, Keele BF, Hahn BH, Shaw GM. 2010. Wide variation in the multiplicity of HIV-1 infection among

injection drug users. J Virol 84:6241–6247.

49. Parrish NF, Gao F, Li H, Giorgi EE, Barbian HJ, Parrish EH, Zajic L, Iyer SS, Decker JM, Kumar A, Hora B, Berg A, Cai F, Hopper J, Denny TN, Ding H, Ochsenbauer C, Kappes JC, Galimidi RP, West AP, Bjorkman PJ, Wilen CB, Doms RW, O'Brien M, Bhardwaj N, Borrow P, Haynes BF, Muldoon M, Theiler JP, Korber B, Shaw GM, Hahn BH. 2013. Phenotypic properties of transmitted founder HIV-1. Proc Natl Acad Sci 110:6626–6633.
50. Iyer SS, Bibollet-Ruche F, Sherrill-Mix S, Learn GH, Plenderleith L, Smith AG, Barbian HJ, Russell RM, Gondim MVP, Bahari CY, Shaw CM, Li Y, Decker T, Haynes BF, Shaw GM, Sharp PM, Borrow P, Hahn BH. 2017. Resistance to type 1 interferons is a major determinant of HIV-1 transmission fitness. Proc Natl Acad Sci 114:E590–E599.
51. Freed EO. 2001. HIV-1 Replication. Somat Cell Mol Genet 26:13–33.
52. Yuan T, Li J, Zhang M-Y. 2013. HIV-1 envelope glycoprotein variable loops are indispensable for envelope structural integrity and virus entry. PLoS One 8:e69789.
53. Wild C, Dubay JW, Greenwell T, Baird T, Oas TG, McDanal C, Hunter E, Matthews T. 1994. Propensity for a leucine zipper-like domain of human immunodeficiency virus type 1 gp41 to form oligomers correlates with a role in virus-induced fusion rather than assembly of the glycoprotein complex. Proc Natl Acad Sci U S A 91:12676–80.
54. Melikyan GB, Markosyan RM, Hemmati H, Delmedico MK, Lambert DM, Cohen FS. 2000. Evidence that the transition of HIV-1 Gp41 into a six-helix bundle, not the bundle configuration, induces membrane fusion. J Cell Biol 151:413–424.
55. Garg H, Blumenthal R. 2008. Role of HIV Gp41 mediated fusion/hemifusion in bystander apoptosis. Cell Mol Life Sci 65:3134–44.
56. Pancera M, Majeed S, Ban Y-EA, Chen L, Huang C -c., Kong L, Kwon YD, Stuckey J, Zhou T, Robinson JE, Schief WR, Sodroski J, Wyatt R, Kwong PD. 2010. Structure of HIV-1 gp120 with gp41-interactive region reveals layered envelope architecture and basis of conformational mobility. Proc Natl Acad Sci 107:1166–1171.
57. Lyumkis D, Julien J-P, de Val N, Cupo A, Potter CS, Klasse P-J, Burton DR, Sanders RW, Moore JP, Carragher B, Wilson IA, Ward AB. 2013. Cryo-EM structure of a fully glycosylated soluble cleaved HIV-1 envelope trimer. Science

(80-) 342:1484–1490.
58. Korkut A, Hendrickson WA. 2012. Structural plasticity and conformational transitions of HIV envelope glycoprotein gp120. PLoS One 7:e52170.
59. Kwong PD, Wyatt R, Robinson J, Sweet RW, Sodroski J, Hendrickson WA. 1998. Structure of an HIV gp120 envelope glycoprotein in complex with the CD4 receptor and a neutralizing human antibody. Nature1998/06/26. 393:648–659.
60. Wilen CB, Tilton JC, Doms RW. 2012. HIV: Cell binding and entry. Cold Spring Harb Perspect Med 2:a006866–a006866.
61. Dismuke DJ, Aiken C. 2006. Evidence for a functional link between uncoating of the human immunodeficiency virus type 1 core and nuclear import of the viral preintegration complex. J Virol 80:3712–20.
62. Matreyek K, Engelman A. 2013. Viral and cellular requirements for the nuclear entry of retroviral preintegration nucleoprotein complexes. Viruses 5:2483–2511.
63. Julien J-P, Lee JH, Cupo A, Murin CD, Derking R, Hoffenberg S, Caulfield MJ, King CR, Marozsan AJ, Klasse PJ, Sanders RW, Moore JP, Wilson IA, Ward AB, Hyun J, Cupo A, Murin CD, Derking R, Hoffenberg S, Wilson IA, Ward AB. 2013. Asymmetric recognition of the HIV-1 trimer by broadly neutralizing antibody PG9. Proc Natl Acad Sci U S A 110:4351–4356.
64. Pancera M, Zhou T, Druz A, Georgiev IS, Soto C, Gorman J, Huang J, Acharya P, Chuang GY, Ofek G, Stewart-Jones GBE, Stuckey J, Bailer RT, Joyce MG, Louder MK, Tumba N, Yang Y, Zhang B, Cohen MS, Haynes BF, Mascola JR, Morris L, Munro JB, Blanchard SC, Mothes W, Connors M, Kwong PD. 2014. Structure and immune recognition of trimeric pre-fusion HIV-1 Env. Nature 514.
65. Gaschen B. 2002. Diversity Considerations in HIV-1 Vaccine Selection. Science (80-) 296:2354–2360.
66. Fischer W, Ganusov V V, Giorgi EE, Hraber PT, Keele BF, Leitner T, Han CS, Gleasner CD, Green L, Lo C-C, Nag A, Wallstrom TC, Wang S, McMichael AJ, Haynes BF, Hahn BH, Perelson AS, Borrow P, Shaw GM, Bhattacharya T, Korber BT. 2010. Transmission of single HIV-1 genomes and dynamics of early immune escape revealed by ultra-deep sequencing. PLoS One 5:e12303.
67. Hu W-S, Hughes SH. 2012. HIV-1 Reverse Transcription. Cold Spring Harb Perspect Med 2:a006882–a006882.
68. O'Neil PK, Sun G, Yu H, Ron Y, Dougherty JP, Preston BD. 2002. Mutational

analysis of HIV-1 long terminal repeats to explore the relative contribution of reverse transcriptase and RNA polymerase II to viral mutagenesis. J Biol Chem 277:38053–38061.

69. Pathak VK, Temin HM. 1990. Broad spectrum of in vivo forward mutations, hypermutations, and mutational hotspots in a retroviral shuttle vector after a single replication cycle: substitutions, frameshifts, and hypermutations. Proc Natl Acad Sci 87:6019–6023.
70. Kim T, Mudry RA, Rexrode CA, Pathak VK. 1996. Retroviral mutation rates and A-to-G hypermutations during different stages of retroviral replication. J Virol 70:7594–602.
71. Burke D. 1997. Recombination in HIV: An Important viral evolutionary strategy. Emerg Infect Dis 3:253–259.
72. Pathak VK, Hu W-S. 1997. “Might as Well Jump!” Template switching by retroviral reverse transcriptase, defective genome formation, and recombination. Semin Virol 8:141–150.
73. Jetzt AE, Yu H, Klarmann GJ, Ron Y, Preston BD, Dougherty JP. 2000. High rate of recombination throughout the human immunodeficiency virus type 1 genome. J Virol 74:1234–40.
74. Rhodes T, Wargo H, Hu WS. 2003. High rates of human immunodeficiency virus type 1 recombination: near-random segregation of markers one kilobase apart in one round of viral replication. J Virol 77:11193–11200.
75. Levy DN, Aldrovandi GM, Kutsch O, Shaw GM. 2004. Dynamics of HIV-1 recombination in its natural target cells. Proc Natl Acad Sci 101:4204–4209.
76. Hemelaar J, Gouws E, Ghys PD, Osmanov S. 2006. Global and regional distribution of HIV-1 genetic subtypes and recombinants in 2004. AIDS 20:W13–W23.
77. Janes H, Herbeck JT, Tovanabutra S, Thomas R, Frahm N, Duerr A, Hural J, Corey L, Self SG, Buchbinder SP, McElrath MJ, O’Connell RJ, Paris RM, Rerks-Ngarm S, Nitayaphan S, Pitisuttihum P, Kaewkungwal J, Robb ML, Michael NL, Mullins JI, Kim JH, Gilbert PB, Rolland M. 2015. HIV-1 infections with multiple founders are associated with higher viral loads than infections with single founders. Nat Med 21:1139–1141.
78. Blackard JT. 2012. HIV compartmentalization: A Review on a clinically important phenomenon. Curr HIV Res 10:133–142.
79. Penton PK, Blackard JT. 2014. Analysis of HIV quasispecies suggests

compartmentalization in the liver. AIDS Res Hum Retroviruses 30:394–402.
80. Stefic K, Chaillon A, Bouvin-Pley M, Moreau A, Braibant M, Bastides F, Gras G, Bernard L, Barin F. 2017. Probing the compartmentalization of HIV-1 in the central nervous system through its neutralization properties. PLoS One 12:e0181680.
81. Gantt S, Carlsson J, Heath L, Bull ME, Shetty AK, Mutsvangwa J, Musingwini G, Woelk G, Zijenah LS, Katzenstein DA, Mullins JI, Frenkel LM. 2010. Genetic analyses of HIV-1 env sequences demonstrate limited compartmentalization in breast milk and suggest viral replication within the breast that increases with mastitis. J Virol 84:10812–10819.
82. Heath L, Fox A, McClure J, Diem K, van 't Wout AB, Zhao H, Park DR, Schouten JT, Twigg HL, Corey L, Mullins JI, Mittler JE. 2009. Evidence for limited genetic compartmentalization of HIV-1 between lung and blood. PLoS One 4:e6949.
83. Ohagen A, Devitt A, Kunstman KJ, Gorry PR, Rose PP, Korber B, Taylor J, Levy R, Murphy RL, Wolinsky SM, Gabuzda D. 2003. Genetic and functional analysis of full-length Human Immunodeficiency Virus Type 1 env genes derived from brain and blood of patients with AIDS. J Virol 77:12336–12345.
84. Power C, McArthur JC, Johnson RT, Griffin DE, Glass JD, Perryman S, Chesebro B. 1994. Demented and nondemented patients with AIDS differ in brain-derived human immunodeficiency virus type 1 envelope sequences. J Virol 68:4643–49.
85. Dunfee RL, Thomas ER, Gorry PR, Wang J, Taylor J, Kunstman K, Wolinsky SM, Gabuzda D. 2006. The HIV Env variant N283 enhances macrophage tropism and is associated with brain infection and dementia. Proc Natl Acad Sci 103:15160–15165.
86. Harrington PR, Schnell G, Letendre SL, Ritola K, Robertson K, Hall C, Burch CL, Jabara CB, Moore DT, Ellis RJ, Price RW, Swanstrom R. 2009. Cross-sectional characterization of HIV-1 env compartmentalization in cerebrospinal fluid over the full disease course. AIDS 23:907–915.
87. Bednar MM, Sturdevant CB, Tompkins LA, Arrildt KT, Dukhovlinova E, Kincer LP, Swanstrom R. 2015. Compartmentalization, viral evolution, and viral latency of HIV in the CNS. Curr HIV/AIDS Rep 12:262–271.
88. Ritola K, Robertson K, Fiscus S a, Hall C, Swanstrom R. 2005. Increased Human Immunodeficiency Virus Type 1 (HIV-1) env compartmentalization in

the presence of HIV-1-associated dementia. J Virol 79:10830–10834.
89. Schnell G, Price RW, Swanstrom R, Spudich S. 2010. Compartmentalization and Clonal Amplification of HIV-1 Variants in the Cerebrospinal Fluid during Primary Infection. J Virol 84:2395–2407.
90. Diem K, Nickle DC, Motoshige A, Fox A, Ross S, Mullins JI, Corey L, Coombs RW, Krieger JN. 2008. Male genital tract compartmentalization of Human Immunodeficiency Virus Type 1 (HIV). AIDS Res Hum Retroviruses 24:561–571.
91. Gupta P, Leroux C, Patterson BK, Kingsley L, Rinaldo C, Ding M, Chen Y, Kulka K, Buchanan W, McKeon B, Montelaro R. 2000. Human Immunodeficiency Virus Type 1 shedding pattern in semen correlates with the compartmentalization of viral quasi species between blood and semen. J Infect Dis 182:79–87.
92. Pillai SK, Good B, Pond SK, Wong JK, Strain MC, Richman DD, Smith DM. 2005. Semen-specific genetic characteristics of Human Immunodeficiency Virus Type 1 env. J Virol 79:1734–1742.
93. Coombs RW, Speck CE, Hughes JP, Lee W, Sampoleo R, Ross SO, Dragavon J, Peterson G, Hooton TM, Collier AC, Corey L, Koutsky L, Krieger JN. 1998. Association between culturable Human Immunodeficiency Virus Type 1 (HIV-1) in semen and HIV-1 RNA levels in semen and blood: Evidence for compartmentalization of HIV-1 between semen and blood. J Infect Dis 177:320–330.
94. Anderson JA, Ping L-H, Dibben O, Jabara CB, Arney L, Kincer L, Tang Y, Hobbs M, Hoffman I, Kazembe P, Jones CD, Borrow P, Fiscus S, Cohen MS, Swanstrom R. 2010. HIV-1 populations in semen arise through multiple mechanisms. PLoS Pathog 6:e1001053.
95. Bull ME, Heath LM, McKernan-Mullin JL, Kraft KM, Acevedo L, Hitti JE, Cohn SE, Tapia KA, Holte SE, Dragavon JA, Coombs RW, Mullins JI, Frenkel LM. 2013. Human Immunodeficiency Viruses appear compartmentalized to the female genital tract in cross-sectional analyses but genital lineages do not persist over time. J Infect Dis 207:1206–1215.
96. Bull M, Learn G, Genowati I, McKernan J, Hitti J, Lockhart D, Tapia K, Holte S, Dragavon J, Coombs R, Mullins J, Frenkel L. 2009. Compartmentalization of HIV-1 within the female genital tract Is due to monotypic and low-diversity variants not distinct viral populations. PLoS One 4:e7122.

97. Kemal KS, Foley B, Burger H, Anastos K, Minkoff H, Kitchen C, Philpott SM, Gao W, Robison E, Holman S, Dehner C, Beck S, Meyer WA, Landay A, Kovacs A, Bremer J, Weiser B. 2003. HIV-1 in genital tract and plasma of women: Compartmentalization of viral sequences, coreceptor usage, and glycosylation. Proc Natl Acad Sci 100:12972–12977.
98. Chaudhary S, Noel RJ, Rodríguez N, Collado S, Munoz J, Kumar A, Yamamura Y. 2011. Correlation between CD4 T cell counts and virus compartmentalization in genital and systemic compartments of HIV-infected females. Virology 417:320–326.
99. Takahata N. 1989. Gene genealogy in three related populations: consistency probability between gene and population trees. Genetics 122:957–966.
100. Doyle JJ. 1992. Gene Trees and Species Trees: Molecular Systematics as One-Character Taxonomy. Syst Bot 17:144–163.
101. Maddison WP. 1997. Gene Trees in Species Trees. Syst Biol 46:523–536.
102. Rosenberg NA. 2002. The Probability of Topological Concordance of Gene Trees and Species Trees. Theor Popul Biol 61:225–247.
103. Poss M, Rodrigo a G, Gosink JJ, Learn GH, de Vange Panteleeff D, Martin HL, Bwayo J, Kreiss JK, Overbaugh J. 1998. Evolution of envelope sequences from the genital tract and peripheral blood of women infected with clade A human immunodeficiency virus type 1. J Virol 72:8240–51.
104. Andreoletti L, Skrabal K, Perrin V, Chomont N, Saragosti S, Gresenguet G, Moret H, Jacques J, Longo JDD, Matta M, Mammano F, Belec L. 2007. Genetic and phenotypic features of blood and genital viral populations of clinically asymptomatic and antiretroviral-treatment-naive clade A Human Immunodeficiency Virus Type 1-infected Women. J Clin Microbiol 45:1838–1842.
105. Sullivan ST, Mandava U, Evans-Strickfaden T, Lennox JL, Ellerbrock T V., Hart CE. 2005. Diversity, divergence, and evolution of cell-free Human Immunodeficiency Virus Type 1 in vaginal secretions and blood of chronically infected women: Associations with immune status. J Virol 79:9799–9809.
106. Philpott S, Burger H, Tsoukas C, Foley B, Anastos K, Kitchen C, Weiser B. 2005. Human Immunodeficiency Virus Type 1 genomic RNA sequences in the female genital tract and blood: Compartmentalization and intrapatient recombination. J Virol 79:353–363.
107. Avise JC, Shapira JF, Daniel SW, Aquadro CF, Lansman RA. 1983.

Mitochondrial DNA differentiation during the speciation process in Peromyscus. Mol Biol Evol 1:38–56.

108. Pamilo P, Nei M. 1988. Relationships between gene trees and species trees. Mol Biol Evol 5:568–583.
109. Leruez-Ville M, Dulioust E, Costabliola D, Salmon D, Tachet A, Finkielsztejn L, De Almeida M, Silbermann B, Sicard D, Jouannet P, Rouzioux C. 2002. Decrease in HIV-1 seminal shedding in men receiving highly active antiretroviral therapy: an 18 month longitudinal study (ANRS EP012). AIDS 16:486–488.
110. Montaner JS, Lima VD, Barrios R, Yip B, Wood E, Kerr T, Shannon K, Harrigan PR, Hogg RS, Daly P, Kendall P. 2010. Association of highly active antiretroviral therapy coverage, population viral load, and yearly new HIV diagnoses in British Columbia, Canada: a population-based study. Lancet 376:532–539.
111. Sáez-Cirión A, Bacchus C, Hocqueloux L, Avettand-Fenoel V, Girault I, Lecuroux C, Potard V, Versmisse P, Melard A, Prazuck T, Descours B, Guergnon J, Viard J-P, Boufassa F, Lambotte O, Goujard C, Meyer L, Costagliola D, Venet A, Pancino G, Autran B, Rouzioux C. 2013. Post-treatment HIV-1 controllers with a long-term virological remission after the interruption of early initiated antiretroviral therapy ANRS VISCONTI study. PLoS Pathog 9:e1003211.
112. Li T, Tubiana R, Katlama C, Calvez V, Mohand HA, Autran B. 1998. Long-lasting recovery in CD4 T-cell function and viral-load reduction after highly active antiretroviral therapy in advanced HIV-1 disease. Lancet 351:1682–1686.
113. Simoni JM, Pearson CR, Pantalone DW, Marks G, Crepaz N. 2006. Efficacy of Interventions in Improving Highly Active Antiretroviral Therapy Adherence and HIV-1 RNA Viral Load. JAIDS J Acquir Immune Defic Syndr 43:S23–S35.
114. Sidibé M, Loures L, Samb B. 2016. The UNAIDS 90-90-90 target: a clear choice for ending AIDS and for sustainable health and development. J Int AIDS Soc 19:21133.
115. Damle S. 2013. HIV/AIDS - Accomplishments and challenges? Contemp Clin Dent 4:419.
116. McNairy ML, El-Sadr WM. 2014. Antiretroviral therapy for the prevention of HIV transmission: What will It take? Clin Infect Dis 58:1003–1011.

117. Kang L. 2012. Overview of achievements and challenges of the fight against AIDS in China. ISRN Immunol 2012:1–6.
118. Gupta RK, Jordan MR, Sultan BJ, Hill A, Davis DHJ, Gregson J, Sawyer AW, Hamers RL, Ndembi N, Pillay D, Bertagnolio S. 2012. Global trends in antiretroviral resistance in treatment-naive individuals with HIV after rollout of antiretroviral treatment in resource-limited settings: a global collaborative study and meta-regression analysis. Lancet 380:1250–1258.
119. Brenner B, Ibanescu R-I, Hardy I, Roger M, Brenner BG, Ibanescu R-I, Hardy I, Roger M. 2017. Genotypic and phylogenetic insights on prevention of the spread of HIV-1 and drug resistance in “real-world” settings. Viruses 10:10.
120. Patel P, Borkowf CB, Brooks JT, Lasry A, Lansky A, Mermin J. 2014. Estimating per-act HIV transmission risk. AIDS 28:1509–1519.
121. Emmanuel F, Archibald C, Razaque A, Sandstrom P. 2009. Factors associated with an explosive HIV epidemic among injecting drug users in Sargodha, Pakistan. JAIDS J Acquir Immune Defic Syndr 51:85–90.
122. Reid SR. 2009. Injection drug use, unsafe medical injections, and HIV in Africa: a systematic review. Harm Reduct J 6:24.
123. Auvert B, Taljaard D, Lagarde E, Sobngwi-Tambekou J, Sitta R, Puren A. 2005. Randomized, controlled intervention trial of male circumcision for reduction of HIV infection risk: The ANRS 1265 trial. PLoS Med 2:e298.
124. Gray RH, Kigozi G, Serwadda D, Makumbi F, Watya S, Nalugoda F, Kiwanuka N, Moulton LH, Chaudhary MA, Chen MZ, Sewankambo NK, Wabwire-Mangen F, Bacon MC, Williams CF, Opendi P, Reynolds SJ, Laeyendecker O, Quinn TC, Wawer MJ. 2007. Male circumcision for HIV prevention in men in Rakai, Uganda: a randomised trial. Lancet 369:657–666.
125. Bailey RC, Moses S, Parker CB, Agot K, Maclean I, Krieger JN, Williams CF, Campbell RT, Ndinya-Achola JO. 2007. Male circumcision for HIV prevention in young men in Kisumu, Kenya: a randomised controlled trial. Lancet 369:643–656.
126. Szabo R. 2000. How does male circumcision protect against HIV infection? BMJ 320:1592–1594.
127. Williams BG, Lloyd-Smith JO, Gouws E, Hankins C, Getz WM, Hargrove J, de Zoysa I, Dye C, Auvert B. 2006. The potential impact of male circumcision on HIV in sub-Saharan Africa. PLoS Med 3:e262.
128. Kashuba ADM, Patterson KB, Dumond JB, Cohen MS. 2012. Pre-exposure

prophylaxis for HIV prevention: how to predict success. Lancet 379:2409–2411.

129. Gomez GB, Borquez A, Case KK, Wheelock A, Vassall A, Hankins C. 2013. The cost and impact of scaling up pre-exposure prophylaxis for HIV prevention: a systematic review of cost-effectiveness modelling studies. PLoS Med 10:e1001401.
130. Long EF, Stavert RR. 2013. Portfolios of Biomedical HIV Interventions in South Africa: A Cost-Effectiveness Analysis. J Gen Intern Med 28:1294–1301.
131. Celum C, Baeten JM, Hughes JP, Barnabas R, Liu A, Van Rooyen H, Buchbinder S. 2013. Integrated Strategies for Combination HIV Prevention. JAIDS J Acquir Immune Defic Syndr 63:S213–S220.
132. Burns DN, Grossman C, Turpin J, Elharrar V, Veronese F. 2014. Role of Oral Pre-exposure Prophylaxis (PrEP) in Current and Future HIV Prevention Strategies. Curr HIV/AIDS Rep 11:393–403.
133. Cohen MS, McCauley M, Gamble TR. 2012. HIV treatment as prevention and HPTN 052. Curr Opin HIV AIDS 7:99–105.
134. Cohen J. 2010. Treatment as prevention. Science (80-) 327:1196–1197.
135. Ehreth J. 2003. The global value of vaccination. Vaccine2003/01/18. 21:596–600.
136. Korber B, Gaschen B, Yusim K, Thakallapally R, Kesmir C, Detours V. 2001. Evolutionary and immunological implications of contemporary HIV-1 variation. Br Med Bull 58:19–42.
137. Haynes BF, Gilbert PB, McElrath MJ, Zolla-Pazner S, Tomaras GD, Alam SM, Evans DT, Montefiori DC, Karnasuta C, Sutthent R, Liao H-X, DeVico AL, Lewis GK, Williams C, Pinter A, Fong Y, Janes H, DeCamp A, Huang Y, Rao M, Billings E, Karasavvas N, Robb ML, Ngauy V, de Souza MS, Paris R, Ferrari G, Bailer RT, Soderberg KA, Andrews C, Berman PW, Frahm N, De Rosa SC, Alpert MD, Yates NL, Shen X, Koup RA, Pitisuttithum P, Kaewkungwal J, Nitayaphan S, Rerks-Ngarm S, Michael NL, Kim JH. 2012. Immune-correlates analysis of an HIV-1 vaccine efficacy trial. N Engl J Med 366:1275–1286.
138. Garrett N, Mngadi K, Naicker N, Morris L. 2017. HIV Vaccine and Passive Immunity Trials, p. 91–104. *In* The CAPRISA Clinical Trials: HIV Treatment and Prevention. Springer International Publishing, Cham.
139. Haynes BF, Mascola JR. 2017. The quest for an antibody-based HIV vaccine.

Immunol Rev 275:5–10.
140. Stephenson KE, Barouch DH. 2013. A global approach to HIV-1 vaccine development. Immunol Rev 254:295–304.
141. Stephenson KE, Barouch DH. 2016. Broadly neutralizing antibodies for HIV eradication. Curr HIV/AIDS Rep 13:31–37.
142. Stephenson KE, D'Couto HT, Barouch DH. 2016. New concepts in HIV-1 vaccine development. Curr Opin Immunol 41:39–46.
143. Gama L, Koup RA. 2018. New-Generation High-Potency and Designer Antibodies: Role in HIV-1 Treatment. Annu Rev Med 69:409–419.
144. Abdool Karim Q, Abdool Karim SS, Frohlich JA, Grobler AC, Baxter C, Mansoor LE, Kharsany ABM, Sibeko S, Mlisana KP, Omar Z, Gengiah TN, Maarschalk S, Arulappan N, Mlotshwa M, Morris L, Taylor D. 2010. Effectiveness and Safety of Tenofovir Gel, an Antiretroviral Microbicide, for the Prevention of HIV Infection in Women. Science (80-) 329:1168–1174.
145. Saunders KO, Wang L, Joyce MG, Yang Z, Balazs AB, Cheng C, Ko S, Kong W, Rudicell RS, Georgiev IS, Duan L, Foulds KE, Donaldson M, Xu L, Schmidt SD, Todd J, Baltimore D, Roederer M, Haase AT, Kwong PD, Rao SS, Mascola JR, Nabel GJ. 2015. Broadly Neutralizing Human Immunodeficiency Virus Type 1 Antibody Gene Transfer Protects Nonhuman Primates from Mucosal Simian-Human Immunodeficiency Virus Infection. J Virol 89:8334–8345.
146. Marcobal A, Liu X, Zhang W, Dimitrov AS, Jia L, Lee PP, Fouts TR, Parks TP, Lagenaur LA. 2016. Expression of Human Immunodeficiency Virus Type 1 Neutralizing Antibody Fragments Using Human Vaginal Lactobacillus. AIDS Res Hum Retroviruses 32:964–971.
147. Addo MM, Yu XG, Rosenberg ES, Walker BD, Altfeld M. 2002. Cytotoxic T-lymphocyte (CTL) responses directed against regulatory and accessory proteins in HIV-1 infection. DNA Cell Biol 21:671–678.
148. Genescà M. 2011. Characterization of an Effective CTL Response against HIV and SIV Infections. J Biomed Biotechnol 2011:103924.
149. Jerne NK. 1955. The natural-selection theory of antibody formation. Proc Natl Acad Sci 41:849–857.
150. Tonegawa S. 1983. Somatic generation of antibody diversity. Nature 302:575–581.
151. Glanville J, Zhai W, Berka J, Telman D, Huerta G, Mehta GR, Ni I, Mei L,

Sundar PD, Day GMR, Cox D, Rajpal A, Pons J. 2009. Precise determination of the diversity of a combinatorial antibody library gives insight into the human immunoglobulin repertoire. Proc Natl Acad Sci 106:20216–20221.

152. Wiley SR, Raman VS, Desbien A, Bailor HR, Bhardwaj R, Shakri AR, Reed SG, Chitnis CE, Carter D. 2011. Targeting TLRs Expands the Antibody Repertoire in Response to a Malaria Vaccine. Sci Transl Med 3:93ra69-93ra69.
153. Alberts B, Johnson A, Lewis J, Raff M, Roberts K, Walter P. 2003. Molecular biology of the cell. 91:401.
154. Schroeder HW, Cavacini L. 2010. Structure and function of immunoglobulins. J Allergy Clin Immunol 125:S41-52.
155. Morris L, Moody TA. 2017. Chapter 1 - Broadly Neutralizing Antibodies A2 - Modjarrad, Kayvon, p. 3–21. *In* Koff, WC (ed.), Human Vaccines. Academic Press.
156. Janeway CA, Travers P, Walport M, Shlomchik M. 2001. Immunobiology: The Immune System In Health And Disease. Immuno Biol 5.
157. Tomaras GD, Yates NL, Liu P, Qin L, Fouda GG, Chavez LL, Decamp AC, Parks RJ, Ashley VC, Lucas JT, Cohen M, Eron J, Hicks CB, Liao H-X, Self SG, Landucci G, Forthal DN, Weinhold KJ, Keele BF, Hahn BH, Greenberg ML, Morris L, Karim SSA, Blattner WA, Montefiori DC, Shaw GM, Perelson AS, Haynes BF. 2008. Initial B-cell responses to transmitted human immunodeficiency virus type 1: virion-binding immunoglobulin M (IgM) and IgG antibodies followed by plasma anti-gp41 antibodies with ineffective control of initial viremia. J Virol 82:12449–12463.
158. Liao H-X, Chen X, Munshaw S, Zhang R, Marshall DJ, Vandergrift N, Whitesides JF, Lu X, Yu J-S, Hwang K-K, Gao F, Markowitz M, Heath SL, Bar KJ, Goepfert PA, Montefiori DC, Shaw GC, Alam SM, Margolis DM, Denny TN, Boyd SD, Marshal E, Egholm M, Simen BB, Hanczaruk B, Fire AZ, Voss G, Kelsoe G, Tomaras GD, Moody MA, Kepler TB, Haynes BF. 2011. Initial antibodies binding to HIV-1 gp41 in acutely infected subjects are polyreactive and highly mutated. J Exp Med 208:2237–2249.
159. Richman DD, Wrin T, Little SJ, Petropoulos CJ. 2003. Rapid evolution of the neutralizing antibody response to HIV type 1 infection. Proc Natl Acad Sci U S A 100:4144–9.
160. Wei X, Decker JM, Wang S, Hui H, Kappes JC, Wu X, Salazar-Gonzalez JF, Salazar MG, Kilby JM, Saag MS, Komarova NL, Nowak MA, Hahn BH, Kwong

PD, Shaw GM. 2003. Antibody neutralization and escape by HIV-1. Nature 422:307–312.

161. Derdeyn CA, Moore PL, Morris L. 2014. Development of broadly neutralizing antibodies from autologous neutralizing antibody responses in HIV infection. Curr Opin HIV AIDS 9:210–216.
162. Moore PL, Gray ES, Choge IA, Ranchobe N, Mlisana K, Abdool Karim SS, Williamson C, Morris L, Karim SSA, Williamson C, Morris L, Team C 002 S. 2008. The c3-v4 region is a major target of autologous neutralizing antibodies in human immunodeficiency virus type 1 subtype C infection. J Virol 82:1860–1869.
163. Moody MA, Gao F, Gurley TC, Amos JD, Kumar A, Hora B, Marshall DJ, Whitesides JF, Xia S-M, Parks R, Lloyd KE, Hwang K-K, Lu X, Bonsignori M, Finzi A, Vandergrift NA, Alam SM, Ferrari G, Shen X, Tomaras GD, Kamanga G, Cohen MS, Sam NE, Kapiga S, Gray ES, Tumba NL, Morris L, Zolla-Pazner S, Gorny MK, Mascola JR, Hahn BH, Shaw GM, Sodroski JG, Liao H-X, Montefiori DC, Hraber PT, Korber BT, Haynes BF. 2015. Strain-Specific V3 and CD4 Binding Site Autologous HIV-1 Neutralizing Antibodies Select Neutralization-Resistant Viruses. Cell Host Microbe 18:354–362.
164. Moore PL, Ranchobe N, Lambson BE, Gray ES, Cave E, Abrahams M-R, Bandawe G, Mlisana K, Abdool Karim SS, Williamson C, Morris L. 2009. Limited neutralizing antibody specificities drive neutralization escape in early HIV-1 subtype C infection. PLoS Pathog 5:e1000598.
165. Gray ES, Madiga MC, Hermanus T, Moore L, Wibmer CK, Tumba NL, Mlisana K, Sibeko S, Karim SSA, Morris L, Gray ES, Madiga MC, Hermanus T, Moore PL, Wibmer CK, Tumba NL, Werner L, Mlisana K, Sibeko S, Williamson C, Karim SSA, Morris L, Team S, Abdool Karim SS, Morris L. 2011. The neutralization breadth of HIV-1 develops incrementally over four years and is associated with CD4+ T cell decline and high viral load during acute infection. J Virol 85:4828–4840.
166. Sather DN, Carbonetti S, Malherbe DC, Pissani F, Stuart AB, Hessell AJ, Gray MD, Mikell I, Kalams SA, Haigwood NL, Stamatatos L. 2014. Emergence of broadly neutralizing antibodies and viral coevolution in two subjects during the early stages of infection with human immunodeficiency virus type 1. J Virol 88:12968–12981.
167. Landais E, Huang X, Havenar-Daughton C, Murrell B, Price MA,

Wickramasinghe L, Ramos A, Bian CB, Simek M, Allen S, Karita E, Kilembe W, Lakhi S, Inambao M, Kamali A, Sanders EJ, Anzala O, Edward V, Bekker L-G, Tang J, Gilmour J, Kosakovsky-Pond SL, Phung P, Wrin T, Crotty S, Godzik A, Poignard P. 2016. Broadly neutralizing antibody responses in a large longitudinal sub-Saharan HIV primary infection cohort. PLOS Pathog 12:e1005369.

168. Rusert P, Kouyos RD, Kadelka C, Ebner H, Schanz M, Huber M, Braun DL, Hozé N, Scherrer A, Magnus C, Weber J, Uhr T, Cippa V, Thorball CW, Kuster H, Cavassini M, Bernasconi E, Hoffmann M, Calmy A, Battegay M, Rauch A, Yerly S, Aubert V, Klimkait T, Böni J, Fellay J, Regoes RR, Günthard HF, Trkola A. 2016. Determinants of HIV-1 broadly neutralizing antibody induction. Nat Med 22:1260–1267.
169. McElrath MJ, Haynes BF. 2010. Induction of immunity to human immunodeficiency virus type-1 by vaccination. Immunity 33:542–554.
170. Scheid JF, Mouquet H, Ueberheide B, Diskin R, Klein F, Oliveira TY, Pietzsch J, Fenyo D, Abadir A, Velinzon K, Hurley A, Myung S, Boulad F, Poignard P, Burton DR, Pereyra F, Ho DD, Walker BD, Seaman MS, Bjorkman PJ, Chait BT, Nussenzweig MC. 2011. Sequence and structural convergence of broad and potent HIV antibodies that mimic CD4 binding. Science (80-) 333:1633–1637.
171. Verkoczy L, Chen Y, Bouton-Verville H, Zhang J, Diaz M, Hutchinson J, Ouyang YB, Alam SM, Holl TM, Hwang KK, Kelsoe G, Haynes BF. 2011. Rescue of HIV-1 broad neutralizing antibody-expressing B cells in 2F5 VH x VL knockin mice reveals multiple tolerance controls. J Immunol 187:3785–3797.
172. Moore PL, Williamson C, Morris L. 2015. Virological features associated with the development of broadly neutralizing antibodies to HIV-1. Trends Microbiol 23:204–211.
173. Bonsignori M, Liao HX, Gao F, Williams WB, Alam SM, Montefiori DC, Haynes BF. 2017. Antibody-virus co-evolution in HIV infection: paths for HIV vaccine development. Immunol Rev 275:145–160.
174. Wibmer CK, Moore PL, Morris L. 2015. HIV broadly neutralizing antibody targets. Curr Opin HIV AIDS 10:135–143.
175. Zolla-Pazner S, Cardozo T. 2010. Structure-function relationships of HIV-1 envelope sequence-variable regions refocus vaccine design. Nat Rev Immunol 10:527–535.

176. Alam SM, Dennison SM, Aussedat B, Vohra Y, Park PK, Fernandez-Tejada A, Stewart S, Jaeger FH, Anasti K, Blinn JH, Kepler TB, Bonsignori M, Liao H-X, Sodroski JG, Danishefsky SJ, Haynes BF. 2013. Recognition of synthetic glycopeptides by HIV-1 broadly neutralizing antibodies and their unmutated ancestors. Proc Natl Acad Sci 110:18214–18219.
177. Walker LM, Simek MD, Priddy F, Gach JS, Wagner D, Zwick MB, Phogat SK, Poignard P, Burton DR. 2010. A limited number of antibody specificities mediate broad and potent serum neutralization in selected HIV-1 infected individuals. PLoS Pathog 6:e1001028.
178. Georgiev IS, Doria-Rose NA, Zhou T, Do Kwon Y, Staupe RP, Moquin S, Chuang G-Y, Louder MK, Schmidt SD, Altae-Tran HR, Bailer RT, McKee K, Nason M, O'Dell S, Ofek G, Pancera M, Srivatsan S, Shapiro L, Connors M, Migueles SA, Morris L, Nishimura Y, Martin MA, Mascola JR, Kwong PD. 2013. Delineating antibody recognition in polyclonal sera from patterns of HIV-1 isolate neutralization. Science (80-) 340:751–756.
179. Kim JH, Excler JL, Michael NL. 2015. Lessons from the RV144 Thai phase III HIV-1 vaccine trial and the search for correlates of protection. Annu Rev Med 66:423–437.
180. Moore PL, Gray ES, Sheward D, Ranchobe N, Lai Z, William J, Nonyane M, Tumba N, Hermanus T, Sibeko S, Mlisana K, Karim SSA, Williamson C, Pinter A, Morris L, Moore PL, Gray ES, Sheward D, Madiga M, Ranchobe N, Lai Z, Honnen WJ, Nonyane M, Tumba N, Hermanus T, Sibeko S, Mlisana K, Karim SSA, Williamson C, Pinter A, Morris L, Study C. 2011. Potent and Broad Neutralization of HIV-1 Subtype C by Plasma Antibodies Targeting a Quaternary Epitope Including Residues in the V2 Loop Potent and Broad Neutralization of HIV-1 Subtype C by Plasma Antibodies Targeting a Quaternary Epitope Including Resid.
181. Wibmer CK, Bhiman JN, Gray ES, Tumba N, Abdool Karim SS, Williamson C, Morris L, Moore PL. 2013. Viral escape from HIV-1 neutralizing antibodies drives increased plasma neutralization breadth through sequential recognition of multiple epitopes and immunotypes. PLoS Pathog 9:e1003738.
182. Doria-Rose NA, Schramm CA, Gorman J, Moore PL, Bhiman JN, DeKosky BJ, Ernandes MJ, Georgiev IS, Kim HJ, Pancera M, Staupe RP, Altae-Tran HR, Bailer RT, Crooks ET, Cupo A, Druz A, Garrett NJ, Hoi KH, Kong R, Louder MK, Longo NS, McKee K, Nonyane M, O'Dell S, Roark RS, Rudicell RS,

Schmidt SD, Sheward DJ, Soto C, Wibmer CK, Yang Y, Zhang Z, Mullikin JC, Binley JM, Sanders RW, Wilson IA, Moore JP, Ward AB, Georgiou G, Williamson C, Abdool Karim SS, Morris L, Kwong PD, Shapiro L, Mascola JR. 2014. Developmental pathway for potent V1V2-directed HIV-neutralizing antibodies. Nature 509:55–62.

183. Pejchal R, Walker LM, Stanfield RL, Phogat SK, Koff WC, Poignard P, Burton DR, Wilson IA. 2010. Structure and function of broadly reactive antibody PG16 reveal an H3 subdomain that mediates potent neutralization of HIV-1. Proc Natl Acad Sci 107:11483–11488.
184. McLellan JS, Pancera M, Carrico C, Gorman J, Julien JP, Khayat R, Louder R, Pejchal R, Sastry M, Dai K, O'Dell S, Patel N, Shahzad-ul-Hussan S, Yang Y, Zhang B, Zhou T, Zhu J, Boyington JC, Chuang GY, Diwanji D, Georgiev I, Kwon YD, Lee D, Louder MK, Moquin S, Schmidt SD, Yang ZY, Bonsignori M, Crump JA, Kapiga SH, Sam NE, Haynes BF, Burton DR, Koff WC, Walker LM, Phogat S, Wyatt R, Orwenyo J, Wang LX, Arthos J, Bewley CA, Mascola JR, Nabel GJ, Schief WR, Ward AB, Wilson IA, Kwong PD. 2011. Structure of HIV-1 gp120 V1/V2 domain with broadly neutralizing antibody PG9. Nature 480:336–343.
185. Julien J-P, Cupo A, Sok D, Stanfield RL, Lyumkis D, Deller MC, Klasse P-J, Burton DR, Sanders RW, Moore JP, Ward AB, Wilson IA. 2013. Crystal structure of a soluble cleaved HIV-1 envelope trimer. Science (80-) 342:1477–1483.
186. Bhiman JN, Anthony C, Doria-Rose NA, Karimanzira O, Schramm CA, Khoza T, Kitchin D, Botha G, Gorman J, Garrett NJ, Karim SSA, Shapiro L, Williamson C, Kwong PD, Mascola JR, Morris L, Moore PL. 2015. Viral variants that initiate and drive maturation of V1V2-directed HIV-1 broadly neutralizing antibodies. Nat Med 21:1332–1336.
187. Pancera M, Shahzad-Ul-Hussan S, Doria-Rose NA, McLellan JS, Bailer RT, Dai K, Loesgen S, Louder MK, Staupe RP, Yang Y, Zhang B, Parks R, Eudailey J, Lloyd KE, Blinn J, Alam SM, Haynes BF, Amin MN, Wang LX, Burton DR, Koff WC, Nabel GJ, Mascola JR, Bewley CA, Kwong PD. 2013. Structural basis for diverse N-glycan recognition by HIV-1-neutralizing V1-V2-directed antibody PG16. Nat Struct Mol Biol 20:804–813.
188. Liao H-X, Bonsignori M, Alam SM, McLellan JS, Tomaras GD, Moody MA, Kozink DM, Hwang K-K, Chen X, Tsao C-Y, Liu P, Lu X, Parks RJ, Montefiori

DC, Ferrari G, Pollara J, Rao M, Peachman KK, Santra S, Letvin NL, Karasavvas N, Yang Z-Y, Dai K, Pancera M, Gorman J, Wiehe K, Nicely NI, Rerks-Ngarm S, Nitayaphan S, Kaewkungwal J, Pitisuttithum P, Tartaglia J, Sinangil F, Kim JH, Michael NL, Kepler TB, Kwong PD, Mascola JR, Nabel GJ, Pinter A, Zolla-Pazner S, Haynes BF. 2013. Vaccine induction of antibodies against a structurally heterogeneous site of immune pressure within HIV-1 envelope protein variable regions 1 and 2. Immunity 38:176–186.

189. Moore PL, Gray ES, Sheward D, Ranchobe N, Lai Z, William J, Nonyane M, Tumba N, Hermanus T, Sibeko S, Mlisana K, Karim SSA, Williamson C, Pinter A, Morris L, Moore PL, Gray ES, Sheward D, Madiga M, Ranchobe N, Lai Z, Honnen WJ, Nonyane M, Tumba N, Hermanus T, Sibeko S, Mlisana K, Karim SSA, Williamson C, Pinter A, Morris L, Study C, Abdool Karim SS, Williamson C, Pinter A, Morris L. 2011. Potent and broad neutralization of HIV-1 subtype C by plasma antibodies targeting a quaternary epitope including residues in the V2 loop. J Virol 85:3128–3141.
190. Moore PL, Sheward D, Nonyane M, Ranchobe N, Hermanus T, Gray ES, Abdool Karim SS, Williamson C, Morris L. 2013. Multiple pathways of escape from HIV broadly cross-neutralizing V2-dependent antibodies. J Virol 87:4882–4894.
191. Klein F, Mouquet H, Dosenovic P, Scheid JF, Scharf L, Nussenzweig MC. 2013. Antibodies in HIV-1 vaccine development and therapy. Science (80-) 341:1199–1204.
192. Pejchal R, Doores KJ, Walker LM, Khayat R, Huang P-S, Wang S-K, Stanfield RL, Julien J-P, Ramos A, Crispin M, Depetris R, Katpally U, Marozsan A, Cupo A, Maloveste S, Liu Y, McBride R, Ito Y, Sanders RW, Ogohara C, Paulson JC, Feizi T, Scanlan CN, Wong C-H, Moore JP, Olson WC, Ward AB, Poignard P, Schief WR, Burton DR, Wilson IA. 2011. A Potent and Broad Neutralizing Antibody Recognizes and Penetrates the HIV Glycan Shield. Science (80-) 334:1097–1103.
193. Mouquet H, Scharf L, Euler Z, Liu Y, Eden C, Scheid JF. 2012. Complex-type N -glycan recognition by potent broadly neutralizing HIV antibodies 1–10.
194. Kong L, Lee JH, Doores KJ, Murin CD, Julien J-P, McBride R, Liu Y, Marozsan A, Cupo A, Klasse P-J, Hoffenberg S, Caulfield M, King CR, Hua Y, Le KM, Khayat R, Deller MC, Clayton T, Tien H, Feizi T, Sanders RW, Paulson JC, Moore JP, Stanfield RL, Burton DR, Ward AB, Wilson IA. 2013. Supersite of

immune vulnerability on the glycosylated face of HIV-1 envelope glycoprotein gp120. Nat Struct Mol Biol 20:796–803.

195. Doores KJ, Kong L, Krumm SA, Le KM, Sok D, Laserson U, Garces F, Poignard P, Wilson IA, Burton DR. 2015. Two Classes of Broadly Neutralizing Antibodies within a Single Lineage Directed to the High-Mannose Patch of HIV Envelope. J Virol 89:1105–1118.
196. Garces F, Lee JH, de Val N, Torrents de la Pena A, Kong L, Puchades C, Hua Y, Stanfield RL, Burton DR, Moore JP, Sanders RW, Ward AB, Wilson IA. 2015. Affinity Maturation of a Potent Family of HIV Antibodies Is Primarily Focused on Accommodating or Avoiding Glycans. Immunity 43:1053–1063.
197. Garces F, Sok D, Kong L, McBride R, Kim HJ, Saye-Francisco KF, Julien J-P, Hua Y, Cupo A, Moore JP, Paulson JC, Ward AB, Burton DR, Wilson IA. 2014. Structural evolution of glycan recognition by a family of potent HIV antibodies. Cell 159:69–79.
198. Simonich CAA, Williams KLL, Verkerke HPP, Williams JAA, Nduati R, Lee KKK, Overbaugh J. 2016. HIV-1 Neutralizing Antibodies with Limited Hypermutation from an Infant. Cell 166:77–87.
199. Walker LM, Huber M, Doores KJ, Falkowska E, Pejchal R, Julien JP, Wang SK, Ramos A, Chan-Hui PY, Moyle M, Mitcham JL, Hammond PW, Olsen OA, Phung P, Fling S, Wong CH, Phogat S, Wrin T, Simek MD, Protocol GPI, Koff WC, Wilson IA, Burton DR, Poignard P. 2011. Broad neutralization coverage of HIV by multiple highly potent antibodies. Nature 477:466–470.
200. Sok D, Doores KJ, Briney B, Le KM, Saye-Francisco KL, Ramos A, Kulp DW, Julien J-P, Menis S, Wickramasinghe L. 2014. Promiscuous glycan site recognition by antibodies to the high-mannose patch of gp120 broadens neutralization of HIV. Sci Transl Med 6:236ra63-236ra63.
201. MacLeod DT, Choi NM, Briney B, Garces F, Ver LS, Landais E, Murrell B, Wrin T, Kilembe W, Liang C-H, Ramos A, Bian CB, Wickramasinghe L, Kong L, Eren K, Wu C-Y, Wong C-H, Kosakovsky Pond SL, Wilson IA, Burton DR, Poignard P. 2016. Early antibody lineage diversification and independent limb maturation lead to broad HIV-1 neutralization targeting the Env high-mannose patch. Immunity 44:1215–1226.
202. Corti D, Langedijk JPM, Hinz A, Seaman MS, Vanzetta F, Fernandez-Rodriguez BM, Silacci C, Pinna D, Jarrossay D, Balla-Jhagjhoorsingh S, Willems B, Zekveld MJ, Dreja H, O'Sullivan E, Pade C, Orkin C, Jeffs SA,

Montefiori DC, Davis D, Weissenhorn W, McKnight Á, Heeney JL, Sallusto F, Sattentau QJ, Weiss RA, Lanzavecchia A. 2010. Analysis of Memory B Cell Responses and Isolation of Novel Monoclonal Antibodies with Neutralizing Breadth from HIV-1-Infected Individuals. PLoS One 5:e8805.

203. Zhou T, Georgiev I, Wu X, Yang ZY, Dai K, Finzi A, Kwon YD, Scheid JF, Shi W, Xu L, Yang Y, Zhu J, Nussenzweig MC, Sodroski J, Shapiro L, Nabel GJ, Mascola JR, Kwong PD, Kwon D, Scheid JF, Shi W, Xu L, Yang Y, Zhu J, Nussenzweig MC, Sodroski J, Shapiro L, Nabel GJ, Mascola JR, Kwong PD. 2010. Structural basis for broad and potent neutralization of HIV-1 by antibody VRC01. Science (80-)2010/07/10. 329:811–817.
204. Wu X, Zhou T, Zhu J, Zhang B, Georgiev I, Wang C, Chen X, Longo NS, Louder M, McKee K. 2011. Focused evolution of HIV-1 neutralizing antibodies revealed by structures and deep sequencing. Science (80-) 333:1593–1602.
205. Bonsignori M, Wiehe K, Grimm SK, Lynch R, Yang G, Kozink DM, Perrin F, Cooper AJ, Hwang KK, Chen X, Liu M, McKee K, Parks RJ, Eudailey J, Wang M, Clowse M, Criscione-Schreiber LG, Moody MA, Ackerman ME, Boyd SD, Gao F, Kelsoe G, Verkoczy L, Tomaras GD, Liao HX, Kepler TB, Montefiori DC, Mascola JR, Haynes BF. 2014. An autoreactive antibody from an SLE/HIV-1 individual broadly neutralizes HIV-1. J Clin Invest 124:1835–1843.
206. Gao F, Bonsignori M, Liao H, Kumar A, Xia S, Lu X, Cai F, Hwang K, Song H, Zhou T, Lynch RM, Alam SM, Moody MA, Ferrari G, Berrong M, Kelsoe G, Shaw GM, Hahn BH, Montefiori DC, Kamanga G, Cohen MS, Hraber P, Kwong PD, Korber BT, Mascola JR, Kepler TB, Haynes BF. 2014. Cooperation of B cell lineages in induction of HIV-1-broadly neutralizing antibodies. Cell 158:481–491.
207. Bonsignori M, Zhou T, Sheng Z, Chen L, Gao F, Joyce MG, Ozorowski G, Chuang GY, Schramm CA, Wiehe K, Alam SM, Bradley T, Gladden MA, Hwang KK, Iyengar S, Kumar A, Lu X, Luo K, Mangiapani MC, Parks RJ, Song H, Acharya P, Bailer RT, Cao A, Druz A, Georgiev IS, Kwon YD, Louder MK, Zhang B, Zheng A, Hill BJ, Kong R, Soto C, Mullikin JC, Douek DC, Montefiori DC, Moody MA, Shaw GM, Hahn BH, Kelsoe G, Hraber PT, Korber BT, Boyd SD, Fire AZ, Kepler TB, Shapiro L, Ward AB, Mascola JR, Liao HX, Kwong PD, Haynes BF. 2016. Maturation Pathway from Germline to Broad HIV-1 Neutralizer of a CD4-Mimic Antibody. Cell 165:449–463.
208. Wu X, Yang Z-Y, Li Y, Hogerkorp C-M, Schief WR, Seaman MS, Zhou T,

Schmidt SD, Wu L, Xu L, Longo NS, Mckee K, Dell SO, Louder MK, Diane L, Feng Y, Nason M, Doria-rose N, Connors M, Kwong PD, Roederer M, Richard T, Nabel GJ, Mascola JR. 2010. Rational design of envelope identifies broadly neutralizing human monoclonal antibodies to HIV-1. Science (80-) 329:856–861.

209. Zhou T, Lynch RM, Chen L, Acharya P, Wu X, Doria-Rose NA, Joyce MG, Lingwood D, Soto C, Bailer RT, Ernandes MJ, Kong R, Longo NS, Louder MK, McKee K, O'Dell S, Schmidt SD, Tran L, Yang Z, Druz A, Luongo TS, Moquin S, Srivatsan S, Yang Y, Zhang B, Zheng A, Pancera M, Kirys T, Georgiev IS, Gindin T, Peng HP, Yang AS, Mullikin JC, Gray MD, Stamatatos L, Burton DR, Koff WC, Cohen MS, Haynes BF, Casazza JP, Connors M, Corti D, Lanzavecchia A, Sattentau QJ, Weiss RA, West AP, Bjorkman PJ, Scheid JF, Nussenzweig MC, Shapiro L, Mascola JR, Kwong PD, Program NCS, Mullikin JC, Gray MD, Stamatatos L, Burton DR, Koff WC, Cohen MS, Haynes BF, Casazza JP, Connors M, Corti D, Lanzavecchia A, Sattentau QJ, Weiss RA, West Jr. AP, Bjorkman PJ, Scheid JF, Nussenzweig MC, Shapiro L, Mascola JR, Kwong PD. 2015. Structural Repertoire of HIV-1-Neutralizing Antibodies Targeting the CD4 Supersite in 14 Donors. Cell 161:1280–1292.
210. Mascola JR, Haynes BF. 2013. HIV-1 neutralizing antibodies : understanding nature ' s pathways. Immunol Rev 254:225–244.
211. Kwong PD, Mascola JR. 2012. Human antibodies that neutralize HIV-1: identification, structures, and B cell ontogenies. Immunity 37:412–425.
212. Lee JH, Leaman DP, Kim AS, Torrents de la Peña A, Sliepen K, Yasmeen A, Derking R, Ramos A, de Taeye SW, Ozorowski G, Klein F, Burton DR, Nussenzweig MC, Poignard P, Moore JP, Klasse PJ, Sanders RW, Zwick MB, Wilson IA, Ward AB. 2015. Antibodies to a conformational epitope on gp41 neutralize HIV-1 by destabilizing the Env spike. Nat Commun 6:8167.
213. Van Gils MJ, Van Den Kerkhof TLGM, Ozorowski G, Cottrell CA, Sok D, Pauthner M, Pallesen J, De Val N, Yasmeen A, De Taeye SW, Schorcht A, Gumbs S, Johanna I, Saye-Francisco K, Liang CH, Landais E, Nie X, Pritchard LK, Crispin M, Kelsoe G, Wilson IA, Schuitemaker H, Klasse PJ, Moore JP, Burton DR, Ward AB, Sanders RW. 2016. An HIV-1 antibody from an elite neutralizer implicates the fusion peptide as a site of vulnerability. Nat Microbiol 2:423–426.
214. McCoy LE, Burton DR. 2017. Identification and specificity of broadly

neutralizing antibodies against HIV. Immunol Rev 275:11–20.

215. Falkowska E, Le KM, Ramos A, Doores KJ, Lee JH, Blattner C, Ramirez A, Derking R, van Gils MJ, Liang CH, McBride R, von Bredow B, Shivatare SS, Wu CY, Chan-Hui PY, Liu Y, Feizi T, Zwick MB, Koff WC, Seaman MS, Swiderek K, Moore JP, Evans D, Paulson JC, Wong CH, Ward AB, Wilson IA, Sanders RW, Poignard P, Burton DR. 2014. Broadly neutralizing HIV antibodies define a glycan-dependent epitope on the prefusion conformation of gp41 on cleaved envelope trimers. Immunity 40:657–668.
216. Huang J, Kang BH, Pancera M, Lee JH, Tong T, Feng Y, Imamichi H, Georgiev IS, Chuang GY, Druz A, Doria-Rose NA, Laub L, Sliepen K, Van Gils MJ, De La Peña AT, Derking R, Klasse PJ, Migueles SA, Bailer RT, Alam M, Pugach P, Haynes BF, Wyatt RT, Sanders RW, Binley JM, Ward AB, Mascola JR, Kwong PD, Connors M. 2014. Broad and potent HIV-1 neutralization by a human antibody that binds the gp41-gp120 interface. Nature 515:138–142.
217. Scharf L, Scheid JF, Lee JH, West AP, Chen C, Gao H, Gnanapragasam PNP, Mares R, Seaman MS, Ward AB, Nussenzweig MC, Bjorkman PJ. 2014. Antibody 8ANC195 Reveals a Site of Broad Vulnerability on the HIV-1 Envelope Spike. Cell Rep 7:785–795.
218. Scharf L, Wang H, Gao H, Chen S, McDowall AW, Bjorkman PJ. 2015. Broadly neutralizing antibody 8ANC195 recognizes closed and open states of HIV-1 Env. Cell 162:1379–1390.
219. Wibmer CK, Gorman J, Anthony CS, Mkhize NN, Druz A, York T, Schmidt SD, Labuschagne P, Louder MK, Bailer RT. 2016. Structure of an N276-dependent HIV-1 neutralizing antibody targeting a rare V5 glycan hole adjacent to the CD4 binding site. J Virol 90:10220–10235.
220. Salzwedel K, West JT, Hunter E. 1999. A conserved tryptophan-rich motif in the membrane-proximal region of the human immunodeficiency virus type 1 gp41 ectodomain is important for Env-mediated fusion and virus infectivity. J Virol 73:2469–2480.
221. Zwick MB. 2005. The membrane-proximal external region of HIV-1 gp41: a vaccine target worth exploring. AIDS 19:1725–1737.
222. Gray ES, Moore PL, Bibollet-Ruche F, Li H, Decker JM, Meyers T, Shaw GM, Morris L. 2008. 4E10-resistant variants in a Human Immunodeficiency Virus type 1 subtype C-infected individual with an anti-membrane-proximal external region-neutralizing antibody response. J Virol 82:2367–2375.

223. Matyas GR, Wieczorek L, Beck Z, Ochsenbauer-Jambor C, Kappes JC, Michael NL, Polonis VR, Alving CR. 2009. Neutralizing antibodies induced by liposomal HIV-1 glycoprotein 41 peptide simultaneously bind to both the 2F5 or 4E10 epitope and lipid epitopes. AIDS 23:2069–2077.
224. Hessell AJ, Rakasz EG, Tehrani DM, Huber M, Weisgrau KL, Landucci G, Forthal DN, Koff WC, Poignard P, Watkins DI, Burton DR. 2010. Broadly neutralizing monoclonal antibodies 2F5 and 4E10 directed against the human immunodeficiency virus type 1 gp41 membrane-proximal external region protect against mucosal challenge by simian-human immunodeficiency virus SHIVBa-L. J Virol 84:1302–1313.
225. Georgiev IS, Rudicell RS, Saunders KO, Shi W, Kirys T, Mckee K, O'Dell S, Chuang G-Y, Yang Z-Y, Ofek G, Connors M, Mascola JR, Nabel GJ, Kwong PD, Dell SO, Chuang G-Y, Yang Z-Y, Ofek G, Mascola JR, Nabel GJ, Peter D, Mckee K, Dell SO, Chuang G-Y, Yang Z-Y, Ofek G. 2014. Antibodies VRC01 and 10E8 Neutralize HIV-1 with High Breadth and Potency Even with Ig-Framework Regions Substantially Reverted to Germline. J Immunol 192:1100–1106.
226. Huang J, Ofek G, Laub L, Louder MK, Doria-rose NA, Longo NS, Imamichi H, Bailer RT, Chakrabarti B, Sharma SK, Alam SM, Wang T, Yang Y, Zhang B, Migueles SA, Wyatt R, Haynes BF, Kwong PD, Mascola JR, Connors M. 2012. Broad and potent neutralization of HIV-1 by a gp41-specific human antibody. Nature 491:406–412.
227. Haynes BF, Fleming J, St. Clair EW, Katinger H, Stiegler G, Kunert R, Robinson J, Scearce RM, Plonk K, Staats HF, Ortel TL, Liao H-X, Alam SM, Clair EWS, Katinger H, Stiegler G, Kunert R, Robinson J, Scearce RM, Plonk K, Staats HF, Ortel TL, Liao H-X, Alam SM. 2005. Cardiolipin Polyspecific Autoreactivity in Two Broadly Neutralizing HIV-1 Antibodies. Science (80-) 308:1906–1908.
228. Frey G, Peng H, Rits-Volloch S, Morelli M, Cheng Y, Chen B. 2008. A fusion-intermediate state of HIV-1 gp41 targeted by broadly neutralizing antibodies. Proc Natl Acad Sci U S A 105:3739–3744.
229. Moldt B, Rakasz EG, Schultz N, Chan-Hui P-Y, Swiderek K, Weisgrau KL, Piaskowski SM, Bergman Z, Watkins DI, Poignard P. 2012. Highly potent HIV-specific antibody neutralization in vitro translates into effective protection against mucosal SHIV challenge in vivo. Proc Natl Acad Sci 109:18921–18925.

230. Juan D De, Pazos F, Valencia A, de Juan D, Pazos F, Valencia A. 2013. Emerging methods in protein co-evolution. Nat Rev Genet 14:249–261.
231. Liao H-X, Lynch R, Zhou T, Gao F, Alam SM, Boyd SD, Fire AZ, Roskin KM, Schramm CA, Zhang Z, Zhu J, Shapiro L, Becker J, Benjamin B, Blakesley R, Bouffard G, Brooks S, Coleman H, Dekhtyar M, Gregory M, Guan X, Gupta J, Han J, Hargrove A, Ho S, Johnson T, Legaspi R, Lovett S, Maduro Q, Masiello C, Maskeri B, McDowell J, Montemayor C, Mullikin J, Park M, Riebow N, Schandler K, Schmidt B, Sison C, Stantripop M, Thomas J, Thomas P, Vemulapalli M, Young A, Mullikin JC, Gnanakaran S, Hraber P, Wiehe K, Kelsoe G, Yang G, Xia S-M, Montefiori DC, Parks R, Lloyd KE, Scearce RM, Soderberg KA, Cohen M, Kamanga G, Louder MK, Tran LM, Chen Y, Cai F, Chen S, Moquin S, Du X, Joyce MG, Srivatsan S, Zhang B, Zheng A, Shaw GM, Hahn BH, Kepler TB, Korber BTMM, Kwong PD, Mascola JR, Haynes BF. 2013. Co-evolution of a broadly neutralizing HIV-1 antibody and founder virus. Nature 496:469–476.
232. Wu X, Zhang Z, Schramm CA, Joyce MG, Do Kwon Y, Zhou T, Sheng Z, Zhang B, O'Dell S, McKee K, Georgiev IS, Chuang G-Y, Longo NS, Lynch RM, Saunders KO, Soto C, Srivatsan S, Yang Y, Bailer RT, Louder MK, Mullikin JC, Connors M, Kwong PD, Mascola JR, Shapiro L, Benjamin B, Blakesley R, Bouffard G, Brooks S, Coleman H, Dekhtyar M, Gregory M, Guan X, Gupta J, Han J, Hargrove A, Ho S, Legaspi R, Maduro Q, Masiello C, Maskeri B, McDowell J, Montemayor C, Park M, Riebow N, Schandler K, Schmidt B, Sison C, Stantripop M, Thomas J, Thomas P, Vemulapalli M, Young A. 2015. Maturation and diversity of the VRC01-antibody lineage over 15 years of chronic HIV-1 infection. Cell 161:470–485.
233. Soto C, Ofek G, Joyce MG, Zhang B, McKee K, Longo NS, Yang Y, Huang J, Parks R, Eudailey J, Lloyd KE, Alam SM, Haynes BF, Mullikin JC, Connors M, Mascola JR, Shapiro L, Kwong PD, NISC Comparative Sequencing Program. 2016. Developmental pathway of the MPER-Directed HIV-1-Neutralizing antibody 10E8. PLoS One 11.
234. Mabvakure BM, Scheepers C, Garrett N, Abdool Karim S, Williamson C, Morris L, Moore PL. 2018. Positive selection at key residues in the HIV Envelope distinguishes broad and strain-specific plasma neutralizing antibodies. J Virol.
235. Grenfell BT, Pybus OG, Gog JR, Wood JLN, Daly JM, Mumford JA, Holmes EC. 2004. Unifying the epidemiological and evolutionary dynamics of

pathogens. Science (80-) 303:327–332.
236. Holmes EC, Grenfell BT. 2009. Discovering the phylodynamics of RNA viruses. PLoS Comput Biol 5:e1000505.
237. Azarian T, Grant LR, Arnold BJ, Hammitt LL, Reid R, Santosham M, Weatherholtz R, Goklish N, Thompson CM, Bentley SD, O'Brien KL, Hanage WP, Lipsitch M. 2018. The impact of serotype-specific vaccination on phylodynamic parameters of Streptococcus pneumoniae and the pneumococcal pan-genome. PLOS Pathog 14:e1006966.
238. Drummond AJ, Rambaut A. 2007. BEAST: Bayesian evolutionary analysis by sampling trees. BMC Evol Biol 7:214.
239. Suchard MA, Rambaut A. 2009. Many-core algorithms for statistical phylogenetics. Bioinformatics 25:1370–1376.
240. Moore PL, Williamson C. 2016. Approaches to the induction of HIV broadly neutralizing antibodies. Curr Opin HIV AIDS 11:569–575.
241. Pegu A, Hessell AJ, Mascola JR, Haigwood NL. 2017. Use of broadly neutralizing antibodies for HIV-1 prevention. Immunol Rev 275:296–312.
242. Hraber P, Seaman MS, Bailer RT, Mascola JR, Montefiori DC, Korber BT. 2014. Prevalence of broadly neutralizing antibody responses during chronic HIV-1 infection. AIDS 28:163–169.
243. Piantadosi A, Chohan B, Panteleeff D, Baeten JM, Mandaliya K, Ndinya-Achola JO, Overbaugh J. 2009. HIV-1 evolution in gag and env is highly correlated but exhibits different relationships with viral load and the immune response. AIDS 23:579–587.
244. Euler Z, van Gils MJ, Bunnik EM, Phung P, Schweighardt B, Wrin T, Schuitemaker H. 2010. Cross-reactive neutralizing humoral immunity does not protect from HIV type 1 disease progression. J Infect Dis 201:1045–1053.
245. Cohen K, Altfeld M, Alter G, Stamatatos L. 2014. Early preservation of CXCR5+ PD-1+ helper T cells and B cell activation predict the breadth of neutralizing antibody responses in chronic HIV-1 infection. J Virol 88:13310–13321.
246. Havenar-Daughton C, Lindqvist M, Heit A, Wu JE, Reiss SM, Kendric K, Bélanger S, Kasturi SP, Landais E, Akondy RS, McGuire HM, Bothwell M, Vagefi PA, Scully E, Tomaras GD, Davis MM, Poignard P, Ahmed R, Walker BD, Pulendran B, McElrath MJ, Kaufmann DE, Crotty S. 2016. CXCL13 is a plasma biomarker of germinal center activity. Proc Natl Acad Sci 113:2702–

2707.

247. Piantadosi A, Panteleeff D, Blish CA, Baeten JM, Jaoko W, McClelland RS, Overbaugh J. 2009. Breadth of neutralizing antibody response to human immunodeficiency virus type 1 is affected by factors early in infection but does not influence disease progression. J Virol 83:10269–10274.
248. Euler Z, van den Kerkhof TLGM, van Gils MJ, Burger JA, Edo-Matas D, Phung P, Wrin T, Schuitemaker H. 2012. Longitudinal analysis of early HIV-1-specific neutralizing activity in an elite neutralizer and in five patients who developed cross-reactive neutralizing activity. J Virol 86:2045–2055.
249. Murphy MK, Yue L, Pan R, Boliar S, Sethi A, Tian J, Pfafferot K, Karita E, Allen SA, Cormier E, Goepfert PA, Borrow P, Robinson JE, Gnanakaran S, Hunter E, Kong XP, Derdeyn CA. 2013. Viral escape from neutralizing antibodies in early subtype A HIV-1 infection drives an increase in autologous neutralization breadth. PLoS Pathog 9:e1003173.
250. Moore PL, Gray ES, Wibmer CK, Bhiman JN, Nonyane M, Sheward DJ, Hermanus T, Bajimaya S, Tumba NL, Abrahams M-R, Lambson BE, Ranchobe N, Ping L, Ngandu N, Karim QA, Karim SSA, Swanstrom RI, Seaman MS, Williamson C, Morris L. 2012. Evolution of an HIV glycan–dependent broadly neutralizing antibody epitope through immune escape. Nat Med 18:1688–1692.
251. Haynes BF, Shaw GM, Korber B, Kelsoe G, Sodroski J, Hahn BH, Borrow P, McMichael AJ. 2016. HIV-Host interactions: Implications for vaccine design. Cell Host Microbe 19:292–303.
252. Escolano A, Steichen JM, Dosenovic P, Burton DR, Schief WR, Nussenzweig MC, Escolano A, Steichen JM, Dosenovic P, Kulp DW, Golijanin J, Sok D. 2016. Sequential Immunization Elicits Broadly Neutralizing Anti-HIV-1 Antibodies in Ig Knockin Mice Article Sequential Immunization Elicits Broadly Neutralizing Anti-HIV-1 Antibodies in Ig Knockin Mice. Cell 166:1445–1458.e12.
253. Dosenovic P, von Boehmer L, Escolano A, Jardine J, Freund NT, Gitlin AD, McGuire AT, Kulp DW, Oliveira T, Scharf L, Pietzsch J, Gray MD, Cupo A, van Gils MJ, Yao KH, Liu C, Gazumyan A, Seaman MS, Bjorkman PJ, Sanders RW, Moore JP, Stamatatos L, Schief WR, Nussenzweig MC, Björkman PJ, Sanders RW, Moore JP, Stamatatos L, Schief WR, Nussenzweig MC. 2015. Immunization for HIV-1 broadly neutralizing antibodies in human Ig knockin mice. Cell 161:1505–1515.

254. Briney B, Sok D, Jardine JG, Kulp DW, Skog P, Menis S, Jacak R, Kalyuzhniy O, de Val N, Sesterhenn F, Le KM, Ramos A, Jones M, Saye-Francisco KL, Blane TR, Spencer S, Georgeson E, Hu X, Ozorowski G, Adachi Y, Kubitz M, Sarkar A, Wilson IA, Ward AB, Nemazee D, Burton DR, Schief WR. 2016. Tailored immunogens direct affinity maturation toward HIV neutralizing antibodies. Cell 166:1459–1470.e11.
255. Sok D, Briney B, Jardine JG, Kulp DW, Menis S, Pauthner M, Wood A, Lee EC, Le KM, Jones M, Ramos A, Kalyuzhniy O, Adachi Y, Kubitz M, MacPherson S, Bradley A, Friedrich GA, Schief WR, Burton DR. 2016. Priming HIV-1 broadly neutralizing antibody precursors in human Ig loci transgenic mice. Science (80-) 353:1557–1560.
256. Tian M, Cheng C, Chen X, Duan H, Cheng HL, Dao M, Sheng Z, Kimble M, Wang L, Lin S, Schmidt SD, Du Z, Joyce MG, Chen Y, DeKosky BJ, Chen Y, Normandin E, Cantor E, Chen RE, Doria-Rose NA, Zhang Y, Shi W, Kong WP, Choe M, Henry AR, Laboune F, Georgiev IS, Huang PY, Jain S, McGuire AT, Georgeson E, Menis S, Douek DC, Schief WR, Stamatatos L, Kwong PD, Shapiro L, Haynes BF, Mascola JR, Alt FW. 2016. Induction of HIV neutralizing antibody lineages in mice with diverse precursor repertoires. Cell 166:1471–1484.e18.
257. Gray ES, Moore PL, Choge IA, Decker JM, Bibollet-Ruche F, Li H, Leseka N, Treurnicht F, Mlisana K, Shaw GM, Karim SSA, Williamson C, Morris L. 2007. Neutralizing antibody responses in acute Human Immunodeficiency Virus type 1 subtype C infection. J Virol 81:6187–6196.
258. Li B, Decker JM, Johnson RW, Bibollet-Ruche F, Wei X, Mulenga J, Allen S, Hunter E, Hahn BH, Shaw GM, Blackwell JL, Derdeyn CA. 2006. Evidence for potent autologous neutralizing antibody titers and compact envelopes in early infection with subtype C human immunodeficiency virus type 1. J Virol 80:5211–8.
259. Anthony C, York T, Bekker V, Matten D, Selhorst P, Ferreria R-C, Garrett NJ, Karim SSA, Morris L, Wood NT, Moore PL, Williamson C. 2017. Cooperation between strain-specific and broadly neutralizing responses limited viral escape and prolonged the exposure of the broadly neutralizing epitope. J Virol 91.
260. Sagar M, Wu X, Lee S, Overbaugh J. 2006. Human Immunodeficiency Virus type 1 V1-V2 envelope loop sequences expand and Add glycosylation sites over the course of infection, and these modifications affect antibody

neutralization sensitivity. J Virol 80:9586–9598.
261. Rong R, Li B, Lynch RM, Haaland RE, Murphy MK, Mulenga J, Allen SA, Pinter A, Shaw GM, Hunter E, Robinson JE, Gnanakaran S, Derdeyn CA. 2009. Escape from autologous neutralizing antibodies in acute/early subtype C HIV-1 infection requires multiple pathways. PLoS Pathog 5:e1000594.
262. van Loggerenberg F, Mlisana K, Williamson C, Auld SC, Morris L, Gray CM, Abdool Karim Q, Grobler A, Barnabas N, Iriogbe I, Abdool Karim SS. 2008. Establishing a cohort at high risk of HIV infection in South Africa: Challenges and experiences of the CAPRISA 002 acute infection study. PLoS One 3:e1954.
263. Karim QA, Kharsany AB, Frohlich JA, Baxter C, Yende N, Mansoor LE, Mlisana KP, Maarschalk S, Arulappan N, Grobler A, Sibeko S, Omar Z, Gengiah TN, Mlotshwa M, Samsunder N, Abdool Karim SS. 2011. Recruitment of high risk women for HIV prevention trials: baseline HIV prevalence and sexual behavior in the CAPRISA 004 tenofovir gel trial. Trials 12:67.
264. Salazar-Gonzalez JF, Bailes E, Pham KT, Salazar MG, Guffey MB, Keele BF, Derdeyn CA, Farmer P, Hunter E, Allen S, Manigart O, Mulenga J, Anderson JA, Swanstrom R, Haynes BF, Athreya GS, Korber BTM, Sharp PM, Shaw GM, Hahn BH. 2008. Deciphering Human Immunodeficiency Virus type 1 transmission and early envelope diversification by single-genome amplification and sequencing. J Virol 82:3952–70.
265. Edgar RC. 2004. MUSCLE: multiple sequence alignment with high accuracy and high throughput. Nucleic Acids Res 32:1792–1797.
266. Hall TA. 1999. BioEdit: a user-friendly biological sequence alignment editor and analysis program for Windows 95/98/NT, p. 95–98. *In* Nucleic acids symposium series.
267. Rambaut A. 2007. Se-Al v. 2.0. Website http//tree bio ed ac uk/software/seal.
268. Tamura K, Stecher G, Peterson D, Filipski A, Kumar S. 2013. MEGA6: Molecular Evolutionary Genetics Analysis Version 6.0. Mol Biol Evol 30:2725–2729.
269. Thomas RH. 2001. Molecular evolution and phylogenetics. Heredity (Edinb) 86:385.
270. Rambaut A, Lam TT, Max Carvalho L, Pybus OG. 2016. Exploring the temporal structure of heterochronous sequences using TempEst (formerly Path-O-Gen). Virus Evol 2:vew007.

271. Guindon S, Delsuc F, Dufayard J-F, Gascuel O. 2009. Estimating Maximum Likelihood Phylogenies with PhyML, p. 113–137. *In* Methods in Molecular Biology.
272. Drummond AJ, Suchard MA, Xie D, Rambaut A. 2012. Bayesian phylogenetics with BEAUti and the BEAST 1.7. Mol Biol Evol 29:1969–1973.
273. Baele G, Lemey P, Bedford T, Rambaut A, Suchard MA, Alekseyenko A V. 2012. Improving the accuracy of demographic and molecular clock model comparison while accommodating phylogenetic uncertainty. Mol Biol Evol 29:2157–67.
274. Baele G, Li WLS, Drummond AJ, Suchard MA, Lemey P. 2013. Accurate model selection of relaxed molecular clocks in bayesian phylogenetics. Mol Biol Evol 30:239–43.
275. Rambaut A, Drummond AJ. 2007. Tracer, version 1.5.
276. Rambaut A, Drummond A. 2009. FigTree v1. 3.1. Website http//tree bio ed ac uk/software/figtree.
277. Pond SLK, Muse S V. 2005. HyPhy: Hypothesis Testing Using Phylogenies, p. 125–181. *In* Statistical Methods in Molecular Evolution. Springer-Verlag, New York.
278. Murrell B, Moola S, Mabona A, Weighill T, Sheward D, Kosakovsky Pond SL, Scheffler K. 2013. FUBAR: A Fast, Unconstrained Bayesian AppRoximation for Inferring Selection. Mol Biol Evol 30:1196–1205.
279. Chaillon A, Samleerat T, Zoveda F, Ballesteros S, Moreau A, Ngo-Giang-Huong N, Jourdain G, Gianella S, Lallemant M, Depaulis F, Barin F. 2014. Estimating the timing of mother-to-child transmission of the Human Immunodeficiency Virus type 1 using a viral molecular evolution model. PLoS One 9:e90421.
280. Novitsky V, Wang R, Margolin L, Baca J, Rossenkhan R, Moyo S, van Widenfelt E, Essex M. 2011. Transmission of single and multiple viral variants in primary HIV-1 subtype C infection. PLoS One 6:e16714.
281. Abrahams MR, Anderson JA, Giorgi EE, Seoighe C, Mlisana K, Ping L-H, Athreya GS, Treurnicht FK, Keele BF, Wood N, Salazar-Gonzalez JF, Bhattacharya T, Chu H, Hoffman I, Galvin S, Mapanje C, Kazembe P, Thebus R, Fiscus S, Hide W, Cohen MS, Karim SA, Haynes BF, Shaw GM, Hahn BH, Korber BT, Swanstrom R, Williamson C. 2009. Quantitating the multiplicity of infection with Human Immunodeficiency Virus Type 1 subtype C reveals a non-

Poisson distribution of transmitted variants. J Virol 83:3556–3567.
282. Lemey P, Kosakovsky Pond SL, Drummond AJ, Pybus OG, Shapiro B, Barroso H, Taveira N, Rambaut A. 2007. Synonymous substitution rates predict HIV disease progression as a result of underlying replication dynamics. PLoS Comput Biol 3:e29.
283. Rong R, Bibollet-Ruche F, Mulenga J, Allen S, Blackwell JL, Derdeyn CA. 2007. Role of V1V2 and other Human Immunodeficiency Virus type 1 envelope domains in resistance to autologous neutralization during clade C infection. J Virol 81:1350–1359.
284. Phogat S, Walker L, Wagner D, Chan-Hui P, Simek M, Phung P, Wrin T, Mitcham J, Investigator P, Kaminsky S, Zamb T, Moyle M, Koff W, Burton D. 2009. Potent and broad neutralizing antibodies from HIV-1 non-clade B infected donor reveal a new HIV-1 vaccine target. Retrovirology 6:O6.
285. Sok D, Pauthner M, Briney B, Lee JH, Saye-Francisco KL, Hsueh J, Ramos A, Le KM, Jones M, Jardine JG, Bastidas R, Sarkar A, Liang CH, Shivatare SS, Wu CY, Schief WR, Wong CH, Wilson IA, Ward AB, Zhu J, Poignard P, Burton DR. 2016. A prominent site of antibody vulnerability on HIV envelope incorporates a motif associated with CCR5 binding and its camouflaging glycans. Immunity 45:31–45.
286. Sajadi MM, Dashti A, Rikhtegaran Tehrani Z, Tolbert WD, Seaman MS, Ouyang X, Gohain N, Pazgier M, Kim D, Cavet G, Yared J, Redfield RR, Lewis GK, DeVico AL. 2018. Identification of near-pan-neutralizing antibodies against HIV-1 by deconvolution of plasma humoral responses. Cell 173:1783–1795.e14.
287. Moore PL, Gorman J, Doria-Rose NA, Morris L, Doria-Rose NA, Morris L. 2017. Ontogeny-based immunogens for the induction of V2-directed HIV broadly neutralizing antibodies. Immunol Rev 275:217–229.
288. Sheward DJ, Marais J, Bekker V, Murrell B, Eren K, Bhiman JN, Nonyane M, Garrett N, Woodman ZL, Abdool Karim Q, Abdool Karim SS, Morris L, Moore PL, Williamson C. 2018. HIV superinfection drives de novo antibody responses and not neutralization breadth. Cell Host Microbe 24:593–599.e3.
289. Powell RLR, Kinge T, Nyambi PN. 2010. Infection by discordant strains of HIV-1 markedly enhances the neutralizing antibody response against heterologous virus. J Virol 84:9415–9426.
290. Cortez V, Odem-Davis K, McClelland RS, Jaoko W, Overbaugh J. 2012. HIV-1

superinfection in women broadens and strengthens the neutralizing antibody response. PLoS Pathog 8:1–11.

291. Doria-Rose NA, Bhiman JN, Roark RS, Schramm CA, Gorman J, Chuang G, Pancera M, Cale EM, Ernandes MJ, Louder MK, Asokan M, Bailer RT, Druz A, Fraschilla IR, Garrett NJ, Jarosinski M, Lynch RM, McKee K, O'Dell S, Pegu A, Schmidt SD, Staupe RP, Sutton MS, Wang K, Wibmer CK, Haynes BF, Abdool-Karim S, Shapiro L, Kwong PD, Moore PL, Morris L, Mascola JR. 2016. New member of the V1V2-directed CAP256-VRC26 lineage that shows increased breadth and exceptional potency. J Virol 90:76–91.
292. Landais E, Murrell B, Briney B, Murrell S, Rantalainen K, Berndsen ZT, Ramos A, Wickramasinghe L, Smith ML, Eren K, de Val N, Wu M, Cappelletti A, Umotoy J, Lie Y, Wrin T, Algate P, Chan-Hui P-Y, Karita E, Ward AB, Wilson IA, Burton DR, Smith D, Pond SLK, Poignard P. 2017. HIV envelope glycoform heterogeneity and localized diversity govern the initiation and maturation of a V2 apex broadly neutralizing antibody lineage. Immunity 47:990–1003.e9.
293. Moore PL, Gray ES, Morris L. 2009. Specificity of the autologous neutralizing antibody response. Curr Opin HIV AIDS 4:358–363.
294. Walker LM, Phogat SK, Chan-Hui P-Y, Wagner D, Phung P, Goss JL, Wrin T, Simek MD, Fling S, Mitcham JL, Lehrman JK, Priddy FH, Olsen OA, Frey SM, Hammond PW, Kaminsky S, Zamb T, Moyle M, Koff WC, Poignard P, Burton DR. 2009. Broad and potent neutralizing antibodies from an African donor reveal a new HIV-1 vaccine target. Science (80-) 326:285–289.
295. Walker LM, Huber M, Doores KJ, Falkowska E, Pejchal R, Julien JP, Wang SK, Ramos A, Chan-Hui PY, Moyle M, Mitcham JL, Hammond PW, Olsen OA, Phung P, Fling S, Wong CH, Phogat S, Wrin T, Simek MD, Koff WC, Wilson IA, Burton DR, Poignard P, Protocol GPI, Koff WC, Wilson IA, Burton DR, Poignard P. 2011. Broad neutralization coverage of HIV by multiple highly potent antibodies. Nature 477:466–470.
296. McLellan JS, Pancera M, Carrico C, Gorman J, Julien JP, Khayat R, Louder R, Pejchal R, Sastry M, Dai K, O'Dell S, Patel N, Shahzad-ul-Hussan S, Yang Y, Zhang B, Zhou T, Zhu J, Boyington JC, Chuang GY, Diwanji D, Georgiev I, Do Kwon Y, Lee D, Louder MK, Moquin S, Schmidt SD, Yang ZY, Bonsignori M, Crump JA, Kapiga SH, Sam NE, Haynes BF, Burton DR, Koff WC, Walker LM, Phogat S, Wyatt R, Orwenyo J, Wang LX, Arthos J, Bewley CA, Mascola JR, Nabel GJ, Schief WR, Ward AB, Wilson IA, Kwong PD, Kwon YD, Lee D,

Louder MK, Moquin S, Schmidt SD, Yang ZY, Bonsignori M, Crump JA, Kapiga SH, Sam NE, Haynes BF, Burton DR, Koff WC, Walker LM, Phogat S, Wyatt R, Orwenyo J, Wang LX, Arthos J, Bewley CA, Mascola JR, Nabel GJ, Schief WR, Ward AB, Wilson IA, Kwong PD. 2011. Structure of HIV-1 gp120 V1/V2 domain with broadly neutralizing antibody PG9. Nature 480:336–343.

297. UNAIDS. 2018. Global HIV & AIDS statistics — 2018 fact sheet | UNAIDS.
298. Gantner P, Ghosn J. 2018. Genital reservoir: A barrier to functional cure? Curr Opin HIV AIDS 13:1.
299. Frost SD, Dumaurier MJ, Wain-Hobson S, Brown AJ. 2001. Genetic drift and within-host metapopulation dynamics of HIV-1 infection. Proc Natl Acad Sci U S A 98:6975–80.
300. Brown RJP, Peters PJ, Caron C, Gonzalez-Perez MP, Stones L, Ankghuambom C, Pondei K, McClure CP, Alemnji G, Taylor S, Sharp PM, Clapham PR, Ball JK. 2011. Intercompartmental recombination of HIV-1 contributes to env intrahost diversity and modulates viral tropism and sensitivity to entry inhibitors. J Virol 85:6024–6037.
301. Zhu T, Wang N, Carr A, Nam DS, Moor-Jankowski R, Cooper DA, Ho DD. 1996. Genetic characterization of human immunodeficiency virus type 1 in blood and genital secretions: evidence for viral compartmentalization and selection during sexual transmission. J Virol 70:3098–107.
302. Byrn RA, Zhang D, Eyre R, McGowan K, Kiessling AA. 1997. HIV-1 in semen: an isolated virus reservoir. Lancet 350:1141.
303. Kariuki SM, Selhorst P, Ariën KK, Dorfman JR. 2017. The HIV-1 transmission bottleneck. Retrovirology 14:22.
304. Chomont N, Hocini H, Grésenguet G, Brochier C, Bouhlal H, Andréoletti L, Becquart P, Charpentier C, de Dieu Longo J, Si-Mohamed A, Kazatchkine MD, Bélec L. 2007. Early archives of genetically-restricted proviral DNA in the female genital tract after heterosexual transmission of HIV-1. AIDS 21:153–162.
305. Wright TC, Subbarao S, Ellerbrock T V, Lennox JL, Evans-Strickfaden T, Smith DG, Hart CE. 2001. Human immunodeficiency virus 1 expression in the female genital tract in association with cervical inflammation and ulceration. Am J Obstet Gynecol 184:279–285.
306. Ellerbrock T V., Lennox JL, Clancy KA, Schinazi RF, Wright TC, Pratt-Palmore

M, Evans-Strickfaden T, Schnell C, Pai R, Conley LJ, Parrish-Kohler EE, Bush TJ, Tatti K, Hart CE. 2001. Cellular replication of Human Immunodeficiency Virus Type 1 occurs in vaginal secretions. J Infect Dis 184:28–36.

307. De Pasquale MP, Leigh Brown AJ, Uvin SC, Allega-Ingersoll J, Caliendo AM, Sutton L, Donahue S, D'Aquila RT. 2003. Differences in HIV-1 pol sequences from female genital tract and blood during antiretroviral therapy. JAIDS J Acquir Immune Defic Syndr 34:37–44.
308. Tirado G, Jove G, Reyes E, Sepulveda G, Yamamura Y, Singh DP, Kumar A. 2005. Differential evolution of cell-associated virus in blood and genital tract of HIV-infected females undergoing HAART. Virology 334:299–305.
309. Adal M, Ayele W, Wolday D, Dagne K, Messele T, Tilahun T, Berkhout B, Mayaan S, Pollakis G, Dorigo-Zetsma W. 2005. Evidence of genetic variability of Human Immunodeficiency Virus Type 1 in plasma and cervicovaginal lavage in Ethiopian women seeking care for sexually transmitted infections. AIDS Res Hum Retroviruses 21:649–653.
310. Kemal KS, Burger H, Mayers D, Anastos K, Foley B, Kitchen C, Huggins P, Schroeder T, Picchio G, Back S, Gao W, Meyer III WA, Weiser B. 2007. HIV-1 drug resistance in variants from the female genital tract and plasma. J Infect Dis 195:535–545.
311. Klein K, Nickel G, Nankya I, Kyeyune F, Demers K, Ndashimye E, Kwok C, Chen P-L, Rwambuya S, Poon A, Munjoma M, Chipato T, Byamugisha J, Mugyenyi P, Salata RA, Morrison CS, Arts EJ. 2018. Higher sequence diversity in the vaginal tract than in blood at early HIV-1 infection. PLOS Pathog 14:e1006754.
312. Sievers F, Wilm A, Dineen D, Gibson TJ, Karplus K, Li W, Lopez R, McWilliam H, Remmert M, Soding J, Thompson JD, Higgins DG. 2014. Fast, scalable generation of high-quality protein multiple sequence alignments using Clustal Omega. Mol Syst Biol 7:539–539.
313. Martin DP, Murrell B, Golden M, Khoosal A, Muhire B. 2015. RDP4: Detection and analysis of recombination patterns in virus genomes. Virus Evol 1:vev003.
314. Xia X. 2018. DAMBE7: New and improved tools for data analysis in molecular biology and evolution. Mol Biol Evol 35:1550–1552.
315. Suchard MA, Lemey P, Baele G, Ayres DL, Drummond AJ, Rambaut A. 2018. Bayesian phylogenetic and phylodynamic data integration using BEAST 1.10. Virus Evol 4:1–5.

316. Parker J, Rambaut A, Pybus OG. 2008. Correlating viral phenotypes with phylogeny: Accounting for phylogenetic uncertainty. Infect Genet Evol 8:239–246.
317. Lemey P, Rambaut A, Drummond AJ, Suchard MA. 2009. Bayesian phylogeography finds its roots. PLoS Comput Biol 5:e1000520.
318. Kass RE, Raftery AE. 1995. Bayes Factors. J Am Stat Assoc 90:773–795.
319. Minin VN, Suchard MA. 2007. Counting labeled transitions in continuous-time Markov models of evolution. J Math Biol 56:391–412.
320. Murrell B, Wertheim JO, Moola S, Weighill T, Scheffler K, Kosakovsky Pond SL. 2012. Detecting Individual Sites Subject to Episodic Diversifying Selection. PLOS Genet 8:e1002764.
321. González-Alonso J, Mortensen SP, Dawson EA, Secher NH, Damsgaard R. 2006. Erythrocytes and the regulation of human skeletal muscle blood flow and oxygen delivery: role of erythrocyte count and oxygenation state of haemoglobin. J Physiol 572:295–305.
322. Roberts L, Passmore J-AS, Mlisana K, Williamson C, Little F, Bebell LM, Walzl G, Abrahams M-R, Woodman Z, Abdool Karim Q, Abdool Karim SS. 2012. Genital Tract Inflammation During Early HIV-1 Infection Predicts Higher Plasma Viral Load Set Point in Women. J Infect Dis 205:194–203.
323. Alonso A, Lasseigne BN, Williams K, Nielsen J, Ramaker RC, Hardigan AA, Johnston B, Roberts BS, Cooper SJ, Marsal S, Myers RM. 2017. aRNApipe: A balanced, efficient and distributed pipeline for processing RNA-seq data in high performance computing environments. Bioinformatics 33:btx023.
324. Ferreira VH, Nazli A, Khan G, Mian MF, Ashkar AA, Gray-Owen S, Kaul R, Kaushic C. 2011. Endometrial epithelial cell responses to coinfecting viral and bacterial pathogens in the genital tract can activate the HIV-1 LTR in an NFκB- and AP-1–dependent manner. J Infect Dis 204:299–308.
325. Deese J, Masson L, Miller W, Cohen M, Morrison C, Wang M, Ahmed K, Agot K, Crucitti T, Abdellati S, Van Damme L. 2015. Injectable progestin-only contraception is associated with increased levels of pro-inflammatory cytokines in the female genital tract. Am J Reprod Immunol 74:357–367.
326. Cohen CR, Lingappa JR, Baeten JM, Ngayo MO, Spiegel CA, Hong T, Donnell D, Celum C, Kapiga S, Delany S, Bukusi EA. 2012. Bacterial Vaginosis associated with increased risk of female-to-male HIV-1 transmission: A prospective cohort analysis among African couples. PLoS Med 9:e1001251.

327. Ghys PD, Fransen K, Diallo MO, Ettiègne-Traoré V, Coulibaly I-M, Yeboué KM, Kalish ML, Maurice C, Whitaker JP, Greenberg AE, Laga M. 1997. The associations between cervicovaginal HIV shedding, sexually transmitted diseases and immunosuppression in female sex workers in Abidjan, Côte d'Ivoire. AIDS 11:F85–F93.
328. McClelland RS, Wang CC, Overbaugh J, Richardson BA, Corey L, Ashley RL, Mandaliya K, Ndinya-Achola J, Bwayo JJ, Kreiss JK. 2002. Association between cervical shedding of herpes simplex virus and HIV-1. AIDS 16:2425–2430.
329. Gatski M, Martin DH, Theall K, Amedee A, Clark RA, Dumestre J, Chhabra P, Schmidt N, Kissinger P. 2011. Mycoplasma genitalium infection among HIV-positive women: prevalence, risk factors and association with vaginal shedding. Int J STD AIDS 22:155–159.
330. Manhart LE, Mostad SB, Baeten JM, Astete SG, Mandaliya K, Totten PA. 2008. High Mycoplasma genitalium organism burden Is associated with shedding of HIV-1 DNA from the cervix. J Infect Dis 197:733–736.
331. Cu-Uvin S, DeLong AK, Venkatesh KK, Hogan JW, Ingersoll J, Kurpewski J, De Pasquale MP, D'Aquila R, Caliendo AM. 2010. Genital tract HIV-1 RNA shedding among women with below detectable plasma viral load. AIDS 24:2489–2497.
332. Coombs RW, Wright DJ, Reichelderfer PS, Burns DN, Cohn J, Cu-Uvin S, Baron PA, Cohen MH, Landay AL, Lewis S, Kovacs A, Women's Health Study 001 Team. 2001. Variation of human immunodeficiency virus type 1 viral RNA levels in the female genital tract: implications for applying measurements to individual women. J Infect Dis 184:1187–91.
333. Homans J, Christensen S, Stiller T, Wang C-H, Mack W, Anastos K, Minkoff H, Young M, Greenblatt R, Cohen M, Strickler H, Karim R, Spencer LY, Operskalski E, Frederick T, Kovacs A. 2012. Permissive and protective factors associated with presence, level, and longitudinal pattern of cervicovaginal HIV shedding. J Acquir Immune Defic Syndr 60:99–110.
334. Hart CE, Lennox JL, Pratt-Palmore M, Wright TC, Schinazi RF, Evans-Strickfaden T, Bush TJ, Schnell C, Conley LJ, Clancy K a, Ellerbrock T V. 1999. Correlation of human immunodeficiency virus type 1 RNA levels in blood and the female genital tract. J Infect Dis 179:871–82.
335. Kovacs A, Wasserman SS, Burns D, Wright DJ, Cohn J, Landay A, Weber K,

Cohen M, Levine A, Minkoff H, Miotti P, Palefsky J, Young M, Reichelderfer P, DATRI Study Group, WIHS Study Group. 2001. Determinants of HIV-1 shedding in the genital tract of women. Lancet 358:1593–601.

336. Kovacs A, Chan LS, Chen ZC, Meyer WA, Muderspach L, Young M, Anastos K, Levine AM. 1999. HIV-1 RNA in plasma and genital tract secretions in women infected with HIV-1. J Acquir Immune Defic Syndr 22:124–31.
337. Andreoletti L, Gresenguet G, Chomont N, Matta M, Quiniou Y, Si-Mohamed A, Belec L. 2003. Comparison of washing and swabbing procedures for collecting genital fluids to assess shedding of Human Immunodeficiency Virus Type 1 (HIV-1) RNA in asymptomatic HIV-1-infected women. J Clin Microbiol 41:449–452.
338. Mitchell C, Hitti J, Paul K, Agnew K, Cohn SE, Luque AE, Coombs R. 2011. Cervicovaginal shedding of HIV Type 1 is related to genital tract inflammation independent of changes in vaginal microbiota. AIDS Res Hum Retroviruses 27:35–39.
339. Faria NR, Suchard MA, Rambaut A, Lemey P. 2011. Toward a quantitative understanding of viral phylogeography. Curr Opin Virol 1:423–429.
340. Blackard JT, Cohen DE, Mayer KH. 2002. Human Immunodeficiency Virus superinfection and recombination: Current state of knowledge and potential clinical consequences. Clin Infect Dis 34:1108–1114.
341. Beerenwinkel N, Günthard HF, Roth V, Metzner KJ. 2012. Challenges and opportunities in estimating viral genetic diversity from next-generation sequencing data. Front Microbiol 3:329.
342. DePristo MA, Banks E, Poplin R, Garimella K V., Maguire JR, Hartl C, Philippakis AA, del Angel G, Rivas MA, Hanna M, McKenna A, Fennell TJ, Kernytsky AM, Sivachenko AY, Cibulskis K, Gabriel SB, Altshuler D, Daly MJ. 2011. A framework for variation discovery and genotyping using next-generation DNA sequencing data. Nat Genet 43:491–8.
343. Mardis ER. 2008. The impact of next-generation sequencing technology on genetics. Trends Genet 24:133–141.
344. Li Y, Chen L. 2014. Big biological data: Challenges and opportunities. Genomics Proteomics Bioinformatics 12:187–189.
345. Shendure J, Ji H. 2008. Next-generation DNA sequencing. Nat Biotechnol 26:1135–45.
346. Oxford Nanopore T. 2018. MinION.

347. Kulkarni P, Frommolt P. 2017. Challenges in the setup of large-scale next-generation sequencing analysis workflows. Comput Struct Biotechnol J 15:471–477.
348. Schmidt B, Hildebrandt A. 2017. Next-generation sequencing: big data meets high performance computing. Drug Discov Today 22:712–717.
349. Landais E, Moore PL. 2018. Development of broadly neutralizing antibodies in HIV-1 infected elite neutralizers. Retrovirology 15:61.
350. van Eeden C, Wibmer CK, Scheepers C, Richardson SI, Nonyane M, Lambson B, Mkhize NN, Vijayakumar B, Sheng Z, Stanfield-Oakley S, Bhiman JN, Bekker V, Hermanus T, Mabvakure B, Ismail A, Moody MA, Wiehe K, Garrett N, Karim SA, Dirr H, Fernandes MA, Sayed Y, Shapiro L, Ferrari G, Haynes BF, Moore PL, Morris L, van Eeden KC; Scheepers, C; Nonyane, M; Lambson, B; Mkhize, N; Richardson, S; Vijayakumar, B; Sheng, Z; Stanfield-Oakley S; Bhiman, J; Bekker, V; Hermanus T; Mabvakure B; Ismail, A; Moody, AM; Wiehe, K; Garrett, N; Abdool Karim, S; Dirr, H; Fernandes, CW. 2018. V2-Directed Vaccine-like Antibodies from HIV-1 Infection Identify an Additional K169-Binding Light Chain Motif with Broad ADCC Activity. Cell Rep 25:3123–3135.e6.
351. Mabvakure BM, Lambson BE, Ramdayal K, Masson L, Kitchin D, Allam M, Karim Abdool S, Williamson C, Passmore J-A, Martin DP, Scheepers C, Moore PL, Harkins GW, Morris L. Evidence for both intermittent and persistent compartmentalization of HIV-1 in the female genital tract.
352. Schramm CA, Sheng Z, Zhang Z, Mascola JR, Kwong PD, Shapiro L. 2016. SONAR: A High-Throughput Pipeline for Inferring Antibody Ontogenies from Longitudinal Sequencing of B Cell Transcripts. Front Immunol 7:372.
353. Altschul SF, Gish W, Miller W, Myers EW, Lipman DJ, Stephen F. Altschul' Webb Miller2 Eugene W. Myers3 and David J. Lipmanl WG. 1990. Basic local alignment search tool. J Mol Biol, 1990th ed. 215:403–410.
354. Olsen GJ, Matsuda H, Hagstrom R, Overbeek R. 1994. fastDNAml: a tool for construction of phylogenetic trees of DNA sequences using maximum likelihood. Comput Appl Biosci CABIOS 10:41–48.
355. Wiehe K, Easterhoff D, Luo K, Nicely NI, Bradley T, Jaeger FH, Dennison SM, Zhang R, Lloyd KE, Stolarchuk C, Parks R, Sutherland LL, Scearce RM, Morris L, Kaewkungwal J, Nitayaphan S, Pitisuttithum P, Rerks-Ngarm S, Sinangil F, Phogat S, Michael NL, Kim JH, Kelsoe G, Montefiori DC, Tomaras GD,

Bonsignori M, Santra S, Kepler TB, Alam SM, Moody MA, Liao HX, Haynes BF. 2014. Antibody light-chain-restricted recognition of the site of immune pressure in the RV144 HIV-1 vaccine trial is phylogenetically conserved. Immunity 41:909–918.

356. Andrews S. 2010. FastQC A Quality Control tool for High Throughput Sequence Data. http://www.bioinformatics.babraham.ac.uk/projects/fastqc/.
357. Zhang J, Kobert K, Flouri T, Stamatakis A. 2014. PEAR: a fast and accurate Illumina Paired-End reAd mergeR. Bioinformatics 30:614–620.
358. Edgar RC. 2010. Search and clustering orders of magnitude faster than BLAST. Bioinformatics 26:2460–2461.
359. Hannon L. 2009. FASTX Toolkit.
360. Excler J-L, Ake J, Robb ML, Kim JH, Plotkin SA. 2014. Nonneutralizing Functional Antibodies: a New “Old” Paradigm for HIV Vaccines. Clin Vaccine Immunol 21:1023–1036.
361. Corey L, Gilbert PB, Tomaras G, Haynes BF, Pantaleo G, Fauci AS. 2015. Immune Correlates of Vaccine Protection Against HIV-1 Acquisition: A Review. Sci Transl Med 7:310rv7-310rv7.
362. Smith SA, Burton SL, Kilembe W, Lakhi S, Karita E, Price M, Allen S, Hunter E, Derdeyn CA. 2016. Diversification in the HIV-1 Envelope Hyper-variable Domains V2, V4, and V5 and Higher Probability of Transmitted/Founder Envelope Glycosylation Favor the Development of Heterologous Neutralization Breadth. PLOS Pathog 12:e1005989.
363. Bekker L-G, Moodie Z, Grunenberg N, Laher F, Tomaras GD, Cohen KW, Allen M, Malahleha M, Mngadi K, Daniels B, Innes C, Bentley C, Frahm N, Morris DE, Morris L, Mkhize NN, Montefiori DC, Sarzotti-Kelsoe M, Grant S, Yu C, Mehra VL, Pensiero MN, Phogat S, DiazGranados CA, Barnett SW, Kanesa-thasan N, Koutsoukos M, Michael NL, Robb ML, Kublin JG, Gilbert PB, Corey L, Gray GE, McElrath MJ. 2018. Subtype C ALVAC-HIV and bivalent subtype C gp120/MF59 HIV-1 vaccine in low-risk, HIV-uninfected, South African adults: a phase 1/2 trial. Lancet HIV 5:e366–e378.
364. Haynes BF. 2015. New approaches to HIV vaccine development. Curr Opin Immunol 35:39–47.
365. de Taeye SW, Moore JP, Sanders RW. 2016. HIV-1 envelope trimer design and immunization strategies to induce broadly neutralizing antibodies. Trends Immunol 37:221–232.

366. Sanders RW, Moore JP. 2017. Native-like Env trimers as a platform for HIV-1 vaccine design. Immunol Rev 275:161–182.
367. Kharsany ABM, Karim QA. 2016. HIV infection and AIDS in sub-Saharan Africa: Current status, challenges and opportunities. Open AIDS J 10:34–48.
368. Scheepers C, Shrestha RK, Lambson BE, Jackson KJL, Wright IA, Naicker D, Goosen M, Berrie L, Ismail A, Garrett N, Abdool Karim Q, Abdool Karim SS, Moore PL, Travers SA, Morris L. 2015. Ability to develop broadly neutralizing HIV-1 antibodies is not restricted by the germline Ig gene repertoire. J Immunol 194:4371–4378.
369. Jones B. 2015. The African Genome Variation Project. Nat Rev Genet 16:68–69.

7 Appendices

The following are included in the appendices:

Appendix A: Mabvakure *et al*, Journal of Virology, 2018 (Proofs)

Appendix B: Van Eeden *et al*, Cell Reports, 2018

Appendix C: Ethics Certificate

Appendix D: Supervisor's letter

Appendix E: Support letter -Publications included in the thesis

Appendix F: Certificate of submission- student

Appendix G: Certificate of submission- supervisor

Appendix H: Turn-it-it Report

7.1 Appendix A: Mabvakure et al, Journal of Virology, 2018 (Proofs)

AMERICAN SOCIETY FOR MICROBIOLOGY | Journal of Virology®

PATHOGENESIS AND IMMUNITY

Positive Selection at Key Residues in the HIV Envelope Distinguishes Broad and Strain-Specific Plasma Neutralizing Antibodies

Batsirai M. Mabvakure,[a,b] Cathrine Scheepers,[a,b] Nigel Garrett,[c] Salim Abdool Karim,[c,d] Carolyn Williamson,[c,e] Lynn Morris,[a,b,c] Penny L. Moore[a,b,c]

[a]Center for HIV and STIs, National Institute for Communicable Diseases, National Health Laboratory Service, Johannesburg, South Africa
[b]School of Pathology, Faculty of Health Sciences, University of the Witwatersrand, Johannesburg, South Africa
[c]Centre for the AIDS Programme of Research in South Africa (CAPRISA), KwaZulu Natal, South Africa
[d]Department of Epidemiology, Columbia University, New York, New York, USA
[e]Division of Medical Virology, Institute of Infectious Disease and Molecular Medicine, University of Cape Town, Cape Town, South Africa

ABSTRACT The development of HIV broadly neutralizing antibodies (bNAbs) has previously been shown to be associated with viral evolution and high levels of genetic diversity in the HIV envelope glycoprotein (Env) glycoprotein. However, few studies have examined Env evolution in those who fail to develop breadth in order to assess whether bNAbs result from distinct evolutionary pathways. We compared Env evolution in eight HIV-1-infected participants who developed bNAbs to six donors with similar viral loads who did not develop bNAbs over three years of infection. We focused on Env V1V2 and C3V4, as these are major targets for both strain-specific nAbs and bNAbs. Overall evolutionary rates (ranging from 9.92×10^{-3} to 4.1×10^{-2} substitutions/site/year) and viral diversity (from 1.1% to 6.5%) across Env, and within targeted epitopes, did not distinguish bNAb donors from non-bNAb donors. However, bNAb participants had more positively selected residues within epitopes than those without bNAbs, and several of these were common among bNAb donors. A comparison of the kinetics of strain-specific nAbs and bNAbs indicated that selection pressure at these residues increased with the onset of breadth. These data suggest that highly targeted viral evolution rather than overall envelope diversity is associated with neutralization breadth. The association of shared positively selected sites with the onset of breadth highlights the importance of diversity at specific positions in these epitopes for bNAb development, with implications for the development of sequential and cocktail immunization strategies.

IMPORTANCE Millions of people are still being infected with HIV decades since the first recognition of this virus. Currently, no vaccine is able to elicit bNAbs that will prevent infection by global HIV strains. Several studies have implicated HIV Env diversity in the development of breadth. However, Env evolution in individuals who fail to develop breadth despite mounting potent strain-specific neutralizing responses has not been well defined. Using longitudinal neutralization, epitope mapping, and sequence data from 14 participants, we found that overall measures of viral diversity were similar in all donors. However, the number of positively selected sites within Env epitopes was higher in bNAb participants than in strain-specific donors. We further identified common sites that were positively selected as bNAbs developed. These data indicate that while viral diversity is required for breadth, this should be highly targeted to specific residues to shape the elicitation of bNAbs by vaccination.

KEYWORDS broadly neutralizing antibodies, envelope glycoprotein, epitope diversity, HIV vaccine, HIV-1 intrapatient evolution, sequential vaccination

Citation Mabvakure BM, Scheepers C, Garrett N, Abdool Karim S, Williamson C, Morris L, Moore PL. 2019. Positive selection at key residues in the HIV envelope distinguishes broad and strain-specific plasma neutralizing antibodies. J Virol 93:e01685-18. https://doi.org/10.1128/JVI.01685-18.

Editor Guido Silvestri, Emory University

Copyright © 2019 Mabvakure et al. This is an open-access article distributed under the terms of the Creative Commons Attribution 4.0 International license.

Address correspondence to Penny L. Moore, pennym@nicd.ac.za.

Received 24 September 2018
Accepted 28 November 2018
Accepted manuscript posted online 19 December 2018
Published

A preventative HIV vaccine is likely to require broadly neutralizing antibodies (bNAbs) that have the ability to neutralize the majority of HIV strains (1, 2). Animal studies have shown that bNAbs are able to protect from virus infection (3); however, no vaccine has thus far been able to elicit bNAbs. In contrast, longitudinal studies have shown that bNAbs develop in approximately 10 to 30% of chronically HIV-infected patients (4–6). These bNAbs target five sites on the HIV trimer, namely, the V1V2 region, the CD4 binding site, the membrane proximal external region (MPER), the V3 glycan supersite, and the gp41-gp120 interface, which includes the fusion peptide (7). Although most donors develop some degree of cross-reactivity over many years of infection (8), some donors have neutralizing antibody responses that remain strain-specific, neutralizing only the infecting strain, despite equivalent duration of infection and high viral loads (9). The reason strain-specific responses fail to mature into bNAbs in all HIV-infected individuals is an important unanswered question in the field.

Large cohort studies have shown that neutralization breadth is associated with several host factors, including ethnicity, HLA genotype, CD4 T cell loss, and circulating T follicular helper cells (4–6, 10–13). However, viral factors, such as subtype, high viral loads, virus diversity, and duration of HIV-1 infection, have been shown to be strongly associated with breadth, highlighting the importance of high levels of antigenic stimulation in driving the maturation of bNAbs (5, 6, 14, 15).

Studies of the developmental pathways of bNAbs have shown that these develop from strain-specific precursors that mature to acquire breadth through a prolonged coevolutionary pathway, with emerging viral mutations driving antibodies to tolerate epitope diversity (2, 16–24). This has led to a vaccine framework that aims to incorporate variation within immunogens, an approach that has recently shown promise in small animal immunization studies (25–30). However, similarly detailed studies of viral evolution in donors who maintain strain-specific responses, especially in the context of high levels of antigenic stimulation, have not been performed. Comparative analyses of virus evolutionary pathways in these individuals and those who develop bNAbs may provide useful insights into the viral characteristics and mechanisms that drive antibody maturation toward breadth.

Robust strain-specific neutralizing responses develop in almost all infected individuals, regardless of whether they later develop breadth, and exert significant selection pressure on the HIV envelope glycoprotein (Env) (31, 32). This results in increased Env diversity and divergence from the transmitted/founder (T/F) virus and selection of neutralization-resistant strains. Escape is associated with increased variable loop length and glycosylation that may shield underlying epitopes, making it difficult to neutralize the virus (33–38). Importantly, many of the strain-specific neutralizing responses target the same overall regions of the HIV envelope as bNAbs, such as the V1V2 and C3V4 regions (24, 39), and drive viral diversification within these regions (24, 38, 40).

To assess why these strain-specific responses fail to acquire breadth, we have used longitudinal single-genome-derived full Env sequencing data and matched antibody kinetics and mapping data from 14 participants in the Centre for the AIDS Programme of Research in South Africa (CAPRISA) 002 cohort, six of whom developed strain-specific antibodies and eight who developed bNAbs. Interestingly, we did not observe significant differences in nucleotide substitution rates in participants with or without bNAbs, and we observed high viral diversity across Env and within epitopes in both groups. However, we identified several sites under positive selection that were common to donors with bNAbs. The tight clustering of these sites on the Env trimer suggests that the precise targeting of viral epitopes by early bNAb precursors, and the subsequent viral escape pathways, determines whether these antibodies acquire breadth. Defining these sites highlights developmental differences in strain-specific antibodies and bNAbs which may inform future vaccine design strategies.

RESULTS

Env diversity and evolutionary rates do not distinguish broad and strain-specific donors with equivalent viral loads. We first assessed viral diversity in 6

TABLE 1 Description of HIV-infected participants, plasma neutralization breadth, set point viral load, and antibody targets

Participant	% neutralization breadth at 3 yr[a]	No. of *env* sequences	No. of time points	Set point viral load at 12 mo postinfection (RNA copies/ml)	Antibody targets[b]	Reference
CAP257	82	63	6	8,260	**V1V2, CD4 binding site**	Wibmer et al. (24)
CAP256	77	171	13	178,000	**V1V2**	Moore et al. (46), Doria-Rose et al. (16)
CAP248	59	120	18	64,600	**gp120-gp41 interface**	Wibmer et al., 2017
CAP177	52	144	7	18,000	**C3V4,** V1V2	Moore et al., 2009, Moore et al. (23)
CAP255	47	148	13	18,200	**C3V4**	Lacerda et al., 2013
CAP206	47	130	9	315,000	**MPER,** C3V4	Moore et al. (46), Unpublished data
CAP357[c]	45	76	7	11,596	Not mapped	Unpublished data
CAP8	42	70	5	39,300	Not mapped	Lacerda et al., 2013
CAP88	8	83	8	38,700	V1V2, C3V4	Moore et al., 2009
CAP228	7	100	6	1,520	V1V2, C3V4	Bhiman et al., unpublished data, Van Eeden and Wibmer et al., unpublished data
CAP229	7	35	3	24,600	V1V2	Unpublished data
CAP225	5	29	3	21,500	Not mapped	Unpublished data
CAP200	2	53	4	108,000	C3V4	Unpublished data
CAP45	0	41	6	556	Not mapped	Moore et al., 2009

[a]Gray et al., 2011.
[b]bNAbs are shown in bold.
[c]Madzhivandhila et al., unpublished data.

non-bNAb donors (with <10% breadth) and 8 bNAb donors (with >40% breadth) over 3 years of infection (Table 1). Since viral diversity is closely linked to viral load, we selected, where possible, non-bNAb participants with similarly high viral loads to the bNAb donors. Two participants, CAP45 and CAP228, who had consistently low viral loads, were also included, as we have extensively characterized their neutralizing response (Fig. 1A). We generated between 29 and 171 single-genome-derived *env* sequences per donor from multiple time points spanning 3 years (Table 1). Maximum likelihood analysis of transmitted/founder viruses from bNAb and non-bNAb donors confirmed that these were phylogenetically unlinked (data not shown).

Since dual infection results in increased genetic diversity, we used BEAST to compare the estimated time of infection (tMRCA), based on the sequence diversity, to the actual time of infection for each donor (41, 42). For dual infection, sequence analysis predicts a greater estimated duration of infection than the known time of infection; this is indicated by orange shading for CAP256 (a donor known to be superinfected), but not for CAP88, in Fig. 1B. Twelve out of 14 data sets had sufficiently strong temporal signals (as determined by the positive slope [rate] and correlation coefficient) for phylogenetic molecular clock analysis. The two donors that were excluded from this analysis were CAP225 and CAP229. BEAST analysis indicated single infections in 6 donors, CAP88, CAP206, CAP248, CAP255, CAP357, and CAP257 (Fig. 1C, bNAb participants are ranked by breadth), consistent with the fact that most HIV-1 infections are initiated by a single variant (43–45). In contrast, we saw evidence of dual infection (indicated by asterisks in Fig. 1C) for 5 donors, CAP200, CAP228, CAP8, CAP177, and CAP256. Dual infections were observed both in individuals with bNAbs (CAP8, CAP177, and CAP256) and without bNAbs (CAP200 and CAP228).

We next compared the levels of overall Env glycoprotein diversity in participants with and without bNAbs at 3 years postinfection (Fig. 1D). We observed the highest overall diversity in CAP228, CAP8, CAP256, and CAP257, with a mean overall diversity >4% amino acid differences per site at 3 years postinfection (Fig. 1D). With the exception of CAP257, this is likely a result of their infection with multiple viruses or variants (Fig. 1C). Among these donors, CAP228, who failed to develop breadth, had relatively high diversity despite low viral loads (Fig. 1A). However, overall, there was no statistically significant difference in Env diversity between participants with or without bNAbs ($P = 0.7758$, two-tailed Mann-Whitney test). Similarly, evolutionary rates across the envelope glycoprotein, based on nucleotide substitution rates estimated in BEAST, showed no significant differences between these groups (data not shown).

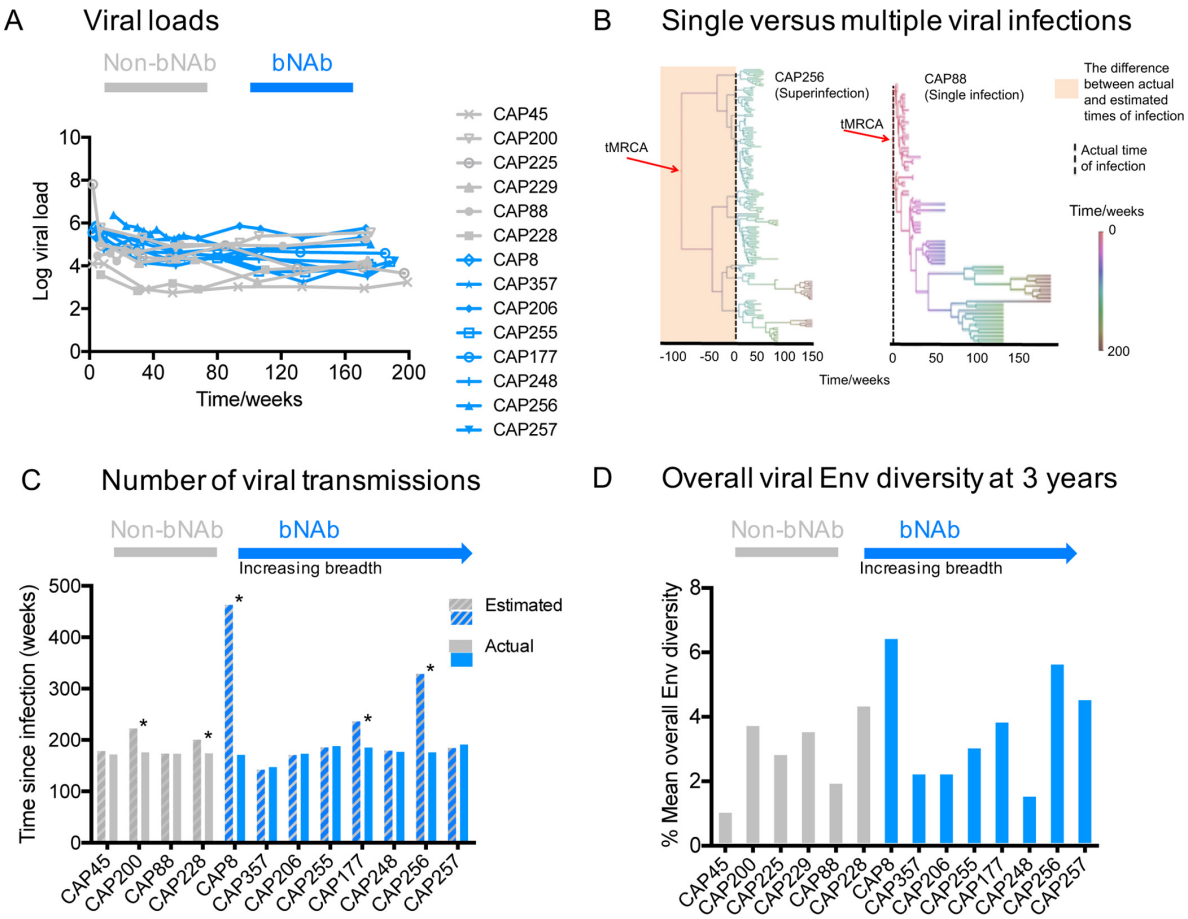


FIG 1 Viral diversity and evolutionary rates are not the sole determinants of bNAb development. (A) Longitudinal viral loads of all participants, with bNAb donors in blue and strain-specific donors in gray. (B) Maximum clade credibility trees of CAP88 and CAP256 constructed in BEAST showing the estimated time of infection (tMRCA) compared to actual time of infection. Similar analyses were performed for all donors. (C) Comparison of estimated and actual time of infection in 12 participants. The asterisk (*) denotes possible dual infection. (D) Comparison of overall amino acid envelope diversity at 3 years in participants with/without bNAbs.

Comparison of Env evolution within antibody-targeted regions between bNAb and non-bNAb donors. We have previously mapped the targets of both strain-specific and bNAb responses in 8 participants to the V1V2 and/or C3V4 regions (Table 1 and Fig. 2A). This enabled us to compare the Env evolutionary pathways between bNAb and non-bNAb donors specifically within these targeted regions. The V1V2 and C3V4 regions were defined according to the HXB2 reference sequence as amino acid positions 131 to 196 and 329 to 418, respectively. We first compared two donors with V1V2-directed bNAbs (CAP256 and CAP257) with four donors who had strain-specific V1V2 responses (CAP88, CAP228, CAP229, and CAP177) (4, 23, 24, 46) (Table 1). Although CAP177 developed bNAbs to the C3V4 region, this participant also had a strain-specific V1V2 response (38) and is included as a strain-specific donor for this epitope. The two participants with V1V2-directed bNAbs (CAP256 and CAP257) had higher levels of V1V2 amino acid diversity (20.7% and 14.5%, respectively) than did non-bNAb participants (ranging from 6.9 to 11.9% amino acid differences per site, though this difference failed to reach significance; Fig. 2B). Similarly, the two bNAb participants showed slightly higher V1V2 evolutionary rates than those of non-bNAb donors; however, these differences were also not significant, with estimates within the same high posterior density intervals (Fig. 2C).

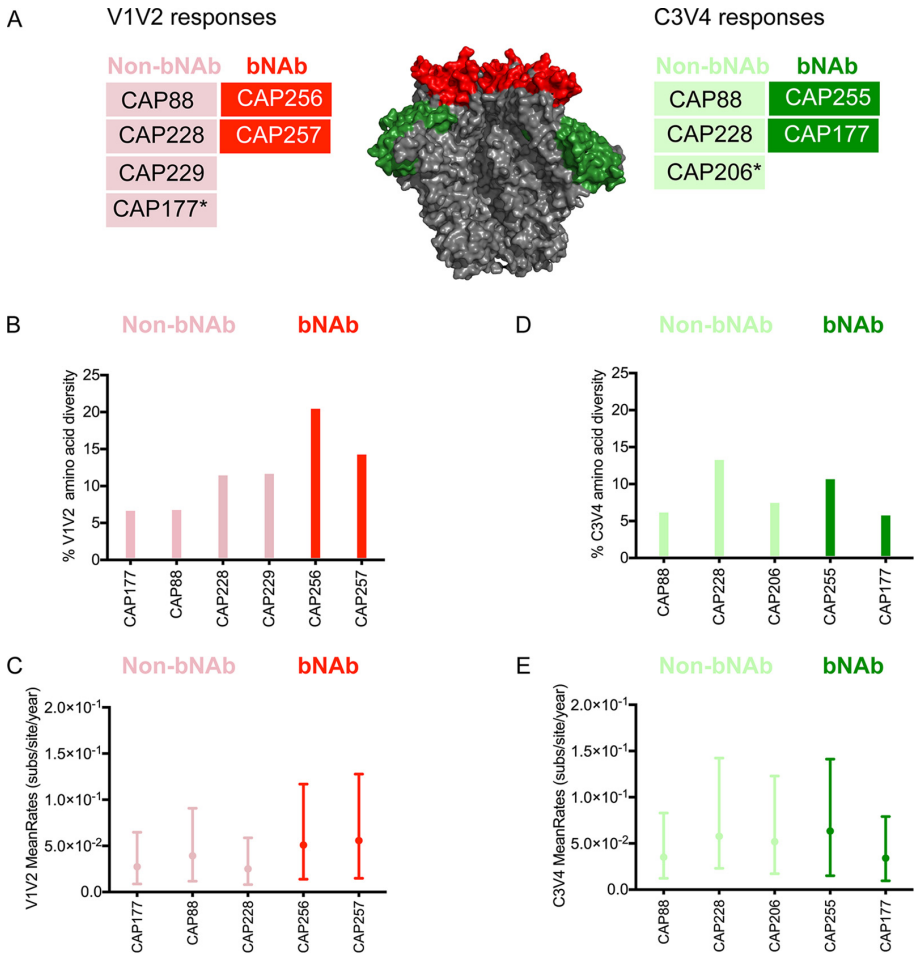


FIG 2 Comparison of diversity and nucleotide substitution rates in epitope regions targeted by participants with/without bNAbs. (A) Schematic of an Env trimer showing the antibody targets defined in previous studies. Participants with V1V2 responses are shown in red, and those with C3V4 responses are shown in green. The dark shading and light shading represent participants with and without bNAbs, respectively. The asterisk (*) denotes participants who have bNAb responses to other regions but also developed strain-specific antibodies to V1V2 or C3V4. Epitope diversity (average number of amino acid differences per site) and nucleotide substitution rates within V1V2 (B and C) and C3V4 (D and E) in participants with/without bNAbs. Broad participants arranged in order of increasing neutralization breadth. subs, substitutions.

We similarly compared two donors who developed N332-specific bNAbs (CAP177 and CAP255; dark green in Fig. 2A) with three donors shown to have strain-specific C3V4 responses (CAP88, CAP206, and CAP228; light green in Fig. 2A). CAP206 developed bNAbs that targeted the MPER (4), but in this analysis, we focused on the strain-specific C3V4 response. bNAb donors showed no difference in overall C3V4 diversity and nucleotide substitution rates from those of the strain-specific C3V4 responders (Fig. 2D and E). Overall, analysis of viral evolution within antibody-targeted regions did not show evidence of unique viral evolutionary pathways for donors with N332-directed bNAbs.

Participants with bNAbs show high levels of positive selection within targeted epitopes. We compared levels of positive selection within epitope regions in bNAb and non-bNAb donors by estimating the ratio of nonsynonymous to synonymous substitutions per site (47). We first compared two approaches for assessing selection, cumulative and a snapshot selection analyses. In the cumulative analysis, all sequences at

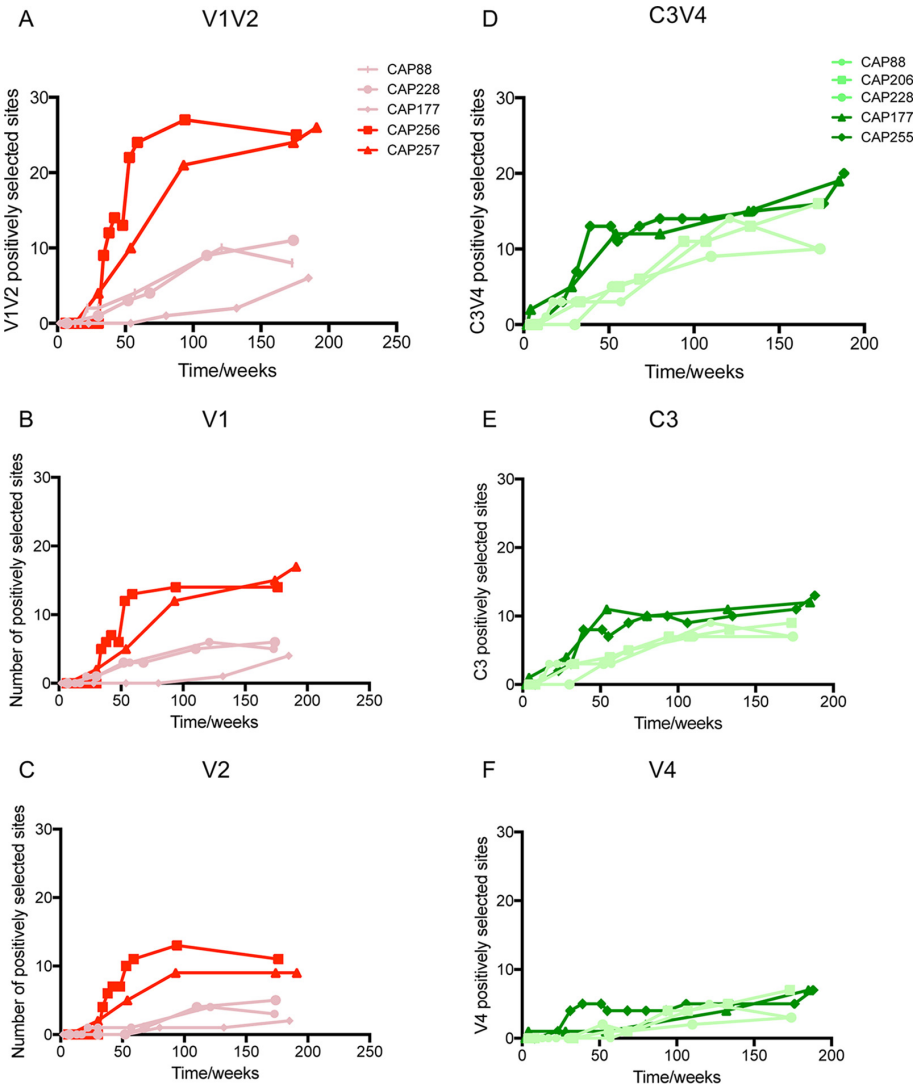


FIG 3 bNAb participants have more sites under positive selection than do non-bNAb individuals. Positive selection was higher in the V1V2 region of participants who developed bNAbs (dark red) than in those with strain-specific neutralizing antibodies (dull red) (A to C). In participants that had C3V4 responses, positive selection was higher in the C3V4 region of bNAb participants (dark green) than in those with strain-specific neutralizing antibodies (dull green) only in the first year of infection (D to F).

preceding time points were combined, whereas in the snapshot analysis, we analyzed sequence data only from two consecutive time points. The cumulative analysis identified the same positively selected sites as the snapshot analysis, as well as additional sites, and unlike the snapshot analysis, it retained all sequences for analysis (data not shown). All subsequent analyses were therefore done using the cumulative analysis.

As expected, the number of positively selected sites increased over time in all participants regardless of whether they developed breadth or not (Fig. 3 A and B). However, participants with V1V2-directed bNAbs exhibited a rapid increase in the number of positively selected sites (peaking at 26 to 27 amino acid residues) in V1V2 within the first year, compared to those without bNAbs (6 to 11 amino acid residues), and this increased level of positive selection persisted over 3 years of infec-

tion (Fig. 3A to C). Similarly, participants with C3V4-directed bNAb responses had an increased number of positively selected sites (peaking at 11 to 12 amino acid residues at 1 year) in C3V4 during the first year of infection compared to that of strain-specific donors (3 to 4 amino acid residues at 1 year) (Fig. 3D to F), though this difference was less pronounced than that observed in V1V2 and was not sustained in later years. These differences between bNAb and non-bNAb donors were driven largely by the C3 region (Fig. 3D to F), consistent with the fact that bNAbs to this region are known to be focused on the N332 glycan in C3, whereas strain-specific antibodies also target the more sequence-variable V4 region (39, 48).

Common sites in the V2, C3, and V3 regions are targeted by bNAbs. We next compared the specific sites under positive selection between participants with and without bNAbs. We focused on the V2 and C3 regions only, as the V1 and V4 regions were highly variable and prone to indels, making accurate identification of specific sites difficult. Using FUBAR, we identified positively selected sites in the V2 and C3 regions in each of the donors and compared their positions between bNAb and non-bNAb donors.

In the V2 region, we identified 28 positively selected sites across all donors with V1V2 responses. bNAb donors CAP256 and CAP257 had 13 and 11 positively selected sites in V2, respectively (Fig. 4A). Of these, six sites (160 and 162, which together form the highly conserved N160 glycosylation sequon, 166, 169, 181, 185, and 187) were common in both V2 bNAb participants (CAP256 and CAP257) (highlighted in red in Fig. 4A and B). Four of these sites were also targeted by at least one of the non-bNAb donors, indicating that strain-specific V1V2 antibodies overlapped bNAbs somewhat. Many of these sites (160/162, 166, 169, and 181) have previously been reported to mediate escape from V1V2-directed bNAbs (17, 49–51), and all are located proximal to one another at the apex of the envelope trimer (Fig. 4B). However, overall positive selection at these sites, particularly at sites 166 and 181, was less common in donors who developed strain-specific V1V2 responses (Fig. 4C).

Similarly, 22 positively selected sites were identified within the C3 region of the 5 participants with C3V4 responses (Fig. 5A**)**. Of these, seven sites (332/334, which form part of mutually exclusive glycosylation sequons, 337, 339, 340, 341, 343, and 344) were common in both participants who developed bNAbs targeting the N332 supersite (highlighted in red in Fig. 5A and B). Unlike the two bNAb participants, donors with strain-specific responses to the C3V4 region did not show evidence of positive selection at residue 332 or 334. In addition to the glycan, residues 341 and 344 were not under positive selection in any of the three strain-specific donors (CAP88, CAP206, and CAP228). Sites 337, 339, 340, and 343, under positive selection in both bNAb donors, were also sometimes under positive selection in participants who failed to develop bNAbs (Fig. 5C). Of note, we identified other common sites (351, 354, 355, 356, and 357) under selection in both CAP177 and CAP206 that clustered together within the C3 region on the envelope trimer (highlighted in yellow in Fig. 5B), suggesting another distinct epitope within this region.

bNAbs to the N332 supersite have also been shown to depend on the ^{324}GDIRQA H^{330} motif at the base of the V3 loop (52). We therefore analyzed selection at these sites. Although none reached statistical significance for positive selection, we observed toggling at sites 325 and 330 in both bNAb donors (CAP177 and CAP255) but also in strain-specific donor CAP228. No such toggling was observed in CAP206 and CAP88 (Fig. 5D). Sites 325 and 330 have been reported to impact neutralization sensitivity when these mutations are accompanied by changes at other sites, such as 301, 324, 326, and 327 (52). The similarity of toggling of V3 residues in CAP177, CAP255, and CAP228 might suggest a similar epitope for both broad antibodies and some strain-specific antibodies, raising the question of how they differ in their mode of recognition.

Increased positive selection is associated with neutralization breadth. We next assessed whether positive selection at key sites increased with the acquisition of breadth. For this analysis, we used previously published data defining the kinetics of

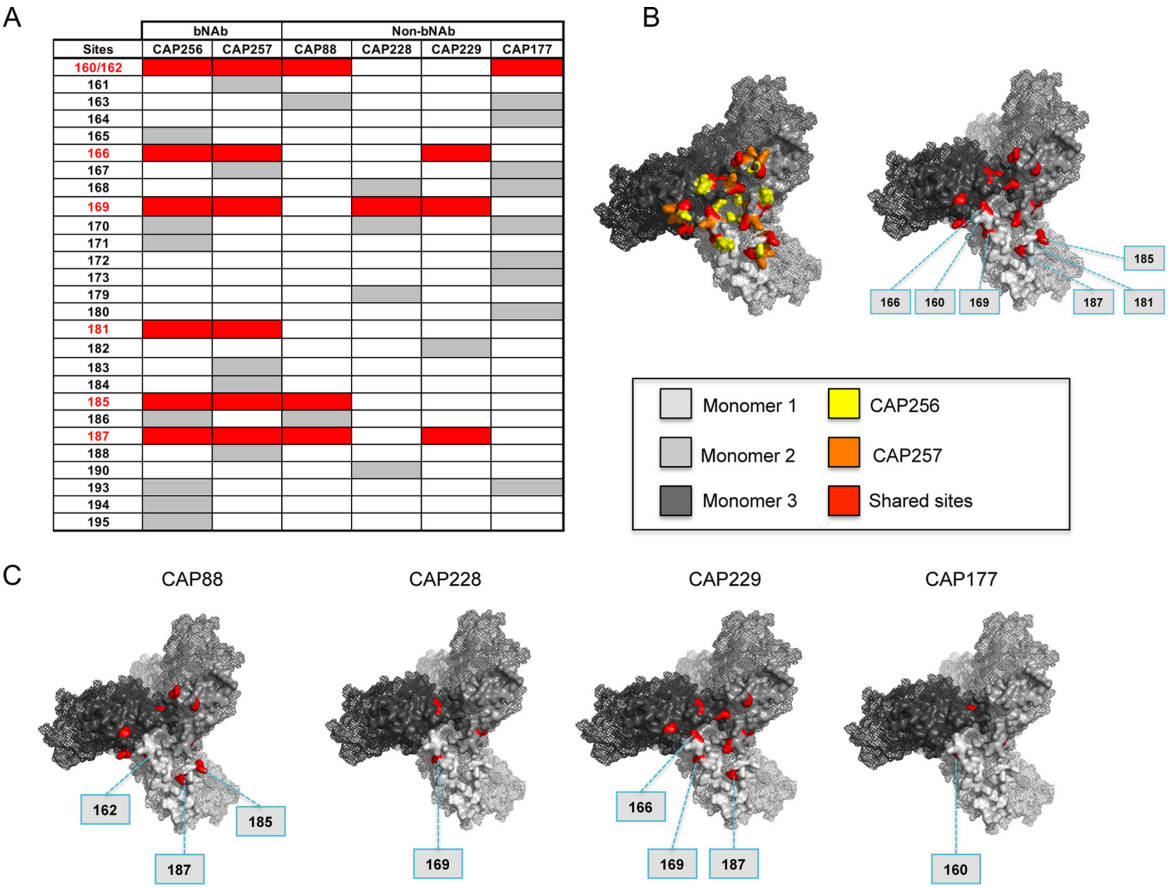


FIG 4 Comparison of positively selected sites in the V2 region of participants with and without bNAbs. (A) Table showing all sites under positive selection in the V2 region of participants with V1V2 responses. Positively selected sites common to both bNAb donors are highlighted in red. (B) Trimer model showing all sites under positive selection in participants with V2 responses (left) and common sites under positive selection in CAP256 and CAP257 (right). (C) bNAb sites also under positive selection in non-bNAb participants.

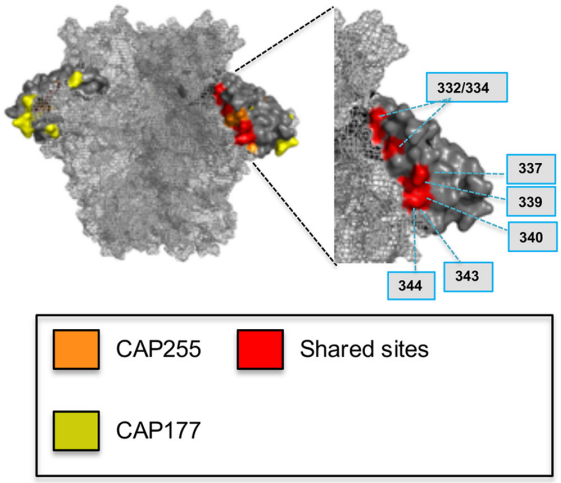
bNAbs (gray shading, Fig. 6) and strain-specific responses to the same epitope (dotted line, Fig. 6) (4, 16, 17, 23, 24, 33, 38). Selected residues are shown compared to the T/F virus, except for CAP256, where the comparison is with the superinfecting virus previously shown to elicit bNAbs in this donor (16). For V1V2 bNAb donors CAP256 and CAP257, we observed positive selection at sites 169/185 and 166/169/181, respectively, prior to the development of breadth, likely driven by the earlier strain-specific responses (Fig. 6A). In CAP256, the positively charged K169 in the superinfecting virus was replaced with a 169Q, which is rare among subtype C viruses but present in the primary infecting virus and confers partial escape from V2 directed antibodies (16). Similarly, in CAP257, a charge-changing minority K169E mutation emerged prior to the development of breadth. In both donors, positive selection expanded to additional sites, including the N160 glycan, with the onset of neutralization breadth. In CAP257, additional mutually exclusive glycans at positions at 185 and 187 were selected for after bNAb development.

In C3V4 donors CAP177 and CAP255, strain-specific responses drove positive selection at two to three sites (N332 and 344 for CAP177 and sites N339 and 340 for CAP255) (Fig. 6B). However, the emergence of bNAbs was associated with a significant increase in levels of positive selection. For CAP177, we have previously shown that the T/F does

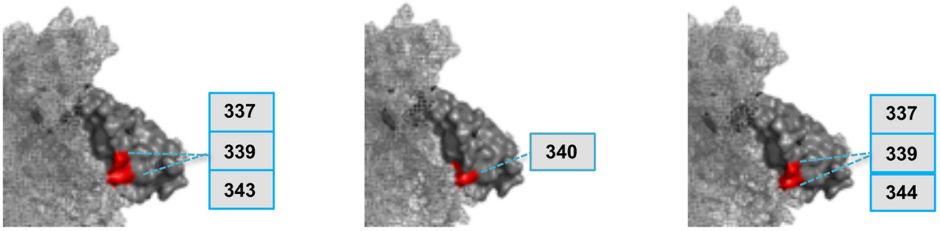
A

	bNAb		Non-bNAb		
Sites	CAP177	CAP255	CAP88	CAP206	CAP228
332/334					
335					
336					
337					
339					
340					
341					
343					
344					
347					
350					
351					
354					
355					
356					
357					
359					
361					
364					
375					
379					

B



C



D

Sites	bNAb		Non-bNAb		
	CAP177	CAP255	CAP88	CAP206	CAP228
325	D325N 54-132wks	D325N 55wks, 176-188wks	-	-	D325N 52-68wks
330	Y330H 54-185wks	H330Y 68-80wks		-	H330Y 52-68wks

FIG 5 Comparison of positively selected sites in the C3 region of participants with and without breadth. (A) Table showing all sites under positive selection in the C3 region of participants with C3V4 responses. Positively selected sites common to both bNAb donors are highlighted in red. (B) Trimer model showing all sites under positive selection in participants with C3V4 responses (left) and common sites under positive selection in bNAb donors CAP256 and CAP257 (right). (C) bNAb sites also under positive selection in non-bNAb participants. (D) Analysis of sites that toggle in the V3 GDIRQAH motif of participants with C3V4-directed responses. wks, weeks.

not contain a N332 glycan, and that early strain-specific nAbs drive selection of this glycan, which is subsequently lost as bNAbs develop (23). For CAP255, the N332 glycan is under positive selection but is nonetheless largely maintained in most late viral variants. Overall, we observed a significant increase in the number of key sites exhibiting selection pressure with the onset of breadth (Fig. 6), both within the V1V2 and C3V4 bNAb donors.

DISCUSSION

The study of virus-antibody coevolution in infected donors who develop bNAbs has provided crucial insights into how viral variation drives antibody lineages toward breadth. However, most studies have focused on bNAb donors, making the assumption that specific Env evolutionary patterns are unique to this group. The existence of infected donors who develop strain-specific responses raises the question of whether viral evolutionary pathways in these donors differ from those of the well-described

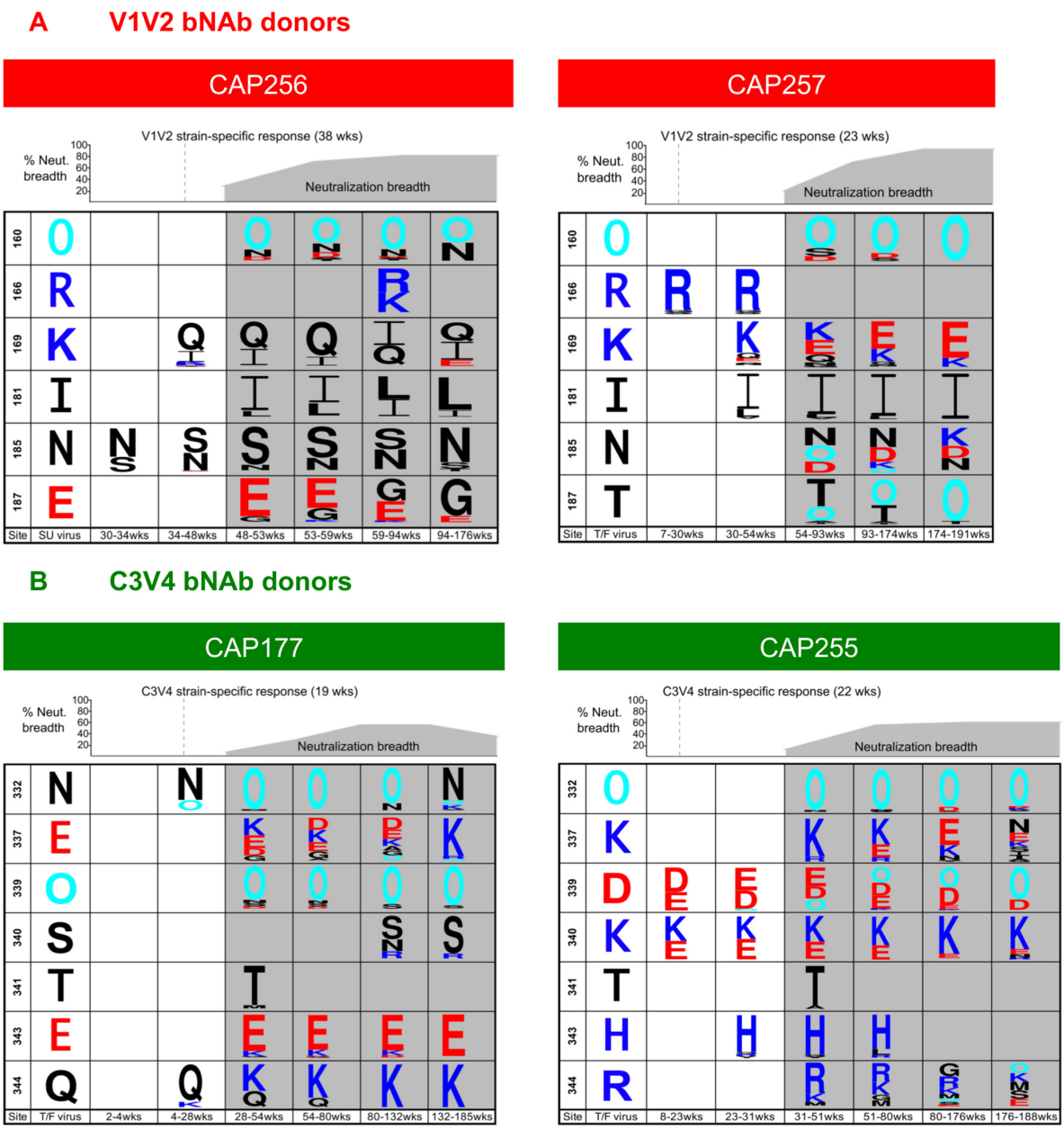


FIG 6 Increased positive selection at the onset of neutralization (Neut.) breadth. Breadth is shown in gray, glycans are shown in cyan, positively charged amino acids are shown in blue, negatively charged amino acids are shown in red, and noncharged amino acids are shown in black. (A) Participants with V1V2-directed bNAbs showed increased numbers of positively selected bNAb sites (160, 166, 169, 181, 185, and 187) as breadth developed. (B) Participants with C3V4 bNAbs accumulated pressure at sites 332, 337, 339, 340, 341, 343, and 344 with breadth.

bNAb donors. The purpose of this study was therefore to compare the evolutionary pathways of HIV envelope glycoproteins in bNAb and non-bNAb donors to identify whether there are distinct viral features that are associated with breadth. Moreover, we used previously generated data describing the targets of both strain-specific antibodies and bNAbs to compare viral evolution specifically within epitope regions. We showed that neither overall viral diversity nor local diversity in targeted regions is sufficient for

breadth. However, we observed higher numbers of positively selected sites in broad neutralizers and identified several positively selected sites in the V2 and C3 regions that were common to bNAb donors and limited or absent in participants without neutralization breadth. Selection pressure at these sites also increased with onset of neutralization breadth, highlighting the role of targeted viral selection in the development of breadth.

One of the consistent factors associated with the development of bNAbs in cohort studies is the presence of high levels of viral diversity (5, 6, 9, 14, 15). However, in this study, Env diversity at 3 years of infection showed no significant differences between bNAb and non-bNAb donors. While this may appear to contradict findings from previous cohort studies, it is important to note that most non-bNAb donors in this study were selected to have a viral load similar to that of the bNAb donors. As viral load and diversity are linked, this analysis indicates that in donors with similar viral loads, overall diversity does not distinguish bNAb donors from non-bNAb donors, and it suggests that additional factors influence breadth. The lack of difference in overall Env diversity between the two groups is consistent, with several studies showing that bNAbs often represent only a small proportion of the overall neutralizing responses in an infected donor (53; see also Kitchin et al., unpublished data). Although most of the strain-specific neutralizing antibodies that comprise the overall neutralizing response fail to acquire breadth (as is evident from the low proportion of bNAb donors in infected cohorts), they nonetheless drive substantial diversity through viral escape. Indeed, there appears to be no clear association between overall autologous neutralization titers and breadth (4, 54, 55).

Superinfection has been reported to be important in the development of breadth, although the effect is small (16, 56, 57). Our study included two donors where the tMRCA and actual time of infection differed significantly, suggesting (CAP8) or confirming (CAP256) superinfection. Both donors developed bNAbs, and CAP256 has been intensively studied over many years (16, 17, 46, 58). However, in CAP256, we have recently demonstrated that superinfection played no significant role in driving breadth (55). Furthermore, within the CAPRISA cohorts, we see no association between superinfection and breadth (55).

Many studies defining viral escape from bNAbs have confirmed that despite the conserved nature of these epitopes, HIV can utilize multiple pathways to escape (36, 37, 59). For this reason, bNAbs fail to confer any clinical benefit to those donors in whom they develop (Gray et al., 2011). In donors who develop breadth, extensive toggling with epitopes has been shown to precede the development of breadth (17, 18). Our observation of higher numbers of positively selected sites in bNAb donors, particularly within the first year of infection, supports the notion that this early but targeted viral toggling is required for the development of breadth, and it differentiates bNAb donors from those who remain strain specific.

The underlying reason for a higher number of positively selected sites in bNAb donors is of interest for HIV vaccine design. This is unlikely to be a consequence of higher titers, as many of the strain-specific responses we have previously mapped have equivalently high titers to bNAb responses (38, 60). The targeting of intrinsically more conserved epitopes may result in fitness costs for some escape mutations and, therefore, higher levels of compensatory mutations at proximal sites (61). Alternatively, it is possible that antibodies with the potential to mature toward breadth bind the virus using a wider "footprint," i.e., they have more contact sites and therefore drive early escape mutations at more sites. The finding that mutations at many of these positively selected sites, when introduced into heterologous viruses, confer neutralization escape from bNAbs may support the latter possibility (4, 17, 18, 49, 50, 52, 62–64).

The identification of sites common to bNAb donors and their clustered location on the trimeric Env may emphasize the importance of accurate early targeting by bNAbs. Studies of the ontogeny of V1V2-directed bNAb donors have confirmed the need for precursors encoding long third complementarity-determining regions to penetrate the glycans shielding the V2 C-strand (where many of the positively selected sites we

identify here are located). The tight clustering of positively selected sites for C3V4 bNAb donors in this study may suggest that initial targeting is important here, too, though mature members of V3-directed bNAb glycan bNAbs show much more promiscuity in their angles of approach as they mature to accommodate sequence variation and glycosylation (65–71).

We selected 14 participants for the in-depth epitope study because of the availability of both sequence and mapping data. Furthermore, both V1V2- and N332-directed bNAbs are among the most common bNAb specificities elicited during infection (4, 6, 54, 72, 73) and may require relatively less somatic hypermutation to acquire breadth (16, 20, 74). This makes these bNAb specificities attractive vaccine targets. However, this study could be improved by performing similar analyses on additional participants with/without bNAbs, including other less commonly elicited bNAb specificities. Since the number of participants in this study was relatively small, larger cohorts and inclusion of donors infected by other viral subtypes would be valuable in the future. We also utilized single-genome Sanger sequencing data with limited depth, whereas sequences generated through next-generation sequencing would give a more accurate representation of viral diversity. Nonetheless, this study has highlighted key differences between the evolutionary pathways of HIV *env* glycoproteins in participants with and without bNAbs. Furthermore, it has clarified that amino acid diversity at key positions within the V1V2 and C3V4 epitopes is more important than overall *env* diversity. The association of shared positively selected sites with the onset of breadth highlights the importance of diversity at these positions in bNAb development. Designing immunogens against HIV that include targeted diversity within these common sites may thus be critical in the maturation of V1V2- and C3V4-directed bNAbs.

MATERIALS AND METHODS

HIV-infected participants and ethics. Longitudinal plasma samples were collected from 14 women chronically infected with HIV-1 subtype C who were enrolled in the CAPRISA 002 and 004 cohorts (75, 76). Of these, 8 participants (CAP8, CAP177, CAP206, CAP357, CAP248, CAP255, CAP256, and CAP257) developed bNAbs (defined by their ability to neutralize more than 40% of a multisubtype panel at 3 years) and 6 donors (CAP45, CAP88, CAP200, CAP225, CAP228, and CAP229) who did not develop bNAbs (with plasma breadth less than 10%) despite approximately 3 years of HIV infection (46) (Madzhivandhila et al., unpublished data).

Evidence of HIV infection in participants was determined using two rapid HIV-1 antibody tests (Determine, Abbott 89 Laboratories, Tokyo, Japan; and Capillus, Trinity Biotech, Jamestown, NY). False positives and negatives were eliminated by using PCR for the negative samples (Ampliscreen version 1.5; Roche Diagnostics, Rotkreuz, Switzerland) and an HIV enzyme immunoassay (EIA; BEP2000; Dade Behring, Marburg, Germany) for positive samples. Viral loads were estimated using the Cobas Amplicor HIV-1 Monitor test version 1.5 (Roche Diagnostics). The time of infection was estimated as either the midpoint between the last antibody-negative visit and first antibody-positive visit, or 14 days prior to an RNA-positive antibody-negative sample.

Written informed consent to conduct research on stored specimens was obtained from all women at enrollment, and ethics clearance for the use of plasma samples was obtained from the Human Research Ethics Committee (Medical) from the University of Witwatersrand (MM040202), the University of Cape Town (025/2004), and CAPRISA at the University of KwaZulu-Natal (E013/04).

Single-genome envelope amplification and sequencing. HIV-1 RNA was isolated from plasma samples, and full-length envelope genes were amplified from single genomes at multiple time points by nested PCR and sequenced using Sanger sequencing, as described previously (77).

Sequence alignments and data partitioning. Sequence alignments were made in MUSCLE (78) and edited manually in BioEdit (79). The longitudinal envelope (*env*) glycoprotein sequences were analyzed as a whole and subsequently partitioned into variable and constant regions for subsequent analyses using the Se-Al software (80). The partitions were defined according to the HXB2 reference sequence as follows: V1, amino acids 131 to 156; V2, amino acids 157 to 196; V3, amino acids 296 to 330; C3, amino acids 329 to 383; and V4, amino acids 385 to 418.

Estimating diversity, nucleotide substitution rates, and time of infection. Overall Env, C3V4, and V1V2 diversity was calculated using the Poisson correction distance model implemented in MEGA6 (81, 82). The model assumes equality of substitution rates among sites and equal amino acid frequencies, while correcting for multiple substitutions at the same site (81, 82). The presence of temporal signal was estimated in TempEst using maximum likelihood trees constructed with PhyML (83, 84). Phylogenetic analyses were performed using BEAST version 1.8.4 (85). Best-fit nucleotide substitution models were selected based on the Bayesian information criteria (BIC) and Akaike information criterion, corrected (AICc) values estimated using the MEGA 6 software (81). The best-fit demographic and clock models were selected with marginal likelihood estimations using the generalized path sampling and stepping stone methods with 50 path steps (86, 87). The Monte Carlo Markov chains were run until convergence

(effective sample sizes, above 200), and the chain lengths were between 50 and 500 million steps. The log files generated were analyzed in Tracer version 1.5, and the maximum clade credibility (MCC) tree annotated in TreeAnnotator and viewed in FigTree (88).

Identifying sites under positive selection pressure. We performed natural selection analyses using Fast Unbiased Bayesian AppRoximation (FUBAR) that identifies codons under positive selection (89, 90). Codons that had more nonsynonymous than synonymous changes and that were supported by a posterior probability greater than 0.9 were reported to be evolving under positive selection. The pipeline for selection analysis involved first analyzing recombination in sequences using the genetic algorithm for recombination detection (GARD) to avoid phylogenetic bias when running FUBAR (89, 90). We employed two strategies in the selection analyses, which we termed the snapshot and cumulative analyses, respectively. The snapshot analysis involved analyzing sequence data from two consecutive time points, whereas the cumulative analysis involved integrating overall sequences in the data set at a particular time point.

Data availability. The accession numbers for the *env* sequences are EF203968 to EF203982, FJ443149 to FJ443158, FJ443223 to FJ443239, FJ443275 to FJ443366, FJ443390 to FJ443393, FJ443492, FJ443499, FJ443501 to FJ443504, FJ443508, FJ443519, FJ444632, GQ999980, GK999985, GK999989, HQ625597 to HQ625599, HQ625604, HQ625605, KC863376 to KC863378, KC863380 to KC863392, KC863446 to KC863456, KC863458 to KC863475, KC863542 to KC863553, KC863555 to KC863568, KC863571, KF241776, KF241779, KF241781, KF241784, KF241786 to KF241788, KF241790, KF241791, KF241793 to KF241796, KF241798, KF996600, KF996601, KF996604 to KF996630, KF996632 to KF996699, KF996700 to KF996716, KT698223 to KT698227, KU198436, MF572810 to MF572829, and MK205449 to MK206213.

ACKNOWLEDGMENTS

We thank donors from the Center for the AIDS Programme of Research in South Africa (CAPRISA) for their commitment to this study, and the CAPRISA staff for managing the cohort. We thank Bronwen E. Lambson, Brent Oosthuysen, Constantinos K. Wibmer, Dale Kitchin, Daniel Sheward, Don Mvududu, Florette Treurnicht, Jinal Bhiman, Maabo Moralo, Melissa-Rose Abrahams, Molati Nonyane, Nonkululeko Ndabambi, Ruwayhida Thebus, Sharon Madzorera, and Talitha York for generating SGAs in previous studies.

We are grateful for funding from the NIH through a U01 grant (AI116086-01), the South African Medical Research Council (MRC) through the SHIP program, and CAPRISA. CAPRISA is funded by the South African HIV/AIDS Research and Innovation Platform of the South African Department of Science and Technology and was initially supported by the U.S. NIAID, NIH, U.S. Department of Health and Human Services grant U19 AI51794. B.M.M. was supported by a bursary from the Poliomyelitis Research Foundation (PRF). P.L.M. is supported by the South African Research Chairs Initiative of the Department of Science and Technology and National Research Foundation of South Africa (grant 98341).

REFERENCES

1. Moore PL, Williamson C. 2016. Approaches to the induction of HIV broadly neutralizing antibodies. Curr Opin HIV AIDS 11:569–575. https://doi.org/10.1097/COH.0000000000000317.
2. Bonsignori M, Liao HX, Gao F, Williams WB, Alam SM, Montefiori DC, Haynes BF. 2017. Antibody-virus co-evolution in HIV infection: paths for HIV vaccine development. Immunol Rev 275:145–160. https://doi.org/10.1111/imr.12509.
3. Pegu A, Hessell AJ, Mascola JR, Haigwood NL. 2017. Use of broadly neutralizing antibodies for HIV-1 prevention. Immunol Rev 275:296–312. https://doi.org/10.1111/imr.12511.
4. Gray ES, Madiga MC, Hermanus T, Moore PL, Wibmer CK, Tumba NL, Werner L, Mlisana K, Sibeko S, Williamson C, Abdool Karim SS, Morris L, CAPRISA 002 Study Team. 2011. The neutralization breadth of HIV-1 develops incrementally over four years and is associated with CD4$^+$ T cell decline and high viral load during acute infection. J Virol 85:4828–4840. https://doi.org/10.1128/JVI.00198-11.
5. Landais E, Huang X, Havenar-Daughton C, Murrell B, Price MA, Wickramasinghe L, Ramos A, Bian CB, Simek M, Allen S, Karita E, Kilembe W, Lakhi S, Inambao M, Kamali A, Sanders EJ, Anzala O, Edward V, Bekker L-G, Tang J, Gilmour J, Kosakovsky-Pond SL, Phung P, Wrin T, Crotty S, Godzik A, Poignard P. 2016. Broadly neutralizing antibody responses in a large longitudinal sub-Saharan HIV primary infection cohort. PLoS Pathog 12:e1005369. https://doi.org/10.1371/journal.ppat.1005369.
6. Rusert P, Kouyos RD, Kadelka C, Ebner H, Schanz M, Huber M, Braun DL, Hozé N, Scherrer A, Magnus C, Weber J, Uhr T, Cippa V, Thorball CW, Kuster H, Cavassini M, Bernasconi E, Hoffmann M, Calmy A, Battegay M, Rauch A, Yerly S, Aubert V, Klimkait T, Böni J, Fellay J, Regoes RR, Günthard HF, Trkola A, the Swiss HIV Cohort Study. 2016. Determinants of HIV-1 broadly neutralizing antibody induction. Nat Med 22:1260–1267. https://doi.org/10.1038/nm.4187.
7. Wibmer CK, Moore PL, Morris L. 2015. HIV broadly neutralizing antibody targets. Curr Opin HIV AIDS 10:135–143. https://doi.org/10.1097/COH.0000000000000153.
8. Hraber P, Seaman MS, Bailer RT, Mascola JR, Montefiori DC, Korber BT. 2014. Prevalence of broadly neutralizing antibody responses during chronic HIV-1 infection. AIDS 28:163–169. https://doi.org/10.1097/QAD.0000000000000106.
9. Moore PL, Williamson C, Morris L. 2015. Virological features associated with the development of broadly neutralizing antibodies to HIV-1. Trends Microbiol 23:204–211. https://doi.org/10.1016/j.tim.2014.12.007.
10. Piantadosi A, Chohan B, Panteleeff D, Baeten JM, Mandaliya K, Ndinya-Achola JO, Overbaugh J. 2009. HIV-1 evolution in gag and env is highly correlated but exhibits different relationships with viral load and the immune response. AIDS 23:579–587. https://doi.org/10.1097/QAD.0b013e328328f76e.
11. Euler Z, van Gils MJ, Bunnik EM, Phung P, Schweighardt B, Wrin T, Schuitemaker H. 2010. Cross-reactive neutralizing humoral immunity does not protect from HIV type 1 disease progression. J Infect Dis 201:1045–1053. https://doi.org/10.1086/651144.
12. Cohen K, Altfeld M, Alter G, Stamatatos L. 2014. Early preservation of CXCR5$^+$ PD-1$^+$ helper T cells and B cell activation predict the breadth of

neutralizing antibody responses in chronic HIV-1 infection. J Virol 88:13310–13321. https://doi.org/10.1128/JVI.02186-14.
13. Havenar-Daughton C, Lindqvist M, Heit A, Wu JE, Reiss SM, Kendric K, Bélanger S, Kasturi SP, Landais E, Akondy RS, McGuire HM, Bothwell M, Vagefi PA, Scully E, IAVI Protocol C Principal Investigators, Tomaras GD, Davis MM, Poignard P, Ahmed R, Walker BD, Pulendran B, McElrath MJ, Kaufmann DE, Crotty S. 2016. CXCL13 is a plasma biomarker of germinal center activity. Proc Natl Acad Sci U S A 113:2702–2707. https://doi.org/10.1073/pnas.1520112113.
14. Piantadosi A, Panteleeff D, Blish CA, Baeten JM, Jaoko W, McClelland RS, Overbaugh J. 2009. Breadth of neutralizing antibody response to human immunodeficiency virus type 1 is affected by factors early in infection but does not influence disease progression. J Virol 83:10269–10274. https://doi.org/10.1128/JVI.01149-09.
15. Euler Z, van den Kerkhof TLGM, van Gils MJ, Burger JA, Edo-Matas D, Phung P, Wrin T, Schuitemaker H. 2012. Longitudinal analysis of early HIV-1-specific neutralizing activity in an elite neutralizer and in five patients who developed cross-reactive neutralizing activity. J Virol 86:2045–2055. https://doi.org/10.1128/JVI.06091-11.
16. Doria-Rose NA, Schramm CA, Gorman J, Moore PL, Bhiman JN, DeKosky BJ, Ernandes MJ, Georgiev IS, Kim HJ, Pancera M, Staupe RP, Altae-Tran HR, Bailer RT, Crooks ET, Cupo A, Druz A, Garrett NJ, Hoi KH, Kong R, Louder MK, Longo NS, McKee K, Nonyane M, O'Dell S, Roark RS, Rudicell RS, Schmidt SD, Sheward DJ, Soto C, Wibmer CK, Yang Y, Zhang Z, NISC Comparative Sequencing, Mullikin JC, Binley JM, Sanders RW, Wilson IA, Moore JP, Ward AB, Georgiou G, Williamson C, Abdool Karim SS, Morris L, Kwong PD, Shapiro L, Mascola JR. 2014. Developmental pathway for potent V1V2-directed HIV-neutralizing antibodies. Nature 509:55–62. https://doi.org/10.1038/nature13036.
17. Bhiman JN, Anthony C, Doria-Rose NA, Karimanzira O, Schramm CA, Khoza T, Kitchin D, Botha G, Gorman J, Garrett NJ, Karim SSA, Shapiro L, Williamson C, Kwong PD, Mascola JR, Morris L, Moore PL. 2015. Viral variants that initiate and drive maturation of V1V2-directed HIV-1 broadly neutralizing antibodies. Nat Med 21:1332–1336. https://doi.org/10.1038/nm.3963.
18. Liao H-X, Lynch R, Zhou T, Gao F, Alam SM, Boyd SD, Fire AZ, Roskin KM, Schramm CA, Zhang Z, Zhu J, Shapiro L, NISC Comparative Sequencing, Becker J, Benjamin B, Blakesley R, Bouffard G, Brooks S, Coleman H, Dekhtyar M, Gregory M, Guan X, Gupta J, Han J, Hargrove A, Ho S, Johnson T, Legaspi R, Lovett S, Maduro Q, Masiello C, Maskeri B, McDowell J, Montemayor C, Mullikin J, Park M, Riebow N, Schandler K, Schmidt B, Sison C, Stantripop M, Thomas J, Thomas P, Vemulapalli M, Young A, Mullikin JC, Gnanakaran S, Hraber P, Wiehe K, Kelsoe G, Yang G, et al. 2013. Co-evolution of a broadly neutralizing HIV-1 antibody and founder virus. Nature 496:469–476. https://doi.org/10.1038/nature12053.
19. Wu X, Zhang Z, Schramm CA, Joyce MG, Do Kwon Y, Zhou T, Sheng Z, Zhang B, O'Dell S, McKee K, Georgiev IS, Chuang G-Y, Longo NS, Lynch RM, Saunders KO, Soto C, Srivatsan S, Yang Y, Bailer RT, Louder MK, NISC Comparative Sequencing, Mullikin JC, Connors M, Kwong PD, Mascola JR, Shapiro L, Benjamin B, Blakesley R, Bouffard G, Brooks S, Coleman H, Dekhtyar M, Gregory M, Guan X, Gupta J, Han J, Hargrove A, Ho S, Legaspi R, Maduro Q, Masiello C, Maskeri B, McDowell J, Montemayor C, Park M, Riebow N, Schandler K, Schmidt B, Sison C, Stantripop M, Thomas J, et al. 2015. Maturation and diversity of the VRC01-antibody lineage over 15 years of chronic HIV-1 infection. Cell 161:470–485. https://doi.org/10.1016/j.cell.2015.03.004.
20. MacLeod DT, Choi NM, Briney B, Garces F, Ver LS, Landais E, Murrell B, Wrin T, Kilembe W, Liang C-H, Ramos A, Bian CB, Wickramasinghe L, Kong L, Eren K, Wu C-Y, Wong C-H, The IAVI Protocol C Investigators & The IAVI African HIV Research Network, Kosakovsky Pond SL, Wilson IA, Burton DR, Poignard P. 2016. Early antibody lineage diversification and independent limb maturation lead to broad HIV-1 neutralization targeting the Env high-mannose patch. Immunity 44:1215–1226. https://doi.org/10.1016/j.immuni.2016.04.016.
21. Gao F, Bonsignori M, Liao H, Kumar A, Xia S, Lu X, Cai F, Hwang K, Song H, Zhou T, Lynch RM, Alam SM, Moody MA, Ferrari G, Berrong M, Kelsoe G, Shaw GM, Hahn BH, Montefiori DC, Kamanga G, Cohen MS, Hraber P, Kwong PD, Korber BT, Mascola JR, Kepler TB, Haynes BF. 2014. Cooperation of B cell lineages in induction of HIV-1-broadly neutralizing antibodies. Cell 158:481–491. https://doi.org/10.1016/j.cell.2014.06.022.
22. Murphy MK, Yue L, Pan R, Boliar S, Sethi A, Tian J, Pfafferot K, Karita E, Allen SA, Cormier E, Goepfert PA, Borrow P, Robinson JE, Gnanakaran S, Hunter E, Kong XP, Derdeyn CA. 2013. Viral escape from neutralizing antibodies in early subtype A HIV-1 infection drives an increase in autologous neutralization breadth. PLoS Pathog 9:e1003173. https://doi.org/10.1371/journal.ppat.1003173.
23. Moore PL, Gray ES, Wibmer CK, Bhiman JN, Nonyane M, Sheward DJ, Hermanus T, Bajimaya S, Tumba NL, Abrahams M-R, Lambson BE, Ranchobe N, Ping L, Ngandu N, Karim QA, Karim SSA, Swanstrom RI, Seaman MS, Williamson C, Morris L. 2012. Evolution of an HIV glycan-dependent broadly neutralizing antibody epitope through immune escape. Nat Med 18:1688–1692. https://doi.org/10.1038/nm.2985.
24. Wibmer CK, Bhiman JN, Gray ES, Tumba N, Abdool Karim SS, Williamson C, Morris L, Moore PL. 2013. Viral escape from HIV-1 neutralizing antibodies drives increased plasma neutralization breadth through sequential recognition of multiple epitopes and immunotypes. PLoS Pathog 9:e1003738. https://doi.org/10.1371/journal.ppat.1003738.
25. Haynes BF, Shaw GM, Korber B, Kelsoe G, Sodroski J, Hahn BH, Borrow P, McMichael AJ. 2016. HIV-host interactions: implications for vaccine design. Cell Host Microbe 19:292–303. https://doi.org/10.1016/j.chom.2016.02.002.
26. Escolano A, Steichen JM, Dosenovic P, Kulp DW, Golijanin J, Sok D, Freund NT, Gitlin AD, Oliveira T, Araki T, Lowe S, Chen ST, Heinemann J, Yao K-H, Georgeson E, Saye-Francisco KL, Gazumyan A, Adachi Y, Kubitz M, Burton DR, Schief WR, Nussenzweig MC. 2016. Sequential immunization elicits broadly neutralizing anti-HIV-1 antibodies in Ig knockin mice. Cell 166:1445.e12–1458.e12. https://doi.org/10.1016/j.cell.2016.07.030.
27. Dosenovic P, von Boehmer L, Escolano A, Jardine J, Freund NT, Gitlin AD, McGuire AT, Kulp DW, Oliveira T, Scharf L, Pietzsch J, Gray MD, Cupo A, van Gils MJ, Yao KH, Liu C, Gazumyan A, Seaman MS, Bjorkman PJ, Sanders RW, Moore JP, Stamatatos L, Schief WR, Nussenzweig MC, Björkman PJ, Sanders RW, Moore JP, Stamatatos L, Schief WR, Nussenzweig MC. 2015. Immunization for HIV-1 broadly neutralizing antibodies in human Ig knockin mice. Cell 161:1505–1515. https://doi.org/10.1016/j.cell.2015.06.003.
28. Briney B, Sok D, Jardine JG, Kulp DW, Skog P, Menis S, Jacak R, Kalyuzhniy O, de Val N, Sesterhenn F, Le KM, Ramos A, Jones M, Saye-Francisco KL, Blane TR, Spencer S, Georgeson E, Hu X, Ozorowski G, Adachi Y, Kubitz M, Sarkar A, Wilson IA, Ward AB, Nemazee D, Burton DR, Schief WR. 2016. Tailored immunogens direct affinity maturation toward HIV neutralizing antibodies. Cell 166:1459.e11–1470.e11. https://doi.org/10.1016/j.cell.2016.08.005.
29. Sok D, Briney B, Jardine JG, Kulp DW, Menis S, Pauthner M, Wood A, Lee EC, Le KM, Jones M, Ramos A, Kalyuzhniy O, Adachi Y, Kubitz M, MacPherson S, Bradley A, Friedrich GA, Schief WR, Burton DR. 2016. Priming HIV-1 broadly neutralizing antibody precursors in human Ig loci transgenic mice. Science 353:1557–1560. https://doi.org/10.1126/science.aah3945.
30. Tian M, Cheng C, Chen X, Duan H, Cheng HL, Dao M, Sheng Z, Kimble M, Wang L, Lin S, Schmidt SD, Du Z, Joyce MG, Chen Y, DeKosky BJ, Chen Y, Normandin E, Cantor E, Chen RE, Doria-Rose NA, Zhang Y, Shi W, Kong WP, Choe M, Henry AR, Laboune F, Georgiev IS, Huang PY, Jain S, McGuire AT, Georgeson E, Menis S, Douek DC, Schief WR, Stamatatos L, Kwong PD, Shapiro L, Haynes BF, Mascola JR, Alt FW. 2016. Induction of HIV neutralizing antibody lineages in mice with diverse precursor repertoires. Cell 166:1471.e18–1484.e18. https://doi.org/10.1016/j.cell.2016.07.029.
31. Gray ES, Moore PL, Choge IA, Decker JM, Bibollet-Ruche F, Li H, Leseka N, Treurnicht F, Mlisana K, Shaw GM, Karim SSA, Williamson C, Morris L, CAPRISA 002 Study Team. 2007. Neutralizing antibody responses in acute human immunodeficiency virus type 1 subtype C infection. J Virol 81:6187–6196. https://doi.org/10.1128/JVI.00239-07.
32. Li B, Decker JM, Johnson RW, Bibollet-Ruche F, Wei X, Mulenga J, Allen S, Hunter E, Hahn BH, Shaw GM, Blackwell JL, Derdeyn CA. 2006. Evidence for potent autologous neutralizing antibody titers and compact envelopes in early infection with subtype C human immunodeficiency virus type 1. J Virol 80:5211–5218. https://doi.org/10.1128/JVI.00201-06.
33. Anthony C, York T, Bekker V, Matten D, Selhorst P, Ferreria R-C, Garrett NJ, Karim SSA, Morris L, Wood NT, Moore PL, Williamson C. 2017. Cooperation between strain-specific and broadly neutralizing responses limited viral escape and prolonged the exposure of the broadly neutralizing epitope. J Virol 91:e00828-17.
34. Sagar M, Wu X, Lee S, Overbaugh J. 2006. Human immunodeficiency virus type 1 V1-V2 envelope loop sequences expand and add glycosylation sites over the course of infection, and these modifications affect antibody neutralization sensitivity. J Virol 80:9586–9598. https://doi.org/10.1128/JVI.00141-06.

35. Wei X, Decker JM, Wang S, Hui H, Kappes JC, Wu X, Salazar-Gonzalez JF, Salazar MG, Kilby JM, Saag MS, Komarova NL, Nowak MA, Hahn BH, Kwong PD, Shaw GM. 2003. Antibody neutralization and escape by HIV-1. Nature 422:307–312. https://doi.org/10.1038/nature01470.
36. Richman DD, Wrin T, Little SJ, Petropoulos CJ. 2003. Rapid evolution of the neutralizing antibody response to HIV type 1 infection. Proc Natl Acad Sci U S A 100:4144–4149. https://doi.org/10.1073/pnas.0630530100.
37. Rong R, Li B, Lynch RM, Haaland RE, Murphy MK, Mulenga J, Allen SA, Pinter A, Shaw GM, Hunter E, Robinson JE, Gnanakaran S, Derdeyn CA. 2009. Escape from autologous neutralizing antibodies in acute/early subtype C HIV-1 infection requires multiple pathways. PLoS Pathog 5:e1000594. https://doi.org/10.1371/journal.ppat.1000594.
38. Moore PL, Ranchobe N, Lambson BE, Gray ES, Cave E, Abrahams M-R, Bandawe G, Mlisana K, Abdool Karim SS, Williamson C, Morris L, the CAPRISA 002 Study, the NIAID Center for HIV/AIDS Vaccine Immunology (CHAVI). 2009. Limited neutralizing antibody specificities drive neutralization escape in early HIV-1 subtype C infection. PLoS Pathog 5:e1000598. https://doi.org/10.1371/journal.ppat.1000598.
39. Moore PL, Gray ES, Choge IA, Ranchobe N, Mlisana K, Abdool Karim SS, Williamson C, Morris L, Karim SSA, Williamson C, Morris L, the CAPRISA 002 Study Team. 2008. The C3-V4 region is a major target of autologous neutralizing antibodies in human immunodeficiency virus type 1 subtype C infection. J Virol 82:1860–1869. https://doi.org/10.1128/JVI.02187-07.
40. Derdeyn CA, Moore PL, Morris L. 2014. Development of broadly neutralizing antibodies from autologous neutralizing antibody responses in HIV infection. Curr Opin HIV AIDS 9:210–216. https://doi.org/10.1097/COH.0000000000000057.
41. Chaillon A, Samleerat T, Zoveda F, Ballesteros S, Moreau A, Ngo-Giang-Huong N, Jourdain G, Gianella S, Lallemant M, Depaulis F, Barin F. 2014. Estimating the timing of mother-to-child transmission of the human immunodeficiency virus type 1 using a viral molecular evolution model. PLoS One 9:e90421. https://doi.org/10.1371/journal.pone.0090421.
42. Novitsky V, Wang R, Margolin L, Baca J, Rossenkhan R, Moyo S, van Widenfelt E, Essex M. 2011. Transmission of single and multiple viral variants in primary HIV-1 subtype C infection. PLoS One 6:e16714. https://doi.org/10.1371/journal.pone.0016714.
43. Bar KJ, Tsao C, Iyer SS, Decker JM, Yang Y, Bonsignori M, Chen X, Hwang K-K, Montefiori DC, Liao H-X, Hraber P, Fischer W, Li H, Wang S, Sterrett S, Keele BF, Ganusov VV, Perelson AS, Korber BT, Georgiev I, McLellan JS, Pavlicek JW, Gao F, Haynes BF, Hahn BH, Kwong PD, Shaw GM. 2012. Early low-titer neutralizing antibodies impede HIV-1 replication and select for virus escape. PLoS Pathog 8:e1002721. https://doi.org/10.1371/journal.ppat.1002721.
44. Keele BF, Giorgi EE, Salazar-Gonzalez JF, Decker JM, Pham KT, Salazar MG, Sun C, Grayson T, Wang S, Li H, Wei X, Jiang C, Kirchherr JL, Gao F, Anderson JA, Ping L-H, Swanstrom R, Tomaras GD, Blattner WA, Goepfert PA, Kilby JM, Saag MS, Delwart EL, Busch MP, Cohen MS, Montefiori DC, Haynes BF, Gaschen B, Athreya GS, Lee HY, Wood N, Seoighe C, Perelson AS, Bhattacharya T, Korber BT, Hahn BH, Shaw GM. 2008. Identification and characterization of transmitted and early founder virus envelopes in primary HIV-1 infection. Proc Natl Acad Sci U S A 105:7552–7557. https://doi.org/10.1073/pnas.0802203105.
45. Abrahams MR, Anderson JA, Giorgi EE, Seoighe C, Mlisana K, Ping L-H, Athreya GS, Treurnicht FK, Keele BF, Wood N, Salazar-Gonzalez JF, Bhattacharya T, Chu H, Hoffman I, Galvin S, Mapanje C, Kazembe P, Thebus R, Fiscus S, Hide W, Cohen MS, Karim SA, Haynes BF, Shaw GM, Hahn BH, Korber BT, Swanstrom R, Williamson C, CAPRISA Acute Infection Study Team, the Center for HIV-AIDS Vaccine Immunology Consortium. 2009. Quantitating the multiplicity of infection with human immunodeficiency virus type 1 subtype C reveals a non-Poisson distribution of transmitted variants. J Virol 83:3556–3567. https://doi.org/10.1128/JVI.02132-08.
46. Moore PL, Gray ES, Sheward D, Ranchobe N, Lai Z, William J, Nonyane M, Tumba N, Hermanus T, Sibeko S, Mlisana K, Karim SSA, Williamson C, Pinter A, Morris L, Moore PL, Gray ES, Sheward D, Madiga M, Ranchobe N, Lai Z, Honnen WJ, Nonyane M, Tumba N, Hermanus T, Sibeko S, Mlisana K, Karim SSA, Williamson C, Pinter A, Morris L, Study C, Abdool Karim SS, Williamson C, Pinter A, Morris L, CAPRISA 002 Study. 2011. Potent and broad neutralization of HIV-1 subtype C by plasma antibodies targeting a quaternary epitope including residues in the V2 loop. J Virol 85:3128–3141. https://doi.org/10.1128/JVI.02658-10.
47. Lemey P, Kosakovsky Pond SL, Drummond AJ, Pybus OG, Shapiro B, Barroso H, Taveira N, Rambaut A. 2007. Synonymous substitution rates predict HIV disease progression as a result of underlying replication dynamics. PLoS Comput Biol 3:e29. https://doi.org/10.1371/journal.pcbi.0030029.
48. Rong R, Bibollet-Ruche F, Mulenga J, Allen S, Blackwell JL, Derdeyn CA. 2007. Role of V1V2 and other human immunodeficiency virus type 1 envelope domains in resistance to autologous neutralization during clade C infection. J Virol 81:1350–1359. https://doi.org/10.1128/JVI.01839-06.
49. Mascola JR, Haynes BF. 2013. HIV-1 neutralizing antibodies: understanding nature's pathways. Immunol Rev 254:225–244. https://doi.org/10.1111/imr.12075.
50. Liao H-X, Bonsignori M, Alam SM, McLellan JS, Tomaras GD, Moody MA, Kozink DM, Hwang K-K, Chen X, Tsao C-Y, Liu P, Lu X, Parks RJ, Montefiori DC, Ferrari G, Pollara J, Rao M, Peachman KK, Santra S, Letvin NL, Karasavvas N, Yang Z-Y, Dai K, Pancera M, Gorman J, Wiehe K, Nicely NI, Rerks-Ngarm S, Nitayaphan S, Kaewkungwal J, Pitisuttithum P, Tartaglia J, Sinangil F, Kim JH, Michael NL, Kepler TB, Kwong PD, Mascola JR, Nabel GJ, Pinter A, Zolla-Pazner S, Haynes BF. 2013. Vaccine induction of antibodies against a structurally heterogeneous site of immune pressure within HIV-1 envelope protein variable regions 1 and 2. Immunity 38:176–186. https://doi.org/10.1016/j.immuni.2012.11.011.
51. Phogat S, Walker L, Wagner D, Chan-Hui P, Simek M, Phung P, Wrin T, Mitcham J, Investigator P, Kaminsky S, Zamb T, Moyle M, Koff W, Burton D. 2009. Potent and broad neutralizing antibodies from HIV-1 non-clade B infected donor reveal a new HIV-1 vaccine target. Retrovirology 6:O6. https://doi.org/10.1186/1742-4690-6-S3-O6.
52. Sok D, Pauthner M, Briney B, Lee JH, Saye-Francisco KL, Hsueh J, Ramos A, Le KM, Jones M, Jardine JG, Bastidas R, Sarkar A, Liang CH, Shivatare SS, Wu CY, Schief WR, Wong CH, Wilson IA, Ward AB, Zhu J, Poignard P, Burton DR. 2016. A prominent site of antibody vulnerability on HIV envelope incorporates a motif associated with CCR5 binding and its camouflaging glycans. Immunity 45:31–45. https://doi.org/10.1016/j.immuni.2016.06.026.
53. Sajadi MM, Dashti A, Rikhtegaran Tehrani Z, Tolbert WD, Seaman MS, Ouyang X, Gohain N, Pazgier M, Kim D, Cavet G, Yared J, Redfield RR, Lewis GK, DeVico AL. 2018. Identification of near-pan-neutralizing antibodies against HIV-1 by deconvolution of plasma humoral responses. Cell 173:1783–1795.e14. https://doi.org/10.1016/j.cell.2018.03.061.
54. Moore PL, Gorman J, Doria-Rose NA, Morris L. 2017. Ontogeny-based immunogens for the induction of V2-directed HIV broadly neutralizing antibodies. Immunol Rev 275:217–229. https://doi.org/10.1111/imr.12501.
55. Sheward DJ, Marais J, Bekker V, Murrell B, Eren K, Bhiman JN, Nonyane M, Garrett N, Woodman ZL, Abdool Karim Q, Abdool Karim SS, Morris L, Moore PL, Williamson C. 2018. HIV superinfection drives de novo antibody responses and not neutralization breadth. Cell Host Microbe 24:593–599.e3. https://doi.org/10.1016/j.chom.2018.09.001.
56. Powell RLR, Kinge T, Nyambi PN. 2010. Infection by discordant strains of HIV-1 markedly enhances the neutralizing antibody response against heterologous virus. J Virol 84:9415–9426. https://doi.org/10.1128/JVI.02732-09.
57. Cortez V, Odem-Davis K, McClelland RS, Jaoko W, Overbaugh J. 2012. HIV-1 superinfection in women broadens and strengthens the neutralizing antibody response. PLoS Pathog 8:e1002611. https://doi.org/10.1371/journal.ppat.1002611.
58. Doria-Rose NA, Bhiman JN, Roark RS, Schramm CA, Gorman J, Chuang G-Y, Pancera M, Cale EM, Ernandes MJ, Louder MK, Asokan M, Bailer RT, Druz A, Fraschilla IR, Garrett NJ, Jarosinski M, Lynch RM, McKee K, O'Dell S, Pegu A, Schmidt SD, Staupe RP, Sutton MS, Wang K, Wibmer CK, Haynes BF, Abdool-Karim S, Shapiro L, Kwong PD, Moore PL, Morris L, Mascola JR. 2016. New member of the V1V2-directed CAP256-VRC26 lineage that shows increased breadth and exceptional potency. J Virol 90:76–91. https://doi.org/10.1128/JVI.01791-15.
59. Moore PL, Sheward D, Nonyane M, Ranchobe N, Hermanus T, Gray ES, Abdool Karim SS, Williamson C, Morris L. 2013. Multiple pathways of escape from HIV broadly cross-neutralizing V2-dependent antibodies. J Virol 87:4882–4894. https://doi.org/10.1128/JVI.03424-12.
60. Gray ES, Moore PL, Bibollet-Ruche F, Li H, Decker JM, Meyers T, Shaw GM, Morris L. 2008. 4E10-resistant variants in a human immunodeficiency virus type 1 subtype C-infected individual with an anti-membrane-proximal external region-neutralizing antibody response. J Virol 82:2367–2375. https://doi.org/10.1128/JVI.02161-07.

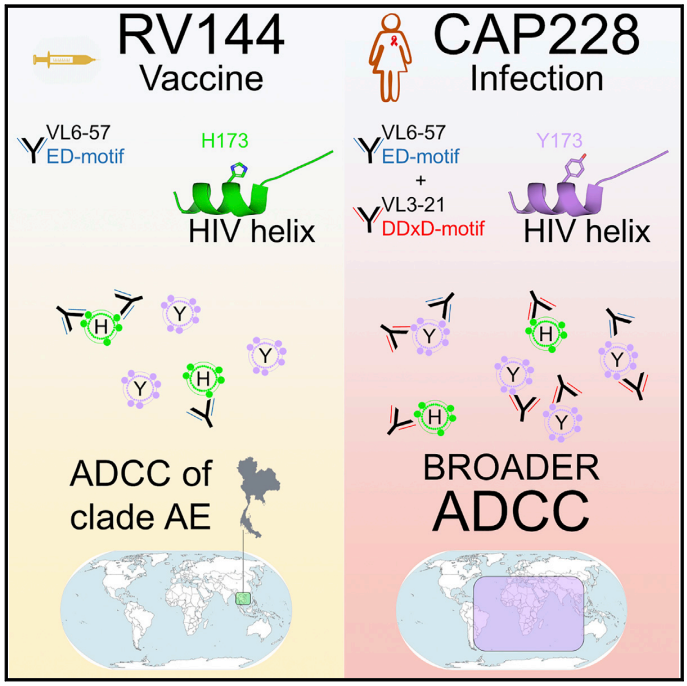
61. Landais E, Murrell B, Briney B, Murrell S, Rantalainen K, Berndsen ZT, Ramos A, Wickramasinghe L, Smith ML, Eren K, de Val N, Wu M, Cappelletti A, Umotoy J, Lie Y, Wrin T, Algate P, Chan-Hui P-Y, Karita E, IAVI Protocol C Investigators, IAVI African HIV Research Network, Ward AB, Wilson IA, Burton DR, Smith D, Pond SLK, Poignard P. 2017. HIV envelope glycoform heterogeneity and localized diversity govern the initiation and maturation of a V2 apex broadly neutralizing antibody lineage. Immunity 47:990.e9–1003.e9. https://doi.org/10.1016/j.immuni.2017.11.002.
62. Moore PL, Gray ES, Morris L. 2009. Specificity of the autologous neutralizing antibody response. Curr Opin HIV AIDS 4:358–363. https://doi.org/10.1097/COH.0b013e32832ea7e8.
63. Walker LM, Phogat SK, Chan-Hui P-Y, Wagner D, Phung P, Goss JL, Wrin T, Simek MD, Fling S, Mitcham JL, Lehrman JK, Priddy FH, Olsen OA, Frey SM, Hammond PW, Protocol G Principal Investigators, Kaminsky S, Zamb T, Moyle M, Koff WC, Poignard P, Burton DR. 2009. Broad and potent neutralizing antibodies from an African donor reveal a new HIV-1 vaccine target. Science 326:285–289. https://doi.org/10.1126/science.1178746.
64. Gray ES, Moody MA, Wibmer CK, Chen X, Marshall D, Amos J, Moore PL, Foulger A, Yu J-S, Lambson B, Abdool Karim S, Whitesides J, Tomaras GD, Haynes BF, Morris L, Liao H-X. 2011. Isolation of a monoclonal antibody that targets the alpha-2 helix of gp120 and represents the initial autologous neutralizing-antibody response in an HIV-1 subtype C-infected individual. J Virol 85:7719–7729. https://doi.org/10.1128/JVI.00563-11.
65. Garces F, Sok D, Kong L, McBride R, Kim HJ, Saye-Francisco KF, Julien J-P, Hua Y, Cupo A, Moore JP, Paulson JC, Ward AB, Burton DR, Wilson IA. 2014. Structural evolution of glycan recognition by a family of potent HIV antibodies. Cell 159:69–79. https://doi.org/10.1016/j.cell.2014.09.009.
66. Kong L, Lee JH, Doores KJ, Murin CD, Julien J-P, McBride R, Liu Y, Marozsan A, Cupo A, Klasse P-J, Hoffenberg S, Caulfield M, King CR, Hua Y, Le KM, Khayat R, Deller MC, Clayton T, Tien H, Feizi T, Sanders RW, Paulson JC, Moore JP, Stanfield RL, Burton DR, Ward AB, Wilson IA. 2013. Supersite of immune vulnerability on the glycosylated face of HIV-1 envelope glycoprotein gp120. Nat Struct Mol Biol 20:796–803. https://doi.org/10.1038/nsmb.2594.
67. Julien J-P, Cupo A, Sok D, Stanfield RL, Lyumkis D, Deller MC, Klasse P-J, Burton DR, Sanders RW, Moore JP, Ward AB, Wilson IA. 2013. Crystal structure of a soluble cleaved HIV-1 envelope trimer. Science 342:1477–1483. https://doi.org/10.1126/science.1245625.
68. Mouquet H, Scharf L, Euler Z, Liu Y, Eden C, Scheid JF, Halper-Stromberg A, Gnanapragasam PN, Spencer DI, Seaman MS, Schuitemaker H, Feizi T, Nussenzweig MC, Bjorkman PJ. 2012. Complex-type N-glycan recognition by potent broadly neutralizing HIV antibodies. Proc Natl Acad Sci U S A 109:E3268–E3277. https://doi.org/10.1073/pnas.1217207109.
69. Pejchal R, Doores KJ, Walker LM, Khayat R, Huang P-S, Wang S-K, Stanfield RL, Julien J-P, Ramos A, Crispin M, Depetris R, Katpally U, Marozsan A, Cupo A, Maloveste S, Liu Y, McBride R, Ito Y, Sanders RW, Ogohara C, Paulson JC, Feizi T, Scanlan CN, Wong C-H, Moore JP, Olson WC, Ward AB, Poignard P, Schief WR, Burton DR, Wilson IA. 2011. A potent and broad neutralizing antibody recognizes and penetrates the HIV glycan shield. Science 334:1097–1103. https://doi.org/10.1126/science.1213256.
70. Walker LM, Huber M, Doores KJ, Falkowska E, Pejchal R, Julien JP, Wang SK, Ramos A, Chan-Hui PY, Moyle M, Mitcham JL, Hammond PW, Olsen OA, Phung P, Fling S, Wong CH, Phogat S, Wrin T, Simek MD, Koff WC, Wilson IA, Burton DR, Poignard P, Protocol GPI, Koff WC, Wilson IA, Burton DR, Poignard P. 2011. Broad neutralization coverage of HIV by multiple highly potent antibodies. Nature 477:466–470. https://doi.org/10.1038/nature10373.
71. McLellan JS, Pancera M, Carrico C, Gorman J, Julien JP, Khayat R, Louder R, Pejchal R, Sastry M, Dai K, O'Dell S, Patel N, Shahzad-Ul-Hussan S, Yang Y, Zhang B, Zhou T, Zhu J, Boyington JC, Chuang GY, Diwanji D, Georgiev I, Do Kwon Y, Lee D, Louder MK, Moquin S, Schmidt SD, Yang ZY, Bonsignori M, Crump JA, Kapiga SH, Sam NE, Haynes BF, Burton DR, Koff WC, Walker LM, Phogat S, Wyatt R, Orwenyo J, Wang LX, Arthos J, Bewley CA, Mascola JR, Nabel GJ, Schief WR, Ward AB, Wilson IA, Kwong PD, Kwon YD, Lee D, Louder MK, et al. 2011. Structure of HIV-1 gp120 V1/V2 domain with broadly neutralizing antibody PG9. Nature 480:336–343. https://doi.org/10.1038/nature10696.
72. Georgiev IS, Doria-Rose NA, Zhou T, Do Kwon Y, Staupe RP, Moquin S, Chuang G-Y, Louder MK, Schmidt SD, Altae-Tran HR, Bailer RT, McKee K, Nason M, O'Dell S, Ofek G, Pancera M, Srivatsan S, Shapiro L, Connors M, Migueles SA, Morris L, Nishimura Y, Martin MA, Mascola JR, Kwong PD. 2013. Delineating antibody recognition in polyclonal sera from patterns of HIV-1 isolate neutralization. Science 340:751–756. https://doi.org/10.1126/science.1233989.
73. Walker LM, Simek MD, Priddy F, Gach JS, Wagner D, Zwick MB, Phogat SK, Poignard P, Burton DR. 2010. A limited number of antibody specificities mediate broad and potent serum neutralization in selected HIV-1 infected individuals. PLoS Pathog 6:e1001028. https://doi.org/10.1371/journal.ppat.1001028.
74. Simonich CAA, Williams KLL, Verkerke HPP, Williams JAA, Nduati R, Lee KKK, Overbaugh J. 2016. HIV-1 neutralizing antibodies with limited hypermutation from an infant. Cell 166:77–87. https://doi.org/10.1016/j.cell.2016.05.055.
75. van Loggerenberg F, Mlisana K, Williamson C, Auld SC, Morris L, Gray CM, Abdool Karim Q, Grobler A, Barnabas N, Iriogbe I, Abdool Karim SS, CAPRISA 002 Acute Infection Study Team. 2008. Establishing a cohort at high risk of HIV infection in South Africa: challenges and experiences of the CAPRISA 002 acute infection study. PLoS One 3:e1954. https://doi.org/10.1371/journal.pone.0001954.
76. Karim QA, Kharsany AB, Frohlich JA, Baxter C, Yende N, Mansoor LE, Mlisana KP, Maarschalk S, Arulappan N, Grobler A, Sibeko S, Omar Z, Gengiah TN, Mlotshwa M, Samsunder N, Abdool Karim SS. 2011. Recruitment of high risk women for HIV prevention trials: baseline HIV prevalence and sexual behavior in the CAPRISA 004 tenofovir gel trial. Trials 12:67. https://doi.org/10.1186/1745-6215-12-67.
77. Salazar-Gonzalez JF, Bailes E, Pham KT, Salazar MG, Guffey MB, Keele BF, Derdeyn CA, Farmer P, Hunter E, Allen S, Manigart O, Mulenga J, Anderson JA, Swanstrom R, Haynes BF, Athreya GS, Korber BTM, Sharp PM, Shaw GM, Hahn BH. 2008. Deciphering human immunodeficiency virus type 1 transmission and early envelope diversification by single-genome amplification and sequencing. J Virol 82:3952–3970. https://doi.org/10.1128/JVI.02660-07.
78. Edgar RC. 2004. MUSCLE: multiple sequence alignment with high accuracy and high throughput. Nucleic Acids Res 32:1792–1797. https://doi.org/10.1093/nar/gkh340.
79. Hall TA. 1999. BioEdit: a user-friendly biological sequence alignment editor and analysis program for Windows 95/98/NT. Nucleic Acids Symp Ser (Oxf):95–98.
80. Rambaut A. 2007. Se-Al v. 2.0. http://tree.bio.ed.ac.uk/software/seal.
81. Tamura K, Stecher G, Peterson D, Filipski A, Kumar S. 2013. MEGA6: Molecular Evolutionary Genetics Analysis version 6.0. Mol Biol Evol 30:2725–2729. https://doi.org/10.1093/molbev/mst197.
82. Thomas RH. 2001. Molecular evolution and phylogenetics. Heredity (Edinb) 86:385. https://doi.org/10.1046/j.1365-2540.2001.0923a.x.
83. Rambaut A, Lam TT, Max Carvalho L, Pybus OG. 2016. Exploring the temporal structure of heterochronous sequences using TempEst (formerly Path-O-Gen). Virus Evol 2:vew007. https://doi.org/10.1093/ve/vew007.
84. Guindon S, Delsuc F, Dufayard J-F, Gascuel O. 2009. Estimating maximum likelihood phylogenies with PhyML. Methods Mol Biol 537:113–137. https://doi.org/10.1007/978-1-59745-251-9_6.
85. Drummond AJ, Suchard MA, Xie D, Rambaut A. 2012. Bayesian phylogenetics with BEAUti and the BEAST 1.7. Mol Biol Evol 29:1969–1973. https://doi.org/10.1093/molbev/mss075.
86. Baele G, Lemey P, Bedford T, Rambaut A, Suchard MA, Alekseyenko AV. 2012. Improving the accuracy of demographic and molecular clock model comparison while accommodating phylogenetic uncertainty. Mol Biol Evol 29:2157–2167. https://doi.org/10.1093/molbev/mss084.
87. Baele G, Li WLS, Drummond AJ, Suchard MA, Lemey P. 2013. Accurate model selection of relaxed molecular clocks in Bayesian phylogenetics. Mol Biol Evol 30:239–243. https://doi.org/10.1093/molbev/mss243.
88. Rambaut A, Drummond A. 2009. FigTree v1. 3.1. http://tree.bio.ed.ac.uk/software/figtree.
89. Pond SLK, Muse SV. 2005. HyPhy: hypothesis testing using phylogenies, p 125–181. *In* Nielsen R (ed), Statistical methods in molecular evolution. Springer-Verlag, New York, NY.
90. Murrell B, Moola S, Mabona A, Weighill T, Sheward D, Kosakovsky Pond SL, Scheffler K. 2013. FUBAR: A Fast, Unconstrained Bayesian AppRoximation for inferring selection. Mol Biol Evol 30:1196–1205. https://doi.org/10.1093/molbev/mst030.

Article

Cell Reports

V2-Directed Vaccine-like Antibodies from HIV-1 Infection Identify an Additional K169-Binding Light Chain Motif with Broad ADCC Activity

Graphical Abstract



Highlights

- V2-specific antibodies from HIV infection that resemble vaccine-elicited antibodies
- Use of a different antibody light chain with a different V2 K169-binding motif
- Show broad ADCC activity against globally relevant V2 immunotypes
- Increases the repertoire of B cells able to respond to RV144 V2 immunogens

Authors

Charmaine van Eeden, Constantinos Kurt Wibmer, Cathrine Scheepers, ..., Barton F. Haynes, Penny L. Moore, Lynn Morris

Correspondence

lynnm@nicd.ac.za

In Brief

V2-directed antibodies from the RV144 vaccine trial correlated with reduced HIV-1 infection risk but exhibited restricted light chain gene usage. Here, van Eeden et al. isolate similar antibodies from an HIV-1-infected individual and identify a third V2-reactive light chain gene, increasing the antibody repertoire potentially elicited by vaccination.

van Eeden et al., 2018, Cell Reports *25*, 3123–3135
December 11, 2018 © 2018 The Authors.
https://doi.org/10.1016/j.celrep.2018.11.058

CellPress

Cell Reports
Article

V2-Directed Vaccine-like Antibodies from HIV-1 Infection Identify an Additional K169-Binding Light Chain Motif with Broad ADCC Activity

Charmaine van Eeden,[1,2,10] Constantinos Kurt Wibmer,[1,2,10,11] Cathrine Scheepers,[1,2] Simone I. Richardson,[1,2] Molati Nonyane,[1] Bronwen Lambson,[1,2] Nonhlanhla N. Mkhize,[1,2] Balakrishnan Vijayakumar,[3] Zizhang Sheng,[4,5] Sherry Stanfield-Oakley,[6] Jinal N. Bhiman,[1,2,11] Valerie Bekker,[1] Tandile Hermanus,[1] Batsirai Mabvakure,[1,2] Arshad Ismail,[1] M. Anthony Moody,[6] Kevin Wiehe,[6] Nigel Garrett,[7] Salim Abdool Karim,[7] Heini Dirr,[3] Manuel A. Fernandes,[8] Yasien Sayed,[3] Lawrence Shapiro,[4,9] Guido Ferrari,[6] Barton F. Haynes,[6] Penny L. Moore,[1,2,7] and Lynn Morris[1,2,7,12,*]

[1]National Institute for Communicable Diseases of the National Health Laboratory Service, Johannesburg 2131, South Africa
[2]Faculty of Health Sciences, University of the Witwatersrand, Johannesburg 2000, South Africa
[3]Protein Structure-Function Research Unit, University of the Witwatersrand, Johannesburg 2000, South Africa
[4]Department of Biochemistry and Molecular Biophysics, Columbia University, New York, NY 10027, USA
[5]Department of Systems Biology, Zukerman Mind Brain Behavior Institute, Columbia University, New York, NY 10027, USA
[6]Duke Human Vaccine Institute, Duke University, Durham, NC 27710, USA
[7]Centre for the AIDS Programme of Research in South Africa (CAPRISA), University of KwaZulu Natal, Durban 4041, South Africa
[8]Molecular Sciences Institute, School of Chemistry, University of the Witwatersrand, Johannesburg 2000, South Africa
[9]Vaccine Research Center, National Institute of Allergy and Infectious Diseases, NIH, Bethesda, MD 20892-9806, USA
[10]These authors contributed equally
[11]Present address: The Scripps Research Institute, La Jolla, CA 92037, USA
[12]Lead Contact
*Correspondence: lynnm@nicd.ac.za
https://doi.org/10.1016/j.celrep.2018.11.058

SUMMARY

Antibodies that bind residue K169 in the V2 region of the HIV-1 envelope correlated with reduced risk of infection in the RV144 vaccine trial but were restricted to two ED-motif-encoding light chain genes. Here, we identify an HIV-infected donor with high-titer V2 peptide-binding antibodies and isolate two antibody lineages (CAP228-16H/19F and CAP228-3D) that mediate potent antibody-dependent cell-mediated cytotoxicity (ADCC). Both lineages use the IGHV5-51 heavy chain germline gene, similar to the RV144 antibody CH58, but one lineage (CAP228-16H/19F) uses a light chain without the ED motif. A cocrystal structure of CAP228-16H bound to a V2 peptide identified a IGLV3-21 gene-encoded DDxD motif that is used to bind K169, with a mechanism that allows CAP228-16H to recognize more globally relevant V2 immunotypes. Overall, these data further our understanding of the development of cross-reactive, V2-binding, antiviral antibodies and effectively expand the human light chain repertoire able to respond to RV144-like immunogens.

INTRODUCTION

The RV144 vaccine trial was the first HIV-1 phase III human clinical trial to show a favorable outcome, with a moderate 31.2% efficacy (Rerks-Ngarm et al., 2009). A case-control analysis showed that cross-reactive binding antibodies specific for the V2 region of HIV-1 Env correlated inversely with infection risk (Haynes et al., 2012). Selection pressure was identified at positions 169 and 181 in V2, and vaccine efficacy was greater (~48%) when viruses had a matched V2 K169 residue (Rolland et al., 2012). The V2 region of Env is also a target for broadly neutralizing antibodies (bNAbs), and position 169 is often a mutational hot spot for antibody escape during infection (Moore et al., 2011; Wibmer et al., 2013, 2015). However, since vaccine-induced V2 antibodies were only able to neutralize sensitive tier 1 virus strains, exactly how these antibodies contributed to reduced infection rates remains unknown (Rerks-Ngarm et al., 2009).

The isolation of several monoclonal antibodies (mAbs) from RV144 vaccinees that bound short, linear, V2 peptide epitopes spanning positions 166–182 of Env, have provided possible mechanisms of action. These V2 mAbs were able to mediate antibody-dependent cell-mediated cytotoxicity (ADCC) against $CD4^+$ target cells coated in Env derived from breakthrough infections (Liao et al., 2013) and subtype AE HIV-1 CM235 infectious molecular clone-infected target cells (Pollara et al., 2014). Among RV144 vaccinees, 28% had V2 plasma antibody levels sufficient to mediate ADCC and tier 1 neutralization (Rerks-Ngarm et al., 2009). Furthermore, the binding footprints of the isolated V2 mAbs overlap with a tripeptide motif in V2 (position 179–181) that binds the adhesion receptor $\alpha4\beta7$ (Arthos et al., 2008). Antibodies that block this interaction have been shown to reduce simian immunodeficiency virus (SIV) acquisition and impact pathogenesis in the non-human primate model (Byrareddy et al., 2014, 2016).

This is an open access article under the CC BY-NC-ND license (http://creativecommons.org/licenses/by-nc-nd/4.0/).

Structural studies of two RV144 mAbs, CH58 and CH59, showed that the V2 peptide epitope can be structurally dynamic, as both mAbs bound to different conformations of V2 (Liao et al., 2013). CH58 recognizes region 167–176 as an α-helix, and 177–181 as coiled loop, while CH59 binds region 168–173 as a coil, and 174–176 as a short 3_{10}-helix. This differs from bNAbs, which only bind conformational epitopes that include region 167–181 as one of five β-strands forming part of a "Greek key motif" in prefusion Env trimers (Gorman et al., 2016; Julien et al., 2013; McLellan et al., 2011; Pancera et al., 2014). Non-neutralizing V2 antibodies previously isolated from HIV-1-infected donors (such as 830A) also bind to this five-stranded β-barrel conformation of V2, but their epitopes are not exposed on native, pre-fusion trimers (Pan et al., 2015). Similarly, the peptide epitopes recognized by vaccine antibodies CH58 and CH59 are likely only present on aberrantly processed or refolded forms of V2, which coexist with native pre-fusion Env on the viral membrane (Moore et al., 2006). These aberrant forms of Env are the primary target for non-neutralizing antibodies mediating Fc effector functions, which may have contributed toward the reduced infection risk in RV144.

In the case of RV144 V2 antibodies, these antiviral activities were largely dependent on the interactions with positively charged residues at positions 168, 169, and 171 in V2. In both CH58 and CH59, this binding was mediated predominantly by a glutamic-aspartic amino acid pair (ED motif) in the antibody CDRL2 (Liao et al., 2013; Wiehe et al., 2014), which is structurally preconfigured in the naive B cell to interact with V2 (Nicely et al., 2015). Recognition of V2 by other RV144 mAbs, as well as mAbs from vaccinated non-human primates, also required the same ED motif, although alternative solutions could potentially be found among other animal models that lack an appropriate ED-motif-encoding genome (Wiehe et al., 2014, 2017). In humans the ED motif is encoded by only two functionally expressed light chain germline genes, IGLV3-10 and IGLV6-57 (Lefranc, 2001). This suggests a limited capacity for eliciting these types of potentially protective antibodies, even though the V2 region is highly immunogenic, and 20%–45% of HIV-1-infected individuals develop V2-specific antibodies (Israel et al., 1997; Kayman et al., 1994).

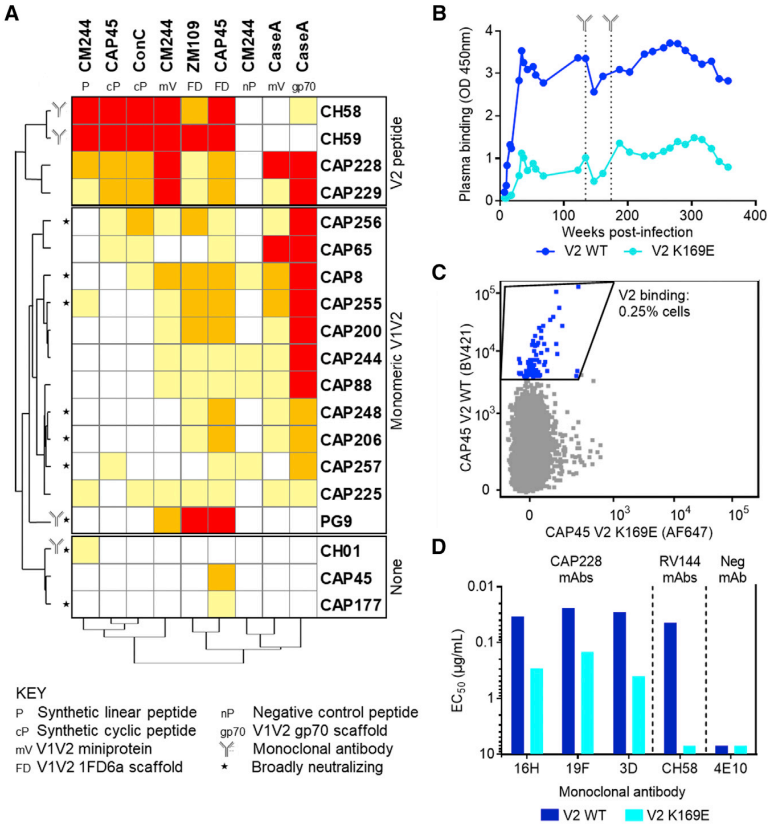
Therefore, to assess whether other modes of K169 binding existed, we screened HIV-positive plasma samples for K169-dependent, V2 peptide-binding antibodies, and isolated three mAbs that resembled the RV144 vaccine antibody CH58. All mAbs used the same IGHV5-51 heavy chain, but two mAbs used a light chain that lacked an ED motif. A crystal structure of one of these, CAP228-16H, bound to a V2 peptide revealed how a surrogate DDxD motif enabled binding to K169. The identification of this third light chain gene capable of binding V2 broadens the K169-binding repertoire that can respond to RV144 vaccine antigens. Overall, these data contribute toward a better understanding of K169-focused V2 epitopes and the antibodies that target them, which are considered essential for RV144 vaccine efficacy.

RESULTS

Isolation of V2 Peptide-Binding Antibodies from HIV-1 Infection

To identify HIV-infected individuals with RV144-like antibodies, we screened plasma samples from 15 HIV-1 subtype C chronically infected donors for binding to V2 peptides and V1V2 scaffolded proteins (Figure 1A). Hierarchical clustering of reactivity was used to divide the plasmas into three groups: V2 peptide-binding antibodies (clustering with CH58/CH59), monomeric V1V2-binding antibodies (clustering with PG9) (Walker et al., 2009), and antibodies with limited-to-no monomer reactivity (clustering with CH01) (Bonsignori et al., 2011). One-half of the selected participants developed bNAbs during infection, including two to the V2 region (CAP256 and CAP257) (Gray et al., 2011a; Moore et al., 2011; Wibmer et al., 2013), but we saw no obvious binding patterns to these antigens among those individuals with bNAbs. Almost all plasma samples contained cross-reactive V1V2 antibodies, with binding to the Case A_gp70 protein used in the RV144 correlates analysis being the strongest. However, only two participants (CAP228 and CAP229) developed high-titer, cross-reactive, V2 peptide-binding antibodies similar to CH58 and CH59 from RV144.

Longitudinal analysis revealed that donor CAP228 developed V2 peptide-binding antibodies within 7 weeks of infection that peaked at 36 weeks and remained high thereafter (Figure 1B). A large proportion of this response was dependent on the K169 residue, as binding to a K169E mutant peptide was considerably reduced. CAP228 did not develop neutralization breadth during this time and maintained relatively low viral loads for 6 years before initiating antiretroviral therapy. To find representative mAbs, memory B cells from CAP228 were isolated at 134 and 174 weeks post-infection (wpi) using a differential K169/E169 single-cell sorting strategy (Figure 1C). A total of 416 cells were sorted, yielding three K169-dependent mAbs that bound strongly to V2 peptides in ELISA (Figure 1D). The mAb CAP228-3D was isolated at 134 wpi only, while a related pair of mAbs, CAP228-16H/19F, were isolated at both 134 and 174 wpi. When compared with RV144-elicited antibodies, the CAP228 antibodies were relatively less affected by the K169E substitution, with only a 10-fold reduction in EC_{50} compared to CH58 where binding was knocked out. CH59 failed to bind the CAP45 V2 peptides (not shown). All three CAP228 mAbs shared a common *IGHV5-51* heavy chain gene also used by CH58 (Figure 1E, highlighted in blue). CAP228-3D also used the same *IGLV6-57* gene as CH58, while CAP228-16H and CAP228-19F (which are clonally related) used a *IGLV3-21* light chain gene (Figure 1E, highlighted yellow and purple). This latter gene does not encode the characteristic K169-binding CDRL2 ED motif but does have potentially analogous negatively charged residues at the same positions (Figure 1E, bolded in red). The CAP228-3D antibody had a 19-amino acid-long CDRH3 (as defined by the ImMunoGeneTics database), like CH58, while the CDRH3s of CAP228-16H/19F were slightly shorter (16 amino acids), although still relatively longer than the average CDRH3 lengths observed for CH59-like antibodies (Wiehe et al., 2014). All three CAP228 antibodies had moderately higher levels of mutation as expected for antibodies from infection when compared to those after immunization. Most noticeably, mutation levels of the two IGLV3-21 light chains (antibodies CAP228-16H/19F) were higher than that of the matched heavy chains, suggesting these may have been under stronger selection pressure.



E

		Heavy Chain				Light Chain			
		IGHV	IGHJ	CDR3 length (amino acid)	Nucleotide mutations	IGLV	IGLJ	CDR2 sequence (amino acid)	Nucleotide mutations
CAP228	**16H**	5-51	3	16 (14)*	4.9%	3-21	2	DDSDRPS	8.3%
	19F	5-51	3	16 (14)*	4.9%	3-21	2	DDSDRPS	8.3%
	3D	5-51	2	19 (17)*	3.9%	6-57	3	EDNQRPS	2.1%
RV144	**CH58**	5-51	4	19 (17)*	1.9%	6-57	3	EDNQRPS	1.8%
	CH59	3-9	4	13 (11)*	2.8%	3-10	3	ED SKRPS	0.9%
	HG107	3-9	4	14 (12)*	2.3%	3-10	1	ED SKRPS	0.9%
	HG120	3-23	4	13 (11)*	1.6%	3-10	2	ED SKRPS	0.9%

*CDR lengths defined by the ImMunoGeneTics database, with Kabat lengths given in parentheses.

Figure 1. Isolation of V2p-Binding Antibodies from an HIV-1-Infected Donor

(A) Hierarchical clustering of 15 HIV-1$^+$ plasma samples and four anti-V2 mAbs (labeled on the right) based on their relative reactivity with V2 peptides or scaffolded V1V2 antigens (labeled on the top and in the key) in an ELISA. Plasma samples and mAbs with broadly neutralizing activity are indicated with an asterisk on the left-hand side. Binding intensity was recorded as a relative percentage of maximum OD, where 0%–40% was colored yellow; 40%–80%, orange; and 80%–100%, red. Samples were grouped either with mAbs CH58 and CH59 (based on strong binding to peptides), with PG9 (based on binding to monomeric V1V2 proteins), or with CH01 (based on no/limited binding to the peptides or proteins).

(B) ELISA binding of CAP228 longitudinal plasma to CAP45 wild-type (blue) and K169E mutant (cyan) cyclical V2 peptides. Absorbance readings are plotted on the y axis against time on the x axis. The 134 and 174 wpi time points at which mAbs were isolated from stored PBMC samples are indicated with the vertical dotted lines.

(C) Fluorescence-activated cell sorting (FACS) plot showing the CAP228 memory B cells (in blue) that bind BV421-labeled wild-type (y axis), but not AF647-labeled K169E mutant (x axis) V2 peptide antigens, and were therefore sorted for PCR amplification.

(D) ELISA titers (y axis) of three CAP228 mAbs (16H, 19F, and 3D) to both wild-type and K169E mutant V2 peptides from CAP45 (colored as in B), when compared to the RV144 vaccine antibody CH58 (CH59 did not bind to CAP45 V2 peptides). 4E10 is an HIV-1 gp41 peptide-reactive mAb used as a negative control (Buchacher et al., 1994).

(E) Properties of the CAP228 antibodies, defined by the International Immunogenetics Database (IMGT), compared to vaccine-elicited mAbs CH58 and CH59. CDRH3 lengths are also given for Kabat numbering in parentheses. Shared immunoglobulin germline genes are shaded, and the negatively charged amino acids in each CDRL2 are shown in red.

The K169-Dependent V2-Binding Repertoire Is Not Limited to ED-Motif Germline Genes

To trace the ontogeny of the CAP228 antibody lineages, gene-specific antibody transcripts were sequenced from peripheral blood mononuclear cells (PBMCs) at seven time points using the Illumina MiSeq platform. Identity-divergence plots revealed clonally related heavy and light chains as early as 7 wpi (Figures S1A and S1B; Table S1), consistent with ELISA mapping data showing the rapid elicitation of V2-binding antibodies (refer to Figure 1B). Despite this rapid emergence, the CAP228 V2 antibodies were not autoreactive and were thus unlikely to have arisen from a preexisting polyreactive B cell pool (Table S2). Although CAP228-16H was isolated at 134 wpi, exact matches to the heavy chain of this antibody were detected at 96, 134, and 174 wpi, while CAP228-16H/19F lineage-related reads were detected at all time points sequenced (Figures S1A, S2A, and S2B). In contrast, while CAP228-3D clonally related light chains were found at all time points sampled, related heavy chains were only detected above threshold at 47 wpi (Figures S1B, S3A, and S3B). CAP228-3D uses the same heavy chain V gene as CAP228-16H/19F and was thus amplified using the same IGHV5-51-specific primers, suggesting a sampling bias that might represent increased clonal expansion of this family when compared to the CAP228-3D lineage. Maximum-likelihood phylogenetic trees for the heavy and light chains of the CAP228-16H/19F lineage (and the light chains of the CAP228-3D lineage) were drawn using the isolated antibody sequences and clonally related MiSeq reads to estimate lineage evolution

(Figures 2A and S1C). The CAP228-16H/19F unmutated precursor (or UCA) was identified directly from these data using the least mutated common ancestor (LMCA) of clonally related heavy and light chain sequences from 7 wpi, while the CAP228-3D precursor was calculated by germline reversion of the heavy chain, matched with the LMCA light chain from related reads at 7wpi.

The predicted CAP228-16H/19F light chain UCA shared the highest similarity with the IGLV3-21*01 allele, which encodes a tyrosine-aspartic acid pair (YD) at position 49–50 in the CDRL2 rather than the two aspartic acids (DD) observed in the CAP228-16H/19F lineage (Figure 2B, highlighted yellow and blue). However, the CAP228-16H UCA also shared a synonymous allelic similarity with the DD-encoding IGLV3-21*02/03 alleles at position 92 (Figure 2B, highlighted red). To confirm the origin of the CAP228-16H/19F lineage, the germline IGLV3-21 alleles of CAP228 were amplified from genomic DNA and sequenced. This revealed two previously undocumented IGLV3-21 alleles, IGLV3-21*01m and IGLV3-21*01mm (using a similar nomenclature for novel alleles as previously reported in Scheepers et al., 2015), where IGLV3-21*01m had a single synonymous mutation at position 92 (also found in IGLV3-21*02/03) when compared to IGLV3-21*01. IGLV3-21*01mm further differed from IGLV3-21*01m by also including the non-synonymous change from Y to D at position 50 (Figure 2B, bottom row). Both the IGLV3-21*01m and IGLV3-21*01mm predicted UCA variants were synthesized and tested for their ability to bind the transmitted/founder autologous CAP228 V1V2 in ELISA. Only the DD-encoding UCA (IGLV3-21*01mm) bound the antigen (Figure 2C, blue and yellow lines). Together with longitudinal lineage sequencing, these data suggested that the IGLV3-21*01mm allele with the CDRL2-encoded DD pair initiated the CAP228-16H/19F lineage.

Since the IGLV3-21*01mm allele has not been previously described, we also assessed the known IGLV3-21*02/03 DD-encoding alleles for their ability to engage V2 antigens. Both the I48V mutation (that reverted the CAP228-16H/19F UCA to IGLV3-21*03) as well as the K17Q mutation (converting the CAP228-16H/19F UCA to IGLV3-21*02) bound equally well to V1V2 (Figure 2C, pink and purple lines), confirming the relevance of both common IGLV3-21 allelic variants in the V2-targeted immune response. An analysis of allelic frequencies using ExaC (Lek et al., 2016) suggested that the average prevalence of DD-encoding IGLV3-21 alleles in the general population is 60% (Figure 2D). The only population group to show significant variation from this was East Asia, where the frequency was much lower at 26%, perhaps explaining why this light chain was not represented among RV144 mAbs. Overall, the identification of this additional light chain broadens the pool of potential K169-dependent V2-reactive human B cells able to respond to RV144 antigens.

The IGLV3-21 CDRL2 DDxD Motif Interacts with V2 Residues K168, K169, and Y173

As CAP228-16H/19F lacked the CDRL2 ED motif used by vaccine-elicited mAbs to interact with positively charged residues at positions K168 and K169/K171, we hypothesized that the negatively charged DD residues in the IGLV3-21 CDRL2 could make similar contacts with V2. To investigate this, a cocrystal structure of CAP228-16H bound to the heterologous CAP45 V2 peptide (residues 164–182) was determined at 2.6 Å resolution (Figure 3; Table 1). CAP228-16H bound to V2 in a conformation that was highly similar to CH58 (Liao et al., 2013), with the N-terminal half of the peptide adopting an α-helical conformation, and the C-terminal portion, a random coil structure (Figure 3A). Electrostatic analysis of the CAP228-16H light chain paratope revealed a large anionic patch, created by three aspartic acid residues in the CDRL2 at positions 50, 51, and 53 (Figure 3B, highlighted in red). This formed a DDxD motif that was analogous to the ED motif of CH58 (positions 50 and 51), but covering a relatively larger surface area.

An aspartic acid at position D51 was common to both antibodies and was similarly oriented to interact with K168 in V2 (Figure 3C, boxed in pale blue). In contrast, the D50 side chain of CAP228-16H is one carbon shorter than the E50 residue of CH58, potentially weakening any interactions with K169. Instead, K169 forms a surrogate water-mediated interaction with the IGLV3-21 germline encoded D53 residue, broadening the K168-K169 recognition motif to include an additional negatively charged side chain (Figure 3C, boxed in pale red). The space vacated by K169 is instead occupied by the CAP45 Y173 residue, which forms a hydrogen bond with D50 that is not seen in the interaction between CH58 and the TH023 V2 peptide, containing the H173 immunotype (Figure 3C, boxed in pale green). To assess the effectiveness of the IGLV3-21 CDRL2 as a surrogate for the IGLV3-10 or IGLV6-57 encoded ED motif, position 50 of the CAP228-16H light chain was mutated to glutamic acid (E) and assessed for binding to the transmitted/founder autologous CAP228 V1V2 in ELISA. The ED version of CAP228-16H bound equally well to V2 when compared with the original DDxD version, with only a twofold difference in EC_{50} (Figure 3D). Altogether, these data confirm the suitability of the DDxD motif to serve as a surrogate for the ED motif, by forming a larger anionic patch that recognizes V2 residues K168 and K169, and also forms interactions with position Y173.

The tyrosines at V2 residues 173 and 177 are also potentially subject to tyrosine sulfation, which is thought to stabilize prefusion trimers through CCR5 mimicry (Cimbro et al., 2014, 2016). Intriguingly, while the V2 peptide used here for structural study was not synthesized with sulfated tyrosines, we observed additional density around position Y177, which we attributed to a sulfate ion that facilitated cocrystallization of the CAP228-16H/V2 complex by also forming important crystal lattice contacts. This sulfate ion was likely from the mother liquor during crystallization but was fortuitously oriented to occupy the same space as a sulfated tyrosine would on virion-associated V1V2 (Figure 3E). An arginine at position 97 in the CAP228-16H CDRH3 interacts directly with the Y177-associated sulfate ion, suggesting that during HIV-1 infection the CAP228-16H lineage could have evolved to recognize a post-translationally sulfated Y177 residue. This ability to bind sulfated tyrosines potentially increases the functional activity of the CAP228 antibodies.

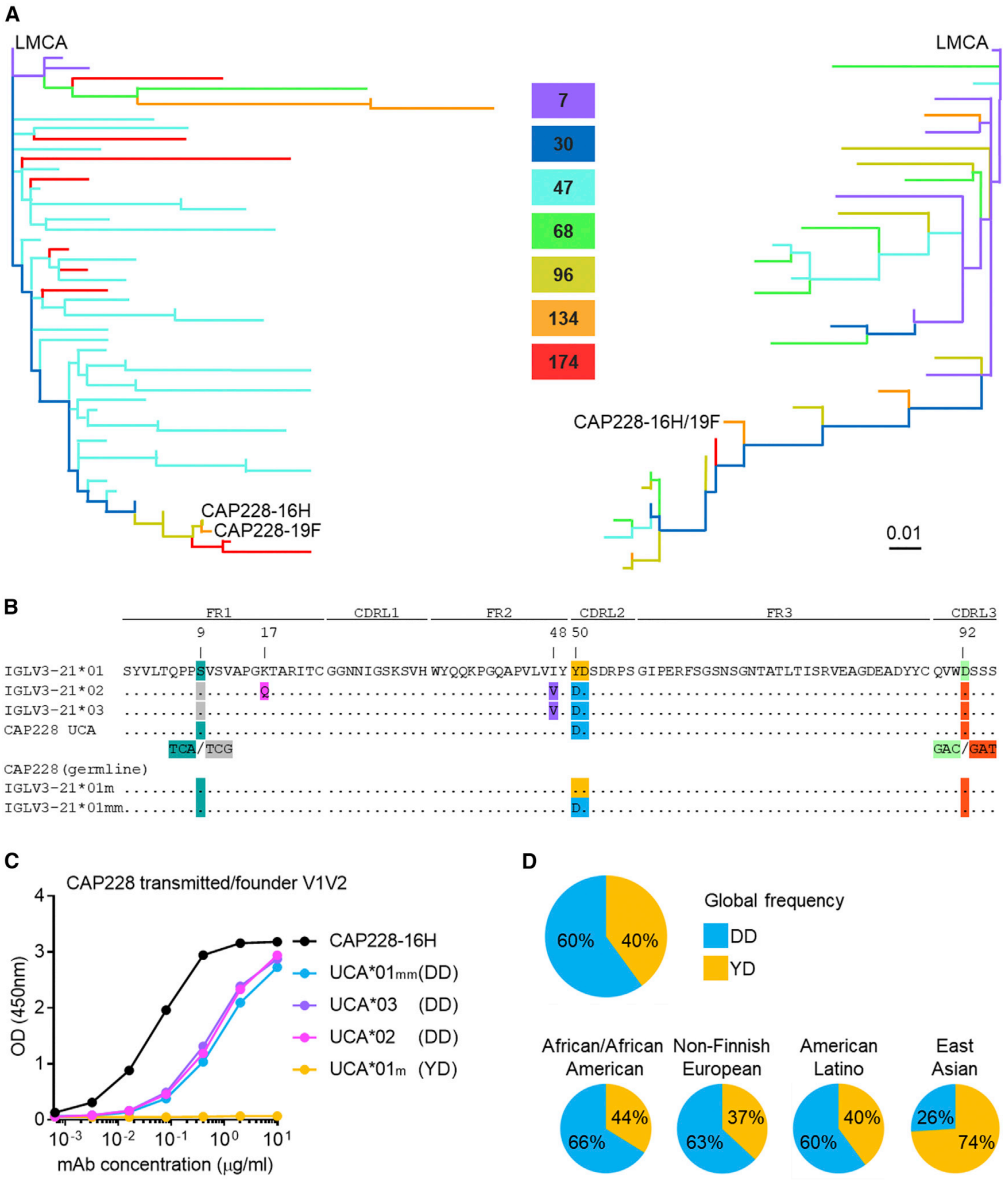


Figure 2. Identification of the CAP228-16H/19F UCA and Lambda IGLV3-21 Germline Alleles

(A) Maximum-likelihood phylogenetic trees of unique, clonally related sequences, for the heavy (left) and light (right) chain of the CAP228-16H/19F lineage, with branches colored by time point. Mature CAP228-16H and 19F mAb sequences are colored according to the time point at which they were first observed rather than when they were isolated. Trees were rooted to the least mutated common ancestor (LMCA).

(B) Alignment of the two additional CAP228 IGLV3-21 germline alleles, when compared to the IMGT sequences for IGLV3-21*01, IGLV3-21*02, and IGLV3-21*03. Shared synonymous (positions 9 or 92) and non-synonymous (positions 17, 48, or 50) differences are indicated with alternate shading, where the key YD/DD amino acid pair is colored yellow or blue, respectively.

(C) ELISA binding of the CAP228-16H/19F UCA to an autologous transmitted/founder CAP228 V1V2 scaffolded protein. Absorbance readings are plotted on the y axis versus mAb concentration on the x axis. Binding of the CAP228-16H/19F UCA allelic variants for IGLV3-21*02 (pink) and IGLV3-21*03 (purple) are also shown. IGLV3-21*01 shares the same amino acid sequence as IGLV3-21*01m.

(D) Frequency of the DD (blue) or YD (yellow) motifs among global allelic variants of the IGVL3-21 germline genes (data from The 1000 Genomes Project [Auton et al., 2015]).

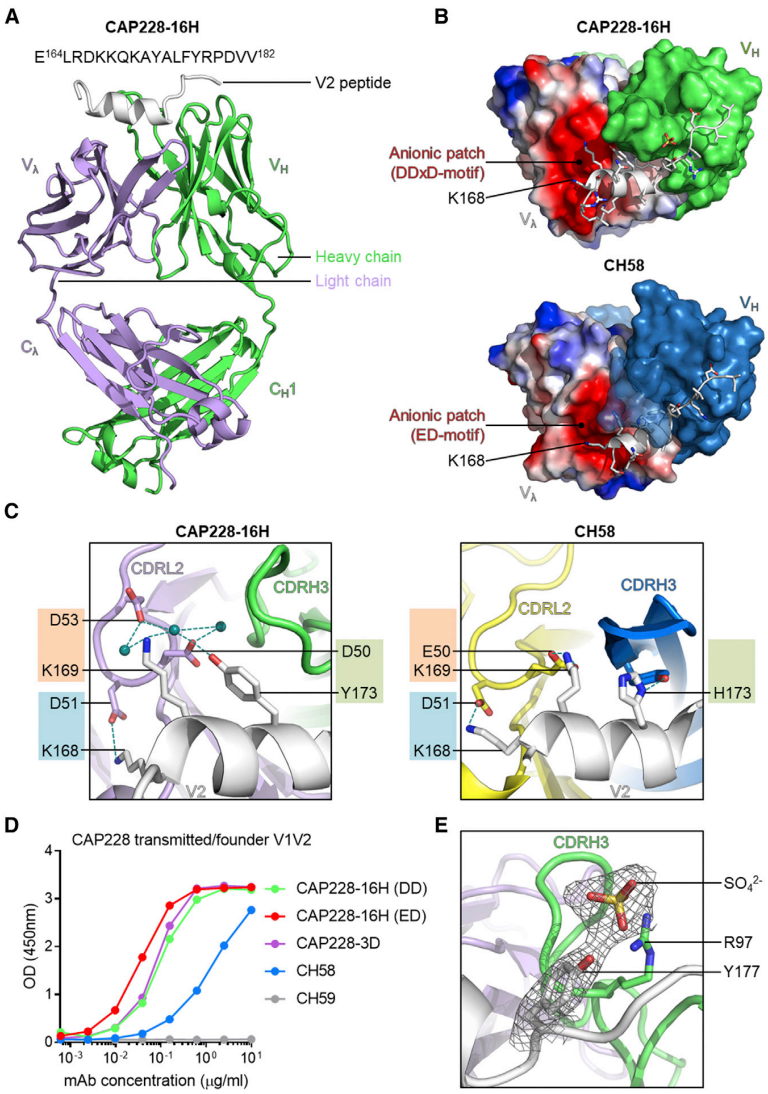


Figure 3. Atomic-Level Details of the Interaction between CAP228-16H and V2 Residues

(A) Cartoon representation of the CAP228-16H Fab domain bound to V2 peptide residues 164–182 from viral strain CAP45 (colored white). Heavy and light chains are colored green and purple, respectively, and the variable or constant domains of the Fab are labeled.

(B) Surface representation of the CAP228-16H paratope (top), compared to the CH58 paratope (bottom), with the bound V2 peptide depicted as a helix-loop cartoon with side chains shown in stick representation. The Fab heavy chains are colored green and blue, respectively, while the light chains are colored per electrostatic potential. The CH58 CDRH3 was colored with partial transparency to show the underlying V2 peptide. The large electronegative patch encoded by the IGLV3-21 DDxD motif, or the IGLV6-57 ED motif, is indicated, and K168 in V2 is labeled for perspective.

(C) A zoom-in of the atomic detail interactions (shown with side chain sticks) between V2 residues K168, K169, and Y/H173, and the CAP228-16H light chain DDxD or CH58 ED motifs. Hydrogen bonds are indicated with the dotted lines, and water molecules are shown with the blue spheres. The antibody CDRL2 and CDRH3 loops, as well as the interacting amino acid pairs being compared, are labeled.

(D) ELISA showing binding to the autologous transmitted/founder CAP228 V1V2 scaffolded protein by wild-type CAP228-16H (green), and the ED-motif mutant (red), when compared to antibodies that naturally contain the ED motif: CAP228-3D (purple), CH58 (blue), and CH59 (gray). Absorbance readings are plotted on the y axis versus mAb concentration on the x axis.

(E) 2Fo-Fc density map of V2 residue Y177, showing the associated sulfate ion, oriented by CAP228-16H CDRH3 residue R97.

CAP228 Antibodies Mediate ADCC against Globally Prevalent V2 Immunotypes

To compare the epitope of CAP228-16H with the vaccine-elicited antibody CH58, the percentage of accessible surface area for each V2 amino acid side chain that was buried within the respective antibody paratopes was compared in the two crystal structures (Figure 4A). The overall binding profile was strikingly similar, with K168, A/V172, Y/H173, L175, F176, K/R178, and D180 burying more than 50% of their accessible surface area in the antibody paratope and forming major components of the antibody binding sites. These positions are among the most conserved in V2 across globally circulating strains in the Los Alamos National Laboratory HIV-1 Env database and are likely important determinants of the cross-reactivity for V2-binding antibodies (Figure 4B). Several key differences between the two antibodies were also apparent (Figure 4A, highlighted in yellow). Unlike CH58, CAP228-16H did not make strong salt bridges with K169 and K171, both of which are relatively more sequence-variable residues across different strains of HIV-1. CAP228-16H and CH58 both sequester the amino acid side chain at position 173 almost in its entirety but make specific hydrogen bonds with the Y173 or H173 immunotypes, respectively. The Y173 immunotype recognized by CAP228-16H is present in the majority (~65%) of HIV-1 circulating strains (Figure 4B, boxed in yellow), with a higher proportion of H173 found among clade AE viruses (55% when compared to 19% globally), which includes the RV144 vaccine strain. Together with a reduced dependence on sequence-variable residues at positions 169 and 171, and the ability to recognize Y177 sulfation, these data suggest that CAP228-like antibodies might better recognize more globally dominant V2 immunotypes.

Table 1. X-Ray Crystallographic Data and Refinement Statistics (Molecular Replacement) for the Antigen-Binding Fragment of CAP228-16H in Complex with HIV-1 CAP45 V2 Peptide

	CAP228-16H Bound to CAP45 V2 Peptide
Data Collection	
Space group	$P2_12_12_1$
Cell dimensions	
a, *b*, *c* (Å)	40.29, 104.27, 212.22
α, β, γ (°)	90, 90, 90
Resolution (Å)	35.37–2.58 (2.67–2.58)[a]
Unique reflections	26,374 (1,710)
R_{merge}	0.158 (0.373)[a]
R_{pim}	0.068 (0.263)[a]
$I / \sigma I$	10.4 (2.1)[a]
CC1/2	0.951 (0.873)[a]
Completeness (%)	90.2 (59.4)[a]
Redundancy	5.1 (2.5)[a]
Refinement	
Resolution (Å)	50–2.58
No. reflections	26,364 (1,709)
R_{work}/R_{free}	0.22/0.25
No. atoms (no H)	6.922
Protein	6,786
Other	20
Solvent	116
B factors (no H)	72
Protein	72
Other	87
Solvent	53
Wilson B-factor (Å^2), all atoms	54
RMS deviations	
Bond lengths (Å)	0.003
Bond angles (°)	0.650
Ramachandran favored %	96.5%
Ramachandran outliers %	0.0%
Rotamer outliers %	0.1%
Molprobity clash score	1.5
MolProbity overall score	1.0
PDB ID	6FY0

[a]Values in parentheses are for highest-resolution shells.

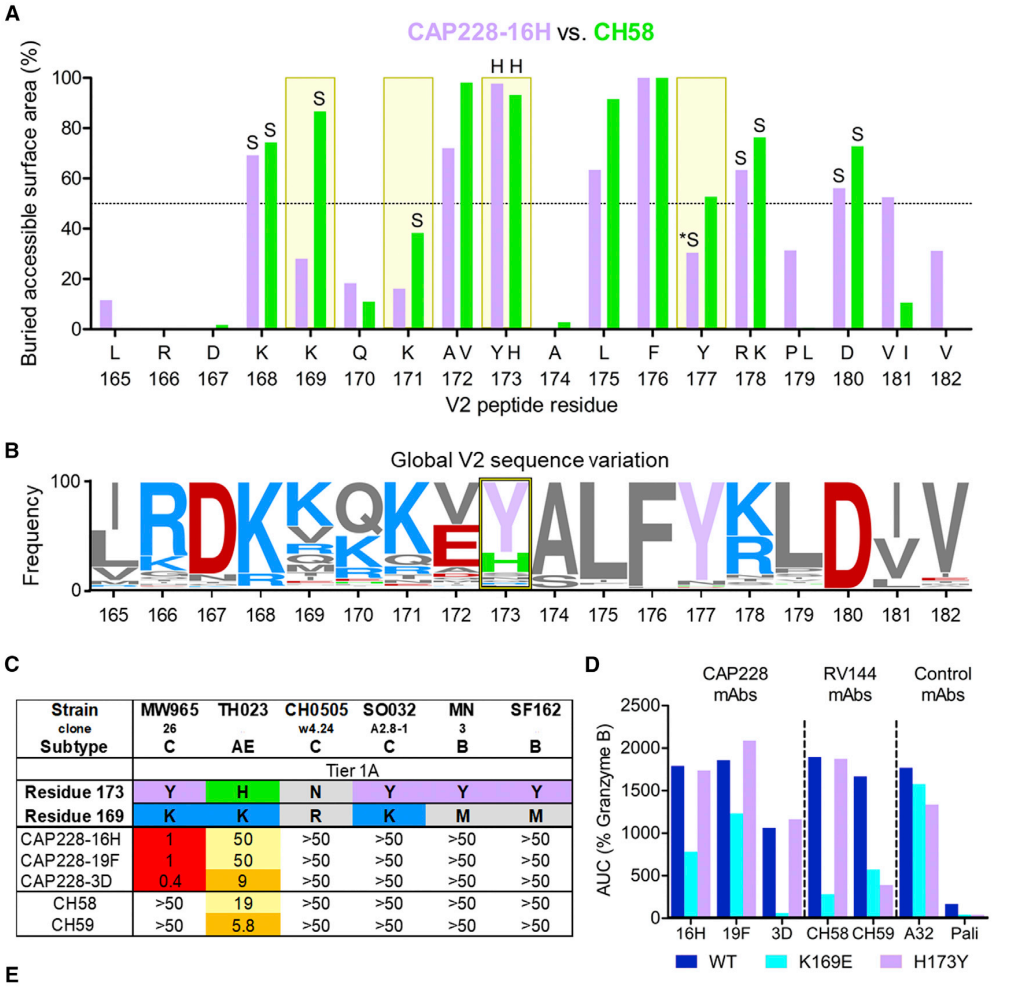
To assess whether these recognition differences might impact on the antiviral activities of V2-binding antibodies, the ability of the CAP228 mAbs to mediate both neutralization and ADCC was tested. Similar to CH58 and CH59, the CAP228 antibodies were able to neutralize some tier 1A viruses in a TZM-bl neutralization assay (Figure 4C), but they displayed marginally more activity than CH58 and were able to neutralize both the subtype AE TH023 (neutralized by CH58 and CH59) and the subtype C MW965 strains (the latter with good potency; highlighted in red on the heatmap). However, none of the other tier 1A viruses, including those from subtype C, was sensitive to any of the CAP228 mAbs. For the most part, tier 1A neutralization sensitivity tracked well with the presence of a lysine at position 169 (with the exception of SO032). These mAbs also failed to neutralize tier 1B and tier 2 viruses, including the CAP228 autologous transmitted/founder virus, consistent with the occlusion of these epitopes on entry-competent envelope trimers (Table S3).

The ADCC activity of the CAP228 mAbs was initially assessed by the granzyme B assay using a clade C gp120 from strain 1086. All mAbs were re-expressed in an IgG1 backbone containing three alanine mutations in the CH2 domain that improves FcγRIIIA engagement (Ferrari et al., 2011; Shields et al., 2001). In this format, all three CAP228 antibodies had ADCC activity comparable to the RV144 vaccine antibodies CH58 and CH59 (Figures 4D and S4A). The mAb A32 is specific for the inner domain of gp120, has a well-defined ADCC activity, and was used as the positive control (Ferrari et al., 2011; Moore et al., 1994), while Palivizumab (a non-HIV respiratory syncytial virus [RSV]-specific mAb) was used as the negative control. The specificity of the ADCC activity was determined by introducing K169E or H173Y mutations into the 1086 gp120. The K169E mutation substantially affected ADCC by CAP228-16H/19F and almost completely abrogated the activity of CAP228-3D and CH58 (Figure 4D, cyan). In contrast, the H173Y mutation only affected the ADCC activity of CH59, consistent with previous studies (Liao et al., 2013). Control antibodies A32 or Palivizumab remained unaffected by either mutation.

The breadth of ADCC mediated by the CAP228 mAbs was then assessed against the autologous transmitted/founder virus, as well as a multiclade panel of viruses (Figures 4E and S4B), using an infected cell assay where cell killing is measured as a reduction of the Luciferase signal provided by the infected cells (Pollara et al., 2014). The CAP228 mAbs showed broader and more potent reactivity compared to CH58 (Figure 4E, highlighted in red), mediating high levels of ADCC against 6 of the 12 infected cell targets including the CAP228 transmitted/founder virus (Figure 4D, labeled in gray). This difference appeared to track with the immunotype at position 173 (Figure 4D, shaded green or purple). RV144 mAb CH58 only showed high ADCC activity against viruses with an H173 immunotype, mediating weak, sporadic killing of cells infected with the Y173 immunotype, while CH59 was intolerant of the Y173 immunotype. The ADCC potency of all five V2 antibodies also tracked well with changes at position 169, where 169R viruses displaying resistance to all mAbs, while 169T or 169I were resistant to the CAP228 mAbs, or CH58 respectively. For a few viruses, ADCC activity was more moderate/weak and differed between the CAP228 mAb lineages. Both CM244 and TV1.21 have longer, more glycosylated V1 loops compared to the other viruses (32- and 37-amino acid loop lengths and 6 glycans, respectively), potentially contributing to resistance. Overall, these data are consistent with ADCC directed to a V2 epitope and demonstrate the ability of CAP228 mAbs to recognize a broader range of viral immunotypes.

DISCUSSION

The V2 region of the HIV-1 envelope has become an increasingly attractive target for vaccine immunogen design following the



Strain	MW965	TH023	CH0505	SO032	MN	SF162
clone	26		w4.24	A2.8-1	3	
Subtype	C	AE	C	C	B	B
	Tier 1A					
Residue 173	Y	H	N	Y	Y	Y
Residue 169	K	K	R	K	M	M
CAP228-16H	1	50	>50	>50	>50	>50
CAP228-19F	1	50	>50	>50	>50	>50
CAP228-3D	0.4	9	>50	>50	>50	>50
CH58	>50	19	>50	>50	>50	>50
CH59	>50	5.8	>50	>50	>50	>50

Strain	1086	CM235	CM244	CAP228	MW965	DU151	TV1	C1080	WITO	CH040	Q23	CH0505
clone	B2	2		2.00.51J	26	2	21	3	c	c	17	w4.24
Subtype	C	AE	AE	C	C	C	C	AE	B	B	A	C
Residue 173	H			Y								N
Residue 169	K								I	T	R	
V1 length*	16	27	32	30	24	16	37	26	23	24	17	16
V1 glycans	3	4	6	4	5	3	6	4	4	4	3	4
CAP228-16H	811	1,264	312	1,435	859	569	14	957	72	0	0	0
CAP228-19F	471	703	55	880	572	497	22	424	22	0	0	7
CAP228-3D	1,002	1,068	0	1,410	680	685	388	907	456	4	0	0
CH58	772	981	268	366	24	115	44	271	12	105	0	0
CH59	1172	617	0	18	0	0	0	0	0	0	0	0
Palivizumab	0	6	0	0	0	0	0	0	0	0	0	0
A32	268	475	196	9	152	278	1038	127	899	120	122	314

* Including V1 flanking glycosylation sequons N130 and N156

Figure 4. Increased V2 Coverage and Broader ADCC Activity Mediated by CAP228 Antibodies

(A) Buried surface area for V2 peptides residues 165–182 (x axis) from the CAP228-16H (purple) or CH58 (green) cocrystal structures, plotted as a percentage of the total accessible surface area (y axis). Significant differences between the two structures are boxed in yellow, and hydrogen bonds (H) or salt bridges (S) are labeled. The potential interaction between CAP228-16H and a sulfated tyrosine at position Y177 is indicated with the asterisk.

(B) Global frequency (y axis) of amino acids occurring at V2 positions 165–182 (x axis), from non-redundant sequences in the HIV-1 Los Alamos National Laboratory (LANL) database, aligned with V2 residues in (A). Positively or negatively charged residues are colored blue or red, while the tyrosines recognized by CAP228-16H, and the histidine required for CH58 are colored purple and green, respectively. Position 173 is boxed in yellow.

(legend continued on next page)

demonstration that V2 peptide-binding antibodies correlated with reduced risk of infection in the RV144 vaccine trial (Haynes et al., 2012; Rerks-Ngarm et al., 2009). Recognition of these V2 antigens relies on key interactions with lysine residues at positions 168, 169, and 171 through a mechanism thought to be restricted to two light chain V genes. Here, we identify an additional light chain V gene that uses an alternative mode of recognition, allowing for broader reactivity with global strains. The CAP228 mAbs, isolated from an HIV-1 subtype C-infected individual, were shown to mediate ADCC against a larger panel of infected cells compared to mAbs from RV144. Overall, these data improve our understanding of the immunogenicity of V2 and suggest an increased ability to respond to RV144-like vaccines in populations where suitable alleles are more prevalent.

The isolation of antibodies from HIV-1-infected donors that resemble those from vaccine recipients suggests that RV144 mAbs were elicited by envelope conformations that are naturally expressed during HIV-1 infection. As such, studies of mAbs isolated from donors such as CAP228 are highly relevant for understanding how the RV144 vaccine may have mediated its protective effect. Screening of donor sera revealed that while antibodies to V2 were commonly elicited during chronic HIV-1 infection, only 2 of 15 individuals showed a dominant binding pattern similar to the RV144 mAbs CH58 and CH59. This suggests that cross-reactive, linear peptide-associated, K169-dependent V2 response may be less immunodominant in HIV infection, and perhaps explains why these types of V2 mAbs have not previously been isolated from infected donors. In keeping with this, we found the frequency of V2 K169-specific B cells in CAP228 to be extremely low, as less than 1 per 25,000 memory B cells were isolated by the differential sorting approach.

Transcripts from both CAP228 mAb lineages were detected at 7 wpi, suggesting that they arose early, and possibly simultaneously, co-incident with the appearance of V2-binding antibodies in CAP228 serum. All of the CAP228 mAbs showed subtle differences in V2 recognition compared to CH58, including being considerably less dependent on residue K169, which may reflect their need to adapt to an evolving viral antigen. Of the two lineages, CAP228-3D was most CH58-like in that it shared the same IGLV6-57 light chain gene. However, transcripts from the CAP228-3D lineage were found at very low frequency, with detection of the heavy chain only at a single time point. In contrast, the CAP228-16H/19F lineage heavy and light chain transcripts were seen across all time points, peaking at 47 weeks (which coincided with high V2-binding titers) and appeared to be the dominant response. Whether or not the unique heavy-light chain pairing of the CAP228-16H/19F lineage conferred a selective advantage is still to be determined. Similarly, the isolation of two lineages that used the same heavy chain as CH58 may indicate a genetic advantage for IGHV5-51 when accessing the V2 site that warrants further study.

Antibodies able to bind V2 peptides were thought to be limited to the two germline light chain genes IGLV3-10 and IGLV6-57 that contain a pre-configured ED motif (Wiehe et al., 2014). Similarly, mAbs isolated from V2-immunized rhesus macaques almost always used the IGLV3-17 light chain gene (an ortholog of human IGLV3-10) (Wiehe et al., 2014). This suggested a potential genetic restriction within the naive repertoire of B cells able to respond appropriately to RV144-type vaccines. Our identification of an additional light chain gene used by CAP228-16H/19F, which can form productive K169-dependent V2 peptide interactions, is therefore encouraging. Structural studies revealed that this IGLV3-21 light chain uses a different CDRL2 arrangement to bind K169, which allowed us to define a DDxD motif that serves as a surrogate for the ED motif. Generation of an artificial antibody with an EDxD motif revealed equivalent binding activity, indicating that this allele can accommodate the traditional mode of binding, although the precise interactions would need to be confirmed structurally. Two of the International Immunogenetics Database (IMGT)-defined IGLV3-21 alleles (*02 and *03) have this DDxD motif, while the third IGLV3-21*01 allele has a YDxD motif. Interestingly, germline sequencing of IGLV3-21 from donor CAP228 revealed two alleles not previously reported in the IMGT database (Lefranc, 2001), indicating the importance of more inclusive online databases. Expression and testing of the UCA with a YDxD motif (IGLV3-21*01m) suggested that B cells carrying this allele were not engaged by the infecting virus. Rather, this lineage derived from the IGLV3-21*01mm allele, containing the DDxD motif. Reversion of IGLV3-21*01mm allele to contain the *02 and *03 amino acid changes showed no impact on binding to the autologous CAP228 transmitted/founder V1V2, suggesting that all DDxD-motif-containing alleles are equally functional in binding V2 antigens. The discovery of two undocumented alleles in one individual, in addition to the three previously reported alleles within the IGLV3-21 light chain gene, suggests that significant variation exists within the African genome as previously demonstrated for the heavy chain variable region (Scheepers et al., 2015). Since the IGLV3-21*01mm allele contains the DDxD motif, this increases the repertoire of light chain alleles able to bind V2 and could increase V2 antibody response rates and titers to RV144-like vaccines.

In addition to expanding the B cell repertoire able to recognize K169, the IGLV3-21 gene used by CAP228-16H also forms productive interactions with Y173 in V2, potentially conferring increased recognition of heterologous strains. Consistent with this, CH58 showed strong ADCC against viruses with the H173 immunotype, while CAP228 antibodies recognized viruses that

(C) Neutralization activity of the CAP228 mAbs, or RV144 mAbs CH58 and CH59, in the TZM-bl assay against six tier 1A HIV-1 viruses. IC_{50} titers ≤ 1 μg/mL were colored red, while those between 1 and 10 μg/mL are shown in orange, and those between 10 and 50 μg/mL are shown in yellow.

(D) Mapping the V2-directed ADCC activity by the granzyme B assay against HIV-1 1086 gp120-coated target cells. Measurements represent area under the curve (y axis) for wild-type (blue), K169E (cyan), or H173Y (purple) mutant proteins.

(E) ADCC activity of the CAP228 mAbs against CEM.NKR-CCR5 cells infected with a multi-clade panel of 12 HIV-1 strains, compared to CH58 and CH59. Data report percentage activity versus antibody concentration as area under the curve, above 15% background, and are displayed as a heatmap with >400 shown in red; 200–400, in orange; 40–200, in yellow; and <40, in white. Relevant genotypic information for each envelope strain is also shown, grouped according to immunotypes at residues 169 and 173. The mAb A32 with known ADCC activity was included as a positive control and showed reactivity against all infected cells, while Palivizumab (a non-HIV antibody) was included as a negative control.

contained both the H173 as well as the more globally prevalent Y173 immunotype. These data suggest that the ability to accommodate Y173, in part due to the use of this newly implicated IGLV3-21 light chain, may have resulted in broader ADCC activity, and these results are consistent with the relatively high ADCC activity previously measured in CAP228 plasma (Richardson et al., 2018). Antibodies from RV144, such as CH58, were elicited by a vaccine strain with an H173 immunotype that is relevant to clade AE (which predominates in Thailand), while CAP228 antibodies were elicited by a virus with a Y173 immunotype. While this likely contributed to selection of IGLV3-21 used by the CAP228-16H/19F mAbs, the CAP228-3D mAb (which uses an ED-motif-containing light chain) was also elicited by the same Y173-containing viruses. CAP228-3D interactions with Y173, like CH58 interactions with H173, are exclusively mediated by the heavy chain CDRH3. Another RV144 mAb, CH59, binds to a different H173-dependent conformation of V2 but failed to bind the peptides or V1V2 scaffolded proteins used in this study with the more globally prevalent Y173 immunotype. The specific immunotype presented at position 173 might therefore form an important antigenic determinant of V2 immunogens. Similarly, the ability to recognize a sulfated Y177 residue, which is less frequently observed on recombinant immunogens when compared with virus associated V1V2 (Cimbro et al., 2014), may improve V2 mAb antiviral activities. These data highlight key differences between CAP228 antibodies elicited through HIV-1 infection, which were more cross-reactive with circulating viral strains than mAbs elicited by vaccination, and suggest that antigens more representative of the diversity within V2 could elicit broader and more potent antiviral responses than RV144.

ADCC activity was identified as a correlate of reduced risk of HIV-1 infection, and mAbs isolated from RV144 also showed strong ADCC activity (Bonsignori et al., 2012; Haynes et al., 2012; Pollara et al., 2014). Here, we show that the ADCC activity of all tested V2 mAbs was largely restricted to cells infected with viruses that harbored a K169 residue in V2, although the IGHV5-51/IGLV3-21 pairing appeared to be more tolerant of variation at this position. This further supports the genetic signature analysis showing that breakthrough infections in RV144 were under-represented for K169 (Rolland et al., 2012). While this has been attributed to a genetic sieve effect, i.e., that K169-containing viruses were selectively prevented from establishing infection, mutations away from K169 may also have arisen post-acquisition as a result of antibody selection pressure, possibly through ADCC (Rolland et al., 2012). Increasingly, the Fc effector functions of antibodies have been shown to play an important role in antiviral efficacy (Horwitz et al., 2017), but direct evidence for viral selection is limited mainly because of the difficulty of analyzing the role of ADCC separately from that of antibody neutralizing activities.

The RV144 trial conducted in Thailand is the only HIV-1 vaccine to show some level of efficacy, and follow-up studies of this approach are being actively pursued. In the HVTN 097 phase I trial, which tested the same vaccines used in RV144, antibody responses to V2 were found to be higher in Africans compared to Thais (Gray et al., 2014). A follow-up study called HVTN 100, that used clade C versions of these vaccine immunogens, passed all pre-set go/no-go criteria, which included V2 antibody response rates (Bekker et al., 2018). These data suggest that V2 antibodies will be readily elicited in HVTN 702, the current efficacy study that is enrolling 5,400 high-risk men and women in South Africa. It remains to be seen whether modifications to the regimen, which includes new HIV-1 clade C components, additional protein boosting, and a different adjuvant, results in higher levels of efficacy compared to RV144. Importantly, this trial will confirm whether or not V2 antibodies with ADCC activity are a correlate of reduced risk of HIV-1 infection. Our data, showing that South Africans can respond to V2 antigens using additional genetic determinants, expands the opportunities for responding favorably to RV144-like vaccines, but confirmation of this awaits isolation and characterization of mAbs from participants in the HVTN 702 trial.

STAR★METHODS

Detailed methods are provided in the online version of this paper and include the following:

- KEY RESOURCES TABLE
- CONTACT FOR REAGENT AND RESOURCE SHARING
- EXPERIMENTAL MODEL AND SUBJECT DETAILS
  - CAPRISA cohort
- METHOD DETAILS
  - Cell sorting
  - Single cell PCR amplification of heavy and light chain variable genes
  - Protein expression and purification
  - ELISA
  - Neutralization assay
  - Infectious molecular clones (IMC)
  - Coated ADCC assay
  - Infected cell ADCC assay
  - Autoreactivity assays
  - Next generation sequencing using the Illumina MiSeq platform
  - Next generation sequencing data processing and analysis
  - Germline gene amplification and cloning
  - X-ray protein crystallography
- DATA AND SOFTWARE AVAILABILITY

SUPPLEMENTAL INFORMATION

Supplemental Information includes four figures and five tables and can be found with this article online at https://doi.org/10.1016/j.celrep.2018.11.058.

ACKNOWLEDGMENTS

We thank the participants of the CAPRISA 002 cohort and the clinical and laboratory staff at CAPRISA for managing the cohort and providing specimens. We are grateful to Jason S. McLellan and Peter D. Kwong (Vaccine Research Centre, NIAID, Bethesda, MD) for providing us with the 1FD6 scaffold, as well as to the MX-2 beamline staff at the Brazilian Synchrotron Light Source (Campinas, Brazil) for help with data collection, Kelsey Hall for technical assistance on the ADCC assay, and Nigel Makoah, Don Mvududu, and Zanele Molaudzi for generating the gp120 proteins. This project was supported by an NIH glycan R01 grant (AI104387); the South African Medical Research Council (MRC); the NHLS Research Trust; South African National Research

Foundation Grant (68898); and the Poliomyelitis Research Foundation (15/83). CAPRISA is funded by the National Institute of Allergy and Infectious Diseases (NIAID), NIH, and U.S. Department of Health and Human Services (Grant AI51794). C.v.E. was supported by the Columbia University-Southern African Fogarty AIDS International Training and Research Program (AITRP) through the Fogarty International Center, National Institutes of Health (Grant 5 D43 TW000231). P.L.M. (Grant 98341) and H.D. (Grant 64788) were supported by the South African Research Chairs Initiative of the Department of Science and Technology and National Research Foundation of South Africa. This research used resources of the Brazilian Synchroton Light Laboratory (LNLS), an open national facility operated by the Brazilian Centre for Research in Energy and Materials (CNPEM) for the Brazilian Ministry for Science, Technology, Innovations and Communications (MCTIC).

AUTHOR CONTRIBUTIONS

Conceptualization, C.v.E., C.K.W., P.L.M., and L.M.; Methodology, C.v.E., C.K.W., Y.S., H.D., L.S., G.F., B.F.H., P.L.M., and L.M.; Cohort, N.G. and S.A.K.; Investigation, C.v.E., C.K.W., M.N., B.L., N.N.M., S.I.R., B.V., S.S.-O., J.N.B., V.B., T.H., A.I., M.A.M., and M.A.F.; Formal Analysis, C.v.E., C.K.W., C.S., S.I.R., Z.S., B.M., and K.W.; Writing – Original Draft, C.v.E., C.K.W., and L.M.; Writing – Review, C.v.E., C.K.W., C.S., S.I.R., H.D., Y.S., L.S., G.F., B.F.H., P.L.M., and L.M.; Supervision and Funding, P.L.M. and L.M.

DECLARATION OF INTERESTS

The authors declare no competing interests.

Received: December 28, 2017
Revised: October 2, 2018
Accepted: November 14, 2018
Published: December 11, 2018

REFERENCES

Adachi, A., Gendelman, H.E., Koenig, S., Folks, T., Willey, R., Rabson, A., and Martin, M.A. (1986). Production of acquired immunodeficiency syndrome-associated retrovirus in human and nonhuman cells transfected with an infectious molecular clone. J. Virol. *59*, 284–291.

Adams, P.D., Afonine, P.V., Bunkoczi, G., Chen, V.B., Davis, I.W., Echols, N., Headd, J.J., Hung, L.W., Kapral, G.J., Gross-Kunstleve, R.W., et al. (2010). PHENIX: a comprehensive Python-based system for macromolecular structure solution. Acta. Crystallogr. D. Biol. Crystallog. *66*, 486–501.

Arthos, J., Cicala, C., Martinelli, E., Macleod, K., Van Ryk, D., Wei, D., Xiao, Z., Veenstra, T.D., Conrad, T.P., Lempicki, R.A., et al. (2008). HIV-1 envelope protein binds to and signals through integrin $\alpha4\beta7$, the gut mucosal homing receptor for peripheral T cells. Nat. Immunol. *9*, 301–309.

Auton, A., Brooks, L.D., Durbin, R.M., Garrison, E.P., Kang, H.M., Korbel, J.O., Marchini, J.L., McCarthy, S., McVean, G.A., and Abecasis, G.R.; 1000 Genomes Project Consortium (2015). A global reference for human genetic variation. Nature *526*, 68–74.

Bekker, L.G., Moodie, Z., Grunenberg, N., Laher, F., Tomaras, G.D., Cohen, K.W., Allen, M., Malahleha, M., Mngadi, K., Daniels, B., et al.; HVTN 100 Protocol Team (2018). Subtype C ALVAC-HIV and bivalent subtype C gp120/MF59 HIV-1 vaccine in low-risk, HIV-uninfected, South African adults: a phase 1/2 trial. Lancet HIV *5*, e366–e378.

Bonsignori, M., Hwang, K.K., Chen, X., Tsao, C.Y., Morris, L., Gray, E., Marshall, D.J., Crump, J.A., Kapiga, S.H., Sam, N.E., et al. (2011). Analysis of a clonal lineage of HIV-1 envelope V2/V3 conformational epitope-specific broadly neutralizing antibodies and their inferred unmutated common ancestors. J. Virol. *85*, 9998–10009.

Bonsignori, M., Pollara, J., Moody, M.A., Alpert, M.D., Chen, X., Hwang, K.K., Gilbert, P.B., Huang, Y., Gurley, T.C., Kozink, D.M., et al. (2012). Antibody-dependent cellular cytotoxicity-mediating antibodies from an HIV-1 vaccine efficacy trial target multiple epitopes and preferentially use the VH1 gene family. J. Virol. *86*, 11521–11532.

Buchacher, A., Predl, R., Strutzenberger, K., Steinfellner, W., Trkola, A., Purtscher, M., Gruber, G., Tauer, C., Steindl, F., Jungbauer, A., et al. (1994). Generation of human monoclonal antibodies against HIV-1 proteins; electrofusion and Epstein-Barr virus transformation for peripheral blood lymphocyte immortalization. AIDS Res. Hum. Retroviruses *10*, 359–369.

Byrareddy, S.N., Kallam, B., Arthos, J., Cicala, C., Nawaz, F., Hiatt, J., Kersh, E.N., McNicholl, J.M., Hanson, D., Reimann, K.A., et al. (2014). Targeting $\alpha4\beta7$ integrin reduces mucosal transmission of simian immunodeficiency virus and protects gut-associated lymphoid tissue from infection. Nat. Med. *20*, 1397–1400.

Byrareddy, S.N., Arthos, J., Cicala, C., Villinger, F., Ortiz, K.T., Little, D., Sidell, N., Kane, M.A., Yu, J., Jones, J.W., et al. (2016). Sustained virologic control in SIV^+ macaques after antiretroviral and $\alpha4\beta7$ antibody therapy. Science *354*, 197–202.

Cimbro, R., Gallant, T.R., Dolan, M.A., Guzzo, C., Zhang, P., Lin, Y., Miao, H., Van Ryk, D., Arthos, J., Gorshkova, I., et al. (2014). Tyrosine sulfation in the second variable loop (V2) of HIV-1 gp120 stabilizes V2-V3 interaction and modulates neutralization sensitivity. Proc. Natl. Acad. Sci. USA *111*, 3152–3157.

Cimbro, R., Peterson, F.C., Liu, Q., Guzzo, C., Zhang, P., Miao, H., Van Ryk, D., Ambroggio, X., Hurt, D.E., De Gioia, L., et al. (2016). Tyrosine-sulfated V2 peptides inhibit HIV-1 infection via coreceptor mimicry. EBioMedicine *10*, 45–54.

DeKosky, B.J., Ippolito, G.C., Deschner, R.P., Lavinder, J.J., Wine, Y., Rawlings, B.M., Varadarajan, N., Giesecke, C., Dörner, T., Andrews, S.F., et al. (2013). High-throughput sequencing of the paired human immunoglobulin heavy and light chain repertoire. Nat. Biotechnol. *31*, 166–169.

Doria-Rose, N.A., Schramm, C.A., Gorman, J., Moore, P.L., Bhiman, J.N., DeKosky, B.J., Ernandes, M.J., Georgiev, I.S., Kim, H.J., Pancera, M., et al.; NISC Comparative Sequencing Program (2014). Developmental pathway for potent V1V2-directed HIV-neutralizing antibodies. Nature *509*, 55–62.

Doria-Rose, N.A., Bhiman, J.N., Roark, R.S., Schramm, C.A., Gorman, J., Chuang, G.Y., Pancera, M., Cale, E.M., Ernandes, M.J., Louder, M.K., et al. (2015). New member of the V1V2-directed CAP256-VRC26 lineage that shows increased breadth and exceptional potency. J. Virol. *90*, 76–91.

Edgar, R.C. (2010). Search and clustering orders of magnitude faster than BLAST. Bioinformatics *26*, 2460–2461.

Edmonds, T.G., Ding, H., Yuan, X., Wei, Q., Smith, K.S., Conway, J.A., Wieczorek, L., Brown, B., Polonis, V., West, J.T., et al. (2010). Replication competent molecular clones of HIV-1 expressing *Renilla* luciferase facilitate the analysis of antibody inhibition in PBMC. Virology *408*, 1–13.

Emsley, P., Lohkamp, B., Scott, W.G., and Cowtan, K. (2010). Features and development of Coot. Acta. Crystallogr. D. Biol. Crystallogr *66*, 486–501.

Ferrari, G., Pollara, J., Kozink, D., Harms, T., Drinker, M., Freel, S., Moody, M.A., Alam, S.M., Tomaras, G.D., Ochsenbauer, C., et al. (2011). An HIV-1 gp120 envelope human monoclonal antibody that recognizes a C1 conformational epitope mediates potent antibody-dependent cellular cytotoxicity (ADCC) activity and defines a common ADCC epitope in human HIV-1 serum. J. Virol. *85*, 7029–7036.

Gorman, J., Soto, C., Yang, M.M., Davenport, T.M., Guttman, M., Bailer, R.T., Chambers, M., Chuang, G.Y., DeKosky, B.J., Doria-Rose, N.A., et al.; NISC Comparative Sequencing Program (2016). Structures of HIV-1 Env V1V2 with broadly neutralizing antibodies reveal commonalities that enable vaccine design. Nat. Struct. Mol. Biol. *23*, 81–90.

Gray, E.S., Moore, P.L., Choge, I.A., Decker, J.M., Bibollet-Ruche, F., Li, H., Leseka, N., Treurnicht, F., Mlisana, K., Shaw, G.M., et al.; CAPRISA 002 Study Team (2007). Neutralizing antibody responses in acute human immunodeficiency virus type 1 subtype C infection. J. Virol. *81*, 6187–6196.

Gray, E.S., Madiga, M.C., Hermanus, T., Moore, P.L., Wibmer, C.K., Tumba, N.L., Werner, L., Mlisana, K., Sibeko, S., Williamson, C., et al.; CAPRISA002 Study Team (2011a). The neutralization breadth of HIV-1 develops

incrementally over four years and is associated with CD4+ T cell decline and high viral load during acute infection. J. Virol. *85*, 4828–4840.

Gray, E.S., Moody, M.A., Wibmer, C.K., Chen, X., Marshall, D., Amos, J., Moore, P.L., Foulger, A., Yu, J.S., Lambson, B., et al. (2011b). Isolation of a monoclonal antibody that targets the alpha-2 helix of gp120 and represents the initial autologous neutralizing-antibody response in an HIV-1 subtype C-infected individual. J. Virol. *85*, 7719–7729.

Gray, G.E., Andersen-Nissen, E., Grunenberg, N., Huang, Y., Roux, S., Laher, F., Innes, C., Gu, N., DiazGranados, C., Phogat, S., et al. (2014). HVTN 097: Evaluation of the RV144 Vaccine Regimen in HIV Uninfected South African Adults. AIDS Res. Hum. Retroviruses *30*, A33–A34.

Guindon, S., Dufayard, J.F., Lefort, V., Anisimova, M., Hordijk, W., and Gascuel, O. (2010). New algorithms and methods to estimate maximum-likelihood phylogenies: assessing the performance of PhyML 3.0. Syst. Biol. *59*, 307–321.

Haynes, B.F., Fleming, J., St Clair, E.W., Katinger, H., Stiegler, G., Kunert, R., Robinson, J., Scearce, R.M., Plonk, K., Staats, H.F., et al. (2005). Cardiolipin polyspecific autoreactivity in two broadly neutralizing HIV-1 antibodies. Science *308*, 1906–1908.

Haynes, B.F., Gilbert, P.B., McElrath, M.J., Zolla-Pazner, S., Tomaras, G.D., Alam, S.M., Evans, D.T., Montefiori, D.C., Karnasuta, C., Sutthent, R., et al. (2012). Immune-correlates analysis of an HIV-1 vaccine efficacy trial. N. Engl. J. Med. *366*, 1275–1286.

Horwitz, J.A., Bar-On, Y., Lu, C.L., Fera, D., Lockhart, A.A.K., Lorenzi, J.C.C., Nogueira, L., Golijanin, J., Scheid, J.F., Seaman, M.S., et al. (2017). Non-neutralizing antibodies alter the course of HIV-1 infection in vivo. Cell *170*, 637–648.e10.

Huson, D.H., and Scornavacca, C. (2012). Dendroscope 3: an interactive tool for rooted phylogenetic trees and networks. Syst. Biol. *61*, 1061–1067.

Israel, Z.R., Gorny, M.K., Palmer, C., McKeating, J.A., and Zolla-Pazner, S. (1997). Prevalence of a V2 epitope in clade B primary isolates and its recognition by sera from HIV-1-infected individuals. AIDS *11*, 128–130.

Julien, J.P., Cupo, A., Sok, D., Stanfield, R.L., Lyumkis, D., Deller, M.C., Klasse, P.J., Burton, D.R., Sanders, R.W., Moore, J.P., et al. (2013). Crystal structure of a soluble cleaved HIV-1 envelope trimer. Science *342*, 1477–1483.

Karasavvas, N., Billings, E., Rao, M., Williams, C., Zolla-Pazner, S., Bailer, R.T., Koup, R.A., Madnote, S., Arworn, D., Shen, X., et al.; MOPH TAVEG Collaboration (2012). The Thai Phase III HIV Type 1 Vaccine trial (RV144) regimen induces antibodies that target conserved regions within the V2 loop of gp120. AIDS Res. Hum. Retroviruses *28*, 1444–1457.

Kayman, S.C., Wu, Z., Revesz, K., Chen, H., Kopelman, R., and Pinter, A. (1994). Presentation of native epitopes in the V1/V2 and V3 regions of human immunodeficiency virus type 1 gp120 by fusion glycoproteins containing isolated gp120 domains. J. Virol. *68*, 400–410.

Lefranc, M.P. (2001). IMGT, the international ImMunoGeneTics database. Nucleic Acids Res. *29*, 207–209.

Lek, M., Karczewski, K.J., Minikel, E.V., Samocha, K.E., Banks, E., Fennell, T., O'Donnell-Luria, A.H., Ware, J.S., Hill, A.J., Cummings, B.B., et al.; Exome Aggregation Consortium (2016). Analysis of protein-coding genetic variation in 60,706 humans. Nature *536*, 285–291.

Li, M., Gao, F., Mascola, J.R., Stamatatos, L., Polonis, V.R., Koutsoukos, M., Voss, G., Goepfert, P., Gilbert, P., Greene, K.M., et al. (2005). Human immunodeficiency virus type 1 env clones from acute and early subtype B infections for standardized assessments of vaccine-elicited neutralizing antibodies. J. Virol. *79*, 10108–10125.

Liao, H.X., Levesque, M.C., Nagel, A., Dixon, A., Zhang, R., Walter, E., Parks, R., Whitesides, J., Marshall, D.J., Hwang, K.K., et al. (2009). High-throughput isolation of immunoglobulin genes from single human B cells and expression as monoclonal antibodies. J. Virol. Methods *158*, 171–179.

Liao, H.X., Bonsignori, M., Alam, S.M., McLellan, J.S., Tomaras, G.D., Moody, M.A., Kozink, D.M., Hwang, K.K., Chen, X., Tsao, C.Y., et al. (2013). Vaccine induction of antibodies against a structurally heterogeneous site of immune pressure within HIV-1 envelope protein variable regions 1 and 2. Immunity *38*, 176–186.

McLellan, J.S., Pancera, M., Carrico, C., Gorman, J., Julien, J.P., Khayat, R., Louder, R., Pejchal, R., Sastry, M., Dai, K., et al. (2011). Structure of HIV-1 gp120 V1/V2 domain with broadly neutralizing antibody PG9. Nature *480*, 336–343.

Montefiori, D.C. (2009). Measuring HIV neutralization in a luciferase reporter gene assay. Methods Mol. Biol. *485*, 395–405.

Moore, J.P., McCutchan, F.E., Poon, S.W., Mascola, J., Liu, J., Cao, Y., and Ho, D.D. (1994). Exploration of antigenic variation in gp120 from clades A through F of human immunodeficiency virus type 1 by using monoclonal antibodies. J. Virol. *68*, 8350–8364.

Moore, P.L., Crooks, E.T., Porter, L., Zhu, P., Cayanan, C.S., Grise, H., Corcoran, P., Zwick, M.B., Franti, M., Morris, L., et al. (2006). Nature of nonfunctional envelope proteins on the surface of human immunodeficiency virus type 1. J. Virol. *80*, 2515–2528.

Moore, P.L., Gray, E.S., Sheward, D., Madiga, M., Ranchobe, N., Lai, Z., Honnen, W.J., Nonyane, M., Tumba, N., Hermanus, T., et al.; CAPRISA 002 Study (2011). Potent and broad neutralization of HIV-1 subtype C by plasma antibodies targeting a quaternary epitope including residues in the V2 loop. J. Virol. *85*, 3128–3141.

Morris, L., Chen, X., Alam, M., Tomaras, G., Zhang, R., Marshall, D.J., Chen, B., Parks, R., Foulger, A., Jaeger, F., et al. (2011). Isolation of a human anti-HIV gp41 membrane proximal region neutralizing antibody by antigen-specific single B cell sorting. PLoS ONE *6*, e23532.

Nicely, N.I., Wiehe, K., Kepler, T.B., Jaeger, F.H., Dennison, S.M., Rerks-Ngarm, S., Nitayaphan, S., Pitisuttithum, P., Kaewkungwal, J., Robb, M.L., et al. (2015). Structural analysis of the unmutated ancestor of the HIV-1 envelope V2 region antibody CH58 isolated from an RV144 vaccine efficacy trial vaccinee. EBioMedicine *2*, 713–722.

Pan, R., Gorny, M.K., Zolla-Pazner, S., and Kong, X.P. (2015). The V1V2 region of HIV-1 gp120 forms a five-stranded beta barrel. J. Virol. *89*, 8003–8010.

Pancera, M., Zhou, T., Druz, A., Georgiev, I.S., Soto, C., Gorman, J., Huang, J., Acharya, P., Chuang, G.Y., Ofek, G., et al. (2014). Structure and immune recognition of trimeric pre-fusion HIV-1 Env. Nature *514*, 455–461.

Pollara, J., Bonsignori, M., Moody, M.A., Liu, P., Alam, S.M., Hwang, K.K., Gurley, T.C., Kozink, D.M., Armand, L.C., Marshall, D.J., et al. (2014). HIV-1 vaccine-induced C1 and V2 Env-specific antibodies synergize for increased antiviral activities. J. Virol. *88*, 7715–7726.

Pollara, L., Hart, F., Brewer, J., Pickeral, B., Packard, B.Z., Hoxie, J.A., Komoriya, A., Ochsenbauer, C., Kappes, J.C., Roederer, M., et al. (2011). High-throughput quantitative analysis of HIV-1 and SIV-specific ADCC-mediating antibody responses. Cytometry A. *79*, 603–612.

Rerks-Ngarm, S., Pitisuttithum, P., Nitayaphan, S., Kaewkungwal, J., Chiu, J., Paris, R., Premsri, N., Namwat, C., de Souza, M., Adams, E., et al.; MOPH-TAVEG Investigators (2009). Vaccination with ALVAC and AIDSVAX to prevent HIV-1 infection in Thailand. N. Engl. J. Med. *361*, 2209–2220.

Richardson, S.I., Chung, A.W., Natarajan, H., Mabvakure, B., Mkhize, N.N., Garrett, N., Abdool Karim, S., Moore, P.L., Ackerman, M.E., Alter, G., and Morris, L. (2018). HIV-specific Fc effector function early in infection predicts the development of broadly neutralizing antibodies. PLoS Pathog. *14*, e1006987.

Rolland, M., Edlefsen, P.T., Larsen, B.B., Tovanabutra, S., Sanders-Buell, E., Hertz, T., deCamp, A.C., Carrico, C., Menis, S., Magaret, C.A., et al. (2012). Increased HIV-1 vaccine efficacy against viruses with genetic signatures in Env V2. Nature *490*, 417–420.

Scheepers, C., Shrestha, R.K., Lambson, B.E., Jackson, K.J., Wright, I.A., Naicker, D., Goosen, M., Berrie, L., Ismail, A., Garrett, N., et al. (2015). Ability to develop broadly neutralizing HIV-1 antibodies is not restricted by the germline Ig gene repertoire. J. Immunol. *194*, 4371–4378.

Schramm, C.A., Sheng, Z., Zhang, Z., Mascola, J.R., Kwong, P.D., and Shapiro, L. (2016). SONAR: a high-throughput pipeline for inferring antibody

ontogenies from longitudinal sequencing of B cell transcripts. Front. Immunol. *7*, 372.

Shields, R.L., Namenuk, A.K., Hong, K., Meng, Y.G., Rae, J., Briggs, J., Xie, D., Lai, J., Stadlen, A., Li, B., et al. (2001). High resolution mapping of the binding site on human IgG1 for Fc gamma RI, Fc gamma RII, Fc gamma RIII, and FcRn and design of IgG1 variants with improved binding to the Fc gamma R. J. Biol. Chem. *276*, 6591–6604.

Tamura, K., Stecher, G., Peterson, D., Filipski, A., and Kumar, S. (2013). MEGA6: Molecular Evolutionary Genetics Analysis version 6.0. Mol. Biol. Evol. *30*, 2725–2729.

Trkola, A., Matthews, J., Gordon, C., Ketas, T., and Moore, J.P. (1999). A cell line-based neutralization assay for primary human immunodeficiency virus type 1 isolates that use either the CCR5 or the CXCR4 coreceptor. J. Virol. *73*, 8966–8974.

van Loggerenberg, F., Mlisana, K., Williamson, C., Auld, S.C., Morris, L., Gray, C.M., Abdool Karim, Q., Grobler, A., Barnabas, N., Iriogbe, I., and Abdool Karim, S.S.; CAPRISA 002 Acute Infection Study Team (2008). Establishing a cohort at high risk of HIV infection in South Africa: challenges and experiences of the CAPRISA 002 acute infection study. PLoS One *3*, e1954.

Walker, L.M., Phogat, S.K., Chan-Hui, P.Y., Wagner, D., Phung, P., Goss, J.L., Wrin, T., Simek, M.D., Fling, S., Mitcham, J.L., et al.; Protocol G Principal Investigators (2009). Broad and potent neutralizing antibodies from an African donor reveal a new HIV-1 vaccine target. Science *326*, 285–289.

Wibmer, C.K., Bhiman, J.N., Gray, E.S., Tumba, N., Abdool Karim, S.S., Williamson, C., Morris, L., and Moore, P.L. (2013). Viral escape from HIV-1 neutralizing antibodies drives increased plasma neutralization breadth through sequential recognition of multiple epitopes and immunotypes. PLoS Pathog. *9*, e1003738.

Wibmer, C.K., Moore, P.L., and Morris, L. (2015). HIV broadly neutralizing antibody targets. Curr. Opin. HIV AIDS *10*, 135–143.

Wiehe, K., Easterhoff, D., Luo, K., Nicely, N.I., Bradley, T., Jaeger, F.H., Dennison, S.M., Zhang, R., Lloyd, K.E., Stolarchuk, C., et al. (2014). Antibody light-chain-restricted recognition of the site of immune pressure in the RV144 HIV-1 vaccine trial is phylogenetically conserved. Immunity *41*, 909–918.

Wiehe, K., Nicely, N.I., Lockwood, B., Kuraoka, M., Anasti, K., Arora, S., Bowman, C.M., Stolarchuk, C., Parks, R., Lloyd, K.E., et al. (2017). Immunodominance of antibody recognition of the HIV envelope V2 region in Ig-humanized mice. J. Immunol. *198*, 1047–1055.

Wu, X., Yang, Z.Y., Li, Y., Hogerkorp, C.M., Schief, W.R., Seaman, M.S., Zhou, T., Schmidt, S.D., Wu, L., Xu, L., et al. (2010). Rational design of envelope identifies broadly neutralizing human monoclonal antibodies to HIV-1. Science *329*, 856–861.

Zhang, J., Kobert, K., Flouri, T., and Stamatakis, A. (2014). PEAR: a fast and accurate Illumina Paired-End reAd mergeR. Bioinformatics *30*, 614–620.

7.3 Appendix C: Ethics Certificate

R14/49 Mr Batsirai Mabvakure et al

HUMAN RESEARCH ETHICS COMMITTEE (MEDICAL)

CLEARANCE CERTIFICATE NO. M160340

NAME: (Principal Investigator)	Mr Batsirai Mabvakure et al
DEPARTMENT:	Virology National Institute for Communicable Diseases
PROJECT TITLE:	Co-evolution of Antibody Lineages and HIV Epitopes
DATE CONSIDERED:	01/04/2016
DECISION:	Approved unconditionally
CONDITIONS:	
SUPERVISOR:	Prof Penny Moore
APPROVED BY:	Professor P. Cleaton-Jones, Chairperson, HREC (Medical)
DATE OF APPROVAL:	04/05/2016

This clearance certificate is valid for 5 years from date of approval. Extension may be applied for.

DECLARATION OF INVESTIGATORS

To be completed in duplicate and **ONE COPY** returned to the Research Office Secretary in Room 10004,10th floor, Senate House/2nd floor, Phillip Tobias Building, Parktown, University of the Witwatersrand. I/We fully understand the the conditions under which I am/we are authorised to carry out the above-mentioned research and I/we undertake to ensure compliance with these conditions. Should any departure be contemplated,from the research protocol as approved, I/we undertake to resubmit to the Committee. **I agree to submit a yearly progress report**. The date for annual re-certification will be one year after the date of convened meeting where the study was initially reviewed. in this case, the study was initially review in March and will therefore be due in the month of March each year.

Principal Investigator Signature Date

PLEASE QUOTE THE PROTOCOL NUMBER IN ALL ENQUIRIES

7.4 Appendix D: Supervisor's letter

NATIONAL INSTITUTE FOR COMMUNICABLE DISEASES
Division of the National Health Laboratory Service

HIV Virology Section, Centre for HIV & STI's
1 Modderfontein Road, Sandringham, 2131, South Africa
Email: pennym@nicd.ac.za
Tel: 011 386 6331

7th January 2019

Postgraduate Office
Faculty of Health Sciences
University of the Witwatersrand

To whom it may concern,

Re: Letter of support – Batsirai Mabvakure, PhD Thesis

I am writing to confirm that the thesis submitted by Batsirai Macdonald Mabvakure satisfies the requirements for a PhD. Batsirai has contributed to four publications, and is first author on three of these. The first has already been published in the *Journal of Virology* [IF: 4.368], and the second is under revision, also at the *Journal of Virology*. This is a joint first author publication, where Batsirai was responsible for performing the bioinformatics analyses and writing of the manuscript. Dr Bronwen Lambson generated viral sequences in this study. The third paper describes design and development of a low-cost bioinformatics cluster for the analysis of large antibody sequence data. This manuscript is in preparation. Batsirai also successfully utilised this cluster to define the ancestor of an unusual antibody lineage, resulting in his co-authorship on a fourth paper, recently published in *Cell Reports* [IF: 8.302]. Citations for all papers are provided below.

- **Mabvakure B.M**, Scheepers C, Garrett N, Abdool Karim S, Williamson C, Morris L and Moore P.L. Positive selection at key residues in the HIV Envelope distinguishes broad and strain-specific plasma neutralizing antibodies. Journal of Virology. 2018 Dec 19. pii: JVI.01685-18. doi: 10.1128/JVI.01685-18
- **Mabvakure B.M.***, Lambson B.E*, Ramdayal K, Masson L, Kitchin D, Ali M, Abdool Karim S, Williamson C, Passmore J, Martin D, Scheepers C, Moore P.L, Harkins G.W and Morris L. Evidence for both intermittent and persistent compartmentalization of HIV-1 in the female genital tract. (Manuscript under revision)
- **Mabvakure B.M**, Rott R, Dobrowsky L, van Heusden P, Scheepers C, Morris L and Moore P.L. Application of a low-cost bioinformatics cluster for the analysis of large antibody sequence data. (Manuscript in preparation)
- Van Eeden C, Wibmer C.K, Scheepers C, Richardson S, Nonyane M, Lambson B, Mkhize N, Vijayakumar B, Sheng Z, Stanfield-Oakley S, Bhiman J, Bekker V, Hermanus T, **Mabvakure B.M**, Arshad Ismail A, Moody A, Wiehe K, Garrett N, Karim SA, Dirr H, Fernandes M, Sayed Y, Shapiro L, Ferrari G, Haynes B, Moore P.L and Morris L. V2-directed vaccine-like antibodies from HIV-1 infection identify a novel K169-binding light chain motif with broad ADCC activity. Cell Reports. 2018 Dec 11;25(11):3123-3135.e6. doi: 10.1016/j.celrep.2018.11.058

All co-authors have agreed to these papers/manuscripts being included as part of Batsirai's PhD thesis.

Yours Sincerely,



Prof. Penny Moore
NRF / DST SARChI Chair and Reader
National Institute for Communicable Diseases & University of the Witwatersrand

7.5 Appendix E: Support letter - Publications included in the thesis

NATIONAL INSTITUTE FOR COMMUNICABLE DISEASES
Division of the National Health Laboratory Service

HIV Virology Section, Centre for HIV & STI's
1 Modderfontein Road, Sandringham, 2131, South Africa
Email: pennym@nicd.ac.za
Tel: 011 386 6331

7th January 2019

Postgraduate Office
Faculty of Health Sciences
University of the Witwatersrand

To whom it may concern

Letter of support: Published Papers included in PhD thesis: Batsirai Mabvakure

We are writing to confirm Batsirai Mabvakure's substantial role in the generation of three multi-authored manuscripts, of which one has been formally published, one is in the revision stage and the other is in preparation. These form part of Batsirai's PhD thesis and all co-authors agree to this.

In the first and third publication (see below), Batsirai performed all the analyses, wrote the manuscripts and generated all the figures. He is therefore first author for both these publications. For the second publication, under revision, Batsirai is joint first author with Dr Bronwen Lambson. Batsirai took a leadership role in this study and was responsible for performing all bioinformatics analyses and for writing of the manuscript. Dr Bronwen Lambson generated viral sequences in this study, and supports the inclusion on this paper as a chapter in Batsirai's thesis.

Batsirai is also a co-author on a fourth manuscript, where he performed the antibody evolutionary analyses using NGS sequences. Although this is shown below for information, this publication does not form one of the chapters of the thesis.

The first-author manuscripts included in this submissions are:

1. **Mabvakure B.M**, Scheepers C, Garrett N, Abdool Karim S, Williamson C, Morris L and Moore P.L. Positive selection at key residues in the HIV Envelope distinguishes broad and strain-specific plasma neutralizing antibodies. Journal of Virology. 2018 Dec 19. pii: JVI.01685-18. doi: 10.1128/JVI.01685-18.

2. **Mabvakure B***, Lambson B.E*, Ramdayal K, Masson L, Kitchin D, Ali M, Abdool Karim S, Williamson C, Passmore J, Martin D, Scheepers C, Moore P.L, Harkins G.W and Morris L. Evidence for both intermittent and persistent compartmentalization of HIV-1 in the female genital tract. (*joint first author) (Manuscript under revision).

3. **Mabvakure B.M**, Rott R, Dobrowsky L, van Heusden P, Scheepers C, Morris L and Moore P.L. Application of a low-cost bioinformatics cluster for the analysis of large antibody sequence data. (Manuscript in preparation).

Co-author contribution (not a chapter in the thesis):

4. Van Eeden C, Wibmer C.K, Scheepers C, Richardson S, Nonyane M, Lambson B, Mkhize N, Vijayakumar B, Sheng Z, Stanfield-Oakley S, Bhiman J, Bekker V, Hermanus T, **Mabvakure B.M**, Arshad Ismail A, Moody A, Wiehe K, Garrett N, Karim SA, Dirr H, Fernandes M, Sayed Y, Shapiro L, Ferrari G, Haynes B, Moore P.L and Morris L. V2-directed vaccine-like antibodies from HIV-1 infection identify a novel K169-binding light chain motif with broad ADCC activity. Cell Reports. 2018 Dec 11;25(11):3123-3135.e6. doi: 10.1016/j.celrep.2018.11.058.

Batsirai's significant contribution to these papers is testament to the high quality of his research.

Yours sincerely

Prof. Penny Moore
Supervisor

Prof. Lynn Morris
Co-supervisor

Dr Cathrine Scheepers
Co-supervisor

Dr Bronwen Lambson
Joint first author, publication under review.

Chairperson: Prof Eric Buch Acting CEO: Dr Karmani Chetty
Physical Address: 1 Modderfontein Road, Sandringham, Johannesburg, South Africa Postal Address: Private Bag X4, Sandringham, 2131, South Africa
Tel: +27 (0) 11 386 6400 Fax: +27 (0) 11 882 0596 www.nicd.ac.za
Practice number: 5200296

7.6 Appendix F: Certificate of submission - student

UNIVERSITY OF THE WITWATERSRAND, JOHANNESBURG | FACULTY OF HEALTH SCIENCES

CERTIFICATE OF SUBMISSION FOR EXAMINATION OF MASTERS RESEARCH REPORT / DISSERTATION OR PHD THESIS SIGNED BY HIGHER DEGREES CANDIDATES

Full name	Batsirai Macdonald Mabvakure		
Student number	892141		
Title of submitted Research Project: **Co-evolution of antibody lineages and HIV epitopes** _____ _____ *NB: If this title is different to your previously approved title, no further action can be taken by the Faculty Office until a change of title has been approved.*			
Contact no	0736146560	E-mail	batsiraim@nicd.ac.za or bmabvakure@gmail.com

1. If you are likely to move in the next 6-12 months, please give the anticipated date of move: 23 February 2019
2. I hereby submit my **PhD thesis for examination**
   (Select whichever is applicable)
3. I have checked all copies of my thesis and declare that no pages are missing or poorly reproduced.
4. I have submitted three bound copies and one copies on CD

5. **I confirm that I have:**
   a) A signed declaration indicating my understanding of the concept of plagiarism and a denial of plagiarism in my research document.
   b) A report from "Turnitin" (or other approved plagiarism detection) software indicating the level of plagiarism in my research document included as an appendix.

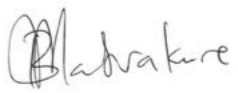
6. **I confirm that I have:**
   a) Not used either human or animal tissue or records **No**
   b) If yes: I have included the ethics waiver letter pertinent to my research as an appendix **NA**
   c) Done research using animals **No**
      If yes: I have included a copy of the animal ethics committee clearance certificate as an appendix in this document **NA**
   d) Done research using human subjects, human tissue or patient records **Yes**
      If yes: I have included a copy of the human ethics clearance certificate as an appendix to the research document **Yes**

7. I understand that I may not graduate unless my University fees have been paid in full.
8. My Supervisor(s) names, departments, telephone numbers and email addresses are as follows:

Name	Prof Penny Moore		
Department	Virology		
Telephone	+27 11 386 6331	E-mail	pennym@nicd.ac.za
Name	Dr Cathrine Scheepers		
Department	Virology		
Telephone	+27 11 386 0443	E-mail	cathrinem@nicd.ac.za
Name	Prof Lynn Morris		
Department	Virology		
Telephone	+27 11 386 6332	E-mail	lynnm@nicd.ac.za

List all publications, which you have published in peer-reviewed journals from your postgraduate research report/dissertation/thesis during the course of your studies in the Faculty of Health Sciences (Include authors, year, title of paper, name of journal, volume number and page numbers). This information is mandatory.

- **Mabvakure B.M**, Scheepers C, Garrett N, Abdool Karim S, Williamson C, Morris L, and Moore P.L. Positive selection at key residues in the HIV Envelope distinguishes broad and strain-specific plasma neutralizing antibodies. Journal of Virology. 2018 Dec 19. pii: JVI.01685-18. doi: 10.1128/JVI.01685-18
- **Mabvakure B**, Lambson B.E, Ramdayal K, Masson L, Kitchin D, Ali M, Abdool Karim S, Williamson C, Passmore J, Martin D, Scheepers C, Moore P.L, Harkins G.W and Morris L. Evidence for both intermittent and persistent compartmentalization of HIV-1 in the female genital tract. (Manuscript under revision)
- **Mabvakure B.M**, Rott R, Dobrowsky L, van Heusden P, Scheepers C, Morris L and Moore P.L. Application of a low-cost bioinformatics cluster for the analysis of large antibody sequence data. (Manuscript in preparation)
- Van Eeden C, Wibmer C.K, Scheepers C, Richardson S, Nonyane M, Lambson B, Mkhize N, Vijayakumar B, Sheng Z, Stanfield-Oakley S, Bhiman J, Bekker V, Hermanus T, **Mabvakure B.M**, Arshad Ismail A, Moody A, Wiehe K, Garrett N, Karim SA, Dirr H, Fernandes M, Sayed Y, Shapiro L, Ferrari G, Haynes B, Moore P.L and Morris L. V2-directed vaccine-like antibodies from HIV-1 infection identify a novel K169-binding light chain motif with broad ADCC activity. Cell Reports. 2018 Dec 11;25(11):3123-3135.e6. doi: 10.1016/j.celrep.2018.11.058



Signature of candidate: _____ Date: 7 January 2019

26/05/2015

7.7 Appendix G: Certificate of submission - supervisors



CERTIFICATE OF SUBMISSION FOR EXAMINATION SIGNED BY SUPERVISORS OF HIGHER DEGREES CANDIDATES

Full name	Batsirai Macdonald Mabvakure		
Student number	892141		
Candidate for the degree of: Doctor of Philosophy (PhD) in Virology *has submitted his thesis* Entitled: Co-evolution of antibody lineages and HIV epitopes			
Contact no	073 614 6560	E-mail	batsiraim@nicd.ac.za / bmabvakure@gmail.com

Mark with an X on appropriate box	Yes	No
Has this thesis/dissertation/research report been submitted with the acquiescence of the supervisor?	X	
To the best of your knowledge are you able to verify that this is the candidate's work, except as otherwise stated by the candidate?	X	
The substance (nor any part of it) has not been submitted in the past nor is being submitted for a degree in any other university?	X	
The candidate has acknowledged wherever any information used in the thesis, dissertation or other work has been obtained by him/her while employed by, or working under the aegis of, any person or organization other than the University or its associated institutions?	X	
Have examiners been nominated and approved?	X	

I certify that this thesis/dissertation/research report has the approval of the Animal Ethics Committee / Committee for Research on Human Subjects and the Number of the Certificate of Approval is: **M160340**

List all publications, which your student has published in peer-reviewed journals from his/her postgraduate research report/dissertation/thesis during the course of his/her studies in the Faculty of Health Sciences (Include authors, year, title of paper, name of journal, volume number and page numbers). This information is mandatory.

- **Mabvakure B.M**, Scheepers C, Garrett N, Abdool Karim S, Williamson C, Morris L, and Moore P.L. Positive selection at key residues in the HIV Envelope distinguishes broad and strain-specific plasma neutralizing antibodies. Journal of Virology. 2018 Dec 19. pii: JVI.01685-18. doi: 10.1128/JVI.01685-18
- **Mabvakure B**, Lambson B.E, Ramdayal K, Masson L, Kitchin D, Ali M, Abdool Karim S, Williamson C, Passmore J, Martin D, Scheepers C, Moore P.L, Harkins G.W and Morris L. Evidence for both intermittent and persistent compartmentalization of HIV-1 in the female genital tract. (Manuscript under revision)
- **Mabvakure B.M**, Rott R, Dobrowsky L, van Heusden P, Scheepers C, Morris L and Moore P.L. Application of a low-cost bioinformatics cluster for the analysis of large antibody sequence data. (Manuscript in preparation)
- Van Eeden C, Wibmer C.K, Scheepers C, Richardson S, Nonyane M, Lambson B, Mkhize N, Vijayakumar B, Sheng Z, Stanfield-Oakley S, Bhiman J, Bekker V, Hermanus T, **Mabvakure B.M**, Arshad Ismail A, Moody A, Wiehe K, Garrett N, Karim SA, Dirr H, Fernandes M, Sayed Y, Shapiro L, Ferrari G, Haynes B, Moore P.L and Morris L. V2-directed vaccine-like antibodies from HIV-1 infection identify a novel K169-binding light chain motif with broad ADCC activity. Cell Reports. 2018 Dec 11;25(11):3123-3135.e6. doi: 10.1016/j.celrep.2018.11.058

Name of Supervisor 1: Prof Penny Moore

Telephone: +27 11 386 6331 Email: pennym@nicd.ac.za



Signature: _____ Date: 7th January 2019

Name of Supervisor 2: Dr Cathrine Scheepers

Telephone: +27 11 386 0443 Email: cathrinem@nicd.ac.za



Signature: _____ Date: 7th January 2019

Name of Supervisor 3: Prof Lynn Morris

Telephone: +27 11 386 6332 Email: lynnm@nicd.ac.za



Signature: _____ Date: 7th January 2019

==========

IMPORTANT NOTICE WITH REGARD TO THE SENATE STANDING ORDERS:

A.22 Submission against advice of Supervisor

If the Supervisor is not prepared to agree to the submission of a thesis, the candidate shall still be entitled, if he or she wishes, to submit it for examination. When a thesis is submitted against the advice of the Supervisor, this should be recorded in the minutes of the Faculty Graduate Studies Committee. In such a case, no internal examiners are appointed but a Supervisor's report will still be required. After the examination process, the external examiner(s) will be advised by the Chairperson of the Faculty Graduate Studies Committee that the thesis was submitted against the advice of the Supervisor.

A.24 *Nomination of Examiners*:

Nomination of examiners should take place at least six weeks before submission of the thesis or dissertation. *(The Postgraduate Office will not accept any submission for examination without the confirmed appointment of the nominated examiners*.)

A.25 Confidentiality of names of examiners (both external and internal)

The names of the examiners should be confidential during the examination process and may only be revealed to the candidate with the acquiescence of the examiner once the final version of the thesis has been submitted to the Faculty and the process has been completed.

26/05/2015

7.8 Appendix H: Turn-it-in Report

Turnitin Originality Report

892141:PhDThesis_Batsirai_chapters_1_3_4_5_turn-it-in.pdf by Batsirai Mabvakure

From Batsirai Mabvakure 892141 (ca654600-c0f7-4b28-9e93-c904e5efd056)

- Processed on 09-Jan-2019 3:29 PM SAST
- ID: 1062500035
- Word Count: 24233

Similarity Index
14%
Similarity by Source

Internet Sources:
9%
Publications:
11%
Student Papers:
5%

sources:

1 1% match (publications)
Charmaine van Eeden, Constantinos Kurt Wibmer, Cathrine Scheepers, Simone I. Richardson et al. "V2-Directed Vaccine-like Antibodies from HIV-1 Infection Identify an Additional K169-Binding Light Chain Motif with Broad ADCC Activity", Cell Reports, 2018

2 < 1% match (Internet from 30-Aug-2017)
http://open.uct.ac.za/bitstream/handle/11427/18609/thesis_hsf_2011_masson_lindi.pdf?sequence=1

3 < 1% match (Internet from 31-Aug-2017)
http://open.uct.ac.za/bitstream/handle/11427/8689/thesis_hsf_2014_bandawe_gp.pdf?sequence=1

4 < 1% match (publications)
Korkut, Anil, and Wayne A. Hendrickson. "Structural Plasticity and Conformational Transitions of HIV Envelope Glycoprotein gp120", PLoS ONE, 2012.

5 < 1% match (Internet from 12-Apr-2015)
http://www.sciencemagazinedigital.org/sciencemagazine/03_october_2014?pg=91

6 < 1% match (student papers from 14-Oct-2011)
Submitted to University of Cape Town on 2011-10-14

7 < 1% match (Internet from 02-Apr-2010)
http://www.direct-ms.org/pdf/ImmunityRegulation/Immune%20review%20NEJM.pdf

All remaining sources showed ≤ **1%** similarity.

Prof. Penny L. Moore

Batsirai M. Mabvakure