To those who have supported me during this eventful journey
I would first and foremost like to acknowledge my husband, Kurt Magalis, for not only his love and support during very long periods of undivided attention to my work but also for his willingness to listen during the periods between. I would also like to thank my parents, David and Gail Rife, who never ceased to praise all of my accomplishments, large or small, during and prior to this journey. I am grateful as well for the National Institutes of Mental Health for funding this project and my committee members – Drs. Mark Wallet, James Maruniak, Robert McKenna, and Luciano Brocchieri - for their never-ending encouragement and words of wisdom. Last but certainly not least, I would like to thank my mentor, Dr. Marco Salemi, who has only strengthened my reverence for science and helped me to grow in all aspects of research by demonstrating patience, perseverance, and passion. None of my work would have been possible without the acknowledged individuals and institutions.
TABLE OF CONTENTS

<table>
<thead>
<tr>
<th>Section</th>
<th>Page</th>
</tr>
</thead>
<tbody>
<tr>
<td>ACKNOWLEDGMENTS</td>
<td>4</td>
</tr>
<tr>
<td>LIST OF TABLES</td>
<td>8</td>
</tr>
<tr>
<td>LIST OF FIGURES</td>
<td>9</td>
</tr>
<tr>
<td>ABSTRACT</td>
<td>13</td>
</tr>
<tr>
<td>CHAPTER</td>
<td></td>
</tr>
<tr>
<td>1 BACKGROUND</td>
<td>15</td>
</tr>
<tr>
<td>The HIV Epidemic</td>
<td>15</td>
</tr>
<tr>
<td>AIDS Progression and NeuroAIDS Development</td>
<td>15</td>
</tr>
<tr>
<td>Pathological Hallmarks of NeuroAIDS</td>
<td>16</td>
</tr>
<tr>
<td>The Proposed Etiological Role of Monocytes in NeuroAIDS</td>
<td>17</td>
</tr>
<tr>
<td>Viral Compartmentalization in the CNS of HAND Patients</td>
<td>18</td>
</tr>
<tr>
<td>SIV/Macaque Model of AIDS and NeuroAIDS</td>
<td>19</td>
</tr>
<tr>
<td>S100B as a Biomarker of SIVE Onset</td>
<td>23</td>
</tr>
<tr>
<td>Significance and Innovation of the Described Studies</td>
<td>24</td>
</tr>
<tr>
<td>2 EVOLUTION OF NEUROADAPTATION IN THE PERIPHERY AND PURIFYING SELECTION IN THE BRAIN CONTRIBUTE TO COMPARTMENTALIZATION OF SIV IN THE BRAIN OF RHESUS MACAQUES WITH SIV- ASSOCIATED ENCEPHALITIS</td>
<td>27</td>
</tr>
<tr>
<td>Introduction</td>
<td>27</td>
</tr>
<tr>
<td>Materials and Methods</td>
<td>29</td>
</tr>
<tr>
<td>Results</td>
<td>41</td>
</tr>
<tr>
<td>Discussion</td>
<td>55</td>
</tr>
<tr>
<td>3 PHYLOGENETICS AND PHYLOANATOMY OF HIV/SIV INTRA-HOST COMPARTMENTS AND RESERVOIRS: THE KEY ROLE OF THE CENTRAL NERVOUS SYSTEM</td>
<td>63</td>
</tr>
<tr>
<td>Evidence of Cellular and Anatomical HIV-1 Reservoirs</td>
<td>63</td>
</tr>
<tr>
<td>HIV-1 Reservoirs and the Challenge of Eradication</td>
<td>63</td>
</tr>
<tr>
<td>HIV-1 Compartmentalization and Metapopulation Structure</td>
<td>65</td>
</tr>
<tr>
<td>Phylogenetic Methods of Determining Viral Compartments and Potential Reservoirs</td>
<td>67</td>
</tr>
<tr>
<td>Testing Viral Compartmentalization</td>
<td>67</td>
</tr>
<tr>
<td>Phyloanatomy</td>
<td>71</td>
</tr>
<tr>
<td>Analyzing the Temporal Structure of Longitudinally Sampled Viral Sequences</td>
<td>74</td>
</tr>
<tr>
<td>Chapter</td>
<td>Title</td>
</tr>
<tr>
<td>---------</td>
<td>----------------------------------------------------------------------</td>
</tr>
<tr>
<td>4</td>
<td>SPATIOTEMPORAL DYNAMICS OF SIMIAN IMMUNODEFICIENCY VIRUS BRAIN INFECTION IN CD8+ LYMPHOCYTE-DEPLETED Rhesus Macaques WITH NEUROAIDS</td>
</tr>
<tr>
<td></td>
<td>Introduction</td>
</tr>
<tr>
<td></td>
<td>Materials and Methods</td>
</tr>
<tr>
<td></td>
<td>Results</td>
</tr>
<tr>
<td></td>
<td>Discussion</td>
</tr>
<tr>
<td>5</td>
<td>SPATIAL AND TEMPORAL CHARACTERISTICS OF SIMIAN IMMUNODEFICIENCY VIRUS DISPERSION DISTINGUISH Rhesus Macaques WITH AND WITHOUT NEUROAIDS</td>
</tr>
<tr>
<td></td>
<td>Introduction</td>
</tr>
<tr>
<td></td>
<td>Materials and methods</td>
</tr>
<tr>
<td></td>
<td>Results</td>
</tr>
<tr>
<td></td>
<td>Discussion</td>
</tr>
<tr>
<td>6</td>
<td>CONTRIBUTION OF INFECTED MONOCYTES TO VIRAL EVOLUTION IN THAI INDIVIDUALS WITH AND WITHOUT HUMAN IMMUNODEFICIENCY VIRUS-ASSOCIATED NEUROCOGNITIVE DISORDERS (HAND)</td>
</tr>
<tr>
<td></td>
<td>Introduction</td>
</tr>
<tr>
<td></td>
<td>Materials and Methods</td>
</tr>
<tr>
<td></td>
<td>Results</td>
</tr>
<tr>
<td></td>
<td>Discussion</td>
</tr>
<tr>
<td>7</td>
<td>SIMIAN IMMUNODEFICIENCY VIRUS INTRA-HOST PHYLOGENOMIC PATTERNS IN CD8+ LYMPHOCYTE-DEPLETED AND NATURALLY PROGRESSING Rhesus Macaques: INSIGHTS INTO DIFFERENCES IN DISEASE PROGRESSION</td>
</tr>
<tr>
<td></td>
<td>Introduction</td>
</tr>
<tr>
<td></td>
<td>Materials and Methods</td>
</tr>
<tr>
<td></td>
<td>Results</td>
</tr>
<tr>
<td></td>
<td>Discussion</td>
</tr>
<tr>
<td>8</td>
<td>CONCLUDING REMARKS</td>
</tr>
</tbody>
</table>

**APPENDIX**

<table>
<thead>
<tr>
<th>Section</th>
<th>Title</th>
<th>Page</th>
</tr>
</thead>
<tbody>
<tr>
<td>A</td>
<td>SUPPORTING INFORMATION FOR CHAPTER 4</td>
<td>167</td>
</tr>
<tr>
<td>B</td>
<td>SUPPORTING INFORMATION FOR CHAPTER 6</td>
<td>172</td>
</tr>
</tbody>
</table>
C  SUPPORTING INFORMATION FOR CHAPTER 7 .................................................. 192

LIST OF REFERENCES .......................................................................................... 196

BIOGRAPHICAL SKETCH .................................................................................. 231
### LIST OF TABLES

<table>
<thead>
<tr>
<th>Table</th>
<th>Page</th>
</tr>
</thead>
<tbody>
<tr>
<td>2-1</td>
<td>Gross pathology of SIVmac251-infected naturally progressing rhesus macaques.</td>
</tr>
<tr>
<td>4-1</td>
<td>Bayesian coalescent estimates of SIV molecular clock and demographic parameters in each infected primate.</td>
</tr>
<tr>
<td>7-1</td>
<td>Analysis of temporal signal for time-stamped phylogenies.</td>
</tr>
<tr>
<td>7-2</td>
<td>Analysis of tissue-dependent compartmentalization of viral RNA sequences.</td>
</tr>
<tr>
<td>A-1</td>
<td>Coalescent demographic model selection.</td>
</tr>
<tr>
<td>A-2</td>
<td>Tree correlation coefficients.</td>
</tr>
<tr>
<td>A-3</td>
<td>Sampling spectra.</td>
</tr>
<tr>
<td>B-1</td>
<td>Analysis of compartmentalization of gp120 sequence data for three HIV-1-infected Thai patients.</td>
</tr>
<tr>
<td>Figure</td>
<td>Description</td>
</tr>
<tr>
<td>--------</td>
<td>-------------</td>
</tr>
<tr>
<td>1-1</td>
<td>Characteristic timeline of AIDS progression and development of HIV-associated neurological disease</td>
</tr>
<tr>
<td>1-2</td>
<td>Prevalence of HIV-associated neurocognitive disorder categories among HIV-1-infected individuals</td>
</tr>
<tr>
<td>1-3</td>
<td>Models of viral population dynamics resulting in compartmentalization of virus in the brain</td>
</tr>
<tr>
<td>1-4</td>
<td>Hypothesis model</td>
</tr>
<tr>
<td>2-1</td>
<td>Spectrum of tissues and cell populations sampled from the rhesus macaques</td>
</tr>
<tr>
<td>2-2</td>
<td>Viral load and CD4+ T-cell population characteristics for naturally progressing (left) and CD8+ lymphocyte-depleted (right) macaques</td>
</tr>
<tr>
<td>2-3</td>
<td>Maximum likelihood (ML) phylogenetic tree reconstruction for each macaque with detectable viral RNA in the brain</td>
</tr>
<tr>
<td>2-4</td>
<td>Percentage of total brain tissue viral RNA sequences within individual CD8+ lymphocyte-depleted and naturally progressing macaques that share 100% sequence identity</td>
</tr>
<tr>
<td>2-5</td>
<td>Analysis of tissue-dependent compartmentalization of viral RNA sequences</td>
</tr>
<tr>
<td>2-6</td>
<td>Genetic distance calculations for viral sequences sampled at necropsy from longitudinally sampled, naturally progressing macaques with sufficiently detectable virus in the brain</td>
</tr>
<tr>
<td>2-7</td>
<td>Mapped amino acid changes contributing to episodic diversifying selection along internal branches of representative phylogenies</td>
</tr>
<tr>
<td>2-8</td>
<td>Site-specific selection over time for SIV sequences within naturally progressing and CD8+lymphocyte-depleted macaques diagnosed with SIVE or without SIVE</td>
</tr>
<tr>
<td>2-9</td>
<td>Percent recombination in necropsy-sampled sequences obtained from individual tissues/cell populations in naturally progressing and CD8+lymphocyte-depleted macaques</td>
</tr>
<tr>
<td>2-10</td>
<td>Quantification of viral genomic RNA and integrated DNA in longitudinally sampled plasma and peripheral cell populations</td>
</tr>
</tbody>
</table>
3-1 HIV-1 maximum likelihood (ML) tree of brain and peripheral tissue derived sequences from a patient with HIV-associated dementia .................................................. 72

3-2 Bayesian maximum clade credibility (MCC) tree highlighting reduced intra-host temporal structure .................................................................................................................. 77

3-3 Result of temporal signal analysis in Path-O-Gen .................................................. 82

3-4 Bayesian maximum clade credibility (MCC) trees highlighting viral migration events from brain into circulation and estimated ancestral state posterior probabilities (PP) using phylodynamics ............................................................................. 83

3-5 Graphical representation of recent advancements in phylogenetic techniques used to assess the contribution of anatomical compartments to viral dispersion and reservoir formation. .................................................................................................................. 84

4-1 Absolute CD8+ T-lymphocytes and NK cells count over time in infected macaques .................................................................................................................. 93

4-2 SIV gp120 Bayesian maximum clade credibility (MCC) trees .................................. 97

4-3 SIV phylogeographic and demographic patterns within primates followed longitudinally .................................................................................................................. 98

4-4 SIV gene flow events from peripheral tissues to the brain ...................................... 101

5-1 Graphical representation of viral dispersion patterns over the course of infection in the context of SIVE for the SIV-infected macaque model of HIV infection and neuroAIDS .................................................................................................................. 115

5-2 Plasma concentrations of s100B at necropsy and peak levels in animals grouped according to SIVE diagnosis and detectable presence of viral RNA in brain tissue sections (DVB) .................................................................................. 117

6-1 Sampled posterior distribution of Bayesian phylogenetic trees for all HIV gp120 sequences derived from plasma and sorted peripheral leukocytes in three individual Thai patients .................................................................................. 128

6-2 Inferred viral dispersion among peripheral blood compartments in three individual HIV-1-infected Thai patients .................................................................................. 131

6-3 Location of amino acid site-specific selective pressure along gp120 according to patient diagnosis and cell type origin with 3D structure mapping of C2 sites specific to peripheral monocytes in HAND patients .................. 132

7-1 Maximum likelihood (ML) phylogenetic tree reconstruction for each macaque. (A) Naturally progressing cohort .................................................................................. 151
Temporal clustering within maximum likelihood phylogenies for longitudinally sampled naturally progressing and CD8+ lymphocyte-depleted macaques… 154

Genetic distance calculations over time for viral sequences derived from longitudinally sampled peripheral cell populations and meninges of naturally progressing and CD8+ lymphocyte-depleted macaques … 156

Genetic distance calculations for viral sequences derived from early infection (21 dpi) and necropsy samples of naturally progressing and CD8+ lymphocyte-depleted macaques … 159

Site-specific selection over time for SIV sequences within individual naturally progressing and CD8+ lymphocyte-depleted macaque tissues/cell populations … 160

Experimental design and sampling timeline … 170

Slatkin-Maddison compartmentalization test … 170

BEAST special history xml additions … 171

HIV-1-infected Thai patient diagnosis, sampling timeline, and sequence information … 181

Neighbor joining (NJ) phylogeny incorporating all HIV-1-infected Thai patients for which viral gp120 sequences were obtained … 182

Likelihood mapping of HIV-1-infected Thai patient-specific replicates of uniform sampling from anatomical locations … 183

Maximum likelihood (ML) phylogenies of viral gp120 sequences for individual HIV-1-infected Thai patients … 184

Viral genetic diversity within plasma and peripheral T-cells and monocytes obtained at specific time points from three individual HIV-1-infected Thai patients … 185

Inferred timing of viral dispersion among discrete anatomical compartments for patient P01 sample replicates … 186

Linear regression analysis of the relationship between genetic divergence and sampling time for P01 sample replicate gp120 sequence data … 188

Amino acid signature pattern analysis for plasma, monocyte- and T-cell-associated gp120 sequences in three HIV-1-infected Thai individuals at the time of cART initiation … 189
C-1  Timeline of infection and sampling for the CD8+ lymphocyte-depleted and naturally progressing macaque cohorts. ................................................................. 192

C-2  Differences in time to disease progression for CD8+ lymphocyte-depleted and naturally progressing macaques classified as rapid or slow progressors. . 193

C-3  Linear regression analysis of temporal signal in maximum likelihood phylogenetic trees for longitudinally sampled naturally progressing and CD8+ lymphocyte-depleted macaques ........................................................................ 194

C-4  Percent recombination in sequences obtained from individual tissues/cell populations in naturally progressing and CD8+ lymphocyte-depleted macaques........................................................................................................ 195
As many as 50% of individuals infected with HIV type I (HIV-1) are estimated to harbor HIV-1 within the brain; however, less than 50% (~30%) of infected individuals will progress to some form of HIV-associated neurocognitive disorder (HAND), a disease for which the etiology currently remains obscure. Immune cells known as monocytes have been proposed to act as Trojan horses in facilitating entry of virus into the brain because of their susceptibility to infection and ability to cross the blood-brain barrier. Despite in vitro evidence of this phenomenon, the differing roles of these cells in brain infection versus neuropathology as well as their interaction with the virus has been a topic of debate owing to limited opportunities for in vivo evidence. The emerging field of Bayesian phyloanatomy provides an in silico solution to characterizing the contribution of individual cell types (and/or tissues) to viral evolution and dispersion among anatomical compartments in vivo, as it allows for the incorporation of trait information in reconstruction of phylogenetic relationships among sampled viral nucleic acid sequences in real time. Understanding the evolutionary factors driving trafficking of infected cells to the brain and the emergence of neurovirulent strains during disease onset and progression is of pivotal importance for developing a realistic model of HAND,
or neuroAIDS, and improvement of drug design strategies. Using isolated peripheral blood cell populations in HIV-1-infected humans with a range of neurocognitive impairment, as well as tissues and/or composite cell populations within and outside the central nervous system (CNS) of SIV-infected macaques, we proposed to identify evolutionary patterns in distinct anatomical locations associated with the onset and progression of neuropathogenesis. Viral genomic RNA sequencing and phyloanatomic analyses revealed differential patterns of evolutionary constraints and viral dispersion in and among separate tissues and cell types, including a distinct pattern in monocytes/macrophages from that of the primary CD4+ T-cell target, that indicate a major role of macrophage-mediated neuroadaptation prior to entry into the brain and replication-mediated virulence.
CHAPTER 1
BACKGROUND

The HIV Epidemic

Since its discovery in 1981, HIV-1 infection has been responsible for the deaths of 36 million people worldwide as of 2012, rendering its associated disease, acquired immunodeficiency syndrome (AIDS), a pandemic for which there is still no cure or effective vaccine (1). Despite the availability of antiretroviral treatments, HIV continues to spread predominantly due to its relatively high evolutionary and reproductive rates as well as recombination frequency (2). HIV evolves at the rate of one nucleotide substitution for every 2-3 newly synthesized genomes (3, 4) and replicates precipitously, giving rise to an average of $\sim 10^{11}$ new virions each day in the infected human host (5, 6). This evolutionary rate is greater than ten times that of the rapidly evolving influenza A virus (7) and, as with influenza, has proven to be the primary challenge in the development of effective vaccine and treatment strategies (2, 8-10).

AIDS Progression and NeuroAIDS Development

The timeline of HIV infection is characterized by three main stages – primary infection, wherein patients present commonly misdiagnosed symptoms typical of a viral infection; the asymptomatic period, during which viremia remains steady at low levels; and AIDS onset and progression, which is characterized by rapid virus accumulation in the blood and corresponding plummeting of CD4+ T-cells, rendering the immune system virtually incapable of fighting infection. During this characteristic timeline, we also see that the virus is continually evolving, sometimes very rapidly, diverging from the original founder virus (11, 12) (Figure 1-1). In addition to challenges for the treatment and prevention of AIDS, the occurrence of HIV-related complications, such as
cardiovascular disease, can add to the complexity of treatment strategies and increase morbidity and mortality rates among infected patients. HIV-associated neurocognitive disorders (HAND), a subset of neuroAIDS, collectively are one such complication that is still not fully understood and yet currently affects 30-50% of HIV-infected patients (13) beginning typically around the onset of AIDS (11).

Pathological Hallmarks of NeuroAIDS

HAND is currently categorized into three separate stages, referred to as asymptomatic neurocognitive impairment (ANI), mild neurocognitive disorder (MND), and HIV-associated dementia (Figure 1-2), which can be identified using clinical diagnostics and a series of neuropsychological tests (14). However, accurate biological markers are still in high demand due to a general lack of understanding of the
underlying mechanisms of disease onset and progression, including entry of the virus into the brain. The development of neuropathology is hypothesized to be associated with the emergence of critical viral variants, termed neurotropic, for their ability to replicate in the unique microenvironment of the brain, and neurovirulent, capable of causing neurological disease. These neurotropic/neurovirulent viral strains have been hypothesized to be attributed to genetic changes within the envelope (Env) glycoprotein, which is responsible for the binding of the host cellular CD4 receptor and CCR5 or CXCR4 co-receptor (among others) for cellular adhesion and entry (15). Changes within the N-terminally cleaved portion of the envelope glycoprotein (Env), referred to as gp120, have been implicated in the onset of neuroAIDS due to the ability of soluble gp120 variants to elicit oxidative stress and apoptosis in cultured neurons and cause neurotoxicity in vivo (16) but also primarily due to their association with macrophage tropism (17, 18), a cellular reservoir and significant correlate of HIV-1-associated neurological disease (19). Macrophages and macrophage-like microglia are the primary cell types infected in the brain, the accumulation of which is a hallmark of AIDS-related neuropathology (20, 21). However, the significant contributions of the several identified subpopulations of the macrophage precursor, the monocyte, as well as the viral evolutionary patterns underlying these contributions remains elusive.

The Proposed Etiological Role of Monocytes in NeuroAIDS

In vitro models have indicated that entry of infected monocytes into the brain may occur via trafficking across the blood brain barrier (BBB) driven by chemotactic recruitment (22-24). Following entry, or neuroinvasion, the spread of the virus and/or shedding of viral proteins then results in cellular activation, leading to persistent inflammation, during which inflammatory cytokines are released as well as other CNS-
specific proteins, acting as neurotoxic factors that are able to destroy up to 40% (25) of neuronal cells (26). In support of this “Trojan Horse” hypothesis, previous studies have reported that bi-phasic expansion of specific monocyte/macrophage subpopulations in HIV-infected humans and SIV-infected macaques as well as turnover of monocytes from the bone marrow of SIV-infected macaques correlate with diagnosis of HAD (27) in humans and the severity of neuropathology in macaques (28, 29). There are currently no definitive observations, however, demonstrating that monocytes migrating to the brain were either infected as monocyte progenitors in the bone marrow or as circulating monocytes in the blood. Thus, the question of when and how neurotropic viral variants emerge prior to neuroinvasion and how this relates to monocyte/macrophage tropism remains unanswered.

**Viral Compartmentalization in the CNS of HAND Patients**

Previous phylogenetic analyses of a single HIV-infected individual with HAD has provided evidence in support of the contribution of infected monocytes in the bone marrow to the progression of neuroAIDS based on the similarity of post-mortem brain and bone marrow-derived viral sequences (30). Although sharing certain genetic similarities, similar studies have characterized this post-mortem virus population in the brain of HAND patients as being phylogenetically distinct, or compartmentalized, from remaining peripheral viral populations (31, 32). Models of viral entry into and replication within the brain that would result in this type of evolutionary characteristic include a single introduction of virus into that particular compartment followed by clonal amplification of the infected cell(s) or long-term, isolated viral replication; however, a more convincing model encompasses multiple introductions of distinct viral subpopulations whose ultimate survival demands specific viral genotypic/phenotypic
features required for both entry into and replication in the unique cell types comprising that compartment (Figure 1-3). Discussed in more detail in Chapter 2, the latter scenario offers a more realistic explanation of neuroinvasion of the brain and was the starting hypothesis as to the mechanism of this evolutionary phenomenon. Expanding on this question at hand, a high-resolution sampling strategy in an ideal animal model and advancements in statistical phylogenetic methods have allowed for not only a more in-depth look into this mechanism but also the identification of the contributing tissue/cell type to brain infection and the timing of this contribution, discussed in Chapters 3-7.

Figure 1-2. Prevalence of HIV-associated neurocognitive disorder categories among HIV-1-infected individuals. Figure was adapted from McArthur et al. (2010) (33).

SIV/Macaque Model of AIDS and NeuroAIDS

Over the last two decades, non-human primate (NHP) models of AIDS have revealed key aspects of HIV-1 pathogenesis, such as early post-infection events and
viral mechanisms of CD4+ T-cell destruction (34-36). Simian immunodeficiency virus (SIV) infection of is a unique model for HIV research in that HIV originated relatively recently from SIV as the result of several cross-species transmission events from non-human African primates to humans (37). SIV belongs to the genus Lentivirus in the Retroviridae family, as does HIV, and consists of a large, widespread group of viruses that can be found naturally in both feral and domesticated non-human primates, such as mangabeys and chimpanzees, and are usually non-pathogenic in their natural hosts. Some pathogenic infections, however, are the product of recent cross-species transmission, as is the case for the experimental transmission from sooty mangabeys (SIVsmm) to macaques (SIVmac), resulting in an AIDS-like illness referred to as simian AIDS (sAIDS). In vivo passaging of the original SIVsmm virus in rhesus macaques gave rise to the commonly used viral swarm, SIVmac251 (38-41), also used in the studies described herein for the infection of rhesus macaques (Macaca mulatta).

The genomic organization of HIV-1 and SIVmac are similar, with the exception of the accessory proteins Vpu and Vpx in HIV and SIV, respectively (42). In particular, a cleaved N-terminal portion of the previously mentioned gp120 resides in relatively similar locations within the two genomes near the 3’ end and was the focus of the studies described herein. This region was chosen based on its critical balance of sequence homology and intra-host genetic variability. The extent of applicability of the described results to HIV-1 intra-host evolution was highly dependent on SIVmac251 and HIV-1 amino acid sequence homology for this particular region. Moreover, most glycosylation sites in the variable domains as well as the GGDPE motif located within the host cellular receptor-binding region (amino acid positions 366-370 in HXB2 and
382-386 in SIVmac251) are also homologous in both genomes. At the same time, this region maintains sufficient intra-host genetic variability during the course of infection that provides informative insight into evolutionary trends. Ultimately, inter-species conservation of particular motifs facilitates alignment of these variable sequences with HIV-1 and enables the comparison of genetic changes that may have biological relevance or may be associated with neuroadaptation.

![Figure 1-3](image)

**Figure 1-3.** Models of viral population dynamics resulting in compartmentalization of virus in the brain.

SIV infection of macaques is also a useful animal model of HIV infection and AIDS because of the physiological and immunological resemblance to humans. In particular, the rhesus macaques of Indian origin used in the described studies are currently the best characterized and most utilized non-human primate model for AIDS.
research (43). Similarly to HIV-1 replication in humans, ongoing SIV replication in these animals results in high viral load and the turnover and progressive loss of CD4+ T cells, albeit at a more rapid rate, resulting in progression to AIDS within 1-3 years of infection in contrast to the 8-10 year disease progression in HIV-1-infected humans not receiving anti-retroviral therapy (44). The SIV/macaque model is also an ideal model of AIDS-related neuropathology, characterized similarly by the accumulation of perivascular macrophages, microglia and macrophage nodules, multinucleated giant cells, and gliosis (44-47). The incidence of this disease is also similar (30%) to that of humans (48). Although this similarity increases the appeal of the SIV/macaque model, the cost of macaques is much greater than for other experimental animal models (44). Studies using HIV-infected Nod/scid-IL-2Rycnull mice reconstituted with human hematopoietic stem cells obtained from cord blood (hNSG mice) show that human macrophages can be detected in the mouse central nervous system (CNS) and that HIV-1 proteins are found in a small number of cells (49), offering the promise of an alternative, less expensive animal model. However, use of humanized mice as a model for HIV neuropathogenesis is severely limited because human microglia, accounting for as many as two-thirds of productively infected cells in the CNS of HIV-infected patients with HIV-associated encephalitis (50), are not reconstituted in these mice (32). Thus, the HIV research field has relied heavily on efforts to minimize the number of NHP animals used by increasing the incidence of disease, including the development of a model in which CD8+ lymphocytes have been depleted, resulting in both increased incidence (~80%) and rapid progression to neuroAIDS, characterized by SIV-associated encephalopathy (51). Studies of this model have provided considerable insight into
HAND with regard to disease progression; however, depletion of the primary source of selective pressure, although transient, alters both the innate and adaptive immune system of the host, limiting potential applicability to the naturally progressing infected host. The effect of this immune modulation on the evolution of the virus was also, therefore, explored (described in Chapter 8) with the use of macaques both with and without CD8+ lymphocyte depletion.

**S100B as a Biomarker of SIVE Onset**

Given earlier NHP studies indicating infection of the brain as early as two weeks post-infection (52, 53), entry of virus from multiple sampling origins was not only possible, but probable. This expected finding imposed the critical question of which migration event was primarily responsible for the introduction of the neurovirulent strain(s) into the brain, requiring further knowledge of the timing of the onset of SIVE. Inflammation in the brain due to the release of neurotoxic cytokines and viral proteins from activated microglia and perivascular macrophages indirectly results in endothelial cell destruction and the breakdown of the blood brain barrier (BBB) (5). Once integrity of the BBB is compromised, proteins normally confined within the CNS are released into circulation (54). One such protein, soluble 100β (s100β), is a multifunctional protein expressed primarily by astrocytes (55) and is elevated in the serum of macaques with SIVE (56) as well in humans with traumatic brain injury (57, 58). We sought to monitor s100B protein levels in the serum over the course of SIV infection in order to determine the time of BBB leakage indicative of SIVE onset, hypothesizing that peaks in serum s100β levels would correlate with peak viral migration time from the bone marrow to the brain (Figure 1-4).
Significance and Innovation of the Described Studies

Given the current lack of successful treatment for HAND, investigating the evolutionary process of HIV in multiple cell types and tissues during and prior to brain infection is essential for the advancement of this field of medicine. Using both sophisticated evolutionary analysis and molecular biology, I have attempted to provide significant insights into the emergence of HIV/SIV neurovirulence within the host, focusing specifically on the following questions:

1. What are the evolutionary characteristics that drive viral sequence compartmentalization within the brain of SIV-infected macaques with SIVE? (Chapter 2)
2. Can epidemiological methods be applied to intra-host viral population studies for the evaluation of anatomical spread? (Chapter 3)
3. What are the peripheral tissue/cell type(s) that contribute to brain infection in SIV-infected macaques? And potentially neuroAIDS? (Chapters 4 and 5)
4. Can biomarkers of encephalitis onset help identify the timing of neuroinvasion and primary contributing tissue/cell population to neurovirulence? (Chapter 5)
5. What is the contribution of monocytes to evolutionary dynamics in HIV-infected humans diagnosed with HAND? (Chapter 6)
6. Are there distinct evolutionary patterns implicated in rapid progression and high incidence of neuroAIDS in the CD8+ lymphocyte-depleted macaque model? (Chapter 7)

The primary innovation used in answering these questions consisted of the application of statistical phylogenetic methods, able to incorporate complex models of evolution, to longitudinal sequences sampled from a variety of tissues and cell types from two well-defined animal models of neuroAIDS. Since the development of the first algorithm to compute a phylogenetic tree from amino or nucleic acid sequences (59), phylogenetic analysis has become a powerful tool for the investigation of evolution at a molecular level. Coalescent theory has built upon this, providing solid theoretical frameworks to study how the ancestral relationships of individuals sampled from a population, inferred by a phylogenetic tree, are influenced by the population
demographic history (60). More recently, the phylodynamic framework has been introduced (61), combining phylogenetic and population genetic analyses based on this coalescent theory, and is well suited to investigate both inter- and intra-host evolution of HIV/SIV (62, 63). Taking this idea a step further, Bayesian phylogeography (64-66) was introduced for the purpose of incorporating spatial information, commonly used in epidemiological studies to infer past transmission events and epidemic spread of a pathogen within and among communities or species (67-69) The host anatomy can be viewed similarly with the virus using the vascular system as means of transportation from one tissue to the next within various vectors, or host cell populations. Application of these analyses at the intra-host level, referred to as phyloanatomy (70), has been limited until now by the quality of the data sets, i.e. the lack of longitudinal data and/or sequences derived from several biological locations (71). However, for the studies described herein, sequences were successfully obtained from several time points as well as a multitude of tissues, thereby allowing for the use of these same tools to track movement of the virus within the host similar to the tracking of a virus in an epidemic spread within a population. Furthermore, similar to sampling for the epidemiological study of a virus, sampling as many locations as possible from the vast anatomical target space of the virus allows for a higher resolution investigation of this intra-host viral dispersion. Therefore, virus was sampled from several regions of the host known to harbor large populations of immune cells, the primary targets of the virus. These included plasma, bone marrow, and lung epithelial lining fluid (in addition to the CNS reservoir), resulting in the most extensive sampling protocol used to date in the field of HIV/SIV evolution.
Figure 1-4. Hypothesis model. A) Early seeding of the brain by monocytes infected with non-neurotropic virus. B) Later seeding of the brain by bone marrow monocytes infected with neurotropic/neurovirulent virus. C) Release of brain-specific proteins, such as s100β, into circulation due to encephalitis-associated increase in BBB permeability.
CHAPTER 2

EVOLUTION OF NEUROADAPTATION IN THE PERIPHERY AND PURIFYING SELECTION IN THE BRAIN CONTRIBUTE TO COMPARTMENTALIZATION OF SIV IN THE BRAIN OF Rhesus Macaques WITH SIV-ASSOCIATED ENCEPHALITIS*

Introduction

Human immunodeficiency virus type 1 (HIV-1)-associated neurocognitive disorders (HAND) remain a persistent complication despite the success of highly active antiretroviral therapy (HAART) in prolonging the progression to AIDS in HIV-1-infected patients (33). Although the post-HAART era has witnessed a decrease in the prevalence of the most severe form of HAND, HIV-associated dementia (HAD), an increase in prevalence has been observed for the milder forms, referred to as asymptomatic neurocognitive impairment (ANI) and mild neurocognitive disorder (MND) (27, 73). Moreover, these assessments have been reported to be an underestimation of the true prevalence, based on a study in which approximately 64% of HIV-1-infected patients classified as “non-complainers,” self-reporting no symptoms of cognitive impairment, tested positive for HAND using comprehensive neurocognitive testing criteria (74). Furthermore, with the life expectancy for patients receiving HAART now projected to be greater than 70 years (75, 76), the reported prevalence estimations can only be expected to increase, whereas our understanding of the critical components of disease pathogenesis, such as the ability of the virus to adapt to the microenvironment of the brain, remains limited.

Several studies have implicated viral evolution as a major proponent of HAND progression. Increased viral diversity and rapid progression to AIDS have been linked

with the development of the neurological complications of HIV infection (neuroAIDS) in HIV-1-infected patients (18). Furthermore, phylogenetically distinct viral subpopulations, or sequence compartmentalization, within the brain has been reported to be associated with neuroAIDS in HIV-1-infected patients and rhesus macaques infected with neurovirulent virus, although this association has been debated (reviewed in Bednar et al. [2015] (32)). Characterized by the significant clustering of sequences within a phylogenetic tree according to anatomical location, viral compartmentalization is indicative of either a single introduction of virus into that particular compartment followed by long-term, isolated viral replication, or multiple introductions of distinct viral subpopulations whose ultimate survival demands specific viral genotypic/phenotypic features required for both entry into and replication in the unique cell types found within that compartment. The former scenario offers validity for modeling brain sequence compartmentalization through the limited transport of molecules across the blood-brain barrier (BBB), which separates the brain from circulating virus (reviewed in Banks, Ercal, and Price [2006] (77)). However, recent phylogenetic and immunological studies have suggested continuous infiltration of the brain by “Trojan Horse” cells (78-80). In this alternative scenario, compartmentalization of virus in the brain is likely the result of successful entry and replication of a viral variant that has passed the selective qualifications associated with the unique microenvironment of the central nervous system (CNS) via evolution in similar peripheral cell types. Extensive evolutionary analysis is needed to determine this distinction; however, the ethical limitation of HIV-1-associated neuropathological studies in humans renders this type of analysis difficult.
As in several other diseases, sampling limitation has created a need for animal models that accurately reproduce pathogenesis. Infection of rhesus macaques with pathogenic simian immunodeficiency virus (SIV) clones and/or viral swarms is a widely used model for HIV/AIDS (44) as well as neuroAIDS (81). SIV-infected macaques exhibit clinical manifestations similar to those of HIV-1-infected humans, albeit on a shorter time scale of approximately 1 to 3 years (44), and the incidence (<25%) of SIV-associated encephalitis (SIVE), the pathological hallmark of neuroAIDS, is similar to that of untreated seropositive patients (<30%) (82). However, low incidence, extreme maintenance cost, and disease timeline associated with the macaque model limits its usefulness in terms of producing rapid and statistically robust results. Such limitations have led to the development of rapid disease models (83-85) that utilize antibody-mediated depletion of the CD8+ lymphocyte arm of the antiviral response (84, 86, 87), or a combinatorial approach to infection using neurovirulent and immunosuppressive virus (88), resulting in an increased incidence (>75%) of SIVE (84, 87, 88). In this study, we have used phylogenetic analysis of viral sequences derived from a variety of tissues/cell types, including the meninges and distinct CNS cortical regions, for both the CD8+ lymphocyte-depleted and non-CD8+ lymphocyte-depleted, or naturally progressing, macaques in order to characterize the evolutionary history of compartmentalized virus in the brains of animals that have developed SIVE despite different immunological backgrounds.

Materials and Methods

Study population. Six CD8+ lymphocyte-depleted (D01 to D06) and 12 naturally progressing (N01 to N12) rhesus macaques of Indian origin between 4 and 11 years of age were intravenously inoculated with the well-characterized SIVmac251 (1 ng SIV
p27) viral swarm (89, 90) and are hereafter referred to as the Mac251-DEP and Mac251-NP cohorts, respectively. Procedures involving the Mac251-DEP animals were performed with the approval of Tulane University’s Institutional Animal Care and Use Committee. For additional information on the treatment and handling of macaques in this cohort as well as gross pathology, see the study of Strickland et al. (41). Procedures on macaques with an intact immune system (Mac251-NP), which were allowed to progress naturally to simian AIDS (SAIDS), were conducted according to the standards of the American Association for Accreditation of Laboratory Animal Care and IACUC protocol 04802, and treatment of these animals was in accordance with the Guide for the Care and Use of Laboratory Animals (91). Further detailed information on the handling and supervisory guidelines for these animals can be found in Lamers et al. (92). Plasma viral loads were monitored as previously described by quantitative PCR (qPCR) methods targeting a conserved sequence in gag (93, 94). All possible measures were taken to minimize discomfort of the animals, and the guidelines for humane euthanasia of rhesus macaques were followed. Animals D01, D02, N06, and N07 (two macaques from each cohort) were euthanized at ~21 days post-infection (dpi) in order to evaluate early evolutionary events leading to brain infection. The remaining animals were euthanized at the onset of SAIDS between 75 and 118 dpi (Mac251-DEP) and 204 and 373 dpi (Mac251-NP), which was confirmed at autopsy (for a brief description of gross pathology reports for the Mac251-NP macaque cohort, refer to Table 2-1). All animals were perfused with heparinized saline prior to euthanasia to clear the CNS vasculature, and SIV-associated encephalitis (SIVE) was determined in an autopsy by a veterinary pathologist who was blind to rhesus macaque conditions
based on the presence of microglial nodules and multinucleated giant cells and confirmed by immunohistochemistry staining for SIV p27, as previously described (82, 95-97). As SIVE is associated with rapid disease progression (82), data are reported in this study for 6 of the 10 longitudinally monitored macaques based on a relatively rapid temporal progression to SAIDS (204 to 373 dpi).

**RNA and DNA extraction.** Viral RNA (and integrated viral DNA when available) was isolated from the SIVmac251 inoculum and cell-free plasma samples (140 to 280 μl) using the QIAamp viral RNA minikit (catalog no. 52904; Qiagen) following the manufacturer’s protocol except that two 50-μl final elutions using buffer AVE were performed. Total RNA and integrated DNA, or genomic DNA (gDNA), were isolated from brain tissue (30 to 50 ng), bone marrow aspirates, and sorted bronchoalveolar lavage (BAL) fluid macrophages and peripheral CD3+ T lymphocytes and CD14+ monocytes (sorted as described in Autissier et al. [2010] (98)) using the AllPrep DNA/RNA minikit (catalog no. 80204; Qiagen) according to the manufacturer's guidelines except for two 50-μl final elutions using RNase-free water. Sequences from plasma, bone marrow, and sorted cell samples were obtained for two or three time points, ~21 days post-infection, ~60 dpi, and necropsy, from four Mac251-DEP macaques and four time points, ~21 dpi, ~90 dpi, ~180 dpi, and necropsy, from six Mac251-NP macaques. Samples from the meninges and three distinct cortical regions (parietal, frontal, and temporal cortex) were also obtained at necropsy for sequencing when available. The 100-μl final volume of RNA generated from each method was concentrated using RNeasy MinElute cleanup kit (catalog no. 74204; Qiagen).
<table>
<thead>
<tr>
<th>Macaque</th>
<th>Diagnosis or outcome</th>
<th>Symptom(s) or gross pathology&lt;sup&gt;a&lt;/sup&gt;</th>
<th>Age (yr)</th>
<th>At time of infection</th>
<th>At necropsy</th>
</tr>
</thead>
<tbody>
<tr>
<td>N02</td>
<td>SAIDS</td>
<td>Lymphopenia</td>
<td>5.4</td>
<td>6</td>
<td></td>
</tr>
<tr>
<td></td>
<td></td>
<td>Opportunistic infection (primarily lungs)</td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td></td>
<td></td>
<td>Opportunistic infection (pancreas and lungs)</td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>N04</td>
<td>SAIDS</td>
<td>Opportunistic infection</td>
<td>5.4</td>
<td>6.1</td>
<td></td>
</tr>
<tr>
<td></td>
<td></td>
<td>(primarily lungs)</td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>N05</td>
<td>SAIDS</td>
<td>Multiple opportunistic infections</td>
<td>6.4</td>
<td>7.4</td>
<td></td>
</tr>
<tr>
<td></td>
<td></td>
<td>in lungs</td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>N06</td>
<td>Euthanized early</td>
<td>No opportunistic infections</td>
<td>4.4</td>
<td>4.4</td>
<td></td>
</tr>
<tr>
<td></td>
<td></td>
<td>Mild inflammation and satellite</td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td></td>
<td></td>
<td>cell proliferation in DRG</td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td></td>
<td></td>
<td>consistent with SIV neuropathy</td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>N07</td>
<td>Euthanized early</td>
<td>No opportunistic infections</td>
<td>4.4</td>
<td>4.4</td>
<td></td>
</tr>
<tr>
<td></td>
<td></td>
<td>Variable DRG inflammation</td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>N09</td>
<td>SAIDS</td>
<td>Lymphoma</td>
<td>4.6</td>
<td>5.4</td>
<td></td>
</tr>
<tr>
<td></td>
<td></td>
<td>Arteriopathy</td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td></td>
<td></td>
<td>Myocardial degeneration and fibrosis</td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td></td>
<td></td>
<td>Gliosis</td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>N10</td>
<td>SAIDS/SIVE</td>
<td>Severe meningoencephalitis</td>
<td>7.7</td>
<td>8.3</td>
<td></td>
</tr>
<tr>
<td></td>
<td></td>
<td>Opportunistic infections, including CMV</td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>N12</td>
<td>SAIDS</td>
<td>Moderate endocardial fibrosis</td>
<td>4.2</td>
<td>4.9</td>
<td></td>
</tr>
<tr>
<td></td>
<td></td>
<td>Severe enteritis and colitis</td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td></td>
<td></td>
<td>Gliosis</td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td></td>
<td></td>
<td>Opportunistic infection</td>
<td></td>
<td></td>
<td></td>
</tr>
</tbody>
</table>

<sup>a</sup>DRG, dorsal root ganglia.

cDNA synthesis. cDNA was generated from the RNA of each sample using the SuperScript III first-strand synthesis system (catalog no. 18080-051; Invitrogen Life Technologies). Viral RNA isolated from both the inoculating viral swarm and plasma samples were reverse transcribed using the provided oligo(dT)<sub>20</sub> primer. Total RNA from peripheral CD3<sup>+</sup> and CD14<sup>+</sup> cells, BAL fluid, and bone marrow aspirate samples was reverse transcribed using the KOUTR primer (5′-TGTAATAAATCCCTTCCAGTCCCCC-3′) (9479 to 9454 bp of SIVmac239) (99), whereas total RNA from brain tissue and meninges was reverse transcribed using the
cULTR primer (5′-ATGGCAGCTTTATTGAAGAGG-3′) (10125 to 10145 bp of SIVmac239) (92), a primer that binds in the SIV 3′ long terminal repeat (3′ LTR) and 5′ LTR regions. A modified cDNA synthesis protocol was followed for the inoculating viral swarm, plasma, and brain tissue RNA to maximize the length of the cDNA (92). Total RNA from the remaining tissues/cell populations was reverse transcribed according to the manufacturer’s recommendations as follows: RNA was incubated at 65°C for 5 min with deoxynucleoside triphosphates (dNTPs) (0.5 mM [each]) and 5 μM KOUTR and then cooled quickly to 4°C. First-strand cDNA synthesis was performed in a 40-μl reaction volume containing 1× reverse transcription buffer (10 mM Tris-HCl [pH 8.4], 25 mM KCl), 5 mM MgCl₂, 10 mM dithiothreitol, 2 U/μl of RNase-OUT (RNase inhibitor), and 10 U/μl SuperScript III reverse transcriptase (RT). The reaction mixture was heated to 50°C for 60 min and then 85°C for 5 min. The reaction mixture was cooled to 37°C, and 0.1 U/μl of Escherichia coli RNase H was added, followed by a 20-min incubation. All cDNA was stored at −20°C until needed.

**Single-genome sequencing.** Viral sequences derived from the Mac251-DEP cohort were obtained using bulk PCR and cloning methods (41, 78). Due to the potential limitations of clonal analysis, a modified single-genome sequencing protocol based on previously published methods (100) was used for all samples obtained from the Mac251-NP cohort as well as the frontal lobes from the brains of macaques in the Mac251-DEP cohort. The sequencing protocol used for the remaining samples obtained from the Mac251-DEP cohort has been previously described (78). cDNA was serially diluted until an average of 30% or less of the nested PCRs were positive. During the first round of PCR, diluted cDNA was amplified in 25-μl reaction mixtures containing 1×
Platinum blue PCR supermix (Invitrogen Life Technologies) and 0.2 μM each primer: SOUTF (5′-GGCTAAGGCTAATACATCTTCTGCATC-3′) and NOUTR (5′-TTTAAGCAAGCAAGCGTGGAG-3′) (coordinates 6565 to 6591 and 10102 to 10122 of SIVmac239, respectively). In the first round of PCR, the following cycling parameters were used: (i) 95°C for 5 min; (ii) 35 cycles, with 1 cycle consisting of 94°C for 1 min, 58°C for 1 min, and 72°C for 4 min; (iii) 72°C for 10 min. If template cDNA was present, a 3.5-kb product was expected containing the complete env and nef gene products.

Second-round gp120 PCR amplification consisted of 2 μl of the first-round PCR added to a 23-μl second-round reaction mixture consisting of 1× Platinum blue PCR supermix (Invitrogen Life Technologies) and 0.2 μM each primer: SINF (5′-GTAAGTATGGGATGTCTTGGGAATCAG-3′) and SINR (5′-GACCCCTCTTTTATTTCTTGAGGTGCC-3′) (coordinates 6598 to 6624 and 8158 to 8184 of SIVmac239, respectively). In the second round of PCR, the following cycling parameters were as follows: (i) 95°C for 5 min; (ii) 35 cycles, with 1 cycle consisting of 94°C for 1 min, 58°C for 1 min, and 72°C for 2 min; (iii) 72°C for 10 min. This second-round PCR generates a 1.5-kb product when positive, containing the entire envelope gp120 sequence with flanking sequence on each end. Second-round gp120 PCRs were visualized on 1% agarose gels stained with ethidium bromide, and reaction mixtures containing single, 1.5-kb products were considered positive and selected for sequencing. The SOUTF, SINF, and SINR primers are based on published oligonucleotide sequences (101), while the remaining primers were designed using Primer3 (102), observing regions of conservation in alignments of published SIVmac251
sequences downloaded from the Los Alamos HIV Sequence Database (http://www.hiv.lanl.gov).

RNA and DNA extractions, cDNA synthesis, and first-round PCR set-up were always conducted in a restricted-access amplicon-free room with separate air-handling and laboratory equipment where no amplified PCR products or recombinant cloned plasmids are allowed and work surfaces and equipment are thoroughly cleaned before and after use with Eliminase (Decon Labs, Inc.). Sequencing was performed on an Applied Biosystems 3730xl DNA analyzer (Life Technologies) at the University of Florida Interdisciplinary Center for Biotechnology Research (UF ICBR).

**Sequence alignment.** Sequence alignment was performed as previously described (92). Briefly, following sequence assembly in Geneious R7 (available from BioMatters, Auckland, New Zealand), sequences (SIVmac239 coordinates 6706 to 8049) were aligned using the Clustal X algorithm (103) implemented in BioEdit (104) (available from http://www.mbio.ncsu.edu/bioedit/bioedit.html); the alignment was further modified by a manual optimization protocol taking into account conserved glycosylation motifs (105). The highly variable region within the V1 domain was removed so as not to confound the genetic analysis. Intra-host recombinants were determined as discussed below and removed prior to sequence analysis. Approximately 20 sequences per tissue per time point were obtained after removal of potential recombinants (Figure 2-1).
Figure 2-1. Spectrum of tissues and cell populations sampled from the rhesus macaques.
**Viral diversity and divergence estimates.** Estimates of mean pairwise viral diversity within individual tissues/cell populations, as well as mean divergence of longitudinally sampled sequences from the inoculating viral swarm were calculated in MEGA v5.2.2 (106) (available from http://www.megasoftware.net) using the maximum composite likelihood model of nucleotide substitution (107) and 1,000 bootstrap replicates. Due to the influence of the bulk PCR/cloning method on sequence heterogeneity (100), mutated sites representing <1% of observed point mutations (estimated PCR error rate) were removed from the final Mac251-DEP alignment, as described by Strickland et al. (89). Furthermore, after removal of these sites, the overall viral diversity and divergence within tissue/cell populations during early and late infection were compared for the two cohorts (data available upon request) to determine whether differences between the two animal models could be explained by the sequencing methodology.

**Compartmentalization analysis.** Compartmentalization analyses results were obtained from two separate compartmentalization tests in HyPhy (108) (available from http://hyphy.org/) in order to evaluate the extent of distinct viral subpopulations within individual tissues. Analyses included both tree- and distance-based methods. Tree correlation coefficients were calculated based on the number of branches (rb) or branch length (r) separating sequences within separate defined compartments (109) in maximum likelihood trees generated using the generalized time reversible (GTR) nucleotide substitution model with gamma-distributed rate variation across sites (GTR+G). Tree reconstruction was performed using RAxML v8.1.23 (110) using 1,000 bootstrap replicates. Statistical significance was determined using a null distribution of
permutated sequences (1,000 permutations), whereby a P value of \( \leq 0.5 \) was considered significant. The Simmonds association index (SAI or AI) was determined for sequence alignments using SIVmac239 as a reference sequence. The SAI represents the mean ratio of 100 bootstrap replicates of the association value, calculated from the test sequences, to those of 10 sample-reassigned controls. The association value (d) is defined as \( d = (1 - f) / (2n - 1) \), where \( n \) is the number of sequences below the node and \( f \) is the frequency of most common sample type (111). Bootstrapping (1,000 replicates) was used as a test of significant compartmentalization, for which support of >80% was considered significant. A more thorough description of such tests can be found in Zarate et al. (2007) (112).

**Selection analysis.** Because statistical measures of metapopulation structure can be affected by selection as well as migration dynamics, an unrestricted branch site random-effects model, referred to as BUSTED (Branch-Site Unrestricted Statistical Test for Episodic Diversification; implemented in the datamonkey webserver [http://datamonkey.org](http://datamonkey.org)), was used to test for gene-wide episodic diversifying selection (113). The analysis was restricted to only internal branches, which are assumed to capture at least one round of virus replication, to mitigate the biasing effects of transient deleterious mutations on the ratio of non-synonymous to synonymous substitution rates estimates along terminal branches, where selection has not had time to fully filter such population level variation (114, 115).

In addition, nucleotide sequence alignments for all tissues/cell populations at individual time points for each macaque were used to determine site-specific selection over the course of infection for both macaque cohorts. The fast, unconstrained
Bayesian approximation for inferring selection (FUBAR) model (116) (implemented in http://datamonkey.org) was used to identify potential individual amino acid sites under selection within viral gp120 sequences for individual macaques as well as for individual tissues/cell types at each time point. Sites with a posterior probability of >0.9 of an increased (diversifying) or decreased (purifying) rate of non-synonymous relative to synonymous substitutions were considered to have experienced a significant level of selective pressure. Macaques were then classified according to SIVE diagnosis or early sacrifice in order to determine similarities and differences among classifications across macaque cohorts.

**Recombination analysis.** Recombination analysis was carried out for sequence populations in each tissue and in each macaque using three methods, as previously described (117): (i) GARD (118), (ii) an algorithm that uses split decomposition networks in combination with the “phi-test” (119, 120) implemented in Splitstree4 software (121) calibrated specifically to identify likely recombinants within intra-host viral populations (119, 122, 123), and (iii) Recombine (124). Recombinant sequences were also identified using bootscanning, implemented in Simplot (125) and RDP (126), with observed trends similar to those obtained with the algorithms listed above (data not shown). Putative recombinant sequences were screened against viral swarm sequences in order to determine whether they had originated from the inoculum.

**Viral RNA and DNA quantification.** Viral RNA and gDNA levels were estimated for longitudinally sampled and postmortem tissues/cell populations of Mac251-NP macaques and postmortem frontal lobes for Mac251-DEP animals. Viral gp120 RNA was quantified using QUALITY (127) based on limiting dilution PCR-positive
amplifications (data available upon request). For time points for which only one dilution was used, which cannot be used for copy number estimation in QUALITY, the following calculation was employed: \(-\ln(1 - \text{proportion of successful PCR wells}) \times (\text{dilution/PCR template volume [2 μl]}),\) assuming a Poisson probability distribution of RNA templates within PCR samples. Integrated viral DNA was quantified using qPCR amplification of Gag \(p27\) using the recommended concentrations of TaqMan universal master mix II (Invitrogen Life Technologies) and probe and the 0.9 μM concentration of primers used by Hofmann-Lehmann \textit{et al.} (128). Briefly described here, the TaqMan probe (SIVmac239 coordinates 1487 to 1506) consisted of 5'-TGTCCACCTGCCATTAAGCCCGA-3', whereas the primer sequences were as follows: 5'-GCAGAGGAGGAAATTACCCAGTAC-3' (sense; SIVmac239 coordinates 1441 to 1464) and 5'-CAATTTTACCCAGGCATTATGTT-3' (antisense; SIVmac239 coordinates 1532 to 1508). Using the Applied Biosystems 7500 FAST PCR system, the cycling parameters were as follows: (i) initial denaturation (95°C for 10 min); (ii) 35 to 40 cycles of 95°C for 15 s; (iii) 60°C for 1 min. Cell-associated viral sequence copy numbers were represented in terms of copies/10,000 cells based on the stock concentration of gDNA and the presumed 6.2 pg gDNA/macaque cell: \((\text{viral copy number/ng gDNA}) \times (62 \text{ ng gDNA/10,000 cells}).\)

**Nucleotide sequence accession numbers.** Sequences were deposited in GenBank (accession numbers JF764947 to JF766081 [Mac251-DEP], KR998525 to KR999900 [Mac251-NP N02, N10, and inoculating viral swarm], KX068225 to KX068624 [Mac251-NP N04-N07, N09, and N12], and KX081185 to KX082629 [remainder of the Mac251-NP cohort]).
Results

SIV longitudinal sampling and disease progression in two macaque cohorts. Viral sequences from the two SIVmac251-infected macaque cohorts (Mac251-DEP and Mac251-NP) were successfully obtained from various tissues/cell populations sampled over time, consisting of plasma, bone marrow, elicited bronchoalveolar lavage (BAL) fluid macrophages, and fluorescence-activated cell-sorted (FACsorted) CD3+ T lymphocytes and CD14+ monocytes from blood (Figure 2-1). Samples from meninges and from the frontal, temporal, and parietal cortexes were also obtained at necropsy in addition to peripheral tissues/cell populations.

As expected based on previously reported incidence, three (D03 to D05) of the four longitudinally monitored Mac251-DEP animals and one (N10) of the 12 animals in the Mac251-NP cohort were diagnosed with SIV encephalitis (SIVE) during autopsy, as assessed by a veterinary pathologist who was blind to rhesus macaque conditions. The remaining Mac251-DEP macaque, D06, was determined not to have SIVE encephalitis (SIVnoE) but still presented with meningitis, which has often been lumped together with SIVE in the spectrum of SIV-associated CNS neuropathology. The Mac251-NP cohort was thus fundamental to this study, as macaque N02 exhibited no degree of CNS neuropathology (SIVE or meningitis) but still presented with sufficient viral RNA in the brain for sequence analysis (~20 sequences/lobe), allowing for the unique opportunity to assess evolutionary patterns along this neuropathological spectrum.

Prior to evolutionary analysis, patterns in viral load and CD4+ T-cell counts were investigated in order to rule out possible confounding factors responsible for differences in neuropathological disease progression (95). Despite inherent immunological differences due to CD8+ lymphocyte depletion of the Mac251-DEP animals, differences
in the magnitude of plasma virus at baseline (0 dpi), early infection (7 dpi), peak viremia (8 to 14 dpi), and set point (35 to 60 dpi), were not specific to the cohort or SIVE diagnosis (Figure 2-2). Similar observations were reported for differences in CD4+ T-cell nadir and counts over time (Figure 2-2).

Figure 2-2. Viral load and CD4+ T-cell population characteristics for naturally progressing (left) and CD8+ lymphocyte-depleted (right) macaques. Viral load was determined as previously described by quantitative PCR methods targeting a conserved sequence in gag (27, 28). CD4+ T-cell counts were determined using flow cytometry, as previously described (32). (a) CD4+ T-cell nadir (lowest reported CD4+ T-cell count) and viral load at baseline (0 to 7 days postinfection [dpi]), initial peak (8 to 14 dpi), and set point (35 dpi for naturally progressing macaques and 60 dpi for CD8+ lymphocyte-depleted macaques) for both macaque cohorts. (b) Viral load over the course of infection. (c) CD4+ T-cell counts over the course of infection.
Sequence compartmentalization and sub-compartmentalization occur solely within the brains of macaques with SIVE and are independent of CD8+ lymphocyte depletion. Maximum likelihood (ML) trees for the Mac251-DEP animals have been reported elsewhere (78) but have been included with those inferred for the Mac251-NP cohort using a similar color scheme for easy comparison (Figure 2-3). Upon preliminary visualization of the ML trees, brain sequences appeared monophyletic for N10 (SIVE) and D04 and D05 (SIVE), whereas brain sequence monophyly was not present for the macaque with meningitis (D06) or the macaque with no CNS neuropathology (N02). Further sub-compartmentalization of sequences belonging to particular cranial lobes was also observed within both cohorts. This phenomenon could not be explained entirely by clonal expansion, as the largest proportion of identical sequences observed was 50% (D05 temporal lobe; Figure 2-4). Results of the quantitative analysis using tree correlation coefficients (TCCs) have been published previously by Strickland et al. (78) for the Mac251-DEP animals and have been reported again here for easy comparison (Figure 2-5). Additional analysis using the Simmonds association index (SAI) was performed in this study to incorporate uncertainty in the ML tree topology (Figure 2-5). Significant brain sequence compartmentalization, as indicated by positive results for both quantitative methods (P value of ≤ 0.05 for TCCs and AI of <0.33 for SAI), confirmed the tree visualizations and was determined to be significantly associated with SIVE diagnosis using Fisher’s exact test (P value of <0.05). Sub-compartmentalization of sequences within individual cranial lobes was assessed using the more sensitive SAI compartmentalization test alone and was observed for all macaques with SIVE (SIVE macaques), with the
exception of D03. This exception was likely due to the relatively short life span of this animal (75 dpi), as the extent of brain sequence compartmentalization has been reported to correlate with length of time of infection (129-131).

The results described above are consistent with previous reports of association between brain sequence compartmentalization and neuroAIDS (reviewed in Bednar et al. [2015] (32)). Furthermore, although only one macaque within the Mac251-NP cohort was diagnosed with SIVE, it is important to note that this macaque (N10) exhibited the highest degree of overall brain compartmentalization (Figure 2-5) and was highly sub-compartmentalized, according to the SAI (Figure 2-5). Thus, even if both models appear to be appropriate for further investigation of this evolutionary phenomenon, increased resolution is observed in the naturally progressing macaques, which is likely due to their longer life spans, as has been previously suggested (129-131).

**Late entry of a uniform viral population is implicated in the contribution to brain sequence compartmentalization in macaques with SIVE.** Because virus can penetrate the macaque brain early on during infection (as early as 21 dpi in the Mac251-DEP animals), we sought to determine whether this early entry event is implicated in evolution of virus in the brain and the significant compartmentalization of viral sequences obtained from necropsy samples. Pairwise genetic distances estimated for tissue-specific sequences relative to other sequences within the given tissue (diversity) or to the viral swarm at the time of inoculation (divergence) often reveal information regarding evolutionary processes such as timing of a migration event and natural selection leading up to clustering patterns such as compartmentalization, given assumptions about the initial virus population size. For example, in the case of early
infection of the brain by few viral variants followed by limited genetic drift, lower levels of sequence diversity within the brain and sequence divergence from the inoculating viral swarm would be expected relative to that of peripheral sequences sampled contemporaneously. Alternatively, late entry of a relatively larger viral population size that has acquired specific substitution(s) critical for migration across the BBB would be characterized by similar divergence but reduced diversity compared to that of peripheral sequences. We hypothesized that the distinct viral subpopulation within the brains of the SIVE animals, resulting in significant compartmentalization, was indeed due to significant differences in diversity and/or divergence, as described in each of these scenarios. Reduced brain sequence diversity and/or divergence relative to peripheral sequences would similarly be expected when comparing SIVE macaque brain virus to that of the SIVnoE animals, for which compartmentalization was not observed, and hence, differences between brain and peripheral sequence diversity and/or divergence in these animals would also not be expected. For this study, necropsy-sampled brain sequence diversity and divergence from the inoculating viral swarm were compared to those of necropsy-sampled peripheral tissues/cell populations for both macaques diagnosed with SIVE and macaques with sufficient (N09 excluded) detected virus in the brain but no SIVE (Figure 2-4). Because meninges have been shown previously to harbor sequences genetically similar to virus both in the brain and periphery (132), as well as distinct from the brain parenchyma (89), sequences from this location were considered separately.
Figure 2-3. Maximum likelihood (ML) phylogenetic tree reconstruction for each macaque with detectable viral RNA in the brain. Mac251-NP (N02, N09, and N10) and Mac251-DEP (D03 to D06) macaques were sampled longitudinally prior to necropsy, whereas D01 and D02 (boxed) were sacrificed at 21 days postinfection (dpi). Trees for D01 to D06 are adapted from Strickland et al. (10), and taxa are colored according to tissue sampling origin with respect to peripheral and cranial lobe locations. For all unrooted ML trees, the GTR + G model of nucleotide substitution was used in RAxML with bootstrapping analysis (1,000 replicates).
Figure 2-4. Percentage of total brain tissue viral RNA sequences within individual CD8+ lymphocyte-depleted and naturally progressing macaques that share 100% sequence identity. *ES, early sacrifice. Macaques D01 and D02 were sacrificed at 21 dpi.

Significant differences were observed between peripheral tissues/cell populations, meninges, and brain sequence diversities for macaque N10 (SIVE), whereby brain sequences within the distinct cortical regions exhibited significantly lower diversity than in meninges ($P < 0.05$), which, in turn, was significantly lower ($P < 0.001$) than combined peripheral tissues/cell populations (Figure 2-6a). However, a significant difference in divergence was not observed for any of these three anatomical classifications within this macaque (Figure 2-6b). Importantly, significant differences in both diversity and divergence between these three tissue compartments were not observed for N02 (Figures 2-6a and b), the Mac251-NP macaque with detectable virus in the brain but no pathological markers of SIVE. Indeed, viral diversity and divergence were also significantly lower in the brain ($P < 0.001$) and meninges ($P < 0.01$) of macaque N10 versus N02 (Figure 2-6a). The difference in diversity was not due to relatively low diversity for N10 sequences, as the difference in viral diversity between the peripheral tissues/cell populations of these two macaques was not significant.
(Figure 2-6a); however, the difference in divergence may have been due to relatively lower N10 sequence divergence from the viral swarm, as the difference in viral divergence between the peripheral tissues/cell populations of these two macaques was approaching significance \((P = 0.09)\) (Figure 2-6b).

Similar results were observed for viral diversity within the Mac251-DEP cohort (Figure 2-6c), wherein brain sequence diversity for macaques D03, D04, and D05 (SIVE) was significantly lower \((P < 0.01)\) than that of D06 (SIVnoE). Unlike N10 in the Mac251-NP cohort, however, brain sequence diversity for D03, D04, and D05 was not significantly lower than that of peripheral tissues/cell populations, although the mean pairwise genetic difference (nucleotide substitutions/site) was almost 1.5-fold, suggesting that, in both models, reduced viral genetic diversity in the brain relative to that of peripheral tissues/cells likely contributes to brain sequence compartmentalization in SIVE macaques. These also exhibited similar divergence estimates between peripheral tissues/cells and the brain (Figure 2-6d). Importantly, similar to Mac251-NP animal N02 (SIVnoE), neither brain sequence diversity nor divergence estimates for Mac251-DEP animal D06 (SIVnoE) significantly differed from that of peripheral sequences.

In order to rule out episodic diversifying selection just prior to brain entry as a driving factor for the observation of brain sequence compartmentalization, an unrestricted branch site random-effects model (BUSTED) was used, revealing evidence of transient diversifying selection events along each of the Mac251-DEP and Mac251-NP macaque phylogenies (likelihood ratio test [LRT] \(P\) values = 0.0), irrespective of SIVE status. The minimum number of amino acid site changes (SIVmac239 gp120
E78N and S128L/I belonging to Mac251-DEP macaque D04) identified by this method were shared among all macaques, with the exception of S128L/I for early sacrificed Mac251-NP macaque N07. These sites were mapped onto individual macaque phylogenies and corresponded to either internal branches that did not give rise to brain sequence lineages or did not directly separate brain sequence clades from sequences belonging to other tissues/cell types (Figure 2-7), indicating that these diversification events were not responsible for the observed brain sequence compartmentalization.

When taking into consideration differences in brain sequence evolution alone as well as brain sequences relative to peripheral sequence evolution between SIVE and SIVnoE macaques, the results suggest that, despite early brain infection in the Mac251-DEP cohort, compartmentalization of brain sequences in SIVE animals is likely the result of tissue-specific purifying selection of a viral subpopulation that has acquired, through evolution in the periphery, a more readily neuroadaptive genotype/phenotype. Given the similar divergence estimates between brain and peripheral sequences in the SIVE animals, this neuroadaptive phenotype is likely associated with facilitated entry during late infection, although sufficient time may not have passed after this entry event to observe replicative advantages in the evolutionary history, as would be characterized by the eventual reduction in divergence.

**Differences in purifying selective pressure may contribute to the evolution of neurovirulent strains and entry of virus into the brains of macaques with SIVE.**

We next sought to determine whether differences in selective pressure over time could be found among the macaques diagnosed with SIVE or without SIVE and whether particular regions of the SIV genome were affected more than others (Figures 2-
8a to c). Macaques were classified as SIVE, SIVnoE, or early sacrificed (ultimate neuropathological outcome unknown), and sites experiencing selective pressure that were shared among at least two macaques within each category, when available, were plotted along the length of the amino acid sequence. Macaques N10 (SIVE) and D06 (SIVnoE) were an exception, as they were the only macaques in their respective categories within their cohort. A high concentration of sites under selection was observed in the first half of the gene for the Mac251-DEP macaques at ∼60 dpi and the last time point (Figure 2-8b), which was similar to the selection pattern observed at ∼180 dpi for the Mac251-NP macaques (Figure 2-8a). There was also an increase in the number of sites experiencing purifying selection over time for the constant region C2 within both cohorts, particularly for SIVE macaques. Because this is a heavily glycosylated region, these sites were investigated for the presence of asparagaine residues prior to the last time point. Although none of the sites corresponded to previously known or potential glycosylation sites, evidence suggests a role for increased C2 site conservation in the ability of the virus to either enter or replicate in the brain, and this should be considered for further investigation with respect to its role in this process. An increase in the number of sites within pV1-V2 regions (pV1 refers to partial V1) under selection was also observed within both cohorts over time for SIVE macaques; however, this trend was observed for SIVnoE macaques as well, suggesting generalized immune-driven evolution. Across the gp120 gene in general, a larger difference was observed in the proportion of sites experiencing either purifying or diversifying selection between the SIVnoE and SIVE macaques within the Mac251-NP cohort compared with the Mac251-DEP animals (Figure 2-8c). Considering the facts
that the SIVnoE macaque within the Mac251-DEP cohort was diagnosed with meningitis whereas the SIVnoE macaque within the Mac251-NP cohort exhibited no degree of CNS neuropathology, the results indicate that selective pressures are relatively similar for virus subpopulations responsible for meningitis and SIVE, which may explain the similar brain sequence divergence between the two groups (SIVE and SIVnoE) within the Mac251-DEP cohort, but not the Mac251-NP cohort.

**Different levels of recombination do not readily distinguish macaques based on SIVE status.** In order to determine whether differences in recombination patterns could be observed between the macaques diagnosed with SIVE or without SIVE, percent recombination was compared for macaques N10 (SIVE) and N02 (SIVnoE) within the Mac251-NP cohort (Figure 2-9a) and macaques D03, D04, and D05 (SIVE) and D06 (SIVnoE) within the Mac251-DEP cohort (Figure 2-9b). Although differences were not significant, the Mac251-NP macaque with SIVE, N10, appeared to have relatively low levels of recombinant sequences in each of the distinct cortical regions examined within the brain, CD14+ monocytes and CD3+ lymphocytes, plasma, and bone marrow compared to the macaque with no CNS neuropathology (N02). In contrast, virus from the Mac251-DEP macaques diagnosed with SIVE (D03, D04, and D05) exhibited similar levels of recombination in the brain and even greater levels in peripheral tissues/cell populations compared to the macaque with meningitis (D06).
Figure 2-5. Analysis of tissue-dependent compartmentalization of viral RNA sequences. aThe designated brain tissue consisted of sequences from the parietal, frontal, and temporal lobes when applicable, whereas peripheral tissues includes sorted, circulating CD3+ T lymphocytes and CD14+ monocytes, BAL macrophages, bone marrow, and plasma. Note that sequences could not be obtained from the temporal lobe of macaque N09. bTree correlation coefficients (TCC) were calculated based on the number of branches (rb) or branch length (r) separating sequences within separate defined compartments. TCC data are taken from Strickland et al. (2014) and reported here for comparison. Statistical significance was determined using a null distribution of permuted sequences (1,000 permutations). A p-value of _0.05 was considered significant. cSimmonds association index (SAI or AI) represents the mean ratio of 100 bootstrap replicates of the association value, calculated from the test sequences, to those of 10 sample-reassigned controls. The association value (d) is defined as \[ d = \frac{(1 - f)}{(2n - 1)} \], where n is the number of sequences below the node and f is the frequency of most common sample type. AI values of <0.33 are shown in boldface type, indicating significant compartmentalization. The bootstrap support (BS) for AI values is also provided. SIVmac239 was used as a reference sequence for SAI determination. BS of >80 was considered significant. d ES, early sacrifice. The animals were sacrificed at 21 dpi. *P value of <0.05 for Fisher’s exact test of association of SIVE with compartmentalization.
Viral genomic RNA and DNA levels in the brain readily distinguish macaques based on SIVE status. A similar number of brain sequences obtained from macaques N10, N02, and D03 to D06 were used for sequence and phylogenetic analyses (Figure 2-1); however, the question remained as to whether or not differences in degrees of viral production in the brain would be observed in SIVE and SIVnoE macaques with detectable viral genetic material in the brain in both cohorts, including the Mac251-NP macaque N09, for which relatively low levels of viral RNA were amplified (Figure 2-1). The difficulty in obtaining sequences for N09 was well represented in the limiting dilution PCR-derived viral RNA quantifications, as less than one viral RNA copy per 10,000 cells was estimated to be present (Figure 2-10). Although not statistically significant, viral RNA levels appeared highly elevated in the frontal lobes of the Mac251-DEP macaques compared with those of the naturally progressing macaques (Figure 2-10). Not surprisingly, however, sections of the frontal lobes from the macaques with SIVE from both cohorts contained greater levels of viral RNA compared with the macaques with detectable viral RNA but no SIVE, including the Mac251-DEP macaque (D06) with meningitis (Figure 2-10).

In order to determine whether viral RNA was a true representation of the presence of virus, integrated viral DNA was quantified using qPCR, including DNA from the frontal lobes of the Mac251-DEP animals. The Mac251-NP macaques for which no viral RNA sequences were obtained from brain tissue using single-genome sequencing also did not present detectable integrated viral DNA and vice versa, suggesting that viral RNA was indicative of viral integration into cells of the brain and that the virus in the brain during the late stage of infection was replicating and not likely dormant,
although this may not be the case for earlier virus sampled at 21 dpi for the CD8+ lymphocyte-depleted macaques. Unlike RNA levels, viral DNA levels were elevated in the frontal lobes of the Mac251-NP macaques compared with those of the Mac251-DEP macaques, which was primarily attributed to levels reported for N10, as SIV DNA levels were below the limit of detection for N02 and N09 (Figure 2-10). Again, not surprisingly, sections of the frontal lobes from the macaques with SIVE from both cohorts contained greater levels of viral DNA compared with the macaques with detectable viral RNA but no SIVE, consistent with previous studies (133) as well as with the hypothesis presented in this study of late viral entry of a larger viral subpopulation characterized by selection on entry.

With the exception of brain tissue, the presence of viral RNA was not representative of DNA integration in all tissues/cell populations (Figure 2-10). Several instances of absent viral RNA for the Mac251-NP macaques were coupled with the presence of detectable viral DNA, consistent with virus in the CNS with no productive infection. There were also several cases of viral RNA for which viral DNA was not readily detected, predominantly in peripheral CD14+ monocytes and bone marrow, but never in CD3+ lymphocytes. This observation is consistent with inefficient integration, or the presence of viral episomal DNA, which has been previously noted (133). However, the limit of detection for qPCR to 35 copies per reaction cannot be ignored, considering also the low level of viral RNA in the majority of these cell populations.

Surprisingly, this phenomenon was not the case for virus in the brain, despite results from a previous study by Pang et al. (134) revealing 6- to 81-fold greater levels of unintegrated versus integrated viral DNA in the brains of HIV-1-infected patients
diagnosed with HAD. The method used for genomic DNA isolation is able to detect smaller fragments of DNA, 15 to 30 kb in length; however, SIV episomal DNA may still go undetected and requires further investigation. With respect to plasma and peripheral cell populations, ranges of infection as indicated by viral RNA and/or DNA quantification also differed drastically not only between macaques within the Mac251-NP macaque cohort but also between different time points within the same macaque, especially in the case of BAL fluid macrophage viral RNA, for which the range was between approximately 250,000 to 570,000 copies/10,000 cells among the macaques at 6 months post-infection.

**Discussion**

Significant compartmentalization of viral sequences found primarily within the brains of macaques with SIVE in this study, regardless of immune modulation via CD8+ lymphocyte depletion during the acute stage of infection, provides further evidence in support of the link between viral evolution and neuroAIDS. Phylogenetic detection of this event following postmortem sampling suggests either a single, early introduction of a small virus subpopulation into the brain followed by limited genetic drift due to the isolated nature of the brain parenchyma or multiple introductions, whereby specific viral genotypic/phenotypic features are advantageous during entry, ultimately resulting in a bottleneck effect and thus reduced genetic diversity.

Although molecular biology techniques are required to definitively determine whether the same virus entering the brain during early infection does not, in fact, give rise to the population observed at necropsy, the results of the present study point to the evolutionary achievement of neurotropism in the periphery prior to late viral entry into the brain, whereby neurotropism is characterized by a metastable genotype/phenotype
conferring the ability to enter and replicate efficiently in brain-specific cells, such as perivascular macrophages (135). Rational inferences from these results also suggest that such a phenotype may require preservation of specific amino acids within the C2 region of gp120, although this observation has limited support given the small number of macaques in this study diagnosed with SIVE and without SIVE in the naturally and rapidly progressing cohorts, respectively.

Our results have also indicated that selective pressures are relatively similar for virus subpopulations associated with SIV-associated meningitis and encephalitis, which may be explained by macrophage tropism, as the primary cell types infected in the brain and meninges are macrophages (135). Although distinct genotypic signatures of macrophage tropism have not been definitively associated with CNS neuropathology, a viral genotype conferring increased replication in this particular cell type may facilitate entry into and replication in macrophage-like cells within the brain, ultimately paving the way for neuroadaptation and the emergence of a neurovirulent viral strain, assuming that the entering virus is not already neurotoxic.

Numerous studies have highlighted the importance of understanding viral evolution in the context of neuropathological disease progression, yet additional factors, such as age and co-infection, should not be ruled out. For example, both the Mac251-NP macaque with SIVE (N10) and Mac251-DEP macaque diagnosed with severe SIVE were the oldest macaques at the time of infection, approximately 8 and 11 years of age, respectively, whereas the average ages at time of infection were approximately 5 years for the Mac251-NP macaques and 8 years for the Mac251-DEP macaques (information available upon request). Several HIV-1 studies have found an association between age
and neuroAIDS status, with an increased incidence of HAND in individuals greater than 50 years of age (136-138). However, the translation to rhesus macaques is unclear, as neuroAIDS status in human subjects was determined based on neurocognitive testing, rather than neuropathological and histological markers. In addition, macaque N10 also presented with detectable cytomegalovirus (CMV) in the brain. Progression to AIDS is often accompanied by infection with opportunistic pathogens, and CMV is among a subset of pathogens able to enter the brain and potentially lead to neuronal dysregulation (139). However, the qualitative and quantitative results of our study suggest that SIV infection, and not CMV, was the driving factor in the development of SIVE. Overall, our data indicate that despite multiple introductions of virus into the brain over the course of infection, brain sequence compartmentalization in macaques with SIV-associated CNS neuropathology is likely the result of late entry of virus that has acquired through evolution in the periphery sufficient adaptation for the distinct microenvironment of the CNS. Further studies investigating the specific peripheral source (e.g., cell population) of the evolutionary phenomenon leading up to this late entry event would not only improve our understanding of the driving force behind neuroAIDS but aid in the targeted prevention of colonization of the CNS by neurovirulent HIV strains.
Figure 2-6. Genetic distance calculations for viral sequences sampled at necropsy from longitudinally sampled, naturally progressing macaques with sufficiently detectable virus in the brain. (a to d) Viral diversity (a and c) and divergence (b and d) were estimated in nucleotide substitutions/site for gp120 sequences collected at necropsy, classified as peripheral tissues/cell populations, meninges, or brain, and were averaged for the naturally progressing macaques N10 and N02 (a and b) and CD8+ lymphocyte-depleted macaques D03, D04, D05, and D06 (c and d). Estimations were calculated in MEGA using the maximum composite likelihood model of nucleotide substitution and 1,000 bootstrap replicates. Error bars represent standard errors. Dashed lines represent average diversity and divergence for sequences from plasma/peripheral cell populations and brain and are colored according to SIVE (black) or SIVnoE (gray) diagnosis. The SIVE category includes macaques N10 (a and b) and D03, D04, and D05 (c and d), whereas the SIVnoE category includes macaques N02 (a and b) and D06 (c and d). Black asterisks indicate intra-host statistical significance for N10 (intra-host significant differences were not observed for N02), whereas dollar signs indicate inter-category (SIVE versus SIVnoE) statistical significance for plasma/peripheral cell populations, meninges, and brain sequences. Mean values that are statistically significantly different by a two-tailed Student's t test indicated as follows: *, P < 0.05; ***, P < 0.001; $, P < 0.05; $$, P < 0.01; $$$, P < 0.001. The standard deviation was calculated using the standard error and square root of the total number of sequences per population. NA, not available.
Figure 2-7. Mapped amino acid changes contributing to episodic diversifying selection along internal branches of representative phylogenies. An unrestricted branch site random-effects model (BUSTED) implemented in the datamonkey webserver was used to test for gene-wide episodic diversifying selection (47) restricted to only internal branches. Amino acid changes at the minimum number of sites contributing to statistically elevated diversifying selection, which were shared by all macaques, were mapped onto the maximum likelihood phylogenetic trees. The phylogenies for two macaques have been chosen as representatives of the two distinct macaque cohorts, macaques D04 and N10, both of which were diagnosed with SIVE. Each of the two amino acid positions (78 and 128, SIVmac239 coordinates) has been represented on separate phylogenies for each macaque, wherein branches are colored according to the amino acid at that position.
Figure 2-8. Site-specific selection over time for SIV sequences within naturally progressing and CD8+ lymphocyte-depleted macaques diagnosed with SIVE or without SIVE. The fast, unconstrained Bayesian approximation for inferring selection (FUBAR) model, implemented in HyPhy, was used to measure selection at individual sites within viral gp120 sequences for individual macaques at each sampled time point over the course of infection. (a and b) Naturally progressing and CD8+ lymphocyte-depleted macaques were classified according to SIVE diagnosis or early sacrifice, and amino acid sites with a posterior probability of >0.9 of diversifying (+) or purifying (−) selection that were shared among macaques within each category are labeled along the length of the study sequence and colored according to their respective macaque classification (SIVE, SIVnoE, or early sacrifice [Early Sac]). Labels are stacked so as to highlight sites under selection that are present in more than one macaque group. Boxed regions correspond to variable loops V1 to V4, with only partial V1 (pV1) available in the alignment. The last time point is an average of sampling times among all macaques prior to euthanization. The data in panels a and b have been quantified in panel c in order to highlight the differences in the number of sites experiencing selection over time for each cohort and classification.
Figure 2-9. Percent recombination in necropsy-sampled sequences obtained from individual tissues/cell populations in naturally progressing and CD8+ lymphocyte-depleted macaques. (a and b) The percentage of total sequences obtained from individual tissues/cell populations at necropsy containing recombination breakpoints were classified as peripheral tissues/cell populations or brain and were averaged for the naturally progressing macaques N10 and N02 (a) and CD8+ lymphocyte-depleted macaques D03, D04, D05, and D06 (b). Dashed lines represent average percent recombination for sequences from plasma/peripheral cell populations and brain and are colored according to SIVE and SIVnoE diagnosis. The SIVE category includes macaques N10 (a) and D03, D04, and D05 (b), whereas the SIVnoE category includes macaques N02 (a) and D06 (b). Significance was determined using Student's t test.
Figure 2-10. Quantification of viral genomic RNA and integrated DNA in longitudinally sampled plasma and peripheral cell populations.
Evidence of Cellular and Anatomical HIV-1 Reservoirs

HIV-1 Reservoirs and the Challenge of Eradication

Thirty years after the discovery of HIV-1 as the causative agent of AIDS, a cure has yet to be found. Highly active anti-retroviral therapy (HAART) has significantly reduced AIDS-related morbidity and mortality, with improved regimens that are more potent, have fewer side effects, and lower medicinal burden. Unfortunately, treatment does not fully restore human health but simply prolongs the onset of this destructive disease and, if interrupted, invariably leads to viral rebound even in patients with low-to-undetectable plasma viremia (140-142). In addition to lifelong treatment, replacing certain drugs within the HAART regimen has also become eminent owing to the emergence of drug resistant viral variants over the course of infection (143). The identification and quantification of critical sources of low-level viremia and viral rebound, as well as the emergence of drug-resistant variants in the face of treatment has been a well-studied area of research since the development of HAART and has expanded with recent advancements in phylogenetic and sequencing techniques. This research has contributed significantly to our understanding of the complexity of HIV-1 infection and replication and, importantly, the ability of the virus to evolve and adapt to both the host immune response and antiretroviral therapy.

HIV-1 latency was initially attributed to proviral DNA in resting memory CD4+ T cells (144), which have since been shown to contribute to low-level persistent viremia during HAART, emergence of antiviral escape variants, and viral rebound after treatment interruption (145-153). These cells do not release infectious virus in the resting state but can do so following cellular activation, which can occur under a variety of conditions (144, 145). Virus production in latently infected cells, occurring in the absence of new rounds of infection of surrounding cells (154-157), is typical of viral reservoirs, defined as “a cell type or anatomical site in association with which a replication-competent form of the virus persists with more stable kinetics than the main pool of actively replicating virus” (158, 159).

Several studies have indicated that in addition to resting naïve and memory CD4+ T cells, viral reservoirs include, but may not be limited to, a recently discovered immature memory T cell population with stem-cell like properties, termed CD4+ T memory stem cells (160-162), as well as monocytes/macrophages (163), tumor-associated macrophages (164), and astrocytes (163). Follicular dendritic cells are an additional, unique reservoir in that they contribute to viral persistence by trapping virus particles on the cell surface, rather than harboring latent provirus, and this interaction can remain stable for as long as nine months (165, 166). Another potential reservoir is gut mucosa, although its contribution to viral rebound after treatment interruption seems to be limited (167). On the other hand, persistence of HIV-1-infection in the spleen (168) has been shown to contribute to viral rebound in patients with incomplete suppression by HAART (148). Moreover, the ability of the virus to switch from the use of the CCR5 co-receptor for entry to CXCR4 in ~50% of patients (169) has been attributed to
increasing replication in resting, naïve CD4+ T cells (170) and subsets of CXCR4-expressing macrophages (171), although higher affinity for the CD4 receptor has specifically been associated with macrophage tropism, as macrophages display a lower CD4 surface density (159). The timing of the switch from CCR5 to CXCR4 is particularly important for CCR5 antagonist monotherapies used in developing countries, as the re-emergence of pre-treatment CXCR4-using variants have been reported (172).

It has been shown that viral sequences obtained from circulating CD4+ T cells during low-level viremia and viral rebound post-therapy withdrawal are phylogenetically distinct from circulating cell-free virus in plasma (173), as well as viral sequences from monocytes (174). Such findings not only implicate monocytes and/or additional reservoir cell populations in the low-level production of virus but also indicate that viral reservoirs are characterized by specific evolutionary dynamics due to their ability to remain hidden from both synthetic and host-mediated antiviral response.

**HIV-1 Compartmentalization and Metapopulation Structure**

Although difficult to sample in humans, several anatomic “compartments” have been identified that can harbor tissue-specific HIV-1 subpopulations phylogenetically distinct from those circulating in peripheral plasma or other tissues (175). According to the accepted definition, an anatomic compartment is “a site for which there is limited exchange of viral genetic information with other sites” (159). Moreover, if the compartment is characterized by suboptimal free drug concentrations, it is referred to as an anatomical sanctuary (159, 176, 177). Since the first demonstration that genetic differences exist between blood- and brain-derived viral sequences (178), a large proportion of viral compartment studies have focused on the central nervous system (CNS). The CNS is often considered an “immune privileged” site due to its low-density T
cell population and weak adaptive immune response attributed to the blood brain barrier (BBB) and blood cerebrospinal fluid (CSF) barrier, which restrict migration of certain cells and other materials (179). This also renders the CNS a potential sanctuary (180), where poor drug penetration owing to tightly regulated anatomical barriers plays a major role in limiting drug efficacy (181).

CNS compartmentalization is likely the result of founder effects, either during primary or late infection, and tissue-specific selective pressures that shape distinct viral populations, followed by restricted viral trafficking (78, 146, 182-187). HIV-1 production in the CNS mainly occurs in perivascular macrophages and macrophage-like microglia (29, 135), and has been associated with both the emergence of drug resistance (188, 189) and AIDS-related neuropathology (62, 190), a spectrum of disorders collectively known as HIV-associated neurocognitive disorders (HAND), ranging from asymptomatic to HIV-associated dementia (HAD) (191). Subjects with and without HAD often harbor viral populations in the CSF that are genetically distinct from virus in the blood (192-197) and exhibit characteristics of macrophage/microglia tropism (198), in which case both latency characterized by low-level replication and anatomical isolation may contribute to the formation of a reservoir during primary infection (199, 200).

The compartmentalization of the virus in the CNS exemplifies HIV-1 metapopulation structure (i.e. division in distinct sub-populations) within different cells and anatomic sites of the infected host. Indeed, tissue-specific viral populations have been detected within female genital tract (183, 201-203), male foreskin (122), and semen (204). Latent viral forms associated with compartmentalization within the female genital tract and/or semen and foreskin of the male may be important in transmission,
since they appear to be preferentially transmitted over more recent variants (205-208). Limited compartmentalization has also been observed in breast milk (209-211), gut mucosa (212), and lung tissues (213). The existence of distinct intra-host viral subpopulations has been confirmed by studies using the SIV-macaque model, which have exhibited compartmentalization in spleen and secondary lymph nodes during primary and early chronic infection (214) as well as in the genital tract (215).

**Phylogenetic Methods of Determining Viral Compartments and Potential Reservoirs**

**Testing Viral Compartmentalization**

The phenomenon of HIV-1 compartmentalization that can ultimately give rise to a reservoir is probably the result of local conditions affecting viral evolution at specific anatomical sites limiting viral trafficking and/or exerting pressures that alter the fitness landscape of the viral population (216, 217). Compartmentalization can be inferred in a phylogenetic tree of viral sequences amplified from different tissues and/or cell types of an infected subject by the presence of a separate monophyletic clade including all or most of the sequences from a specific site (2). In addition, sequences sampled from a viral reservoir are expected to lack temporal structure – i.e. sequences sampled at different time points during the infection will intermix since they persist in the reservoir – and show less mean divergence from the most recent common ancestor (MRCA), i.e. the root node of the tree, than other sequences in the tree (175). Confidence for the monophyletic clades is usually assessed through the use of bootstrapping (218) or Bayesian posterior probabilities estimated by Markov chain Monte Carlo (MCMC) methods (219-221). Bootstrapping is a technique based on the evaluation of distance methods (e.g. neighbor-joining) or maximum likelihood trees inferred from alignment
replicates, usually 500 to 1000, built by randomly sampling (with replacement) sites from the original alignment. A bootstrap value represents the proportion of trees in which a particular group of sequences clustered together, with values >80% considered as a significant support for the cluster. Alternatively, Bayesian posterior probabilities are inferred from a posterior distribution of possible phylogenies obtained using an MCMC algorithm that incorporates prior knowledge of evolutionary model parameters (222). Monophyletic clades present in >90% of the posterior distribution of trees are considered significantly supported.

Phylogenetic analysis of HIV/SIV sequences sampled from post mortem brain tissues and longitudinal plasma samples and/or peripheral tissues has shown a clear-cut example of compartmentalization in the CNS, wherein all brain-derived sequences cluster within a distinct and well-supported monophyletic clade (Figure 3-1) (62, 187, 190, 223). However, real data sets may also display partial compartmentalization, characterized by a certain degree of intermixing between sequences obtained from different cell or tissue types, such as, for example, sequences derived from different anatomic sites in the brain (Figure 3-1). Therefore, quantitative methods for assessing compartmentalization have been developed that take into account both intra-compartmental genetic distances and distances from sequences outside of the proposed compartment (refer to Zárate et al., [2007] (112) for a detailed review). The most commonly used methods can be divided into two main categories: (1) tree-based and (2) distance-based methods. Three well-known tree-based methods are the Slatkin-Maddison (SM) test (109), Tree Correlation Coefficients (TCC) (224), and Simmonds Association Index (SAI) (111). The SM test determines the minimum number of
migration events between pre-assigned compartments (i.e. group of viral sequences amplified from different sites) consistent with the topology of the phylogenetic tree and based on the maximum parsimonious reconstruction of the tissue/cell type of origin of ancestral sequences (i.e. the internal nodes of the tree). Statistical support is based on the comparison of observed migrations with a null distribution of randomly intermixed sequences (panmictic population). Significantly less migrations than 99.9% of the null distribution indicate compartmentalization. TCCs (r, and r_b) represent correlated distances between two sequences in a phylogenetic tree with the information about whether or not they were isolated from the same compartment. The distance between two sequences can be either the number of tree branches separating the sequences (r_b) or the cumulative genetic distance between the sequences (r). Statistical significance is achieved by estimating the distribution of these coefficients by permuting sequences between compartments. The SAI assesses the degree of population structure in the phylogenetic tree by weighting the contribution of each internal node based on its depth in the tree and evaluating the significance of the observed value using a bootstrap sample both over the population structure and tree topology. Distance-based methods include Wright’s measure of population subdivision (F_{ST}) (183, 225, 226), nearest neighbor statistic (S_{nn}) (227), and analysis of molecular variance (AMOVA) (228). F_{ST} compares the mean pairwise genetic distance between two sequences sampled from different pre-assigned compartments to the mean distance between sequences sampled from the same compartment, in which statistical significance is derived via population-structure randomization. S_{nn} is simply a measure of how often the nearest neighbors of each sequence are isolated from the same or different assigned
compartments. AMOVA calculates an association based on the genetic diversity of the sequences between and within compartments and is an extension of FST, in which the distances are restricted to Euclidean and the variability is calculated from the sum of the squared distances between the sequences.

Comparison of different compartmentalization tests based on 92 separate HIV-1 data sets and 1,500 simulated data sets not surprisingly resulted in disagreement between tree-based and distance-based methods (112). Tree-based methods are expected to be more robust, since they include phylogenetic information otherwise ignored. However, they can also be misleading if branches are unresolved and have low support, since in such cases they lead to an overestimation of topological distance, or degree of segregation, between sequences. Furthermore, each method can be affected by sampling bias, for example underrepresentation of sequences from a particular compartment, which is often the case for difficult-to-sample anatomical locations prior to autopsy, such as the brain, or sites/cell populations with low-level replication, such as peripheral monocytes. The largest change (>70%) in the proportion of sequences classified as compartmentalized with skewed sampled size was observed for the TCC and SM tests, with SAI next in line (56%), indicating an increased effect of sampling underrepresentation on the power of tree-based over distance-based compartmentalization detection methods. Also, as with other phylogenetic analyses, recombination can greatly bias the results. Introduction of recombinant sequences results in reduced power, particularly for tree-based methods, to detect compartmentalization due to increased intermixing between clades of the tree, thereby increasing phylogenetic uncertainty. Finally, since the development and improvement of
sequencing methods using single genome amplification (SGA) (100, 229-232), comparisons of SGA with sequencing methods that rely on earlier PCR/cloning techniques have shown that the latter can introduce significant errors that influence the degree of diversity and thus compartmentalization (100, 233), although the general validity of this conclusion has been questioned (234). In summary, given the potential limitations of each method, compartmentalization should always be tested using more than one algorithm and discrepancies should be evaluated by taking into account the specific assumptions of each test in relation with the actual data set under investigation.

**Phyloanatomy**

As discussed above, restricted exchange of genetic information with other sites is the defining property of a viral compartment. Recently, Bayesian phylogeography methods, originally developed to study the spread of viral epidemics (64, 235), have been successfully applied to the study of viral trafficking – also referred to as viral gene flow (migration) – between different anatomical sites and cell populations (78). Intra-host phylogeography, or phyloanatomy, can provide significant insights into the establishment and maintenance of potential viral reservoirs, and it is based on the incorporation of spatial information (sampled tissue) into the evolutionary analysis of an organism that is evolving according to a population structure, or coalescent, model. Earlier implementations of phyloanatomy based on maximum parsimony have also been utilized for the investigation of HIV-1 intra-host evolution (187), and have been described elsewhere (109, 236-238), but are not the focus of this review.
Figure 3-1. HIV-1 maximum likelihood (ML) tree of brain and peripheral tissue derived sequences from a patient with HIV-associated dementia. HIV-1 gp120 sequences were amplified from tissues sampled at necropsy and are a subset of the sequences described in Salemi et al. (2005) (62). The ML tree was inferred with the HKY+G model. Values along the branches represent bootstrap support (500 replicates); only values >80% are shown. Branch lengths are scaled in nucleotide substitutions per site according to the scale bar at the bottom. Sequences from different brain (frontal lobe, spinal chord, occipital lobe) and peripheral (lymph node, seminal vesicle) tissues are indicated in blue and red, respectively, and by different symbols according to the legend in the figure. Brain sequences are completely compartmentalized (bootstrap support 100%). Sub-compartmentalization of individual CNS as well as peripheral tissues is also evident, although not all supported by bootstrapping.
By using the Bayesian phylogeography framework implemented in the BEAST software package (239, 240), information on intra-host migration rates, origin and timing of specific migration events, as well as the number of migrations (“Markov jumps”) and time spent in a given location (“Markov rewards”) can be inferred from an alignment of heterochronous sequences from various anatomical locations (71). Such migrations, supported in the form of Bayesian posterior probabilities, are interpreted as the probability that sequences from one anatomical location gave rise to, or shared ancestral history with, sequences in another location at a given point in time. Although seemingly complex, the use of the continuous-time Markov Chain (CTMC) model to trace transitions between discrete (or continuous) geographical traits along branches of a phylogenetic tree is analogous to the use of the general time-reversible nucleotide substitution model of transitions between nucleotide character traits along the same tree, both of which are inherent viral evolutionary processes naturally embedded in the genealogy.

Unfortunately, like other epidemiological studies, even for a moderately small number of spatial locations, most migration events might not even be sampled. Therefore, recent improvements of the Bayesian phylogeographic methodology, using hierarchical modeling, have incorporated joint analysis of paralleled migration processes, such as similar data sets from multiple patients, for a more statistical approach to gene flow analysis. Referred to as the Bayesian hierarchical phylogenetic model (HPM), this framework allows for a more accurate interpretation of summary data (241) than would be possible for consensus (242, 243) or combined (244, 245) approaches. Essentially, viral sequence data from multiple subjects can be used in a
single analysis, akin to a strict combined approach, while simultaneously allowing for different phylogenetic parameters in the individual paralleled partitions, e.g. patients, as in the consensus approach. Results from the individual partitions are combined to provide overall summaries of each parameter, such as migration rates. The overall model and individual partition models are fitted simultaneously, enabling the overall model to feed back information, in the form of a prior, into the estimation of the individual paralleled partitions. This feedback mechanism results in a more precise estimate of each evolutionary parameter. Details of the application of HPM methodology to HIV-1 phyloanatomy are outlined for both inter- and intra-host data sets in Cybis et al. (2013) (235). Although the study assessed viral gene flow only between CD8+ and CD4+ T cell compartments, the application of this statistical method to phyloanatomy has the potential to greatly impact future investigations of HIV-1 intra-host.

Analyzing the Temporal Structure of Longitudinally Sampled Viral Sequences

Compartmentalization and restricted viral migration are necessary but not sufficient conditions to infer the presence of a viral reservoir. The flow of virus and/or infected cells between compartments can mask low-level viral persistence, as is the case for the memory CD4+ T-cell population in lymphoid tissues during chronic infection (150, 214). Mannioui et al. (2009) (214) used molecular biology techniques to detect in vivo viral replication during low-level viremia, represented by the presence of episomal DNA, indicative of recent infection (246) and, thus, persistent viral replication within a population. However, evidence is also needed of the contribution of a potential reservoir to the reemergence of archival variants (247) and/or viral rebound that could be determined based on shared ancestry between sequences during rebound and sequences obtained prior to treatment withdrawal. From a phylogenetic perspective, the
clustering pattern of a viral reservoir would be characterized by lack of significant
cumulative genetic divergence between sequences sampled at different time points
(175). To which extent real phylogenies differ from perfectly temporally structured trees
– representing the progressive increase in viral diversity and divergence from the MRCA
during the course of the infection (12, 248) – can be accurately quantified (249). In
particular, the clustering of sequences sampled late in infection with earlier ones would
reduce temporal structure and indicate the reemergence of latent viruses (Figure 3-2).
Similar to the previously mentioned SM test of compartmentalization, topological
temporal structure can be assessed by measuring the extent of clustering whereby
sequences are categorized as discrete states corresponding to sampling time, rather
than anatomical location (249). The number of state changes within the inferred
phylogeny is compared to that of a null distribution, for which the temporal states have
been randomized. The final result is referred to as the temporal clustering statistic,
ranging from 0 (absence of temporal structure) and 1 (perfect temporal structure).

Intra-Host Evolutionary Rate Variation

In addition to temporal structure, analyzing evolutionary rate variation among
phylogenetic lineages can assess temporal signal. In the case of latent virus
reactivation, shared direct ancestry (clustering) between sequences sampled during
early and late infection will significantly reduce the evolutionary rate during that period of
time (2, 250). In contrast, migration of virus between anatomical sanctuaries may
increase the evolutionary rate. Both incidences act to reduce the ‘clock-like’ signal
associated with the molecular clock hypothesis, which assumes a linear increase in the
number of substitutions with time along all phylogenetic lineages (251). Temporal signal
can be quantified using a simple regression of root-to-tip genetic distances in the
phylogeny against sampling time (Figure 3-3) with the program Path-O-Gen (v1.3; available from http://tree.bio.ed.ac.uk/). Path-O-Gen-derived R² values indicate ‘clock-likeness’ of the sequence data, wherein high R² values (>0.5) are indicative of sequences that evolve under a strict molecular clock.

Immonen & Leitner (2014) (252) have recently developed a method to quantify the impact of potential reservoirs on viral evolutionary rates by identifying lineages within a serially sampled phylogeny that evolve at a significantly lower rate than at least one other lineage in the tree (252). The method assumes that HIV mutates according to a Poisson process with a single molecular clock rate when replicating but does not mutate when latent. For sequence pairs that share a common ancestry, latency is calculated based on the Poisson probability of accumulating the number of mutations observed on the shorter lineage d_short, given the evolutionary rate of the longer lineage d_long. Thus, latency is defined when the shorter lineage provides a Pr[(λk)/(k!e^λ)] ≤ 0.05, where λ=d_long*L and k<=d_short*L for sequence length L. However, because hypermutation, recombination, and selection can have substantial effect on evolutionary rates, sequences experiencing these evolutionary events should be removed prior to analysis and only 3rd codon positions analyzed.
Figure 3-2. Bayesian maximum clade credibility (MCC) tree highlighting reduced intra-host temporal structure. The coalescent framework implemented in BEAST was used for a serially sampled SIV-infected macaque dataset to obtain a posterior distribution of trees scaled in substitutions/site, from which the MCC tree was derived using Tree Annotator. Branches are colored according to sampling time point in days post-infection (dpi), with temporally clustered sequences (>3) collapsed for illustrative purposes. Three separate instances of shared ancestry between sequences sampled relatively far apart (92 and 223 dpi) have been highlighted, with corresponding clade posterior probabilities reported to the right.
Recent Evidence of CNS as a Key Viral Reservoir

Each of the phylogenetic methods described above is not sufficient alone to identify with certainty a viral reservoir. Additionally, over- or under-sampling of sequences from specific cell types or tissues, as well as the number and time intervals of available longitudinal sequences, can greatly influence the results. Experimental design and optimization of sampling schemes is crucial for any phylogeny-based investigation of viral reservoirs (253). However, when properly designed phylogenetic analyses are used in combination to and complemented with studies of viral replication kinetics within specific tissues and/or cell populations, they can be powerful tools for the advancement of reservoir investigation and drug treatment strategies.

Such an approach, applied to the studies of HIV/SIV brain infection and neuroAIDS, has recently provided several lines of evidence that the CNS, along with its cellular components, are a key viral reservoir, actively contributing to both the re-seeding of peripheral virus and emergence of drug resistance. For example, parenchymal-derived sequences from patients diagnosed with HAD display reduced evolutionary rates (190), which may be explained by the low-level persistent expression of unintegrated viral DNA in CNS-resident macrophages for prolonged periods of time (at least 30 days) (254). Levels of unintegrated viral DNA relative to the integrated form are 10-fold elevated in brain of patients with HAD (133, 134), and CNS-resident macrophage populations that harbor latent, integrated forms of viral DNA (perivascular macrophages and microglia) have relatively low turnover rates, ranging from months to years. As previously mentioned, anatomical barriers such as the BBB and blood-CSF barrier inhibit viral gene flow and drug penetration, resulting in isolated, or compartmentalized, viral replication. Viral compartmentalization in the CNS may be a
critical phenomenon from the standpoint of neurological impairment, but migration of the virus, although limited, in and out of this compartment is of more importance when considering the reservoir potential of the CNS. Clear evidence of reseeding peripheral tissues by CNS-derived viral variants was found for HIV-1-infected patients, as it can be inferred from the emergence of peripheral sequences from CNS-associated ancestral nodes in Bayesian genealogical reconstructions (Figure 3-4) (190). Furthermore, use of the well-characterized SIV\textsubscript{mac}251-infected rhesus macaque model of neuroAIDS demonstrated that this recirculation was not simply limited to terminal illness but can also happen during early stages of infection (Figure 3-4) (78). Evolution of drug resistance has also been identified in the CNS compartment of treated patients (135, 189), which poses a significant challenge to treatment optimization, since specific antiretroviral regimens are based on resistance testing of circulating plasma virus.

Based on the accessibility of CSF relative to parenchymal brain tissue, researchers have proposed that sampling of viral sequences from the CSF could provide genetic information on viral populations in the brain reservoir as early as primary infection (199, 200). However, other studies have indicated that CSF viral sequences are not representative of the viral population dynamics observed in the parenchyma (130, 255), wherein individual parenchymal lobe tissue may harbor distinct viral sub-populations (62, 187). Moreover, extensive sequence analyses in SIV-infected macaques have identified heterogeneous populations within the CNS, defined also by genetically distinct sub-compartments, such as meninges and parenchyma (78), basal ganglia, cerebellum, and hippocampus (256) (Figure 3-1). Although subpopulations of virus found in the CSF are likely derived from regions within the brain, trafficking of virus
between brain parenchyma compartments and peripheral blood is not well understood. This lack of knowledge, together with evidence of the complexity of viral genetic composition within the brain not accurately mirrored in the CSF underlines the significant limitation of using sequences from the CSF to characterize the CNS reservoir to a sufficient depth or pre-screen for drug-resistant variants within tissue sanctuaries. Therefore, although phylogenetic methods have identified the CNS as a key viral reservoir and potentially crucial barrier to treatment and eradication, more studies are urgently needed to improve our understanding of the contribution of individual CNS components in HIV reservoir establishment and maintenance.

**Future Directions**

The past few years have witnessed significant advances in the development of novel and powerful phylogenetic methods applicable to the study of intra-host viral compartments and reservoirs of fast evolving viruses (represented in Figure 3-5), such as SIV/HIV (62, 78, 164, 187) and HCV (257, 258), and made clear the central role played by CNS infection in HIV persistence and rebound. The use of such analysis tools in AIDS research, however, is limited by obvious ethical restraints. HIV evolutionary dynamics of deep tissues are extremely difficult to investigate in vivo since the required sampling often impractical, when not unethical. In addition, most of the findings mentioned in this review highlight the importance of investigating the establishment of viral reservoirs during primary infection, which is difficult to detect in humans. On the other hand, non-human primate models provide a viable alternative. SIV infection of rhesus macaques (Macaca mulatta) is one of the most widely used models for HIV-related pathogenesis studies (43). Infected macaques exhibit similar clinical manifestations as HIV-1 infection, albeit on a shorter time scale of approximately one to
three years (44). More importantly, they offer the unique opportunity for a wide range of infected tissue and cell population sampling that is critical to HIV-1 reservoir studies and can be performed in the presence or absence of treatment. Much can, and still needs to be, learned from this model with respect to viral latency and the contribution of viral evolutionary and population dynamics to reservoir formation.
Figure 3-3. Result of temporal signal analysis in Path-O-Gen. (A) A maximum likelihood tree was inferred using RaxML for serially sampled sequences from a SIV-infected macaque dataset. Branches are colored according to sampling time, with interior branches treated as temporal ancestral states inferred using maximum parsimony in MacClade (available from http://macclade.org). (B) A linear regression model was used to measure the temporal signal, or “clocklikeness,” in Path-O-Gen for root-to-tip genetic divergence (y-axis) for each sequence (yellow dot) in the phylogeny above against sampling time (x-axis). Colored boxes again correspond to sampling time for visual comparison.
Figure 3-4. Bayesian maximum clade credibility (MCC) trees highlighting viral migration events from brain into circulation and estimated ancestral state posterior probabilities (PP) using phyloanatomy. The phylogeographic framework implemented in BEAST was used to obtain a posterior distribution of trees from which MCC trees were derived using Tree Annotator. The trees in the figure include sequences from an HIV-1-infected patient \((190)\) (left panel) and sequences from two SIV-infected CD8+ lymphocyte-depleted macaques (middle and right panels) \((78)\). Branches are colored according to the sampled tissue/cell location (exterior) or highest PP location of the reconstructed ancestral sequence (interior), with blue and red representing sequences derived from brain and peripheral tissues, respectively. Collapsed branches represent monophyletic clades consisting of sequences from the same tissue/cell type and sampling time point. Cross-sectional sequences from the HIV-1-infected patient (left panel) and one of the SIV-infected macaques (right panel) were taken at autopsy/necropsy, the time in days post-infection (dpi) depicted for the macaque. Longitudinal sequences were derived from the macaque depicted in the middle panel prior to necropsy following onset of SAIDS \((78)\). Stars represent inferred viral entry into the periphery from the brain and are located at the midpoint of the branches along which these events occur. PP of the ancestral state at the preceding node is reported at the bottom for each migration event and can be observed using FigTree.
Figure 3-5. Graphical representation of recent advancements in phylogenetic techniques used to assess the contribution of anatomical compartments to viral dispersion and reservoir formation.
CHAPTER 4
SPATIOTEMPORAL DYNAMICS OF SIMIAN IMMUNODEFICIENCY VIRUS BRAIN INFECTION IN CD8+ LYMPHOCYTE-DEPLETED RHESUS MACAQUES WITH NEUROAIDS

Introduction

HIV-1-associated neurocognitive disorders (HAND) (259), ranging from mild neurological symptoms to the most severe form, HIV-associated dementia (HAD) (27), are serious complications of HIV-1 infection. Since the development of combination antiretroviral therapy (cART) (260), the incidence of the milder forms of HAND has increased to approximately 25% (261). In addition, HAND remains highly prevalent, with an estimated 69% of the infected population exhibiting some degree of neurological impairment (74), while treatment using neuroprotective or anti-inflammatory has been unsuccessful (27). Therefore, understanding the underlying pathogenic mechanism behind this disease and its progression is of critical importance to the development of new diagnostics and treatment approaches.

Previous studies of HIV-1 infection in the brain have shown that HIV-1 and its non-human primate counterpart, simian immunodeficiency virus (SIV), enter the brain by way of infected monocytes, which act to replenish perivascular macrophages located at the blood brain barrier, as elegantly reviewed by (262). Infection of these cells often allows for the productive infection of nearby microglia and, to a smaller extent, astrocytes, resulting in uncontrolled inflammation and a self-propagating immune reaction in response to the buildup of neurotoxic macrophage byproducts. While HIV-1 does not replicate in either neurons or oligodendrocytes, these cells are directly affected

by neuroinflammation and neurotoxic viral proteins produced by infected cells, resulting in the neurodegeneration that is characteristic of the cognitive, motor, and behavioral impairment observed for neuroAIDS patients (263). Studies show that most cARTs do not entirely clear the CNS of HIV-1, likely due to poor drug penetration (264). Thus, progressive neuronal loss/dysfunction and CNS inflammation is identified even in well-suppressed patients on long-term cART (265).

The tissue source of infected cells in the periphery responsible for HIV-1 seeding of the brain is still unclear. Moreover, it is not known whether sustained viral replication in the CNS, eventually leading to HAND, is the result of an initial event occurring during primary infection or the continuous influx of infected monocytes/macrophages over the course of the disease. Many studies have shown compartmentalization of distinct HIV-1 subpopulations within brain and cerebrospinal fluid of infected patients (164, 266-270) as well as of SIV in the rhesus macaque (Macaca mulatta) model (132, 271-274). However, it has been difficult to establish if compartmentalization is due to the separate evolution of virus in the CNS or the evolution of neurotropic viral variants in peripheral tissues that are increasingly neuroinvasive. Early viral infection of the brain has been described (275, 276), whereas other research has shown that late neuroinvasion of HIV-1 is also possible (277). Additional studies suggest that neuropathology begins outside of the CNS compartment independent of virus that might already be present in the brain (80). Such hypotheses are difficult to test in human subjects, since the initial infection date is usually unknown, and, for ethical reasons, it is impossible to collect longitudinal samples of lymphoid and non-lymphoid tissues. Alternatively, experimental infection of transiently CD8+ lymphocyte-depleted rhesus macaques with the genetic
heterogeneous SIVmac251 viral swarm constitutes a well-established rapid model of neuroAIDS \((79, 82, 278)\). In this model, approximately 77% of the animals develop SIV-associated encephalitis (SIVE) and/or meningitis within 90-120 days post-infection (dpi) \((79)\). The histopathology and neuropathology of the macaques closely resembles the development of HAD in humans as both diseases are characterized by the accumulation of perivascular macrophages, microglia and macrophage nodules, multinucleated giant cells, and gliosis \((45, 278-281)\). In addition, the SIV/macaque model produces immunological patterns similar to those produced by HIV-1 infection in humans, including the expansion of monocyte subpopulations \((282-285)\). However, little is known about how viral evolution and population dynamics relate to brain infection and the onset and progression of encephalitis in this model of neuroAIDS.

Bayesian phylogenetics and phylogeography are powerful tools used to study viral population dynamics and gene flow (migration) because they can integrate spatial, temporal, and demographic inferences by simultaneously accounting for both phylogenetic and mapping uncertainty \((66, 286, 287)\). These methods have been used primarily to investigate epidemic spread \((67-69, 287-290)\), whereas application at the intra-host level is limited by the quality of the data sets, i.e. the lack of longitudinal data and/or sequences derived from several biological locations \((71)\). In the present study, we have successfully obtained sequences sampled at different time points, from early infection to terminal illness, from several lymphoid and non-lymphoid tissues, thereby enabling the use of Bayesian methods to track the evolution and spread of the virus within the macaque host similar to the tracking of a virus in an epidemic outbreak within a population. The results provide for the first time an in depth characterization of the
intra-host viral evolutionary patterns in this macaque model and demonstrates the applicability of Bayesian phylogeography to intra-host studies investigating the interplay between viral evolution and pathogenesis.

**Materials and Methods**

**Study population.** These studies were performed with the approval of the Tulane University's Institutional Animal Care and Use Committee. Macaques were housed at the Tulane National Primate Research Center in accordance with standards of the American Association for Accreditation of Laboratory Animal Care and Tulane IACUC protocol #3497. All possible measures were taken to minimize discomfort of the animals, and the guidelines for humane euthanasia of rhesus macaques were followed. Ketamine (10-20 mg/kg, IM) was used for anesthetization prior to euthanization by intravenous injection of pentobarbital overdose and exsanguination.

The study design timeline is outlined in Figure A-1 and has previously been described in detail (271). A total of six rhesus macaques of Indian origin were intravenously inoculated with the SIVmac251 (1 ng of SIV p27) viral swarm at approximately 95 months of age. In order to achieve rapid disease progression and increased incidence of SIVE, animals were CD8+ lymphocyte-depleted via administration of 10 mg/kg mouse-rhesus chimeric anti-CD8a antibody, cM-T807, at 6, 8, and 12 dpi. Antibody cMT807 was provided by the NIH Non-human Primate Reagent Resource (RR016001, A1040101). CD8+ lymphocyte depletion was monitored by flow cytometry prior to antibody treatment and weekly thereafter, as previously described (80, 98). Animals D01 and D02 were euthanized at 21 dpi to evaluate early evolutionary events. Animals D03, D04, and D05 were euthanized at the onset of SAIDS at 75 dpi, 91 dpi, and 95 dpi, respectively with confirmed SIVE at autopsy. D06 was euthanized at
118 dpi due to SAIDS and showed perivascular cuffing and inflammation of meninges at autopsy (291).

**Sample collection and sequence generation.** As previously described (271), plasma, CD3\(^+\) lymphocytes, CD14\(^+\) monocytes, unelicited bronchoalveolar lavage fluid, lymph node, and bone marrow were collected at three time points (T1 = 21 days, T2 = 61 days, T3 = 75-118 days), while meninges and brain tissues from parietal, frontal and, temporal lobe were collected at necropsy. SIV gp120 sequences (SMM239 coordinates 6706-8049) from each sampled tissue were obtained as previously described (271) and are summarized in Table A-3 (GenBank accession numbers JF764947-JF766081 and JQ608488-JQ609071). Sequences were aligned using the Clustal (103) algorithm implemented in BioEdit (292); the alignment was further modified by a manual optimization protocol taking into account conserved glycosylation motifs (105). The highly variable region of the V1 domain was removed so as not to confound the genetic analysis (293). All alignments were gap-stripped for further analysis. Intra-host recombinants were determined using SplitsTree (294) as previously described (164) and were omitted from the dataset.

**Bayesian phylodynamic analysis.** SIV intra-host population dynamics, depicted in relative changes of the viral effective population size (\(N_e\)) over time, were investigated using the BEAST software package (239, 295). Model testing was performed for both parametric and non-parametric demographic coalescent models, enforcing either a strict or relaxed molecular clock. A detailed description of the analysis is given in the Supplementary Methods.
Compartmentalization and intra-host Bayesian phylogeography. SIV subpopulation(s) compartmentalization in brain tissues was evaluated by a modified version of the Slatkin-Maddison test for intra-host viral gene flow (62, 296), implemented in MacClade v4 (297), by using the posterior distribution of trees obtained from the BEAST analysis. This analysis was followed by a tree correlation coefficient test (298) implemented in HyPhy (108) for the MCC trees for each animal. Statistical significance for the tree correlation coefficient test was determined based on a null distribution of coefficients generated using 1,000 permutations.

Temporal and spatial gene-flow patterns of SIV among different tissues were inferred using the discrete phylogeography coalescent framework also implemented in BEAST. Two non-parametric demographic models (BSP and Bayesian Skyride; see Supplementary Methods) were used, enforcing a relaxed molecular clock (selected as the best fitting model) and a discrete phylogeographic parameter representing the sampled tissue and/or cell type. An additional computational tool was implemented in BEAST r5017 to allow a special history log file to be generated reporting within a specified period (migration period) the inferred transition along the branches at different time points, as drawn from their posterior distribution. The code to generate the special history report is shown in Figure A-3 and complete .xml files used for the study are available upon request.

Statistical analysis. The Wilcoxon Rank Sum test was used to assess the significance of the correlation of the timing of viral gene flow (migrations) from/to different tissues and the peak in Ne. Statistical significance was assumed with a p-value
<0.05. All tests were performed in using the SAS/statistical analysis software (SAS Inc, Cary, NC) and R v2.15.2 (http://www.R-project.org/) (299).

**Results**

**SIV infection and persistent CD8\(^+\) lymphocyte depletion.** Six rhesus macaques were inoculated with the SIVmac251 viral swarm at approximately 95 months of age (Table 4-1) and subsequently inoculated with an anti-CD8 antibody (271). Two macaques, referred to as D01 and D02, were sacrificed at 21 dpi, while macaques D03-D06 were allowed to progress until the onset of AIDS-related symptoms. Administration of anti-CD8 antibody results in a persistent (>28 days) depletion of CD8\(^+\) T-lymphocytes and CD8\(^+\) natural killer (NK) cells during early infection, leading to rapid disease progression and increased incidence of SIVE and/or meningitis as compared with naturally progressing macaques. CD8\(^+\) T-lymphocyte and NK cell counts over time were monitored by flow cytometry prior to and after antibody administration (Figure 4-1). Both cell types were depleted by 8 days post-infection (dpi). While considered a persistent depletion, a rebound of CD8\(^+\) T-lymphocytes was observed at 42 dpi for D06 and 56 dpi for D03-D05, whereas a rebound of NK cells was first detected at 21 dpi, although absolute counts of both cell types did not reach levels observed prior to antibody administration. CD8\(^+\) lymphocyte depletion was sufficient to produce rapid progression (75-118 dpi) to simian AIDS (SAIDS) in all four animals in contrast with naturally progressing macaques, which live on average 1-3 years: D03 was sacrificed at 75 dpi, D04 at 91 dpi, D05 at 95 dpi and D06 at 118 dpi. As previously shown, animal D03-D05 developed SIVE, whereas D06 exhibited perivascular cuffing and meningitis (271).
Table 4-1. Bayesian coalescent estimates of SIV molecular clock and demographic parameters in each infected primate.

<table>
<thead>
<tr>
<th>Primate ID</th>
<th>Time of infection (years)</th>
<th>Length of Infection (days)</th>
<th>TMRCA (days)*</th>
<th>Baseline Ne †</th>
<th>Endpoint Ne †</th>
<th>Growth Rate ‡</th>
<th>Evolutionary Rate §</th>
</tr>
</thead>
<tbody>
<tr>
<td>D03</td>
<td>9.9</td>
<td>75</td>
<td>-4</td>
<td>87</td>
<td>1877</td>
<td>3.64 $10^{-2}$</td>
<td>3.46 $10^{-4}$</td>
</tr>
<tr>
<td></td>
<td>(4, -13)</td>
<td>(19, 327)</td>
<td>(909, 4452)</td>
<td>(2.65 – 4.73 $10^{-2}$)</td>
<td>(2.87 – 4.03 $10^{-4}$)</td>
<td></td>
<td></td>
</tr>
<tr>
<td>D04</td>
<td>10.9</td>
<td>91</td>
<td>-14</td>
<td>262</td>
<td>2714</td>
<td>3.13 $10^{-2}$</td>
<td>2.03 $10^{-4}$</td>
</tr>
<tr>
<td></td>
<td>(-5, -25)</td>
<td>(75, 786)</td>
<td>(807, 6740)</td>
<td>(2.37 – 3.89 $10^{-2}$)</td>
<td>(1.67 – 2.42 $10^{-4}$)</td>
<td></td>
<td></td>
</tr>
<tr>
<td>D05</td>
<td>5.6</td>
<td>95</td>
<td>-28</td>
<td>859</td>
<td>676</td>
<td>1.39 $10^{-2}$</td>
<td>2.53 $10^{-4}$</td>
</tr>
<tr>
<td></td>
<td>(-15, -44)</td>
<td>(304, 2415)</td>
<td>(307, 4463)</td>
<td>(0.89 – 1.87 $10^{-2}$)</td>
<td>(2.03 – 3.01 $10^{-4}$)</td>
<td></td>
<td></td>
</tr>
<tr>
<td>D06</td>
<td>5.7</td>
<td>118</td>
<td>-36</td>
<td>887</td>
<td>1250</td>
<td>1.43 $10^{-2}$</td>
<td>2.35 $10^{-4}$</td>
</tr>
<tr>
<td></td>
<td>(-21, -53)</td>
<td>(364, 2090)</td>
<td>(415, 3657)</td>
<td>(1.10 – 1.85 $10^{-2}$)</td>
<td>(1.95 – 2.75 $10^{-4}$)</td>
<td></td>
<td></td>
</tr>
</tbody>
</table>

*Median estimates of time to the most recent common ancestor (TMRCA) of SIV in each infected primate. Negative values indicate that the TMRCA does not correspond to the time of inoculation, but rather to the coalescence time of the most recent common ancestor of the SIVmac251 infecting swarm (pre-transmission interval).

†Baseline and Endpoint Ne refer to the median estimates of the effective population size at viral load set point (approximately 8 days post-infection) and time of euthanization, respectively.

‡Ne growth rate parameters (days⁻¹) inferred by enforcing an exponential population growth demographic model.

§SIV gp120 intra-host evolutionary rate estimates in nucleotide substitutions per site per day
Figure 4-1. Absolute CD8+ T-lymphocytes and NK cells count over time in infected macaques. Absolute CD8+ T-lymphocyte (top panel) and NK (bottom panel) cell counts (y-axis) were monitored periodically using flow cytometry. Time (x-axis) is scaled in days post infection (dpi). Lines are colored according to each of the six macaques (figure legend to right).

**Bayesian coalescence analysis of SIVmac251 in infected rhesus macaques.**

The temporal dynamics of SIV intra-host evolution in the infected animals were investigated using the Bayesian coalescence framework implemented in BEAST (295,
In the four macaques sampled longitudinally, Bayes factors (BF) strongly supported a relaxed molecular clock ($\ln BF > 50$) (301), which assumes lognormal distributed evolutionary rates along the viral genealogies. The result was also confirmed with the evaluation of the 95% high-posterior-density (95% HPD) intervals of the coefficient of variation (CoV), which is the evolutionary rate variance scaled by the associated mean. These intervals did not include zero in any of the six SIV data sets (Table A-1), indicating significant deviation of the intra-host SIV evolutionary rates from a strict molecular clock. Several demographic models were compared in order to investigate the relative changes of the viral effective population size ($Ne$) over time. In each infected animal, $\ln BF$ strongly supported the Bayesian Skyride model, indicating that a non-parametric change of $Ne$ over time most accurately described SIV intra-host demographic history (Table A-1, Figure 4-1).

Despite the complexity of the viral demographic history, as indicated by the non-parametric change in $Ne$ over time, similarities for various coalescence parameter estimates were observed among the longitudinally monitored macaques. Analyses of these parameter estimates included comparisons of time to the most recent common ancestor (TMRCA), net $Ne$ change, and $Ne$ growth rate (Table 4-1). In each of the longitudinally sampled animals, TMRCA dated back approximately 4-36 days prior to infection with overlapping 95% HPD intervals. This is expected because the time at the root of the viral genealogy should not correspond to the time of inoculation but rather to the pre-transmission interval, i.e. the coalescence time of the most recent common ancestor of the SIVmac251 infecting swarm (302). In macaques D03 and D04, baseline $Ne$ values calculated at 8 dpi, the time of set point viral load (271), corresponded to
approximately 175 unique viral variants. This finding indicated the presence of a relatively homogeneous SIV population, the overall diversity of which increased by an average of approximately 15% at endpoint (Table 4-1), prior to euthanization at the onset of AIDS. Differences in baseline and endpoint Ne were less distinguished in D05 and D06 (average of 0.1%). A slower viral population growth rate in these macaques was observed in comparison to D03 and D04 (Table 4-1). Nevertheless, Ne 95% HPD intervals were similar for all data sets, suggesting a generalized SIV coalescence pattern represented within each infected macaque (Figure 4-1). In addition, viral evolutionary rates were also similar (Table 4-1); median rates averaged approximately 2.6x10^{-4} nucleotide substitutions per site per day, i.e. 0.10 nucleotide substitutions per site per year. This value is at least one log higher than the SIV evolutionary rate estimated in natural hosts (303) but is likely the result of the impaired innate immune response due to CD8^+ T-lymphocyte and NK cell depletion. It is important to note that despite the relatively short time of infection (75-118 dpi), viral sequences sampled at different time points (21 dpi, 60 dpi and at necropsy) displayed a statistically significant number of nucleotide differences, classifying SIV in the context of this model as a measurably evolving population (304) and allowing for the study of SIV intra-host demographic history and its relationship to brain infection.

Bayesian phylogeography of SIVmac251 brain infection. The spatial dynamics of SIV intra-host evolution in the infected animals was investigated using sequence compartmentalization analyses and the Bayesian phylogeographic framework. SIV migration patterns based on tree topology appeared similar among all macaques, with brain-derived sequences tending to cluster in multiple but independent
monophyletic clades (Figure 4-2). Significant compartmentalization of viral subpopulations infecting the brain for all macaques was confirmed using the Slatkin-Maddison (109) (Figure A-2) and tree correlation coefficient test (296) (Table A-2). The genealogies consistently indicated multiple gene flow events seeding the brain of each animal, the majority of which were traced back to a common ancestor after 21 dpi, although a few brain sequences shared common ancestry with peripheral sequences even prior to viral swarm inoculation (Figures 4-2 and 4-3a). The latter finding indicated the presence of a small number of early brain entry events (≤21 dpi), although an accurate migration time is difficult to infer due to increasing HPD intervals near the root of the tree. Nevertheless, early viral entry into the brain is consistent with previous studies (271) and was also confirmed by the analysis of viral sequences isolated from several brain tissue samples of D01 and D02 sacrificed at 21 dpi.

In order to study in detail the timing of viral gene flow events with respect to initial infection, we compared the 95% confidence intervals (CI) of the distributions of migration times between different tissues/cell types inferred from the posterior distribution of trees. SIV intra-host phylogeographic patterns were similar among macaques and revealed an intriguing temporal order: the initial migration among peripheral tissues occurred within 47-54 dpi, followed by migration from peripheral tissues to brain 53-66 dpi, from brain back to peripheral tissues within 59-84 dpi, and ended with migrations confined to sub-compartments within the brain around 62-92 dpi (Figure 4-3a). The median time of viral migration from the periphery to the brain significantly correlated with the time at which peak viral Ne was observed for three of the four longitudinally monitored macaques (Figure 4-3b).
Figure 4-2. SIV gp120 Bayesian Maximum Clade Credibility (MCC) Trees. The MCC tree for each of the primates followed longitudinally (D03-D06) was estimated from the posterior distribution of trees obtained with a Bayesian coalescent framework enforcing a relaxed molecular clock. Branch lengths are scaled in time (days) and colored to represent peripheral (red) and brain (blue) tissues. Gray branches represent coalescent events of the viral swarm (pre-transmission interval) prior to inoculation. Viral migration events between the peripheral and brain tissues are represented as color changes at individual nodes.
Figure 4-3. SIV phylogeographic and demographic patterns within primates followed longitudinally. (a) Viral gene flow (migration) events among different tissues within infected primates. Box plots represent the 95% confidence intervals of the distribution of migration times, in days post-infection (dpi), estimated by phylogeographic analysis. Broken lines represent outliers. Migrations within peripheral tissues (P to P) are displayed in yellow, from peripheral tissues to brain (P to B) in green, from brain to peripheral tissues (B to P) in blue, and within brain tissues (B to B) in pink. Time intervals were inferred from a posterior distribution of trees obtained using a Bayesian coalescent framework enforcing a lognormal relaxed molecular clock with a Bayesian Skyride demographic prior and discrete phylogeographic parameter. The tissues are ordered according to the median migration time given by the number adjacent the black vertical bar inside each box. (b) Bayesian Skyride Plots with SIV migration times from peripheral tissues to the brain below. Bayesian Skyride plots show the change in virus effective population size ($N_e$) over time in dpi. The white line represents the median $N_e$ with the surrounding blue area representing 95% HPD intervals. Arrows indicate the time of the highest peak in $N_e$ during the course of the infection. The dashed line highlights the fluctuations of the $N_e$ over time. Correlation between periphery to brain viral migration times and the $N_e$ peak in the Skyride plots was analyzed using the Spearman's rank order (correlation coefficient $\rho$ and $p$-value are given at the bottom of each plot).
The temporal pattern of viral migrations from individual peripheral tissues to the brain was also assessed (Figure 4-4). SIV strains infecting the brain originated from different tissues and/or cell types even prior to 3 weeks post-infection. Early viral entry into the CNS from plasma, peripheral blood CD3+ and CD14+ cells, and cells within the bronchoalveolar lavage fluid (305) and bone marrow was already evident in the two animals sacrificed at 21 dpi, with a median migration time from periphery to brain occurring around 12 dpi (Figure 4-4a). Interestingly, SIV phylogeographic analysis in macaques sampled longitudinally revealed that the majority of migration events from the periphery to the brain occurred later in infection (>12 dpi) and followed a specific and consistent temporal order (Figure 4-4b). Initial viral migrations to the brain originated from plasma and CD14+ cells, trailed by contribution from BAL macrophages, CD3+ cells and lymph nodes, and followed, in all cases, by a final migration of viral strains from the bone marrow within 62-85 dpi for all longitudinally monitored macaques. In contrast, for the two macaques euthanized at 21 dpi, brain infection from peripheral tissues did not exhibit a particular temporal order (Figure 4-4a).

**Discussion**

In the present study, longitudinal samples derived from a multitude of peripheral tissues as well as post-mortem brain tissue were used to evaluate the contribution of intra-host evolution and gene flow to the establishment and maintenance of SIV infection in the brain of the well-established SIV/macaque rapid model of neuroAIDS. Results of Bayesian phylodynamic and phylogeographic analyses were similar among macaques with varying forms of neuropathology, providing well-supported inferences into the interplay between viral evolution and disease progression.
Figure 4-4. SIV gene flow events from peripheral tissues to the brain. Box plots show the 95% confidence intervals of the distribution of migration times (days post infection), estimated by phylogeographic analysis, from peripheral tissues to brain. Broken lines represent outliers. Time intervals were inferred from posterior distribution of trees obtained with a Bayesian coalescent framework enforcing a relaxed molecular clock and a discrete phylogeographic parameter. The color of the bar indicates a specific tissue or cell type: plasma (red), CD3+ T cells from peripheral blood (yellow), CD14+ monocytes from peripheral blood (green), bronchial lavage (orange), lymph nodes (brown), and bone marrow (cyan). The tissues are ordered according to the median time of brain entry (black vertical bar), with the earliest at the top and the latest at the bottom. The thickness of the bar is proportional to the percentage of the observed migrations (shown to the right of each bar for primate D03-D06). (A) Primates D01-D02 euthanized at 21 dpi. (B) Primates D04-D06 followed longitudinally until development of SIVE.
Results of the Bayesian phylodynamic analysis indicated a similar evolutionary pattern in all of the longitudinally sampled macaques, consisting of a steady increase in effective population size ($N_e$) that culminated at around 55 dpi. An exception to the similarities in evolutionary patterns was the increased viral population growth rate observed for macaques D03 and D04 as compared with D05 and D06. These results may be related to the difference in the age of the macaques at the time of virus inoculation, with an increased rate of disease progression reported to correlate with increase in age at seroconversion (306). Individual phylogenetic tree topologies for each of the macaques also appeared similar, with significantly compartmentalized virus within the brain, particularly within the combined region of the temporal and parietal lobes, comprising multiple monophyletic clades seeded by peripheral tissues. This contribution from all of the peripheral tissues to virus in the brain was also observed for the gene flow analysis and occurred throughout the course of infection for all four longitudinally sampled macaques. The majority of these migrations occurred during 53-66 dpi, which correlated significantly with the temporal peak in $N_e$ for three of the four longitudinally monitored macaques. This correlation suggests a relationship between viral migration to the brain and an increase in viral diversity. On the other hand, the median time of viral migration from each of the individual tissues appeared to be temporally ordered, with migration from the bone marrow consistently occurring last. Overall, the combined results of the phylodynamic and phylogeographic analyses indicate two important events during rapid neuroAIDS progression in these animals: 1) a peak in viral diversity related to the migration of virus from peripheral tissues to the brain, consistently followed by 2) viral migration from the bone marrow to the brain.
The observation of both temporal and spatial patterns with respect to the establishment and maintenance of SIV in the brain is informative considering the previous historical roadblocks in assessing early HIV evolution in humans, in which the timing of infection and viral source is often unknown and the collection of longitudinal biopsies for experimental research would be unethical. Furthermore, due to the growing body of research incorporating SIV infection in the CD8+ lymphocyte-depleted rhesus macaque model, studies concerning how the virus evolves in this model are increasingly critical to understanding infection and immune dynamics (79, 284, 307-312). In our investigation, we were able to show that, even within the relatively short lifespan of these macaques, meaningful estimates of evolutionary rates and population dynamics over time could be obtained. While increased viral evolutionary rate, massive infection, and poor control of viral replication were expected given the impaired innate immune response (80, 284, 308, 313 2005), the consistency of demographic patterns among different animals strengthens confidence in the results.

The rapid disease model also provided a unique platform for assessing the application of an adapted Bayesian phylogeographic framework to the study of intra-host viral evolution. The relatedness of intra-host sequences often hinders the ability to determine accurate phylogeographic inferences (71); however, the use of an extensive sequence data set comprised of longitudinally collected sequences from several different tissues enabled us to infer viral transition times and pathways among different tissues and/or cell types within the macaque hosts, as well as the contribution of these patterns to varying forms of AIDS-related neuropathology.
NeuroAIDS has long been considered a macrophage-mediated disease, characterized by HIV/SIV infection of brain-derived macrophages, resulting in activation and the production of inflammatory cytokines and other soluble proteins that impact the overall integrity of CNS cells (314). Studies have shown that there is an increased turnover of CNS-derived activated perivascular macrophages as well as an increase in the rate and magnitude of monocyte trafficking from the bone marrow in animals that develop SIVE (79) [Burdo, 2010]. Data shown here provide further support for these findings and critical insight into the temporality of this occurrence. The late migration time of virus from the bone marrow to the brain observed in this study is consistent with the late clinical onset of HIV-associated neurocognitive disorders, commonly associated with the onset of AIDS (315). Although early viral entry and continual seeding of the brain from peripheral tissues may contribute to neuropathology, our findings suggest a more central role for SIV-infected cells within the bone marrow in the progression or even onset of neuroAIDS. In addition, once the virus had entered the brain, our data not only revealed compartmentalization of virus within this tissue but also viral gene flow back to peripheral tissues later in infection, an observation highlighting the significance of the CNS as a potential viral reservoir capable of reseeding the body during cART.

In conclusion, the present work provided a better understanding of neuroAIDS pathogenesis in SIV-infected CD8+ lymphocyte-depleted macaques and revealed that SIV-infected cells within the bone marrow may contribute to neuropathogenesis. Targeting infection and migration of bone marrow cells should be a focus in future studies related to treatment and prevention of HIV/SIV-associated neurological impairment.
CHAPTER 5
SPATIAL AND TEMPORAL CHARACTERISTICS OF SIMIAN IMMUNODEFICIENCY VIRUS DISPERSION DISTINGUISH Rhesus Macaques With and Without NeuroAIDS

Introduction

HIV infection confers high risk of encephalopathy, neurodegeneration, and opportunistic infection of the central nervous system (CNS), collectively referred to as neuroAIDS. Despite the success of combined antiretroviral therapy (cART) in suppressing viral load in the peripheral blood (316, 317), neuroAIDS in the form of HIV encephalopathy (HIVE) has increased in prevalence to an estimated 50% of infected patients (13) and can result in devastating deterioration of motor and cognitive function (14, 33, 318). The role of inflammation, the host immune system (319), and CNS cART suboptimalinity (260, 320, 321) in the escalation of neuropathology is well appreciated; however, the etiology of this disease, affecting only a fraction of the infected population, is not well understood. Evolution and infection dynamics of the virus within the host have been linked to neuroAIDS progression, with distinct genotypic (70) and phenotypic (322-324) characteristics described for virus isolated from the CNS, suggesting that HIV-1 evolutionary adaptation to the CNS plays a pivotal role in onset of neuronal damage. Slower decay kinetics of virus in the CNS of patients with HIV-associated dementia (HAD) compared to asymptomatic patients has been associated with long-lived cell types (192), suggesting neuroadaptation is linked with the ability to replicate in cell populations other than the primary targets - CD4+ T-cells – such as monocytes/macrophages (M/M).

Additional evidence in support of the role of infected M/M and microglia in HIVE has been presented, though several questions remain as to the origins of CNS infection,
particularly the brain. An increased rate of monocyte egress from the bone marrow, increased numbers of a subset of monocytes in the blood highly susceptible to HIV infection, and the accumulation of productively infected macrophages in the CNS support the idea that migration of monocytes across the blood-brain barrier (BBB) is responsible (325, 326). Elevated monocyte transmigration across the BBB in vitro has even been observed in response to HIV infection (22-24), although direct observation of the transmigration of infected monocytes has yet to be observed in vivo and the reliability of in vitro BBB models established. Helms et al. (327) defines an ideal model as one that incorporates a complex array of parameters, which can differ between studies due to temperature, cell-handling conditions (328), and even absence of communication with other tissue-specific cells (320). Moreover, a complete understanding of the in vivo model, particularly in humans, and, hence, full list of ideal markers is lacking (327).

In contrast to in vitro models, an epidemiological approach to phylogenetic analyses within the host, referred to as phyloanatomy (70), offers an in silico solution to understanding both viral evolution and dissemination of virus among various anatomical compartments. Derived from the similar framework used to study regional and global migration of pathogens during an epidemic, phyloanatomy assumes a network of isolated tissues tied together via the vascular system and circulating immune cells carrying the virus. Using serial sampling during the course of infection and assuming a clock-like model of evolution, this method is also able to resolve the timing of relevant evolutionary and epidemiological events. The brain both harbors detectable virus (52, 53) and exhibits inflammation and structural changes during early HIV infection (329);
however, both phylogenetic (72) and molecular (326) analyses have indicated that virus
is reintroduced to the CNS terminally with AIDS. Symptoms of HAND are also most
commonly observed in advanced stages of HIV infection (330). Therefore, the question
of which is responsible for neurovirulence - early viral entry followed by isolated
replication or late entry of a neuroadapted viral variant - is difficult to answer without a
complete picture of disease progression and intra-host molecular processes.

Pathological evaluation of HAND in humans is only possible from individuals who
succumb to disorder, which only provides a snap-shot of end-stage pathology, while
mechanisms that contribute to the complications remain elusive (81). Consequently,
studies using cerebrospinal fluid as a surrogate have abounded but the utility debated
due to the physical and HIV phylogenetic distinctions from brain parenchyma (70, 331,
332). Serial sampling from all possible tissue sources of infection in humans is an
additional hurdle, yet less than 2% of the total body lymphocyte population resides in
the peripheral blood, thus highlighting the role of HIV-1 infection in anatomical sites
and their potential role in seeding the brain during HIV infection. Sampling from these
sites at multiple, pre-determined time points is more readily available for animal
models, which have provided valuable insight into the plasticity of tissue infection (43,
333). The most commonly used animal model of HIV infection and neuroAIDS is the
SIV-infected macaque, which exhibits remarkable pathological similarities to HIV-
infection and disease progression in untreated humans (334). Previous SIV sequence
analyses performed by our lab have indicated a potential role for multiple tissues in
seeding the brain (78) and a prominent role for peripheral tissues in the evolution of a
neuroadapted virus (72). The specific tissue(s) responsible for brain infection during
early and end-stage disease have not yet been identified in the context of neuropathology but are readily discernable using a combination of molecular biology and statistical phyloanatomy techniques (287).

Identifying viral dissemination patterns among the plethora of infected compartments within the encephalopathic host is a critical step toward identifying biological markers that could help predict the risk for development of cognitive impairment. We proposed to investigate these patterns in this study by employing the Bayesian phyloanatomy framework for investigation of the significant source of brain infection in rhesus macaques with and without SIVE among a variety of tissues and peripheral cell populations. A multiple tissues have been implicated in seeding the brain with virus during end-stage disease (78), we also proposed the use of a previously identified biomarker, s100B, of BBB deterioration in humans (54, 335, 336) and SIVE lesion formation in macaques (56) to identify the precise tissue origin responsible for evolution of the neurovirulent viral variant. Results of this study provided evidence of the link between viral evolutionary and epidemiological patterns and AIDS-related neuropathology, including the differential role of infected macrophages and peripheral T-cells in seeding of both the CNS and periphery.

**Materials and methods**

**Study Population.** Two macaque cohorts were used in this study, which included four CD8+ lymphocyte-depleted (D03-D06) and eight naturally progressing, or non-CD8+ lymphocyte-depleted (N02, N09, N10), Indian rhesus macaques (Macaca mulatta), hereafter referred to as the Mac251-DEP and Mac251-NP, respectively. Both cohorts were infected intravenously with the viral swarm SIVmac251 (1 ng SIV p27) (90). CD8+ lymphocyte depletion was achieved by subcutaneous administration of anti-
CD8 antibody cM-T807 (6, 8, and 12 days post-infection [dpi]) (41). All animals were euthanized at the onset of SAIDS (75-300 dpi), the criteria for which included: 1) weight loss >15% body weight in 2 weeks or >30% body weight in 2 months, 2) documented opportunistic infection, 3) persistent anorexia >3 days without explicable cause, 4) severe intractable diarrhea, progressive neurological signs, or significant cardiac and/or pulmonary signs, as previously described (80). Seven of the animals harbored detectable viral sequences within analyzed frontal, parietal, and/or temporal cortex tissue sections of the brain parenchyma, as identified using single genome amplification of viral genomic RNA (described below). Pathological diagnosis of SIV encephalitis (SIVE) was determined post mortem by a veterinary pathologist and included the presence of microglial nodules and multi-nucleated giant cells confirmed by immunohistochemistry staining for SIV p27, as described previously (82, 95-97). Animals diagnosed with SIVE included D03, D04, D05, and N10, whereas animals with detectable brain viral RNA but no detectable SIVE (referred to as SIVnoE + DVB) included D06, N02, and N09. Plasma s100B concentrations at necropsy were determined for additional SIV-infected animals with no detectable viral RNA in the brain or SIVE, which included N01, N03-N05, and N12; however, viral RNA sequences for these animals were not included in the phyloanatomy analysis, as the focus of this study was origin of brain infection.

**Ethical guidelines.** Procedures involving the Mac251-DEP animals were performed with the approval of the Tulane University’s Institutional Animal Care and Use Committee (IACUC). The treatment and handling of macaques in this cohort have been described previously (41). Procedures involving the Mac251-NP animals, housed
at the New England Primate Research Center, were conducted according to the standards of the American Association for Accreditation of Laboratory Animal Care and IACUC protocol #04802, and treatment of all animals was in accordance with the Guide for the Care and Use of Laboratory Animals of the Institute of Laboratory Animal Resources (337). Further detailed information on the handling and supervisinal guidelines for the Mac251-NP cohort can be found in (92). All possible measures were taken to minimize discomfort of the animals, and the guidelines for humane euthanasia of rhesus macaques were followed.

**Sample collection and sequencing.** Plasma viral loads were monitored as previously described by quantitative PCR (qPCR) methods targeting a conserved sequence in the group antigen gene (gag) (93, 94). Plasma, FAC-sorted peripheral CD3+ T-lymphocytes and CD14+ monocytes, unelicited bronchoalveolar lavage fluid (BAL) macrophages, and bone marrow (BM) aspirates were collected two to three time points – 21 days post infection (dpi), 60 dpi, and necropsy – from the four Mac251-DEP macaques and at four time points – 21 dpi, 90 dpi, 180 dpi, and necropsy – from the three Mac251-NP macaques. Viral genomic RNA was extracted from the longitudinal samples, as well as from the meninges and brain tissue sections, as described previously (41, 72, 78, 92). Viral RNA envelope glycoprotein gp120 sequences derived from the Mac251-DEP cohort were obtained using bulk PCR and cloning methods (41, 78). Due to the potential limitations of clonal analysis, a modified single genome sequencing protocol based on previously published methods (100) was used for all samples obtained from the Mac251-NP cohort as well as frontal lobe from the Mac251-DEP cohort. Sequences were aligned as previously described (78), and approximately
20 gp120 sequences per tissue per time point were obtained after removal of potential recombinants. Detailed information regarding sample collection, sequencing protocols, and the sequence alignment procedure have been reported previously (41, 72).

**Nucleotide sequence accession.** All of the sequences used in this study have been described previously (72, 78, 92) and are accessible in GenBank (accession numbers JF765272-JF766081 [Mac251-DEP], KR999328-KR999727 [Mac251-NP N02 and N10], and KX081254- KX081353; KX081479-KX081498; KX081619-KX081702; KX081840-KX081862; KX082029-KX082107; KX082229-KX082252; KX082428-KX082531 [Mac251-NP N09]. N03 sequences are currently under submission to GenBank.

**Bayesian phyloanatomy.** Bayesian genealogical tree reconstruction for individual macaque-specific gp120 sequence alignments was performed using BEAST v1.8.3 (239, 240) available from http://beast.bio.ed.ac.uk/), assuming an uncorrelated relaxed molecular clock model of evolutionary rate variation across branches (338) and Bayesian Skyride demographic model (295, 339). Prior information can be observed from the representative xml found in https://github.com/rifebd88/SIV_Phyloanatomy. A subset (500) of systematically drawn trees from the resulting posterior tree distribution for each macaque was obtained for use as empirical data for further analysis, owing to the computational complexity of integrating over the possible tree space for a large number of taxa (340). The tree subsets and script used to generate them can also be found at https://github.com/rifebd88/SIV_Phyloanatomy. Macaque-specific gp120 sequence alignments were categorized according to SIVE diagnosis (SIVE and SIVnoE) and treated as individual partitions according to a hierarchical phylogenetic model.
(HPM) in subsequent Bayesian analysis. An asymmetric transition rate matrix within the Bayesian stochastic search variable selection (BSSVS) model allowed for inferred directionality of viral dissemination patterns occurring at significantly non-zero rates between discrete sampled anatomical compartments. Within the HPM, an epoch model (340) was used to infer spatiotemporal dissemination patterns during intervals of time over the duration of the evolutionary history of the viral population (assumed constant over time). Time intervals included the first 21 dpi (approximating acute infection), last 21 days prior to necropsy (approximating AIDS onset), and the duration between (approximating asymptomatic infection). Prior information can be observed from the representative xml found in https://github.com/rifebd88/SIV_Phyloanatomy.

Effective Markov chain Monte Carlo sampling (341) for all Bayesian analyses was assessed by calculating the effective sample size (ESS) for each estimated parameter. ESS values > 200, calculated in Tracer (342) (available from http://beast.bio.ed.ac.uk/Tracer), were considered suitable indicators of effective sampling. Bayes factor (BF) support (BF>3) (287) for non-zero transition rates of SIV viral lineages among discrete anatomical locations within the host was assessed at the hierarchical level using SPREAD (343) (latitude and longitude coordinate designations of ‘1’ for all anatomical locations).

**ELISA detection of plasma s100B.** Frozen (-80C) plasma samples collected prior to (when available) and throughout the duration of infection for each macaque corresponded to 5-7 and 12-19 time points for the Mac251-DEP and Mac251-NP animals, respectively. Samples were thawed on ice to room temperature, and duplicate 100 ul aliquots (diluted if necessary) were used in the quantitative analysis of s100B
concentration. Each aliquot underwent a single freeze-thaw cycle. ELISAs were performed using the commercially available, solid-phase “Monkey s100 Calcium Binding Protein B (S100B)” kit (BlueGene Biotech, #E09S0042) according to the manufacturer’s protocol. Results were compared using one-way ANOVA with Tukey’s post-hoc multiple comparisons test between SIVE, SIVnoE+DVB, and SIVnoE animals. A p-value≤0.05 was considered significant.

Results

SIV dispersion paths among tissues and cell populations vary over the course of infection. Bayesian analysis of viral diffusion over the course of infection in SIVE and SIVnoE+DVB animals revealed that the virus may be more readily accessible to certain tissues depending on the duration of infection. In both animal groups, viral dissemination during the early phase of infection (first 21 dpi) was primarily driven by exchange between PBMCs (sorted monocytes and T-cells) and peripheral blood and from bone marrow to the peripheral T-cells, with few exceptions (Figures 5-1A and B). One notable exception was the early circulation of virus among the individual brain cortices, or lobes, of the SIVE animals, suggesting earlier brain infection in these animals than in SIVnoE+DVB animals. As with all dispersion pathways depicted in Figure 5-1 and discussed herein, this circulation was considered significant (BF>3), characterized by flow of virus from the parietal to the frontal cortex, although the significant anatomical source could not be identified.

Following the period of early infection, viral dispersion patterns in the periphery diverged, differing from early infection for both animal groups. During the designated asymptomatic time interval, SIVnoE+DVB animal sequences exhibited limited dissemination between, as well as to, PBMCs and peripheral blood (Figure 5-1A).
Significant exchange between PBMCs and peripheral blood continued into asymptomatic infection, however, for SIVE animals, with the additional contribution of virus from broncho-alveolar lavage macrophages (BAL) (Figure 5-1B). Viral dissemination among the three parenchymal cortices and meninges was detected during asymptomatic infection for both groups, with no significant identifiable anatomical origin.

The late infection period, nearing the onset of AIDS, was similarly distinct from both early and asymptomatic infection for both animal groups. A loss of significant exchange of virus between peripheral monocytes and T-cells was detected for both groups during this time interval. Despite similarly limited contribution of peripheral tissues and cell types to infection of this cell population, an increased number of dispersion pathways originating from peripheral T-cells, particularly to the brain, was observed for the SIVnoE+DVB animals as compared to early infection (Figure 5-1A). Alternatively, for the SIVE animals, a larger transmission network was observed compared to earlier time periods (Figure 5-1B). These transitions consisted largely of dispersion from the lungs and into the brain, bone marrow, PBMCs, and peripheral blood. Also in contrast to earlier time intervals, meninges contributed significantly back to the periphery in both animal groups, specifically the peripheral blood and lungs in the SIVnoE+DVB and SIVE animals, respectively (Figures 5-1A and B). It is important to note, however, that, as brain and meningeal sequences were only available upon necropsy, the exchange of cranial and meningeal viral populations with those of peripheral tissues may be underestimated for earlier time points. Nearly uniform sampling over time for the remaining tissues and cell populations offers greater
confidence in the indication of a highly dynamic viral population over the course of SIV infection.

Figure 5-1. Graphical representation of viral dispersion patterns over the course of infection in the context of SIVE for the SIV-infected macaque model of HIV infection and neuroAIDS. The Bayesian stochastic search variable selection model (BSSVS), assuming asymmetric diffusion among discrete anatomical compartments, was implemented in BEAST v1.8.3. The hierarchical phylogenetic model was used to infer spatiotemporal trends in dispersion across animals (A) with SIVE (n=4) and (B) with detectable virus in isolated brain sections (DVB) but no SIVE (n=3). The BSSVS model was allowed to differ between designated time intervals corresponding to early infection (21 days post-infection), AIDS onset (21 days prior to necropsy), and asymptomatic infection (time span between early infection and AIDS onset). Arrows indicate directionality of significant (Bayes Factor >3) diffusion and are colored according to time interval(s) of occurrence. (C) Similar patterns (SHARED, grey) and patterns unique to SIVE (grape) or SIVnoE+DVB (light green) animals have been highlighted and expanded for easy comparison.

**AIDS-related neuropathogenesis in SIV-infected macaques is characterized by distinct viral dispersion patterns.** Although differing in timing, the SIVE and
SIVnoE+DVB animal groups shared similar viral dispersion pathways, including the significant exchange of virus among peripheral blood and cranial compartments, independently, and between peripheral blood and bone marrow (Figure 5-1C). However, notable differences in not only the origin of brain viral sequences but of viral dispersion within the periphery were observed. Significant contribution of peripheral tissues to monocyte, T-cell, and peripheral blood compartments was observed during this period in the SIVE animals, derived almost exclusively from BAL lineages. This family of macrophages was also the exclusive contributor among sampled locations to cranial virus in SIVE animals, with this exclusion reserved for peripheral T-cells in SIVnoE+DVB animals, both of which occurred during late infection (Figures 5-1A and B).

**s100B is a poor predictor of SIVE diagnosis in small populations.** The measurement of plasma s100B concentrations over time for each animal was originally designed in order to determine the primary contributing tissue to the presence of neurovirulent virus in the brain using comparison of peak s100B concentration to the median time of tissue- or cell-specific viral dispersion to the brain. However, upon validation of the use of necropsy-sampled s100B as a marker of SIVE diagnosis, results revealed no significant differences (ANOVA, p=0.94) among SIVE, SIVnoE+DVB, and SIVnoE animals (Figure 5-2A). In order to exclude the possibility of early entry of neurovirulent virus into the brain and subsequent decrease in peripheral s100B concentrations, peak values were used, although similarly reflecting no significant differences (ANOVA, p=0.81) among animal groups (Figure 5-2B). As the sample size rendered virtually no statistical power, the reliability of the comparison described above
for inferring the origin of neurovirulent virus was in question and, therefore, the
comparison abandoned.

Figure 5-2. Plasma concentrations of s100B at necropsy and peak levels in animals
grouped according to SIVE diagnosis and detectable presence of viral RNA in
brain tissue sections (DVB). Plasma concentrations of s100B were
determined using ELISA for animals over multiple time points prior to and
during SIV infection. (A) Necropsy-sampled plasma s100B concentrations for
the three animal groups. (B) Peak s100B concentrations for the three animal
groups labeled according to macaque ID and time interval of occurrence
(P=pre-infection, E=early infection, A=asymptomatic infection, O=AIDS
onset). Bars represent mean. Statistical significance was assessed using
one-way ANOVA and Tukey’s post-hoc multiple comparisons test.

Discussion
Bayesian phyloanatomy offers a powerful alternative to limited in vitro models of
intra-host viral trafficking by incorporating a framework for epidemiological inferences of
the origins of viral sequence populations. Furthermore, the hierarchical phylogenetic
model within the Bayesian phyloanatomy framework provides a statistical approach to
identifying trends in epidemiological and evolutionary patterns across limited sample
sizes, such as the number of infected hosts. We, therefore, proposed the application of
these methods for a better understanding of the etiological origins of the
phylogenetically distinct HIV population within the brain of individuals with HIV-
associated encephalopathy (HIVE) (70). This study describes the most diverse spatial
and temporal sampling scheme used for phyloanatomic inference with the aim of identifying the timing and mode of brain infection and SIV-associated encephalitis (SIVE) in the SIV-infected macaque model of HIV infection and neuroAIDS.

Results from this study indicate that viral dispersion patterns among peripheral blood compartments, bone marrow, lung, and the CNS are similar during early infection among macaques, but diverge significantly prior to AIDS onset in two macaque groups differing in neuropathological diagnosis. These results are consistent with previous reports of indistinguishable gp120 genotypes among macaques at early sampling time points (92). Following early infection, not only do the terminal origins of brain infection differ between groups, but viral dispersion among peripheral tissues and cell populations readily distinguish the two groups, suggesting the role of alternative evolutionary and epidemiological processes in asymptomatic and end-stage disease.

In light of diverse tissue sampling, alveolar macrophages appear to play a significant role in viral spread throughout an encephalopathic host, as well as seeding of the brain during the late phase of infection. Given this finding, in combination with multiple lines of evidence implicating bone-marrow derived monocytes in CNS inflammation (29, 326), we postulate that trafficking of infected alveolar macrophages, whether it be a direct or indirect route, to the CNS begets an inflammatory cascade that results in recruitment of monocytes and, thus, increased number of target cells for HIV. Although found to migrate (344), as well as transport pathogens (345), from the lungs to secondary lymphatics, observation of a non-lymphatic or even CNS end-point in the migratory route of lung macrophages has not been reported. Hence, the literal translation of migration of these infected cells, or corresponding viral products, to
several regions throughout the body may be better explained by a wide-spread macrophage-tropic genotype, consistent with previous studies linking neurotropism with this genotype and/or phenotype (70, 192, 322-324), particularly during the terminal phase of infection (326, 346). Regardless of the mechanism, this dispersion and/or purely phylogenetic pattern was in stark contrast to the significant, and exclusive, contribution of peripheral T-cells to brain infection in animals without SIVE. Although infection of both cell types has been observed in the brain parenchyma of HIV- and SIV-infected patients (347), the results of this study suggest that infiltration and/or productive infection of T-cells in the CNS is not sufficient for severe neuronal injury, previously linked to increased CNS viral burden (72, 255).

Despite their contribution to brain viral lineages in animals without SIVE, the limited exchange of peripheral T-cell virus with remaining tissues and cell populations during asymptomatic infection and AIDS onset suggests a significantly reduced T-cell population size, a shift in T-cell phenotype less susceptible to infection, or the involvement of tissues and cell populations not sampled in this study. As previous studies have linked reduced T-cell counts with neuroAIDS progression (348), rather than neuroprotection, we believe this explanation unlikely. Similarly, although studies have linked T-cell phenotypic change with AIDS progression (e.g., (349)), little is known regarding its role in neuroprotection. Alternatively, the representation of all infected tissues and cell populations in this study is far from ideal, despite its advantage over previous studies. It is important to note, however, that absentee origins in this study due to sampling limitations, such as that for early brain and bone marrow infection, is more greatly appreciated than false positive origins, leading the research community in the
direction of additional sampling, rather than in vivo studies in validation of unreliable results. Further investigation incorporating phylogenetic and molecular tracking of additional tissues, such as the lymphatic and gastrointestinal systems, and more reliable markers of SIVE-mediated neuronal injury are essential for a thorough understanding of viral spread pathways within the HIV/SIV-infected host. We believe these additional studies to be justified, as results described herein indicate that these pathways provide a clear window into the role of viral population dynamics in AIDS-related disease progression.
CHAPTER 6
CONTRIBUTION OF INFECTED MONOCYTES TO VIRAL EVOLUTION IN THAI INDIVIDUALS WITH AND WITHOUT HUMAN IMMUNODEFICIENCY VIRUS-ASSOCIATED NEUROCOGNITIVE DISORDERS (HAND)

Introduction

Human immunodeficiency virus (HIV)-associated neurocognitive disorders (HAND) are estimated to persist in up to 50% of infected patients (350) in the face of combined antiretroviral therapy (33), with little clarity regarding underlying etiological mechanism(s) (351). Because of their susceptibility to infection and ability to cross the blood-brain barrier (29), monocytes have been proposed to act as “Trojan Horses”, facilitating entry of virus into the brain. Yet, the contribution of monocytes to ongoing viral infection and evolution, be it in the brain or peripheral tissues, has been heavily debated owing to the limited representation of in vitro model systems (352). The emerging field of Bayesian phyloanatomy provides an in silico solution to characterizing the contribution of individual cell types (and/or tissues) to viral evolution and dissemination (70, 287). However, the inclusion of monocyte populations in phyloanatomic models has not yet been undertaken as few studies have reported successful isolation of viral genomic RNA, indicative of replicating virus, from patient-derived monocytes (353). This study presents evidence of not only productive patient monocyte infection but the role of monocytes, distinct from peripheral T-cells, in ongoing HIV evolution, dissemination, and neuropathology. Despite the low prevalence of monocyte infection in vivo (354), results revealed a significant contribution of monocytes, but not T-cells, to plasma viral lineages in all patients with detectable monocyte infection. Furthermore, in the absence of significant dissemination of virus from peripheral blood and T-cells to the monocyte population in HAND patients,
monocyte-derived virus was characterized by elevated viral diversity and differing patterns of selection pressure. This study demonstrates the power of Bayesian phyloanatomy in understanding the plasticity of cellular infection dynamics among human hosts and warrants consideration of the significant role of monocytes in ongoing viral evolution and neuropathology using more sophisticated in vivo methods in humans and reliable animal models.

Despite the difficulties in obtaining viral genomic RNA from circulating monocytes in HIV-1-infected patients, cells of the monocyte/macrophage (M/M) lineage have been reported to be major sources of viral evolution in patients both during and following suppressive combined antiretroviral therapy (cART) (174, 353, 355-358). Differential roles of these cells have also been linked to HAND symptoms and SIV-associated neuropathology in the macaque model of neuroAIDS (80, 276, 359, 360). The relationship between monocyte infection and brain entry as well as evolution of a neurovirulent viral variant in vivo has been difficult to establish. However, results of numerous phylogenetic studies have indicated that differences in viral evolutionary dynamics are potentially responsible for the development of HAND (18, 32, 72, 361, 362). Based on these findings, we sought to investigate the evolutionary dynamics and genotypic characteristics not only in the primary target of HIV (CD4+ T-lymphocytes) but, more importantly, in the cell population implicated in AIDS-related neuropathology (CD14+ monocytes) from three late-infection individuals presenting with or without some form of HIV-associated neurocognitive disorder (HAND) at the time of cART initiation and, for one of the patients, following one year of cART adherence.
**Materials and Methods**

Peripheral blood samples were obtained from twenty-two HIV subtype AE-infected individuals enrolled in the SEARCH 007 study in Bangkok, Thailand (363). According to the Thai Ministry of Health guidelines (363), patients initiated therapy upon enrollment, having presented with symptomatic infection or a CD4+ T-cell count below 350 cells/mm. Upon enrollment, patients also underwent clinical assessment and neuropsychological testing for HIV-associated neurocognitive disorders (HAND), of which 15 (59%) were diagnosed with some form of HAND. Monocytes and T-cells were identified using markers of CD14 and CD3, respectively, and sorted using fluorescence-activated cell sorting (364) from individual samples. Viral env gp120 RNA sequences were obtained using reverse transcription polymerase chain reaction (PCR) and a modified limiting-dilution two-round PCR approach (“single genome sequencing”) (100) in order to prevent PCR-mediated resampling and recombination. Sampling, sequencing, and neuropsychological evaluations were repeated during a 12-month clinical follow-up. Monocyte virus sequences (>3) were only detected for three patients (P01, P02, P13) for which there were also T-cell and plasma virus sequences (refer to Figure B-1 for sequence number). Because monocyte sequences were detected additionally at the 12-month follow up time only for P01, multiple time-point Bayesian phylogenetic analysis was performed only for this patient in order to reduce the effect of sampling bias on evolutionary reconstruction of viral sequences from the other two patients. Additional precautions were taken to increase the reliability of inference from the sequence data, including but not limited to the following: (1) manual optimization of sequence alignments according to positional homology (105), (2) removal of potential recombinant sequences (121), and (3) assessment of sequence cross-contamination
Furthermore, because sampling was not uniform across sampling locations, three replicates of random sampling (without replacement) according to the minimal number of available sequences in one of the three locations was performed for each of the three patients (with equal representation of months 0 and 12 for P01) in order to incorporate the majority of the total number of sequences in the Bayesian phylogenetic analysis. Following random sampling, we assessed patient- and replicate sample-specific phylogenetic resolution. Also prior to Bayesian analysis, maximum likelihood tree reconstruction and corresponding tree correlation coefficients (TCC) were used to assess the influence of population structure, or the subdivision of sequences based on compartment, on phylogenetic relationships. For this study, TCCs represent the relationship between population isolation and the distance within the tree, with population subdivision defined in terms of either space or time and tree distance measured according to the number of branches (rb) or cumulative genetic distance along the branches (r) separating two sequences (224). Sufficient phylogenetic signal (Figure B-3) was determined with no consistent detectable degree of origin-specific population structure across samples for all three patients (Figure B-4 and Table B-1). Spatiotemporal reconstruction of viral dispersion using a continuous time Markov chain (CTMC) model of asymmetric transitions, or jumps, between environments (T-cell, monocyte, plasma) and the Bayesian stochastic search variable selection (BSSVS) model for identification of significant dispersal pathways (71). Additional assumptions included an uncorrelated relaxed molecular clock model of evolutionary rate variation across branches (338) and constant population size over time. A mean rate prior of 6.82E-04 substitutions/site/month, based on previous estimates (365), was assumed
with lognormal distribution. The hierarchical phylogenetic model (HPM) was used in BEAST v1.8.3 to assess the stability of viral dispersion across the three replicates of uniform sampling for each patient (241).

**Results**

**HIV-1-infected patients exhibit varying degrees of productive monocyte infection and sequence diversity relative to peripheral blood and T-cells.** HIV-1 genomic RNA was successfully PCR-amplified in monocytes from ten of twenty-two patients, with diagnoses ranging from cognitively normal (NL) to the most severe form of HAND - HIV-associated dementia (HAD). Successful amplification suggested productive infection of this peripheral cell population in at least a subset of HIV-infected patients regardless of HAND status (Figure B-1). Three of these patients (P01, P02, and P13), also ranging in diagnosis, presented with a sufficient number (>3) of single genome amplified monocyte, T-cell, and plasma viral glycoprotein (gp120) sequences for reliable phylogenetic inference. In addition to successful amplification, evidence implicating monocytes in the production of virus within the human host was provided in the posterior distribution of time-scaled Bayesian phylogenetic trees, wherein monocyte ancestral lineages were observed for all three patients, and to a greater degree in the patient with HAD - P13 (Figure 6-1). Monocyte viral lineages did not tend to cluster together monophyletically for any of the three patients in the time-scaled (Figure 6-1) nor distance-based maximum likelihood (ML) phylogenies (Figure B-4). This finding was confirmed using quantitative analyses of the genetic distance between sampling origin-specific sequences within the ML trees (Table B-1). Despite the absence of distinct clustering patterns, significant differences in diversity between sequences derived from monocytes and those of plasma and T-cells were observed, with directionality (relative
increase or decrease) differing depending on the patient and time of sampling (Figure B-5). Significantly reduced viral diversity within the monocyte sequences relative to both plasma and CD3+ T-lymphocyte sequences was observed for P02 (NL) and P01 (MND) during the first clinical visit, although not to the extent that would result in phylogenetic distinction. Contrastingly, monocyte sequences sampled from P13 (HAD) and P01 (MND) during the first and last clinical time point (12 months), respectively, were characterized by significantly increased viral diversity relative to plasma and greater than or equal to that of T-cells, suggesting that an increased viral effective population size in monocytes, associated with increased transmission dynamics (366), may be prevalent in patients with HIV-related neuropathology.

**Differing HIV-1-associated neuropathologies are characterized by unique monocyte contributions to viral dispersion within the peripheral blood.** The combination of prevalent paraphyletic and polyphyletic clustering of sequences according to sampling origin in the patient phylogenies (Figure 1) suggested that we could virtually exclude undetected intermediary subpopulation transmission links that might arise due to unavailability of additional sampling locations. This finding ultimately allowed for reliable inferences of direct transmission events between peripheral blood compartments using phyloanatomic reconstruction (367). The role of the monocyte ancestral lineages in viral dispersion for each of the three patients, represented within the Bayesian trees, was assessed using the Bayesian stochastic search variable selection (BSSVS) diffusion model (Figure 6-2). Despite differences in sampling time intervals and HAND severity among the patients, significant contribution of monocytes to cell-free viral lineages in the plasma, consistent with earlier studies (368), and of
plasma virus to peripheral T-cell lineages were consistently supported (BF>3). These three patients differed only in the role of plasma and peripheral T-cells in dissemination to monocyte viral sequences, with contribution from both populations in P02 (NL), plasma virus alone in P01 (MND), and neither population in P13 (HAD). Temporal reconstruction of viral dispersion indicated that the median time of transmission for each of the well-supported dispersion patterns occurred prior to the late administration of cART, although extended well into the first year post-therapy (Figure B-6). It is important to note, however, that, as P01 was the only patient with multiple time points (0 and 12 months post-cART) of sampling for all three anatomical sources, a generalizable inference of the timing of the three well-supported dispersion patterns could not be reliably obtained for all HIV-infected patients. Moreover, despite temporal clustering of sequences for all three replicate samples (see Methods) belonging to P01 (Table B-1), only one of the replicates exhibited measurable evolutionary divergence of sequences from the inferred evolutionary root of the ML tree (Figure B-7). Even in the presence of sampling variation, temporal inferences did not appear to differ with regards to the three well-supported dispersion patterns (Figure B-6), indicating the robustness of the BSSVS approach. The results of the phyloanatomy analysis thus provided evidence for patient-, and potentially, neuropathology-specific variation in spatial viral dispersion patterns between peripheral blood targets of HIV, with preliminary evidence to suggest their occurrence primarily but not exclusively prior to cART.
Figure 6-1. Sampled posterior distribution of Bayesian phylogenetic trees for all HIV gp120 sequences derived from plasma and sorted peripheral leukocytes in three individual Thai patients. A sample of 1,000 trees from the posterior distribution was obtained using an uncorrelated relaxed molecular clock model of evolutionary rate variation across branches (338) and constant population size over time. A mean rate prior of 6.82E-04 substitutions/site/month, based on previous estimates (365) with lognormal distribution, was used to scale the trees in time. Patients P01, P02, and P13 were diagnosed as having normal cognition (NL), mild neurocognitive disorder, or HIV-associated dementia (HAD).

**Distinct selective pressures characterize viral evolution monocytes of patients with HAND.** Differences at the level of amino acids between virus from plasma, T-cells, and monocytes were detected (Figure B-8), providing further evidence in support of different evolutionary histories. These differences were widely dispersed
across the gp120 primary sequence, suggesting that cell-specific adaptation is not the consequence of gene region-specific diversifying selection. Taking a more in-depth look into selection using the site-specific FUBAR selection analysis, variation in amino acids experiencing selective pressure distinguished not only T-cell from monocyte viral populations but also patients with (P001 and P013) and without (P002) HAND (Figure 6-3a). When both T-cell and monocyte sequences were considered, a greater number of total negatively selected sites were detected for HAND–associated sequences when compared to NL, whereas more total positively selected sites were observed for NL-associated sequences when compared to HAND, indicating distinct selective pressures associated with disease progression. None of the negatively or positively selected sites observed for monocyte sequences was duplicated between HAND and NL patients, suggesting this selective pressure is associated with monocyte infection. Nine of the negatively selected sites and two of the positively selected sites for monocyte sequences from patients diagnosed with HAND were unique as compared to T-cell sequences from both HAND and NL patients, implicating these sites as potentially important regulators in the infection of monocytes that contribute to brain infection and neuropathology.

**Purifying selection of the C2 region of gp120 is implicated in HAND progression.** Five of the nine amino acid residues under purifying selection and unique to monocytes in patients diagnosed with HAND were in close proximity and located within the C2 region. Genetic markers within this region have been previously linked with neuroAIDS progression (361) as well as suppression of infectivity of HIV-1 (362) and other pathogens (369, 370) in the oral mucosa due to functional sequence
homology with the CD36/LIMPII TSP1-binding motif (362), termed CLESH-1 (371). These residues were mapped onto the 20 Å resolution HIV-1 GP120 cryo-electron microscopy structure (PDB 3J70) (372) and corresponded to highly solvent-accessible positions outside of the CD4 receptor-binding site (Figure 6-3b), suggesting a biological function distinct from attachment, but not necessarily fusion, to the cell surface.

**Discussion**

Although genetic distinction of monocyte-derived viral sequence populations from remaining peripheral blood compartments has been described previously for patients on cART (174, 355-358), we have presented similar evidence in the context of genomic RNA, indicative of replicating virus, as recent concerns have been raised regarding the bias of DNA phylogenetic analysis toward the inclusion of defective viral products, even during acute infection (373). Furthermore, we have demonstrated the power of Bayesian phyloanatomy in the characterization of the inferred evolutionary and transmission dynamics of this compartment, potentially attributed to the unique selective pressures accompanying monocyte infection (261) and development of the macrophage-mediated neuropathological disease referred to as HAND (325). Although the relationship between viral effective population size and transmission dynamics in the context of intra-host infection of tissues has not been directly assessed, this study provides the necessary justification for a more sophisticated approach to modeling the complex role of alternative tissues and cell types as HIV targets in a patient- and time-specific context.
Figure 6-2. Inferred viral dispersion among peripheral blood compartments in three individual HIV-1-infected Thai patients. The Bayesian stochastic search variable selection (BSSVS) (64) model of asymmetrical transition rates among discrete anatomical locations was utilized in BEAST (239, 240) to reconstruct simultaneously the viral dispersion and gp120 evolutionary histories for each of the three patients (P01, P02, P13). Patient gp120 sequence data were re-sampled thrice (with replacement) according to the minimum number of sequences across the three anatomical compartments, and the hierarchical phylogenetic model (241) was used to infer trends among sample replicates for each patient. Transition rates highly supported using (Bayes Factor [BF]>3) among these replicates are depicted as arrows with arrow width proportional to the BF value (legend in center). Patient diagnoses (in parenthesis) indicate normal cognition (NL), mild neurocognitive disorder, or HIV-associated dementia (HAD).
Figure 6-3. Location of amino acid site-specific selective pressure along gp120 according to patient diagnosis and cell type origin with 3D structure mapping of C2 sites specific to peripheral monocytes in HAND patients. (A) Sites under positive (+) and negative (-) selection were identified using the FUBAR model implemented in the datamonkey web-based server (www.datamonkey.org). Individual sites are represented by squares and are colored according to the cell type and diagnosis (NL or HAND) associated with selection. Amino acids comprising variable loops V1-4, as defined previously, are shaded accordingly. Five sites unique to monocytes in HAND patients in close proximity within the C2 region were mapped onto the CD4 (green) -bound cryo-electron microscopy structure (PDB 3J70) and were colored according to atomic element (carbon=yellow, oxygen=red, nitrogen=blue). Four distinct rotational angles are depicted with direction of rotation indicated by the center grey arrows.
CHAPTER 7
SIMIAN IMMUNODEFICIENCY VIRUS INTRA-HOST PHYLODYNAMIC PATTERNS IN CD8+ LYMPHOCYTE-DEPLETED AND NATURALLY PROGRESSING Rhesus Macaques: Insights into Differences in Disease Progression

Introduction

Viral evolutionary changes often have a profound impact on human immunodeficiency type 1 (HIV-1) disease progression (reviewed in (375)). The most well characterized changes involve a shift in tropism between the CCR5 and CXCR4 co-receptor usage and adaptation within the host to specific tissues and cell types, such as macrophages and T-lymphocytes. Tropism shifts have been linked to faster acquired immune deficiency syndrome (AIDS) progression (306) as well as the onset and progression of AIDS-related neuropathology (32). Understanding the evolutionary processes and dynamics of the viral population and subpopulations within an infected host is crucial for clarifying the complex interplay between the virus and various aspects of the host immune system. However, access to tissues and purified leukocytes harboring viral subpopulations is often difficult in the human host, especially at regular intervals over the entire disease course.

Infection of rhesus macaques with pathogenic simian immunodeficiency virus (SIV) is a widely used model for HIV-1 infection and AIDS progression (44). SIV-infected macaques exhibit similar clinical manifestations as untreated HIV-1-infected humans, albeit on a shorter time scale of approximately one to three years (44). Unfortunately, low incidence, maintenance costs, and disease timeline associated with the macaque model limits its usefulness in terms of producing rapid and statistically sound results. Such limitations have led to the development of rapid disease models, such as the use of antibody-mediated depletion of the CD8+ lymphocyte arm of the
antiviral response (83-85), which results in a compressed timeline of disease progression of six months or less compared to one to three years in non-depleted animals (84, 86, 87).

Rapid models of AIDS progression have provided invaluable insight into pathoimmunological (376-378) and neuropathological (379-381) mechanisms of HIV infection. Moreover, the role of the cellular immune response, specifically CD8+ lymphocytes, in controlling viral replication has become increasingly evident based on results obtained from these models and HIV-infected patients (376, 382). CD8+ lymphocytes, which include cytotoxic T-cells and natural killer cells, are able to suppress viral binding and transcription and eliminate the infected cell population (383), with the exception of macrophages, which are resistant to such lysis (384). CD8+ T-lymphocyte immune responses to HIV, however, are unable to eradicate infection due to fast mutation and replication rate of the virus resulting in continuous emergence of adapted populations, as well as progressive impairment of the immune system. Given their impact on viral infectivity, particularly during the early stages of infection (385, 386), it is not surprising that the level of HIV-specific recognition of virus by CD8+ lymphocytes has been associated with slower disease progression (387) and has been proposed as an addition to current combined antiviral therapeutic strategies (388, 389). Despite knowledge of the critical role of both viral evolutionary changes and the CD8+ lymphocyte response in HIV disease progression, few studies have focused on the impact of cellular immune response on viral intra-host population dynamics, or phylodynamics (61), much less for viral sub-populations within tissues, which vary in immune cell composition. For this study we utilized a phylodynamic approach to garner
more detailed insight as to the relationship of CD8+ lymphocytes and viral evolution at the level of individual infected tissues and peripheral cell populations sampled longitudinally from SIV-infected macaques with or without CD8+ lymphocytes depletion. Evolutionary tree topology for heterochronous (longitudinally sampled) sequences provides important information regarding viral population and subpopulation dynamics, including adaptation, population subdivision, and level of selection pressure (2, 365), by reflecting the interaction of the virus with the environment (61). CD8+ lymphocytes exert significant environmental pressures on the virus population and, consequently, virus evolution, and these pressures can vary among the diverse group of infected immune cell types as well as tissues comprised of highly dissimilar immune infrastructures (375).

Significant differences in the evolutionary history of the virus were, therefore, hypothesized to exist between the two macaque models, not only at the level of the host but also specific tissues and/or cell populations, which would potentially explain differences in disease progression. Results show that while SIV evolutionary and selection patterns in non-CD8+ lymphocyte-depleted animals were characteristic of sequential population turnover and continual viral adaptation, these patterns were less clear in the CD8+ lymphocyte-depleted model. Therefore, although the CD8+ lymphocyte immune response effectively prolongs early disease progression, these cells act to continually modulate the course of viral evolution and the emergence of higher fitness viral populations over time.

**Materials and Methods**

**Study Population.** Two macaque cohorts were used in this study, which included 6 CD8+ lymphocyte depleted (D01-D06) and 12 naturally progressing, or non-CD8+ lymphocyte-depleted (N01-N12), Indian rhesus macaques (*Macaca mulatta*),
hereafter referred to as the Mac251-DEP and Mac251-NP, respectively. Both cohorts were infected intravenously with the viral swarm SIVmac251 (1 ng SIV p27) (90). CD8+ lymphocyte depletion was achieved by subcutaneous administration of anti-CD8 antibody cM-T807 (6, 8, and 12 days post-infection [dpi]) (41). Two animals from each cohort were euthanized at 21 dpi in order to evaluate early evolutionary events within and related to the brain. The remaining animals were euthanized at the onset of SAIDS (75-118 dpi) (please refer to S1 Fig for timeline). Criteria for development of AIDS included: 1) weight loss >15% body weight in 2 weeks or >30% body weight in 2 months, 2) documented opportunistic infection, 3) persistent anorexia >3 days without explicable cause, 4) severe intractable diarrhea, progressive neurological signs, or significant cardiac and/or pulmonary signs, as previously described (80). A pathological diagnosis of SIV encephalitis (SIVE) was determined post mortem by a veterinary pathologist and included the presence of microglial nodules and multinucleated giant cells and was confirmed by immunohistochemistry staining for SIV p27, as previously described (82, 95-97).

**Ethical guidelines.** Procedures involving the Mac251-DEP animals were performed with the approval of the Tulane University’s Institutional Animal Care and Use Committee (IACUC). The treatment and handling of macaques in this cohort have been described previously (41). Procedures involving the Mac251-NP animals, housed at the New England Primate Research Center, were conducted according to the standards of the American Association for Accreditation of Laboratory Animal Care and IACUC protocol #04802, and treatment of all animals was in accordance with the Guide for the Care and Use of Laboratory Animals of the Institute of Laboratory Animal
Resources (337). Further detailed information on the handling and supervisinal guidelines for the Mac251-NP cohort can be found in (92). All possible measures were taken to minimize discomfort of the animals, and the guidelines for humane euthanasia of rhesus macaques were followed.

**Sample collection and sequencing.** Plasma viral loads were monitored as previously described by quantitative PCR (qPCR) methods targeting a conserved sequence in the group antigen gene (*gag*) (93, 94). Plasma, FAC-sorted peripheral CD3+ T-lymphocytes and CD14+ monocytes, unelicited bronchoalveolar lavage fluid (BAL) macrophages, and bone marrow (BM) aspirates were collected two to three time points – 21 days post infection (dpi), 60 dpi, and necropsy – from 4 Mac251-DEP macaques and at four time points – 21 dpi, 90 dpi, 180 dpi, and necropsy – from six Mac251-NP macaques (Figure C-1). Viral genomic RNA (and cellular genomic DNA when available) was extracted from these samples and the SIVmac251 inoculum as previously described (41, 72, 78, 92). Viral genomic material was also extracted from meninges and brain tissue sections from the parietal, frontal, and temporal lobes at necropsy when available. Viral RNA and DNA levels were quantified as described previously (72). Viral RNA envelope glycoprotein *gp120* sequences derived from the Mac251-DEP cohort were obtained using bulk PCR and cloning methods (41, 78). Due to the potential limitations of clonal analysis, a modified single genome sequencing protocol based on previously published methods (100) was used for all samples obtained from the Mac251-NP cohort as well as frontal lobe from the Mac251-DEP cohort. Sequences were aligned as previously described (78), and approximately 20 *gp120* sequences per tissue per time point were obtained after removal of potential
recombinants. Detailed information regarding sample collection, sequencing protocols, and the sequence alignment procedure have been reported previously (41, 72). All sequences used in this study are accessible in GenBank (accession numbers JF765272-JF766081 [Mac251-DEP], KR998525-KR999900; KX081185-KX081229; KX081254-KX082629 [Mac251-NP and inoculating viral swarm]).

**Recombination analysis.** Recombination analysis was carried out for sequence populations in each tissue and in each macaque using three methods, as previously described (117): 1) GARD (118), 2) an algorithm that uses split decomposition networks in combination with the 'PHI test' (119, 120) implemented in Splitstree4 software (121) (available from http://www.splitstree.org) calibrated specifically to identify likely recombinants within intra-host viral populations (119, 122, 123), and 3) Recombine (124). Recombinant sequences were also identified using bootscanning, implemented in Simplot (125) (available from http://sray.med.som.jhmi.edu/SCRoftware/simplot/) and RDP (126) (available from http://web.cbio.uct.ac.za/~darren/rdp.html), with observed trends similar to those obtained with the algorithms listed above (data available upon request). Putative recombinant sequences were screened against viral swarm sequences in order to determine if they had originated from the inoculum and were analyzed separately.

**Viral diversity and divergence estimates.** Estimates of mean pairwise viral diversity within individual tissues/cell populations, as well as mean divergence of longitudinally sampled sequences from the inoculating viral swarm, were calculated in MEGA v5.2.2 (106) (available from http://www.megasoftware.net) using the maximum composite likelihood model of nucleotide substitution (107) and 1,000 bootstrap
replicates. Due to the influence of the bulk PCR/cloning method on sequence heterogeneity (100), mutated sites representing <1% of observed point mutations (estimated PCR error rate) were removed from the Mac251-DEP alignments prior to phylogenetic analysis, as described in (89). After removal of these sites, overall viral diversity and divergence within tissue/cell populations during early and late infection were compared for the two cohorts to determine if differences between the two animal models could be explained by the sequencing methodology.

**Phylogenetic and molecular clock analysis.** Evaluation of phylogenetic resolution satisfying resolved phylogenetic relationships among SIV sequences from each macaque was performed using likelihood mapping (390) implemented in IQ-TREE (391) (available from http://www.IQ-TREE.org/), searching for all possible quartets using the best nucleotide substitution model selected by a hierarchical likelihood ratio test. Absence of substitution saturation, which decreases the phylogenetic information contained in the sequences, was also assessed using DAMBE6 (392) (available from http://dambe.bio.uottawa.ca/DAMBE/). Maximum likelihood (ML) reconstruction of phylogenetic trees for both macaque cohorts was performed in RAxML v8.0.25 (110) (available from http://sco.h-its.org/exelixis/web/software/raxml/index.html) using the general time-reversible (GTR) model of nucleotide substitution (393) with gamma-distributed rate variation across sites and 1,000 bootstrap replicates. Trees were viewed and modified in FigTree v1.4.0 (available from http://tree.bio.ed.ac.uk/software/figtree/) and can be found in previous published studies (72, 78).

Temporal resolution for each macaque phylogeny was assessed using a linear regression of root-to-tip genetic distances inferred from the ML trees against sampling
time in the program Path-O-Gen v1.3 (394), now TempEst (395) (available from http://tree.bio.ed.ac.uk/software/tempest/). Trees were rooted using either the known infecting viral swarm sequences or by selecting the root resulting in the best root-to-tip correlation (both methods produced identical results). Deviation from a strict molecular clock model was assessed based on the coefficient of determination ($R^2$) values from the regression analysis and coefficient of variation for evolutionary rates assuming a relaxed clock model in BEAST v1.8.0 (239) (available from http://beast.bio.ed.ac.uk). The HKY model of nucleotide substitution (396) with gamma-distributed rate variation across sites (4 categories) and Bayesian Skyride and/or Bayesian Skyline Plot (BSP) demographic model (295, 339) were used in parameter estimation. For further information on prior distributions, XMLs are available from the authors upon request.

Model testing was performed based on Bayes Factor (BF) comparison (397) using the harmonic mean estimator in Tracer v1.5 (available from the BEAST software package). Ln(BF)>6 was considered significant evidence in favor of the more complex model.

Statistical differences in $R^2$ values and mean evolutionary rates reported in this study for the Mac251-NP animals and previously for the Mac251-DEP animals (78) were determined using a two-tailed Student’s t-test (unpaired).

The degree of topological temporal structure within the ML phylogenies was assessed using a temporal clustering analysis performed in MacClade v4.08 (398) (available from http://macclade.org/download.html), as previously described (249). Briefly, sequences were assigned states according to sampling time, and the number of state changes within the maximum likelihood phylogeny (inferred without assuming a molecular clock) was compared to that of a null distribution of 1,000 replicates, for which
the character states have been randomized (while the phylogeny itself is held constant). The final result is referred to as the temporal clustering (TC) statistic, ranging from 0-1, with 0 indicating absence of temporal structure (i.e. intermix of sequences sampled at different time points), and 1 perfect temporal structure (i.e. all sequences from the same sampling time cluster together and are the direct ancestors of sequences from the following sampling time).

**Metapopulation structure analysis.** Metapopulation structure (i.e. the existence of distinct sub-populations), or compartmentalization, analyses results were obtained from two separate compartmentalization tests implemented in HyPhy (108) (available from http://hyphy.org/) in order to evaluate the extent of distinct viral subpopulations within individual tissues. Analyses included both tree- and distance-based methods. Tree correlation coefficients were calculated based on the number of branches (rb) or branch length (r) separating sequences within separate defined compartments (109) in the RAxML-derived ML trees. Statistical significance was determined using a null distribution of permutated sequences (1,000 permutations), whereby a p-value of ≤0.05 was considered significant. The Simmonds association index (SAI or AI) was determined for sequence alignments using SIVmac239 as a reference sequence. The SAI represents the mean ratio of 100 bootstrap replicates of the association value, calculated from the test sequences, to those of 10 sample-reassigned controls. The association value (d) is defined as \( d = (1 - f)/2n - 1 \), where n is the number of sequences below the node and f is the frequency of most common sample type (111).

Bootstrapping (1000 replicates) was used as a test of significant compartmentalization,
for which support of >80% was considered significant. A more thorough description of such tests can be found in (112).

**Selection analysis.** Because statistical measures of metapopulation structure can be affected by selection as well as migration dynamics, an unrestricted branch-site random effects model, referred to as BUSTED (Branch-Site Unrestricted Statistical Test for Episodic Diversification; implemented in the https://datamonkey.org), was used to test for gene-wide episodic diversifying selection (113). The analysis was restricted to only internal branches, which are assumed to capture at least one round of virus replication, to mitigate the biasing effects of transient deleterious mutations on the ratio of non-synonymous to synonymous substitution rates estimates along terminal branches, where selection has not had time to fully filter such population level variation (114, 115).

In addition, nucleotide sequence alignments for all tissues/cell populations at individual time points for each macaque were used to determine site-specific selection over the course of infection for both macaque cohorts. The fast, unconstrained Bayesian approximation for inferring selection (FUBAR) model (116) (implemented in http://datamonkey.org) was used to identify potential individual amino acid sites under selection within viral gp120 sequences for individual macaques as well as for individual tissues/cell types at each time point. Sites with posterior probability >0.9 of an increased (diversifying) or decreased (purifying) rate of non-synonymous relative to synonymous substitutions were considered to have experienced a significant level of selective pressure. Macaques were then classified according to SIVE diagnosis or early sacrifice
in order to determine similarities and differences among classifications across macaque cohorts.

**Viral RNA and DNA quantification.** Viral genomic RNA and DNA levels were estimated for longitudinally sampled and *post mortem* tissues/cell populations of Mac251-NP macaques as well as *post mortem* frontal lobe for Mac251-DEP animals. A detailed description of the methods can be found in (72). In brief, viral *gp120* RNA was quantified using QUALITY (127) (implemented at https://indra.mullins.microbiol.washington.edu/quality/) based on limiting dilution PCR positive amplifications (data available upon request). For time points in which only one dilution was used, the following calculation was employed: `-ln[1-proportion of successful PCR wells]*Dilution/PCR Template Volume (2 ul), assuming a Poisson probability distribution of RNA templates within PCR samples. Integrated viral DNA was quantified using qPCR amplification of Gag *p27*. Cell-associated viral sequence copy numbers were represented in terms of copies/10,000 cells based on the stock concentration of gDNA and the presumed 6.2 pg gDNA/macaque cell: (viral copy number/ng gDNA)*(62 ng gDNA/10,000 cells).

**Statistical analysis.** Statistical significance for differences in evolutionary parameters between cohorts was determined using either Welch’s unpaired t-test (unequal variance) or Mann-Whitney U test based on Shapiro-Wilk test of normality (α=0.05).

**Results**

**CD8+ lymphocyte-depleted and naturally progressing macaque cohorts** were characterized by sufficient phylogenetic resolution and significant differences in time to SAIDS. Longitudinal viral sequences were successfully obtained
from various tissue and cell populations – consisting of bone marrow, bronchoalveolar lavage fluid (BAL) macrophages, sorted peripheral CD3+ T-lymphocytes and CD14+ monocytes, and plasma - from 4 Mac251-DEP and ten Mac251-NP macaques over the course of infection (S1 Fig). Additionally, sequences from meninges and three distinct lobes within the brain (frontal, temporal, and parietal cortices were obtained at necropsy. Results of likelihood mapping and substitution saturation analysis for all sequence data sets indicated sufficient resolution for further phylogenetic analysis (data available upon request).

Six of the ten longitudinally sampled Mac251-NP animals were chosen for phylogenetic analysis based on significantly (p-value<0.05) rapid disease progression relative to the remaining four macaques (Figure C-2). This distinction was made to reduce variability resulting from additional potential factors, such as host genetic variation (399) driving differences in disease progression within this cohort. The temporal range in days post-infection (dpi) of AIDS onset in Mac251-NP macaques (204-373 dpi) indicated a significantly slower progression (p-value<0.01) than in Mac251-DEP animals (75-118 dpi). As reported previously (72), patterns in viral load and CD4+ T-cell counts were also investigated in each animal in order to rule out possible confounding factors (such as set point viral load and CD4+ T-cell nadir) responsible for differences in disease progression between the Mac251-DEP and Mac251-NP cohorts. These findings were critical for reliable inferences of the impact of CD8+ lymphocyte depletion on viral evolution in the context of differential disease progression.
Temporal signal and clustering contribute to differences in macaque cohort tree topologies despite similar evolutionary rates. Differences in viral population dynamics between the two cohorts were measured in terms of overall evolutionary rate, rate variation among lineages, and temporal patterns in the underlying tree topology. The level of temporal resolution as well as the “clocklike” nature of the evolutionary rate for macaque-specific SIV evolution was assessed using regression analysis of the distance within the maximum likelihood (ML) tree (Figure 7-1) between the tip of each collected taxon and the inferred root against time of sample collection. Sufficient temporal resolution, indicated by a positive slope for the regression line, was present for all of the macaques, indicating measurable evolution of the SIV population from the time of inoculation to necropsy. The resemblance of SIV evolutionary rate to that of a strict molecular clock (more linear divergence indicated by a greater R² value), however, was significantly greater in the Mac251-NP macaques than in the Mac251-DEP macaques (Figure C-3 and Table 8-1). Despite a more clocklike nature in the Mac251-NP animals, the relaxed clock evolutionary model was deemed appropriate for all animals in both cohorts using model testing in BEAST (BF>20) (397), and Bayesian estimates of mean evolutionary rates did not significantly differ between the two cohorts. However, it is important to note that, consistent with the differences in clocklike signal, the amount of rate variation among intra-host viral lineages, as measured by the coefficient of variation, was significantly greater for the Mac251-DEP macaques (Table 8-1). In addition to a more linear divergence pattern, increased temporal structure, or clustering (TC), characterized by the clustering of sequences within the phylogeny based on time of sampling (249), was observed for the Mac251-NP macaques, as measured
quantitatively using the TC statistic (Figure 7-2). These results indicate important differences in overall viral population dynamics in the two hosts, with viral evolution in the naturally progressing macaques exhibiting more clock-like behavior and increased temporal structure representative of sequential population turnover.

Reduced temporal structure in CD8+ lymphocyte-depleted macaques was not attributed to increased tissue/cell reservoir formation. Lack of, or low-level, temporal structure, as was observed for the Mac251-DEP animals, can often be explained by the reemergence of (or re-activation) of archival genomes from long-lived cell populations, or reservoirs (2, 70, 175). One reasonable indication of a potential tissue/cellular reservoir for virus is the presence of metapopulation structure, or compartmentalization, within a phylogeny, which is characterized by the clustering of sequences based on their respective tissue/cellular origins (70). A few of these viral lineages may escape their residual compartment, giving rise to a peripherally sampled virus (plasma- or PBMC-derived) that is more closely related to older sequences associated with the former compartment than the more recently sampled peripheral virus, thereby reducing the level of temporal clustering. However, differences in tissue/cell-specific clustering patterns were not observed between the ML phylogenetic trees belonging to the two cohorts (Figure 7-1), which were confirmed using both distance- and tree-based quantitative tests of compartmentalization (Table 8-2).

Sequences from individual tissues and/or leukocyte populations appeared to cluster together for 2-3 macaques within each cohort, but these monophylies, indicating phylogenetically distinct viral subpopulations, were not consistent across all macaques (Figure 7-1). For example, monophyletic clustering was observed, but only for
macaques N09 and D06. Quantitative compartmentalization analysis was, therefore, performed on the basis of two hypotheses: 1) each tissue/cell type harbors a distinct SIV subpopulation (complete compartmentalization hypothesis) and 2) only one tissue/cell type (and not necessarily the same in each animal) harbors an SIV subpopulation distinct from the panmictic population of viral strains infecting the remaining tissues/cell types (single compartment hypothesis). The analysis to test the first hypothesis revealed several Mac251-DEP and Mac251-NP animals with significant tissue-specific viral populations using tree correlation coefficients (TCC), which measures both the number of branches and distance along the branches separating sequences from defined tissue/cell types “compartments” (109). None of the macaques in either cohort exhibited significant tissue compartmentalization using the Simmonds Association Index (SAI) (Table 8-2), which was used to account for uncertainty in the tree topology using bootstrapping (112). These results were not surprising given the appearance of clustering for only certain tissues/cell populations within the ML trees.

The second quantitative analysis to assess whether at least one tissue in each animal harbored a distinct SIV subpopulation (single compartment hypothesis) using the more sensitive SAI alone revealed that, aside from brain sequence compartmentalization (discussed in detail in (72)), BAL macrophage and/or meninges sequences were significantly compartmentalized in five different macaques across both cohorts (Table 8-2). Interestingly, the meninges were only significantly compartmentalized among the Mac251-NP cohort for macaque N10 with SIV-associated encephalitis (SIVE). Virus from the meninges of macaque D03 (also SIVE+) were also significantly compartmentalized (Table 8-2), suggesting that meninges sequences can be genetically
distinct from both peripheral and parenchymal brain sequences, consistent with an earlier study by Matsuda et al. (130). Additionally, this compartmentalization may be associated with SIVE, although the unavailability of meninges sequences for the remaining SIVE+ Mac251-DEP animals prohibited a robust statistical approach to this hypothesis.

The occurrence of distinct viral subpopulations in meninges and lung, independent of CD8+ lymphocyte depletion, was not surprising, considering that productive HIV and SIV infection of meninges (79) and lungs and/or epithelial lining fluid has predominantly been associated with cells of monocyte/macrophage lineage (47, 400, 401), rather than T-lymphocytes. This finding is also consistent with previous reports of monocyte/macrophage viral compartmentalization and their role as a potential reservoir (70, 163).

Reduced viral divergence but not diversity is associated with CD8+ lymphocyte depletion-mediated rapid progression to SAIDS. When within-tissue diversity and divergence from the viral swarm were investigated over time, distinct differences were observed between the two cohorts (Figure 7-3). The Mac251-NP cohort displayed similar rates of increased viral diversity and divergence until approximately 90 dpi, when the rate of viral diversity accumulation for most tissue and cell types significantly decreased (p-value≤0.05) until 182-189 dpi. Subsequently, the rate of diversity accumulation remained relatively low (~1.12E-05 substitutions/site/day) until the last time point (average of ~270 dpi across macaques). Accumulation of substitutions compared with the viral swarm (i.e., divergence) increased at a more rapid rate (p-value=0.07) until 180 dpi, when it significantly declined (p-value≤0.05) to ~3.36E-
05 substitutions/site/day. Despite the overall decline, divergence was significantly increased (p-value≤0.05) at the last time point for virus from CD14+ monocytes and cell-free virus in plasma. An increase (p-value=0.07) was also observed for BAL macrophage sequences.

No consistent pattern was apparent for individual tissues or cell populations within the Mac251-DEP macaques (Figure 7-3A). However, variation across macaques for the majority of sampled tissues and cell populations (CD14+ monocytes, BAL macrophages, and CNS cortical tissue) was much greater relative to the other cohort. A significantly increased (p-value≤0.05) rate of sequence divergence was observed during the last sampling time interval in contrast with the Mac251-NP cohort (Figure 7-3B), primarily due to elevated BAL macrophages (p-value≤0.001), CD14+ monocyte (p-value=0.06), and plasma (p-value=0.07) sequence divergence, which closely resembled the evolutionary events during the last sampling time interval in the Mac251-NP cohort.

Consistent with the differing patterns in viral divergence over time between the two cohorts, divergence estimates did not significantly differ between the two cohorts for individual tissues or cell types at 21 dpi but differed significantly (p-values≤0.05) at necropsy (Figure 7-4A). Increased viral divergence was observed in the Mac251-NP cohort, typically observed for the longer-living Mac251-NP animals (data not shown). BAL macrophage and meninges sequences were exceptions to this pattern, for which divergence was greater in the Mac251-DEP animals. Interestingly, these animals also developed SIVE or meningitis within four months of the infection (78). Differences in viral diversity and divergence patterns between the two cohorts did not appear to be imposed by cloning-induced errors (after PCR error correction), as viral diversity
estimates for both cohorts at both 21 dpi and necropsy did not differ significantly, with the exception of significantly increased (p-value ≤ 0.05) viral diversity within Mac251-NP necropsy-sampled bone marrow (Figure 7-4B).

**CD8+ lymphocyte-depleted animals exhibit a compressed timeline of selective pressure, with differing contributions from anatomic compartments.** As selective pressure is an important driving force behind changes in viral diversity and divergence, site-specific selection analysis was utilized to determine the proportion of viral amino acid sites experiencing either purifying or diversifying selection at each sampling time point for macaques within both cohorts (Figure 7-5A), and individual tissues and/or cell populations (Figure 7-5B). Although purifying selection appeared to dominate for both cohorts, the Mac251-DEP macaques harbored a greater proportion of sites under diversifying selection compared to the Mac251-NP macaques at all time points (Figure 7-5A). Significant differences (p-values < 0.05) between the two cohorts for both diversifying and purifying selection were observed by 92 dpi, with the difference in diversifying selection potentially explained by viral evolution in the CD14+ monocyte population (p-value < 0.08), which may have contributed to the similar pattern observed for virus in the plasma (p-value < 0.05) (Figure 7-5B). Despite the lack of significance until 92 dpi, the proportion of sites experiencing either diversifying or purifying selection was significantly greater (p-values ≤ 0.05) in the CD3+ T-lymphocyte population at 21 dpi for the Mac251-DEP as compared with Mac251-NP animals.
Figure 7-1. Maximum likelihood (ML) phylogenetic tree reconstruction for each macaque. (A) Naturally progressing cohort. The following macaques were sampled longitudinally: N02, N04, N09, N05, N10, and N12, whereas N06 and N07 were sampled at a single time point (21 days post-infection). Trees for N02, N09, and N10 are adapted from Rife et al., 2016 (72). (B) CD8+ lymphocyte-depleted cohort. Trees for this cohort are adapted from Strickland & Rife et al. (2014) (78). The following macaques were sampled longitudinally: D03, D04, D05, and D06, whereas D01 and D02 (boxed) were sampled at a single time point (21 days post-infection). For all ML trees, the GTR + G model of nucleotide substitution was used in RAxML with bootstrapping (1,000 replicates) analysis.
Table 7-1. Analysis of temporal signal for time-stamped phylogenies.

<table>
<thead>
<tr>
<th>Macaque ID</th>
<th>$R^2$</th>
<th>Mean evolutionary rate</th>
<th>CV (95% HPD)</th>
</tr>
</thead>
<tbody>
<tr>
<td>Naturally Progressing</td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>N02</td>
<td>0.63</td>
<td>3.17E-04</td>
<td>0.58</td>
</tr>
<tr>
<td>N04</td>
<td>0.51</td>
<td>3.39E-04</td>
<td>0.77</td>
</tr>
<tr>
<td>N05</td>
<td>0.63</td>
<td>2.85E-04</td>
<td>0.67</td>
</tr>
<tr>
<td>N09</td>
<td>0.7</td>
<td>2.55E-04</td>
<td>0.58</td>
</tr>
<tr>
<td>N10</td>
<td>0.74</td>
<td>2.27E-04</td>
<td>0.68</td>
</tr>
<tr>
<td>N12</td>
<td>0.66</td>
<td>2.30E-04</td>
<td>0.76</td>
</tr>
<tr>
<td>Average:</td>
<td>0.65</td>
<td>2.76E-04</td>
<td>0.67</td>
</tr>
<tr>
<td>SD:</td>
<td>0.08</td>
<td>4.62E-05</td>
<td>8.29E-02</td>
</tr>
<tr>
<td>CD8+ Lymphocyte-Depleted</td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>D03</td>
<td>0.08</td>
<td>3.46E-04</td>
<td>1.13</td>
</tr>
<tr>
<td>D04</td>
<td>0.15</td>
<td>2.03E-04</td>
<td>0.99</td>
</tr>
<tr>
<td>D05</td>
<td>0.13</td>
<td>2.53E-04</td>
<td>1.14</td>
</tr>
<tr>
<td>D06</td>
<td>0.07</td>
<td>2.35E-04</td>
<td>0.84</td>
</tr>
<tr>
<td>Average:</td>
<td>0.11</td>
<td>2.59E-04</td>
<td>1.03</td>
</tr>
<tr>
<td>SD:</td>
<td>0.04</td>
<td>6.14E-05</td>
<td>1.41E-01</td>
</tr>
<tr>
<td>$^e$ p-value:</td>
<td>7.80E-07</td>
<td>6.70E-01</td>
<td>0.01</td>
</tr>
</tbody>
</table>

$^a$Data are presented for the six longitudinally sampled naturally progressing macaques in this study and the four longitudinally sampled macaques from a previous study (78).

$^b$R$^2$ values were calculated from the regression between root-to-tip distance in the RAxML-derived maximum likelihood trees and the date of sampling of each sequence using Path-O-Gen v1.3 (available from tree.bio.ed.ac.uk/software/pathogen/).

$^c$The mean evolutionary rate and associated $^d$median coefficient of variation (CV) with corresponding 95% high posterior density (HPD) intervals, were determined in BEAST v1.8.0 (available from http://beast.bio.ed.ac.uk) using the relaxed clock model.

$^e$A two-tailed Student’s T-test was used to determine significant differences in $R^2$, mean evolutionary rate, and CV, whereby a p-value of $\leq 0.05$ was considered significant.
Figure 7-2. Temporal clustering within maximum likelihood phylogenies for longitudinally sampled naturally progressing and CD8+ lymphocyte-depleted macaques. The temporal clustering statistic was determined using MacClade v4.08 based on the comparison of the temporal clustering tree score for the given maximum likelihood tree to a null distribution of 1,000 replicates. Error bars represent standard deviation among all macaques. ***p-value≤0.001 using the Student’s T-test.
### Table 7-2. Analysis of tissue-dependent compartmentalization of viral RNA sequences.

<table>
<thead>
<tr>
<th>Macaque ID</th>
<th>Individual Tissues&lt;sup&gt;a&lt;/sup&gt;</th>
<th>CD3+</th>
<th>CD14+</th>
<th>Plasma</th>
<th>Bone marrow</th>
<th>BAL</th>
<th>Meninges</th>
</tr>
</thead>
<tbody>
<tr>
<td></td>
<td>TCC&lt;sup&gt;b&lt;/sup&gt;</td>
<td>P(rb)</td>
<td>P(r)</td>
<td>SAI (BS)&lt;sup&gt;c&lt;/sup&gt;</td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>N02</td>
<td>0.00</td>
<td>0.00</td>
<td>0.66 (100)</td>
<td>0.71 (100)</td>
<td>0.81 (94)</td>
<td>0.76 (99)</td>
<td>0.86 (85)</td>
</tr>
<tr>
<td>N09</td>
<td>0.00</td>
<td>0.00</td>
<td>0.54 (100)</td>
<td>0.71 (100)</td>
<td>0.54 (100)</td>
<td>0.75 (98)</td>
<td>0.41 (100)</td>
</tr>
<tr>
<td>N10</td>
<td>0.00</td>
<td>0.00</td>
<td>0.62 (100)</td>
<td>0.77 (99)</td>
<td>0.76 (94)</td>
<td>0.81 (98)</td>
<td>0.84 (93)</td>
</tr>
<tr>
<td>N12</td>
<td>0.00</td>
<td>0.00</td>
<td>0.71 (100)</td>
<td>0.80 (97)</td>
<td>0.78 (99)</td>
<td>0.78 (98)</td>
<td>0.56 (100)</td>
</tr>
<tr>
<td>N06</td>
<td>0.35</td>
<td>0.49</td>
<td>0.66 (100)</td>
<td>0.68 (88)</td>
<td>0.98 (39)</td>
<td>0.65 (97)</td>
<td>0.67 (82)</td>
</tr>
<tr>
<td>N07</td>
<td>0.96</td>
<td>0.94</td>
<td>0.65 (100)</td>
<td>0.71 (100)</td>
<td>0.80 (96)</td>
<td>0.67 (100)</td>
<td>0.75 (97)</td>
</tr>
<tr>
<td>N04</td>
<td>0.00</td>
<td>0.00</td>
<td>0.60 (100)</td>
<td>0.76 (100)</td>
<td>0.85 (69)</td>
<td>0.71 (100)</td>
<td>0.56 (100)</td>
</tr>
<tr>
<td>N05</td>
<td>0.00</td>
<td>0.00</td>
<td>0.59 (100)</td>
<td>0.86 (97)</td>
<td>0.78 (99)</td>
<td>0.78 (98)</td>
<td>0.56 (100)</td>
</tr>
<tr>
<td>N06</td>
<td>0.00</td>
<td>0.00</td>
<td>0.58 (100)</td>
<td>0.85 (68)</td>
<td>0.82 (96)</td>
<td>0.75 (98)</td>
<td>0.41 (100)</td>
</tr>
</tbody>
</table>

<sup>a</sup>Plasma, sorted CD3+ T-lymphocyte and CD14+ monocyte, BAL macrophage, bone marrow, meninges, and cranial lobe (not shown) compartments were analyzed either simultaneously (“Individual Tissues”) or individually.

<sup>b</sup>Tree correlation coefficients (TCC) were calculated based on the number of branches (rb) or branch length (r) separating sequences within separate defined compartments. Statistical significance was determined using a null distribution of permuted sequences (1,000 permutations). A p-value of ≤0.05 was considered significant.

<sup>c</sup>Simmonds association index (SAI, or AI) represents the mean ratio of 100 bootstrap replicates of the association value, calculated from the test sequences, to those of 10 sample-reassigned controls. The bootstrap support (BS) for AI values is also provided. SIVmac239 was used as a reference sequence for SAI determination. BS > 80 was considered significant.

<sup>d</sup>Animals sacrificed at 21 dpi (Early Sacrifice)
Figure 7-3. Genetic distance calculations over time for viral sequences derived from longitudinally sampled peripheral cell populations and meninges of naturally progressing and CD8+ lymphocyte-depleted macaques. Viral diversity (solid) and divergence (dashed) macaques were estimated in nucleotide substitutions/site for gp120 sequences derived from various cell populations over time for four different time points in days post-infection (dpi) and meninges at necropsy. Divergence was estimated as the genetic distance from the SIVmac251 inoculating viral swarm sequences. Average viral divergence and diversity within meninges and brain compartments over time for the CD8+ lymphocyte-depleted macaques are highlighted as an inset. The depicted last time point is an average of sampling times among all macaques prior to euthanization. Error bars represent standard deviation among all macaques. *p-value≤0.05, **p-value≤0.01, ***p-value≤0.001 using the Welch’s unpaired t-test. (A) naturally progressing and (B) CD8+ lymphocyte-depleted macaques.

Differences in temporal patterns of recombination rates do not readily distinguish macaques based on cohort. Intra-host recombination is another important feature of HIV-1 and SIV evolution driving viral diversity and thus the
emergence of viral immune escape variants (402-404). Because recombinant sequences were removed prior to the phylogenetic analyses, the level of recombination over time was assessed in all sequences for both cohorts to determine if this was an influential factor in disease progression. Longitudinal patterns of intra-host recombination were previously investigated using the Mac251-DEP cohort (117), the results of which revealed that recombination occurs as early as 21 dpi in the previously mentioned sampled tissues/cell populations within these animals (Figure C-4A). Similarly, recombinant sequences were found at 21 dpi in all tissues and cell populations in the Mac251-NP cohort (Figure C-4B). No association was observed between recombination rates and tissue of origin for the Mac251-DEP cohort, as previously reported (117). However, the pattern over time for the Mac251-NP macaques was highly similar to sequences derived from bone marrow, plasma, CD3+ T-lymphocytes, and BAL macrophages between 21 and 182 dpi. During this time interval, an increase in the percent recombinant sequences was observed between 21 and 92 dpi, followed by a decrease until 182 dpi. Subsequently, an increase was again observed in all cells and tissues, with the exception of the bone marrow. Recombination within sorted CD14+ monocytes was distinctly different, characterized by a decrease until 92 dpi, followed by an increase until 182 dpi. The results suggest that recombination rates may be tissue/cell specific during conventional progression, which may not be observable with limited resolution longitudinal sampling or in a rapid model of disease progression.

**Discussion**

Early modulation of the cell-mediated arm of the host immune system in the CD8+ lymphocyte-depleted, or Mac251-DEP, macaques is thought to play a major role
in evolutionary processes of the virus, both at the population and tissue/cell sub-population levels. This complex interplay between the host’s primary immune defenses and viral evasion strategies can be observed as an evolutionary fingerprint in the phylogenetic reconstruction of serially sampled viral lineages. Analysis of the phylogenetic tree topology was, therefore, critical in providing information as to the extent of this relationship.

The virus-host interaction is evident in HIV-1 intra-host genealogies and represented by sequential population turnover and evidence of continual viral adaptation to the multitude of host immune responses (2, 12, 405). One of the central findings in this report is that the characteristic staircase tree topology, indicative of population turnover driven by continuous immune selection, was only evident in naturally progressing macaques, although this was not to the degree usually observed in plasma samples of HIV-1-infected patients (12). The increased temporal structure in HIV-1-infected patients may, in part, be due to the inherently different infection timeline between the two hosts, but also putatively the incorporation of sequences from anatomical locations other than the blood, highlighting the importance of the macaque model.

The dissipation of temporal structure in the Mac251-DEP relative to the Mac251-NP animals was not supported by the explanation of the reemergence of (or reactivation) of archival genomes from viral reservoirs (70, 175), but instead may simply be explained by insufficient time to observe viral population turnover. Achaz et al. (406) have reported that complete within-host viral population turnover can take up to twenty-two months in HIV-1-infected humans. With this translation to SIV-infected macaques
irresolute to date, this study has provided valuable information as to an appropriate sampling strategy (approximately three month intervals) for informative evolutionary analysis, for which a comparable strategy may not be realistically obtainable in rapid macaque models.

Figure 7-4. Genetic distance calculations for viral sequences derived from early infection (21 dpi) and necropsy samples of naturally progressing and CD8+ lymphocyte-depleted macaques. Distances were estimated in nucleotide substitutions/site for gp120 sequences collected at 21 dpi and necropsy, or the last time point (LTP) and were averaged across naturally progressing macaques (grey) and CD8+ lymphocyte-depleted macaques (black). Black asterisks (*) indicate a statistically significant elevation in viral diversity and/or divergence for naturally progressing macaques relative to CD8+ lymphocyte-depleted macaques, whereas grey asterisks indicate statistical significance for the reverse. Error bars represent standard deviation among all macaques. *p-value<0.05, **p-value<0.01, ***p-value<0.001 according to the Welch’s unpaired t-test. (A) Within-tissue viral divergence from the inoculating viral swarm. (B) within-tissue viral diversity
Figure 7-5. Site-specific selection over time for SIV sequences within individual naturally progressing and CD8+ lymphocyte-depleted macaque tissues/cell populations. The fast, FUBAR model implemented in HYPHY was used to measure selection at individual sites within viral gp120 sequences for individual macaques at each time point. The proportion of sites with a posterior probability >0.9 of an increased (diversifying selection) or decreased (purifying selection) rate of non-synonymous relative to synonymous substitutions are reported as averages across naturally progressing (open triangles) and CD8+ lymphocyte-depleted (filled triangles) macaques for individual tissues/cell populations (colored accordingly) at each sampled time point. Error bars represent standard deviation among all macaques. *p-value≤0.05, **p-value≤0.01 according to the Welch’s unpaired t-test. (A) Sites under diversifying selection. (B) sites under purifying selection.
**Figure 1.**

A. 

**Diversifying Selection**

- Proportion of Sites vs. Time (dpi)

**Purifying Selection**

- Proportion of Sites vs. Time (dpi)

B. 

- **Bone Marrow**
  - Proportion of Sites vs. Time (dpi)

- **CD3+**
  - Proportion of Sites vs. Time (dpi)

- **CD14+**
  - Proportion of Sites vs. Time (dpi)

- **Plasma**
  - Proportion of Sites vs. Time (dpi)

- **BAL**
  - Proportion of Sites vs. Time (dpi)

- **Meninges, Frontal, Parietal, & Temporal Lobes**
  - Proportion of Sites vs. Time (dpi)

\(\hat{\Delta} CD8+ Lymphocyte-Depleted\)

\(\triangle\) Naturally Progressing

**NA**

**NA**
Differences in viral population structure and dynamics over time were also manifested in temporal patterns of viral diversity and divergence as well as indications of immune selective pressure. Analyses of viral diversity and divergence in HIV-infected patients have been used in an attempt to distinguish disease progression rates, often producing conflicting results (e.g., (407, 408)). The most well-known study of evolutionary patterns associated with moderate disease progression was reported by Shankarappa et al. (12), in which HIV-1-infected patients exhibited a consistent three-phase pattern characterized by the rate of change in both viral diversity and divergence: an early phase, characterized by a linear increase in both diversity and divergence; an intermediate phase, characterized by a continued increase in divergence but stabilization or decline in diversity; and a late phase, characterized by the stabilization of divergence and continued stability or decline in diversity, which has been explained statistically by reduced selective pressure due to deterioration of the immune response. Despite the fact that the utilized sampling strategy was based on estimated time of viral population turnover rather than changes in viral load and CD4+ T-cell counts, as in the Shankarappa study, we have shown for the first time a strikingly similar three-phase pattern for naturally progressing, SIV-infected macaques. Furthermore, we were able to dissect these patterns based on individual tissue and cellular locations, revealing a deviation from this trend in CD14+ monocytes, BAL macrophages, and plasma, for which genetic divergence from the viral swarm continued to increase during the last sampling time interval for both macaque cohorts. This similarity suggests that, even when CD8+ lymphocytes are transiently depleted, the “cellular exhaustion” hypothesis is supported, wherein depletion of the primary CD4+ T-cell target and infection of
alternative, slower replicating cell types, such as monocytes and macrophages (47, 79, 400, 401, 409), contribute to the overall diversity/divergence stabilization (410).

Due to limited sampling time point resolution for the Mac251-DEP animals, comparison to the Shankarappa three-phase model was unresolved, although it is important to note the increased variation among the macaques as well as individual tissues/cell types. Because immune selective pressure exerted on the virus by CD8+ lymphocytes is known to be an important environmental factor in driving viral diversity and divergence (63, 411-413), this increased variation could be explained by an altered fitness landscape that is essentially flatter, as opposed to immune selection-driven movement through a hilly landscape toward particular phenotypes that are more successful with respect to immune evasion. This flattened fitness landscape would also explain reduced sequence divergence and compressed timeline of site-specific selection patterns in the rapidly progressing animals. In other words, our findings indicate that, in depleted animals, dramatically reduced selective pressure from the CD8+ lymphocytes allows the virus to move more easily through the flatter landscape and to adapt more quickly to CD4+ T-cell depletion as well as the eventual CD8+ lymphocyte rebound, resulting in faster disease onset. Such a rapid adaptation to CD4+ T-cell loss is also consistent with the observed elevated selective pressure on the sequenced CD3+ T-cell virus in the Mac251-DEP animals as compared with the Mac251-NP cohort.

In summary, the results described in this study indicate a critical role for the cell-mediated immune response in shaping viral evolution within and among tissues and cell populations. Deviation from an apparent viral subpopulation homeostasis due to an
altered viral fitness landscape may be responsible for driving the rapid disease progression observed in the CD8+ lymphocyte-depleted model and implies applicability and relevance of this model to the study of rapid progression among HIV-1-infected patients.
CHAPTER 8
CONCLUDING REMARKS

The temporal and spatial origins of the neurovirulent virus in the brain of HIV-infected patients with AIDS-related neuropathology has been heavily debated owing to the timing of clinically relevant symptoms and presence of virus in the brain as early as two weeks post-infection (52, 53, 78). This debate presented a unique opportunity for the first combined epidemiological and phylogenetic approach (Chapter 3) (70) in investigating patterns of HIV and SIV evolution and dispersion among anatomical compartments in diagnosed and control subjects. Results of this project indicated that late entry (nearing AIDS onset) of peripherally neuroadapted viral variant(s) into the brain was responsible for the unique viral sequence population observed within this compartment and associated with neuroAIDS diagnosis (Chapter 2) (72). Furthermore, the in-depth sampling strategy included in this study, involving several anatomical tissues and cell populations, revealed a complex pattern of viral dispersion among these locations that may be linked with disease progression and neuroadaptation (Chapters 4-6). A small, but important, part of this complex pattern, was the finding of the significant role of infected lung macrophages in seeding the brain of SIV-infected macaques diagnosed with SIVE (Chapter 5). Previously hypothesizing a direct contribution of bone marrow-derived monocytes to brain infection (Chapter 1), results of this project point to an alternative explanation - a recruitment of these monocytes in response to an inflamed brain parenchyma already populated with infected, migratory macrophages, the origin of which may be lungs or similar macrophage-rich tissue. Further studies involving additional tissues not sampled as part of this study, in combination with molecular tracking techniques previously described for similar investigations (326), will
be required to test this hypothesis, and I am confident the results of this project fully justify this investigative path. Similarly, the detection of a distinct selective pressure in HAND patients (Chapter 6) and SIVE+ macaques (Chapter 2) (72) placed on the C2 region of Env encourages in vitro analysis of the role of this region in macrophage-tropism, as well as entry and replication in the BBB-protected brain parenchyma.

In addition to providing a unique perspective on HIV CNS infection, this study highlighted the power of statistical phylogenetic methods in capturing not only the evolutionary history of the virus within the host but also the complexity of movement and the selective pressures placed on the virus from the uniquely different host anatomical microenvironments and levels of immune response (Chapters 5-7), promoting its use in future investigations of tissue infection in the context of anti-retroviral therapy and even the expansion of this toolset outside the realm of chronic microbial infections.
Phylodynamic and phylogeographic inferences of viral population dynamics and demographic history are based on the coalescent framework implemented in the BEAST software package (239, 295). For the phylodynamic analysis, both parametric and non-parametric demographic models of changes in effective population size (Ne) over time were tested in BEAST v1.7.1 under the HKY+G model of nucleotide substitution (selected through a hierarchical likelihood ratio test). Parametric models included constant population size and exponential growth, whereas the non-parametric models consisted of the Bayesian Skyline and Bayesian Skyride plots. A strict or relaxed molecular clock (lognormal distributed rates) model was also tested for each demographic coalescent prior. Non-parametric demographic priors enforcing a relaxed molecular clock (selected as the best fitting model) were implemented in BEAST r5017 with the addition of discrete geographical priors for the Bayesian phylogeography analysis. For both analyses, three independent Markov Chains Monte Carlo (MCMC) were run for $5 \times 10^8$ generations for each animal data set of aligned sequences, with sampling every 50,000 generations. Posterior probabilities were then calculated after the appropriate burn-in (10% of total tree population) using the program Tracer v1.5 (available from http://beast.bio.ed.ac.uk/). Multiple independent runs were combined when needed to enable proper mixing of the MCMC, which was determined by visual examination of traces and ESS values (>200) for each estimated parameter. The marginal likelihoods obtained from the path-sampling and stepping stone algorithms (414) for the six competing models (demographic following either a strict or relaxed molecular clock) were compared using Bayes factors (BF) to select the best fitting
model. LnBF>3 was used as a cut-off indicating strong evidence in favor of the alternative model (397, 415).

Table A-1. Coalescent demographic model selection.

<table>
<thead>
<tr>
<th>Primate ID (#Taxa&lt;sup&gt;a&lt;/sup&gt;)</th>
<th>Demographic model&lt;sup&gt;b&lt;/sup&gt;</th>
<th>CoV&lt;sup&gt;c&lt;/sup&gt;</th>
<th>mLik&lt;sup&gt;d&lt;/sup&gt;</th>
<th>ln (BF)&lt;sup&gt;e&lt;/sup&gt;</th>
</tr>
</thead>
<tbody>
<tr>
<td>D01 (230)</td>
<td>Con</td>
<td>1.51 (1.04, 2.06)</td>
<td>-6131</td>
<td>Exp</td>
</tr>
<tr>
<td></td>
<td>Exp</td>
<td>1.36 (1.04, 1.76)</td>
<td>-6057</td>
<td>Con</td>
</tr>
<tr>
<td></td>
<td>BSP</td>
<td>0.99 (0.72, 1.35)</td>
<td>-6047</td>
<td>Exp</td>
</tr>
<tr>
<td></td>
<td>BSR</td>
<td>1.44 (1.12, 1.82)</td>
<td>-6526</td>
<td>BSP</td>
</tr>
<tr>
<td>D02 (223)</td>
<td>Con</td>
<td>1.57 (1.05, 2.26)</td>
<td>-5941</td>
<td>Exp</td>
</tr>
<tr>
<td></td>
<td>Exp</td>
<td>1.37 (1.02, 1.79)</td>
<td>-5881</td>
<td>Con</td>
</tr>
<tr>
<td></td>
<td>BSP</td>
<td>1.28 (0.85, 1.72)</td>
<td>-5877</td>
<td>Exp</td>
</tr>
<tr>
<td></td>
<td>BSR</td>
<td>1.53 (1.16, 1.98)</td>
<td>-6269</td>
<td>BSP</td>
</tr>
<tr>
<td>D03 (289)</td>
<td>Con</td>
<td>1.29 (0.92, 1.72)</td>
<td>-6813</td>
<td>Exp</td>
</tr>
<tr>
<td></td>
<td>Exp</td>
<td>0.96 (0.69, 1.26)</td>
<td>-6754</td>
<td>Con</td>
</tr>
<tr>
<td></td>
<td>BSP</td>
<td>0.89 (0.63, 1.18)</td>
<td>-6753</td>
<td>Exp</td>
</tr>
<tr>
<td></td>
<td>BSR</td>
<td>1.13 (0.84, 1.48)</td>
<td>-7181</td>
<td>BSP</td>
</tr>
<tr>
<td>D04 (285)</td>
<td>Con</td>
<td>1.29 (0.97, 1.68)</td>
<td>-6782</td>
<td>Exp</td>
</tr>
<tr>
<td></td>
<td>Exp</td>
<td>0.94 (0.71, 1.19)</td>
<td>-6705</td>
<td>Con</td>
</tr>
<tr>
<td></td>
<td>BSP</td>
<td>0.95 (0.72, 1.20)</td>
<td>-6683</td>
<td>Exp</td>
</tr>
<tr>
<td></td>
<td>BSR</td>
<td>0.99 (0.77, 1.25)</td>
<td>-7073</td>
<td>BSP</td>
</tr>
<tr>
<td>D05 (303)</td>
<td>Con</td>
<td>1.01 (0.75, 1.32)</td>
<td>-8038</td>
<td>Exp</td>
</tr>
<tr>
<td></td>
<td>Exp</td>
<td>0.87 (0.64, 1.13)</td>
<td>-7966</td>
<td>Con</td>
</tr>
<tr>
<td></td>
<td>BSP</td>
<td>0.84 (0.60, 1.11)</td>
<td>-7949</td>
<td>Exp</td>
</tr>
<tr>
<td></td>
<td>BSR</td>
<td>1.14 (0.88, 1.44)</td>
<td>-8346</td>
<td>BSP</td>
</tr>
<tr>
<td>D06 (288)</td>
<td>Con</td>
<td>0.84 (0.62, 1.11)</td>
<td>-8305</td>
<td>Exp</td>
</tr>
<tr>
<td></td>
<td>Exp</td>
<td>0.92 (0.69, 1.17)</td>
<td>-8233</td>
<td>Con</td>
</tr>
<tr>
<td></td>
<td>BSP</td>
<td>0.81 (0.59, 1.05)</td>
<td>-8207</td>
<td>Exp</td>
</tr>
<tr>
<td></td>
<td>BSR</td>
<td>0.84 (0.60, 1.11)</td>
<td>-8581</td>
<td>BSP</td>
</tr>
</tbody>
</table>

<sup>a</sup>Total number of sequences analyzed for each primate.

<sup>b</sup>Four demographic models were evaluated: constant population size (Con), exponential population growth (Exp), Bayesian skyline plot (BSP), and Bayesian Skyride (BSR).

<sup>c</sup>The coefficient of variation (CoV) is the evolutionary rate variance (estimated by a Bayesian relaxed molecular clock) scaled by the associated mean. 95% high posterior density (HPD) intervals are given in parenthesis. 95%HPD intervals not including zero indicate significant deviation from a strict molecular clock.

<sup>d</sup>Marginal likelihood estimates obtained with the stepping-stone method described in Baele et. al (2013) (414)

<sup>e</sup>Natural log of Bayes factors comparing each model pair
### Table A-2. Tree correlation coefficients.

<table>
<thead>
<tr>
<th>Macaque ID</th>
<th>All brain</th>
<th>Frontal</th>
<th>Frontal/Parietal</th>
<th>Parietal</th>
<th>Parietal/Temporal</th>
<th>Temporal</th>
<th>Frontal/Temporal</th>
</tr>
</thead>
<tbody>
<tr>
<td></td>
<td>P(r)(^a)</td>
<td>P(r)(^b)</td>
<td>P(r)</td>
<td>P(r)</td>
<td>P(r)</td>
<td>P(r)</td>
<td>P(r)</td>
</tr>
<tr>
<td>D03</td>
<td>0.00</td>
<td>0.00</td>
<td>0.00</td>
<td>0.28</td>
<td>0.00</td>
<td>0.00</td>
<td>0.00</td>
</tr>
<tr>
<td>D04</td>
<td>0.00</td>
<td>0.00</td>
<td>0.00</td>
<td>0.83</td>
<td>0.00</td>
<td>0.71</td>
<td>0.00</td>
</tr>
<tr>
<td>D05</td>
<td>0.00</td>
<td>0.00</td>
<td>0.00</td>
<td>0.00</td>
<td>0.00</td>
<td>0.00</td>
<td>0.00</td>
</tr>
<tr>
<td>D06</td>
<td>0.00</td>
<td>0.01</td>
<td>0.00</td>
<td>0.77</td>
<td>0.00</td>
<td>0.02</td>
<td>0.00</td>
</tr>
</tbody>
</table>

\(^a\)Correlation coefficient measuring the number of branches separating the sequences for each brain tissue compartment or combined compartments within the maximum clade credibility (MCC) tree. Values less than 0.33 were bolded, indicative of compartmentalization.

\(^b\)Correlation coefficient measuring the cumulative genetic distance between individual sequences for each brain tissue compartment(s). Values less than 0.33 were bolded, indicative of compartmentalization.

### Table A-3. Sampling spectra.

<table>
<thead>
<tr>
<th>Macaque ID</th>
<th>Total(^a) sequences</th>
<th>21 dpi(^b)</th>
<th>61 dpi(^b)</th>
<th>Necropsy (21-118 dpi)(^c)</th>
</tr>
</thead>
<tbody>
<tr>
<td></td>
<td></td>
<td>P</td>
<td>O</td>
<td>P</td>
</tr>
<tr>
<td>D01</td>
<td>230</td>
<td>NA</td>
<td>NA</td>
<td>NA</td>
</tr>
<tr>
<td>D02</td>
<td>223</td>
<td>NA</td>
<td>NA</td>
<td>NA</td>
</tr>
<tr>
<td>D03</td>
<td>289</td>
<td>17</td>
<td>23</td>
<td>16</td>
</tr>
<tr>
<td>D04</td>
<td>285</td>
<td>20</td>
<td>20</td>
<td>19</td>
</tr>
<tr>
<td>D05</td>
<td>303</td>
<td>25</td>
<td>18</td>
<td>12</td>
</tr>
<tr>
<td>D06</td>
<td>288</td>
<td>23</td>
<td>21</td>
<td>17</td>
</tr>
</tbody>
</table>

\(^a\)The table shows the number of gp120 sequences obtained from each sampled tissue/time point of the six macaques, D01-D06. For labeling purposes, each tissue was assigned to a one-letter code that is used in the sequence names deposited in GenBank (P - plasma; O - monocyte; U - BAL; K - T-cell; J - lymph node; B - bone marrow; Z - meninges; F - frontal lobe; A - parietal lobe; T - temporal lobe).

\(^b\)Collection time of longitudinal samples in days post infection (dpi). NA = not available.

\(^c\)Necropsy samples were collected at 21 dpi (D01, D02), 75 dpi (D03), 91 dpi (D04), 95 dpi (D05) and 118 dpi (D06), respectively. NA = not available.
Figure A-1. Experimental design and sampling timeline. Six macaques, D01-D06, were inoculated intravenously and treated with an anti-CD8 antibody at 6 dpi and two follow-up treatments at 8 and 12 dpi. Sample collection included plasma, CD3+ lymphocytes, CD14+ monocytes, unelicited bronchial lavage, lymph node, and bone marrow. Meninges and brain tissues, including temporal, frontal and parietal lobe, were collected at necropsy.

Figure A-2. Slatkin-Maddison compartmentalization test. The bubble chart depicts the overall proportion of viral gene flow (migration) events during the infection between peripheral tissues and brain inferred from the SIV genealogies of primates D03-D06. Statistical significance was assessed by comparing the number of observed migrations to the brain to a null distribution obtained from 10,000 trees generated by random-joining splitting of the original MCC tree.
Recording migration times along the branch:
   <markovJumpsTreeLikelihood id="geoTreeLikelihood"
   logCompleteHistory="true"
   saveCompleteHistory="true" useUniformization="true">
   <patterns idref="geoPatterns"/>
   <treeModel idref="treeModel"/>
   <siteModel idref="geoSiteModel"/>
   <svsGeneralSubstitutionModel idref="originModel"/>
   </markovJumpsTreeLikelihood>

Logging migrations times based on the historyFilter time range:
   <log fileName="specialHistory.txt" id="specialHistory" logEvery="10000">
   <completeHistoryLogger>
   <markovJumpsTreeLikelihood idref="geoTreeLikelihood"/>
   <historyFilter maxTime="95" minTime="0"/>
   </completeHistoryLogger>

Figure A-3. BEAST special history xml additions. Xml code added for generating the special history report.
APPENDIX B
SUPPORTING INFORMATION FOR CHAPTER 6

Study population. Fifty HIV-1-infected, cART-naive volunteers were enrolled into the SEARCH 007 study (NCT00777426) at the Thai Red Cross AIDS Research Center in Bangkok, Thailand. All volunteers met the Thai Ministry of Health guidelines to initiate therapy based on having symptomatic HIV infection or a CD4+ T-cell count below 350 cells/mm$^3$ (363). Cognitively normal (NL) subjects were matched to HAND patients by age (within a decade), education (less than a high school degree, high school degree, some college, college degree), gender, and CD4+ T-cell count. In addition, 10 HIV-uninfected controls (CL) were enrolled. HIV-infected volunteers were categorized as NL or having HAND, with nine meeting criteria for Asymptomatic Neurocognitive Impairment (ANI), nine for Mild Neurocognitive Disorder (MND), and nine having HIV-Associated Dementia (HAD) based on a clinical assessment/neuropsychological testing. Clinical examination included a neuropsychological testing battery (417), the International HIV Dementia Scale (418), a macroneurological examination, and brain magnetic resonance imaging/spectroscopy (MRI/MRS). Final diagnoses were assigned by consensus conference that included an HIV neurologist and neuropsychologist using established criteria (419). HIV-uninfected participants were selected using the same inclusion and exclusion criteria, with the exception of CD4+ T-cell count. These exclusion criteria included previous exposure to anti-retroviral therapy, positive Hepatitis C serology, and presence of factors that could cause cognitive abnormalities (e.g., past head injury, learning disabilities, major depression, illicit drug use, active opportunistic infection, past or current CNS infection). Individuals were required to have a negative urine drug test prior to enrollment. Lumbar puncture and
brain MRI were performed, if indicated, to exclude central nervous system (CNS) opportunistic infection. Signed informed consent was obtained for all participants. Volunteers returned to the clinic at three-month intervals for the duration of one year. Peripheral mononuclear cells (PBMCs) were collected and neuropsychological testing performed at baseline and twelve months following enrollment/cART initiation. Data were generated from baseline and at twelve months post-cART initiation. Groups were analyzed based on HAND diagnosis at the time of baseline assessment (420).

Cell sorting. Cryopreserved PBMCs were quickly thawed in a 37°C water bath before being transferred to a 50ml conical tube containing 40ml RPMI with 20% FBS pre-warmed at 37°C. Cells were washed twice and transferred to a FACS tube and stained for 15 minutes at room temperature with an antibody cocktail consisting of anti-CD14-Pacific Blue (clone M5E2), anti-CD3-Alexa Fluor 700 (clone SP34-2), anti-CD20-Cy7-APC (clone B27) and anti-CD16-Cy7-PE (clone 3G8) (all from BD Pharmingen, San Jose, CA), anti-HLA-DR-ECD (clone L243, Beckman Coulter, Miami, FL), and Live/Dead Aqua (Invitrogen, Eugene, OR). All antibodies were titrated to determine optimal concentrations. Antibody-capture beads (CompBeads, BD Biosciences) were used for single-color compensation controls for each reagent used in the study, with the exception of cells being used for anti-CD3 and Live/Dead Aqua. After staining, cells were washed once, filtered and resuspended in 1ml PBS. The BD FACSaria cytometer (BD Biosciences, San Jose, CA) was set up with a pressure of 20 psi and a 100-um nozzle was used. Instrument calibration was checked daily by use of rainbow fluorescent particles (BD Biosciences). After acquiring unstained and single-color control samples to calculate the compensation matrix, we acquired 1 x 10^6 events
in order to set up the sorting gating strategy. CD14+ monocytes population were
gated first based on FSC and SSC parameters, after which we excluded 1) dead cells
by gating out Aqua+ cells and 2) unwanted cells by gating out CD3+ and CD20+ cells
and then gated on HLA-DR+ cells. From the HLA-DR+ population, a dot plot of CD14
vs. CD16 was used to make a sorting gate, which included all monocytes except the
CD14-CD16- subset. For CD3+ T-lymphocyte sorting, FSC and SSC parameters were
used to gate lymphocytes, dead cells were excluded by using Aqua staining, and
CD14+ cells were also excluded. Following this procedure, the CD3+ T-lymphocytes
were gated based on CD3 expression and negativity for CD16. Post-sort purity were
checked for each sample, and both CD14+ and CD3+ sorted subpopulations were
>98% pure. After cell sorting, the highly pure cell populations were washed with PBS
twice and all liquid was aspirated. Cells were then stored as dry pellets at -80°C.

RNA extraction and cDNA synthesis. Cell-free viral RNA was extracted from
patient plasma using the Qiagen QIAamp Viral RNA Mini Kit, whereas sorted PBMC-
associated RNA and DNA were processed using the Qiagen Allprep DNA/RNA Mini kit
according to the manufacturer's protocols. Viral RNA was then reverse transcribed into
cDNA according to the manufacturer’s protocol using the SuperScript® III First-Strand
Synthesis System kit (Invitrogen). The following primer was used for reverse
transcription: ‘K-env-R1’ 5’-CCAATCAGGGAAGAAGCCTG-3′ (HXB2 coordinates
8736–8716) (421).

Single genome amplification and sequencing. HIV-1 env gp120 sequences
were amplified from viral cDNA and genomic DNA (gDNA) using a modified limiting-
dilution two-round PCR approach (‘single genome sequencing’) based on previously
published methods (100) in order to prevent PCR-mediated resampling and recombination. The following primers were used for both rounds of PCR: ‘polenv_AE’ 5’-GAGCAGAAGACAGTGGAAATGA-3’ (HXB2 coordinates 6207-6228; modified from Tuttle et al., [2002] (422) for subtype AE) and ‘192H’ 5’-CCATAGTGCTTCCTGCTGCT-3’ (HXB2 coordinates 7815-7796; modified from Maureen Goodenow for subtype AE). PCR reactions consisted of 2 minutes at 94° C for 1 cycle, 30 seconds at 94° C, 30 seconds at 58° C, and 3 minutes at 72° C for 40 cycles, then 10 minutes at 72° C using the Platinum® Blue PCR SuperMix (Invitrogen). Amplicons were then visualized using 1% agarose gel electrophoresis with an Amplisize™ Molecular Ruler 50-2,000 base pair (bp) ladder (Bio-Rad). Sequencing was performed using an Applied Biosystems 3730xl DNA Analyzer (Life Technologies) at the University of Florida Interdisciplinary Center for Biotechnology Research genomics core facility.

RNA and DNA extractions, cDNA synthesis and first round PCR set-up were conducted in a restricted-access amplicon-free room with separate air-handling, with laboratory equipment where no amplified PCR products or recombinant cloned plasmids were allowed and where work surfaces and equipment were thoroughly cleaned before and after use with Eliminase® (Decon Labs, Inc.).

**Sequence alignment and analysis.** Individual nucleotide sequences were assembled using Geneious vR6 (423) (available from http://www.geneious.com) and are available in GenBank (Need to submit). Sequence chromatograms were visualized for the investigation of sites assigned multiple nucleotide identities for removal of potential PCR errors. Sequences were aligned using the Clustal algorithm (103) implemented in BioEdit v7.1.11 (104) (available from http://www.mbio.ncsu.edu/bioedit/bioedit.html)
followed by manual optimization of positional homology (105) and removal of gap-filled regions within the hypervariable V1V2 domains. The final alignment resulted in 1,068 nucleotides spanning position 6381-7580 of the HXB2 reference strain. Putative intra-host recombinants were determined using SplitsTree4 software (121) and removed prior to phylogenetic analysis. Neighbor-joining (NJ) tree reconstruction was then performed using MEGA v5.2.2 (106) (available from http://www.megasoftware.net/) with the HKY model of nucleotide substitution (396) and gamma-distributed rate variation across sites. Pairwise deletion was used for treatment of gaps within the alignment. Branch support was assessed by bootstrapping (1,000 replicates). Sequences from all participants were included in the NJ tree in order to evaluate patient subtype and the extent of sequencing cross-contamination based on patient-specific clustering patterns.

Evolutionary analysis was performed for participants from whom a sufficient number of monocyte-derived sequences were available to produce adequate phylogenetic signal for the monocyte compartment (P001, P002, and P013). It is important to note, however, that sequences from two separate time points (0 and 12 months) were analysed for P001, who maintained MND diagnosis throughout the study and did not suppress viral load, despite the initiation of cART upon enrolment. Viral genetic diversity was represented by pairwise genetic distances estimated for sequences derived from cell-free virus in the plasma and from sorted peripheral CD3+ T-lymphocytes and CD14+ monocytes within the three previously described patients (P001, P002, and P013). This estimation was performed in R (APE package (424) using the TN93 nucleotide substitution model (416) with gamma-distributed rate variation across sites ($\alpha=0.1$). A viral epidemiology signature pattern analysis (VESPA) was used
to detect distinct frequency variation in particular amino acids between plasma, CD3+ T-lymphocyte, and CD14+ monocyte viral sequences for P001, P002, and P013. The fast, unconstrained Bayesian approximation (FUBAR) (116) selection analysis implemented on the datamonkey web-based server (www.datamonkey.org) was used to infer differences in selective pressure at individual gp120 codon sites between anatomical compartments and patients diagnosed with or without HAND.

**Maximum likelihood tree reconstruction and compartmentalization analysis.** Because viral population structure, such as that dependent on anatomical location or cell type, can affect patterns of polymorphism that contribute to significant genetic variation or that mimic selection (425), the extent of this structure was assessed both qualitatively and quantitatively within each patient genealogy. Maximum likelihood (ML) tree reconstruction was performed for each of the three patients using all available gp120 sequences in order to assess clustering patterns according to anatomical location and time of sampling and was performed in IQ-TREE (391) using the best-fit evolutionary model according to the Bayesian Information Criterion. Tree correlation coefficients (TCC) were estimated to provide a quantitative assessment of compartmentalization, representing the relationship between population isolation and the distance within the tree, with population subdivision defined in this study in terms of either space or time and tree distance measured according to the number of branches (rb) or cumulative genetic distance along the branches (r) separating two sequences (112, 224). Anatomical compartmentalized structure was also analyzed in order to determine if within-host epidemiological linkages in the form of transmission, or dispersion, between peripheral cell populations and plasma could be resolved reliably.
Clustering patterns within the ML and Bayesian (see below) phylogenies based on anatomical sampling origin, particularly a mixture of paraphyletic and polyphyletic clades, were used to exclude the possibility of significant intermediary viral subpopulations or common sources of virus when interpreting the results of the Bayesian phyloanatomy analysis in BEAST (367).

**Bayesian phyloanatomy.** Because sampling was not uniform across sampling locations for these three patients, three replicates of random sampling (without replacement) according to the minimal number of available sequences in one of the three locations (P001=15, P002=20, P013=10) was performed for each of the three patients (with equal representation of months 0 and 12 for P01) in order to reduce the impact of spatial sampling bias while still incorporating the information from all sequences in the Bayesian phylogenetic analysis. Phylogenetic signal was determined prior to Bayesian analysis for each of the patient-specific sample replicates using likelihood mapping (390) implemented in TreePuzzle v5.2 (426) (available from http://www.tree-puzzle.de/), the results of which indicated sufficient signal for phylogenetic analysis (data available upon request). Temporal signal was also assessed for individual sample replicates of P01, consisting of sequences sampled at multiple time points, using linear regression analysis of the relationship of genetic distance from each taxa to the root of the ML tree and the time of sampling using TempEst (395) (available from http://tree.bio.ed.ac.uk/software/tempest/).

Bayesian genealogical tree reconstruction utilizing the coalescent framework was performed using BEAST v1.8.3 (239, 240) (available from http://beast.bio.ed.ac.uk/), assuming a relaxed molecular clock (338) Bayesian skyline demographic model (295),...
and the Bayesian stochastic search variable selection model (BSSVS) (64) of asymmetric transition rates among discrete anatomical locations. Using an asymmetric transition rate matrix within the BSSVS model allowed for inferred directionality of relative rates of viral dispersion between sampled anatomical compartments (plasma, CD3+ T-lymphocytes, and CD14+ monocytes). The hierarchical phylogenetic model was used to summarize trends across the three sample replicates for each patient (241). As mentioned previously, because monocyte sequences were unobtainable at later sampling time points for P001 and P002, only sequences from the first visit (V1) for these two patients were used for phyloanatomic analysis. Consequently, evolutionary rates were fixed at 6.82E-04 substitutions/site/month, based on previous estimates (365), with lognormal distribution. Detailed information regarding additional evolutionary parameters and associated priors used in BEAST analysis is available upon request.

Effective Markov chain Monte Carlo sampling for both Bayesian analyses was assessed by calculating the effective sample size (ESS) for each estimated parameter. ESS values > 200, calculated in Tracer (342) (available from http://beast.bio.ed.ac.uk/Tracer), were considered suitable indicators of effective sampling and were used to calculate the standard deviation of dispersion rate estimates. Trees sampled (1,000) from the posterior distribution were visualized simultaneously and branch density assessed using DensiTree (available from https://www.cs.auckland.ac.nz/~remco/DensiTree/), which high-density areas indicative of increased certainty of clustering patterns.

**Statistical analysis.** Significant differences for compartment-specific viral diversity (pairwise genetic distance) between patients were determined using a non-parametric pairwise multiple comparisons analysis based on rank sums (Dunn test
package in R) with Bonferroni p-value correction following rejection of the D'Agostino Pearson test of normality. Bayes Factors (BF) were calculated according to Lemey et al. (2009) (64) for patient-specific transition rates between each compartment using the MCMC rate posterior odds output from BEAST. BSSVS transition rates with BF>3 were considered to be well supported (287). Statistical significance was determined using a null distribution of permuted sequences (1,000 permutations). A p-value of ≤0.5 was considered significant.
Figure B-1. HIV-1-infected Thai patient diagnosis, sampling timeline, and sequence information. Patients enrolled in the SEARCH 007 study were diagnosed as having normal cognition (NL), asymptomatic neurocognitive impairment (ANI), mild neurocognitive disorder (MND), or HIV-associated dementia (HAD) according to clinical assessment and neuropsychological testing.

<table>
<thead>
<tr>
<th>PATIENT</th>
<th>DIAGNOSIS</th>
<th>0 MONTHS</th>
<th>3 MONTHS</th>
<th>6 MONTHS</th>
<th>9 MONTHS</th>
<th>12 MONTHS</th>
</tr>
</thead>
<tbody>
<tr>
<td></td>
<td>PLASMA</td>
<td>CD3+</td>
<td>CD14+</td>
<td>PLASMA</td>
<td>CD3+</td>
<td>CD14+</td>
</tr>
<tr>
<td>P01</td>
<td>MND</td>
<td>17</td>
<td>15</td>
<td>19</td>
<td>15</td>
<td>21</td>
</tr>
<tr>
<td>P02</td>
<td>NL</td>
<td>28</td>
<td>31</td>
<td>20</td>
<td>28</td>
<td>32</td>
</tr>
<tr>
<td>P03</td>
<td>ANI</td>
<td>19</td>
<td>14</td>
<td>3</td>
<td></td>
<td></td>
</tr>
<tr>
<td>P04</td>
<td>HAD</td>
<td>23</td>
<td>21</td>
<td>3</td>
<td></td>
<td></td>
</tr>
<tr>
<td>P07</td>
<td>ANI</td>
<td>18</td>
<td>30</td>
<td>2</td>
<td></td>
<td></td>
</tr>
<tr>
<td>P08</td>
<td>ANI</td>
<td>21</td>
<td>17</td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>P10</td>
<td>NL</td>
<td>12</td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>P13</td>
<td>HAD</td>
<td>17</td>
<td>48</td>
<td>10</td>
<td></td>
<td></td>
</tr>
<tr>
<td>P15</td>
<td>HAD</td>
<td>27</td>
<td></td>
<td>5</td>
<td></td>
<td></td>
</tr>
<tr>
<td>P19</td>
<td>NL</td>
<td>18</td>
<td>28</td>
<td>3</td>
<td></td>
<td></td>
</tr>
<tr>
<td>P20</td>
<td>MND</td>
<td>22</td>
<td></td>
<td>31</td>
<td>2</td>
<td>19</td>
</tr>
<tr>
<td>P21</td>
<td>HAD</td>
<td>28</td>
<td>17</td>
<td>3</td>
<td>8</td>
<td></td>
</tr>
<tr>
<td>P23</td>
<td>NL</td>
<td>19</td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>P24</td>
<td>NL</td>
<td>12</td>
<td>12</td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>P26</td>
<td>NL</td>
<td>21</td>
<td>25</td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>P28</td>
<td>NL</td>
<td>25</td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>P29</td>
<td>HAD</td>
<td>13</td>
<td>9</td>
<td>1</td>
<td></td>
<td></td>
</tr>
<tr>
<td>P31</td>
<td>HAD</td>
<td>18</td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>P33</td>
<td>HAD</td>
<td>15</td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>P34</td>
<td>NL</td>
<td>34</td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>P35</td>
<td>HAD</td>
<td>27</td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>P39</td>
<td>NL</td>
<td>16</td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
</tbody>
</table>
Figure B-2. Neighbor joining (NJ) phylogeny incorporating all HIV-1-infected Thai patients for which viral gp120 sequences were obtained. NJ tree reconstruction was performed in MEGA v5.2.1 (106) according to the HKY (396) evolutionary model with gamma-distributed rate variation across sites. Branches are colored according to patient, and patients representing HIV-1 subtypes B and AE are outlined.
Figure B-3. Likelihood mapping of HIV-1-infected Thai patient-specific replicates of uniform sampling from anatomical locations. Patient gp120 sequence data were re-sampled thrice (with replacement) according to the minimum number of sequences in one of the three anatomical compartments. Likelihood mapping (390) was performed for each patient (P01, P02, P13) sample replicate (S1-3) in IQ-TREE (391) using the best-fit evolutionary model according to the Bayesian Information Criterion. Triangular compartments correspond to percent of sequence quartets that were unresolved (center), partially resolved (center edges), or fully resolved (corners) in the phylogenetic tree. Greater than 20% fully resolved quartets was considered sufficient for reliable evolutionary inferences.
Figure B-4. Maximum likelihood (ML) phylogenies of viral gp120 sequences for individual HIV-1-infected Thai patients. ML trees were reconstructed in IQ-TREE (391) using the best-fit evolutionary model according to Bayesian Information Criterion. Taxa are shaped according to the time of clinical visit in months post-diagnosis (mpd), or post-cART initiation, and are colored according to isolation origin, (plasma or sorted peripheral blood mononuclear cell [PBMC]).
Figure B-5. Viral genetic diversity within plasma and peripheral T-cells and monocytes obtained at specific time points from three individual HIV-1-infected Thai patients. Viral genetic diversity, represented as the pairwise genetic distances between sequences belonging to the same anatomical compartment, was estimated in R (ape package) using the TN93 evolutionary model \( (416) \). The number of sequences analyzed for each patient-specific compartment is depicted above. Patient (P01, P02, P13), diagnosis (NL=normal cognition, MND=minor neurocognitive impairment, HAD=HIV-associated dementia), and visit number (V1 = 0 months post-cART, V5 = 12 months post-cART) are also indicated. Statistical differences were determined using a non-parametric multiple comparisons test (Dunn package in R) with Bonferroni p-value correction. *p-value<0.05 **p-value<0.01 ***p-value<0.001
Figure B-6. Inferred timing of viral dispersion among discrete anatomical compartments for patient P01 sample replicates. Patient gp120 sequence data were re-sampled thrice (with replacement) according to the minimum number of sequences in one of the three anatomical compartments. The timing, in months post-diagnosis (mpd), of viral dispersion was inferred for individual P01 sample replicates (1-3) using the Bayesian coalescent (295) and phyloanatomy (70) frameworks. *Bayes factor support (>3) indicating a significantly non-zero rate of transition between designated discrete anatomical locations within the Bayesian phylogeny, as determined using the Bayesian stochastic search variable selection model (64) of asymmetric transition rates.
Figure B-7. Linear regression analysis of the relationship between genetic divergence and sampling time for P01 sample replicate gp120 sequence data. Patient gp120 sequence data were re-sampled thrice (with replacement) according to the minimum number of sequences in one of the three anatomical compartments, taking into consideration both sampled time points – 0 and 12 months post-cART. Linear regression analysis was performed in TempEst (395) in order to assess the relationship of genetic distance between each taxa (yellow dot) and the root of the corresponding maximum likelihood tree (y-axis) with the time of sampling (x-axis). Correlation coefficient (r) is indicated for each replicate sample (S1-3).
Figure B-8. Amino acid signature pattern analysis for plasma, monocyte- and T-cell-associated gp120 sequences in three HIV-1-infected Thai individuals at the time of cART initiation. Amino acid differences are measured as differences in the majority of amino acid present for patient- and anatomical compartment-specific sequence alignments at the time of cART initiation relative to the CM240_AE reference sequence. Differences are denoted by the majority amino acid present for a particular alignment. Amino acids comprising variable loops V1-4, as defined previously (374), are shaded accordingly.
Table B-1. Analysis of compartmentalization of *gp120* sequence data for three HIV-1-infected Thai patients.

<table>
<thead>
<tr>
<th>Patient ID</th>
<th>Replicate/Sample(\text{a})</th>
<th>Time (\text{rb})</th>
<th>Location (\text{r})</th>
</tr>
</thead>
<tbody>
<tr>
<td>P01</td>
<td>S1 0.03</td>
<td>0.02</td>
<td>0.77 0.7</td>
</tr>
<tr>
<td></td>
<td>S2 0.003</td>
<td>0.001</td>
<td>0.94 0.9</td>
</tr>
<tr>
<td></td>
<td>S3 0.02</td>
<td>0.009</td>
<td>0.4 0.36</td>
</tr>
<tr>
<td></td>
<td>All sequences 0.08</td>
<td>0.005</td>
<td>0.26 0.4</td>
</tr>
<tr>
<td>P02</td>
<td>S1</td>
<td></td>
<td>0.01 0.02</td>
</tr>
<tr>
<td></td>
<td>S2</td>
<td></td>
<td>0.06 0.11</td>
</tr>
<tr>
<td></td>
<td>S3</td>
<td></td>
<td>0.02 0.06</td>
</tr>
<tr>
<td></td>
<td>All sequences</td>
<td></td>
<td>0.35 0.19</td>
</tr>
<tr>
<td>P13</td>
<td>S1 0.28</td>
<td>0.21</td>
<td></td>
</tr>
<tr>
<td></td>
<td>S2 0.03</td>
<td>0.04</td>
<td></td>
</tr>
<tr>
<td></td>
<td>S3 0.43</td>
<td>0.34</td>
<td></td>
</tr>
<tr>
<td></td>
<td>All sequences</td>
<td></td>
<td>0.52 1</td>
</tr>
</tbody>
</table>

\(\text{a}\)patient *gp120* sequence data were re-sampled thrice (with replacement) according to the minimum number of sequences in one of the three anatomical compartments. Analysis was also performed for alignments consisting of all sequences for each patient.

\(\text{b}\)correlation coefficient (109) statistic (p-value) representing the number of branches (rb) or branch length (r) separating, within the maximum likelihood phylogeny, sequences from separate compartments defined according to sampling time or anatomical sampling origin/location. Statistical significance was determined using a null distribution of permutated sequences (1,000 permutations). A p-value of ≤0.5 was considered significant.
Figure C-1. Timeline of infection and sampling for the CD8+ lymphocyte-depleted and naturally progressing macaque cohorts. Six CD8+ lymphocyte-depleted and 12 naturally progressing macaques were inoculated (orange) with a SIVmac251 viral swarm at day 0 post-infection (DPI). Depletion of CD8+ cells was obtained using an anti-CD8 (α-CD8) monoclonal antibody at 3 time points between 6 and 12 dpi (blue), as described previously (78). Tissue samples, excluding brain and meninges, were serially sampled at depicted dpi (pink), whereas brain and meninges were sampled post-mortem (green). Two macaques from each cohort (“+2”) were sacrificed early at 21 dpi (pink/green), simulating early brain sampling.
Figure C-2. Differences in time to disease progression for CD8+ lymphocyte-depleted and naturally progressing macaques classified as rapid or slow progressors. The naturally progressing animals were distinguished based on time to progression relative to average (398 days post-infection (dpi)). Significant differences in the means (reported with standard deviation [±]) were determined using the Welch’s t-test. *p-value≤0.05 **p-value≤0.01.
Figure C-3. Linear regression analysis of temporal signal in maximum likelihood phylogenetic trees for longitudinally sampled naturally progressing and CD8+ lymphocyte-depleted macaques. Linear regression between root-to-tip genetic divergence (y-axis) in the RAxML-derived maximum likelihood trees (Figure 7-1) and sampling time point (x-axis) of each sequence (yellow dots) using $R^2$ values for each linear regression are given in Table 1. Longitudinal data are presented for (A) six naturally progressing macaques and (B) four CD8+ lymphocyte-depleted macaques.
Figure C-4. Percent recombination in sequences obtained from individual tissues/cell populations in naturally progressing and CD8+ lymphocyte-depleted macaques. Recombinant sequences were identified as previously described (78) and screened against viral swarm sequences in order to remove recombinants derived from initial inoculum. The percent of total sequences obtained from individual tissues/cell populations at each time point containing recombination breakpoints were averaged across macaques in each cohort. The last time point is an average of sampling times among all macaques prior to euthanization. (A) naturally progressing SIV-infected macaques. (B) CD8+ lymphocyte-depleted macaques.
LIST OF REFERENCES


40. Murphey-Corb MM LN; Rangan SRS; Baskia GB; Gormus BJ; Wolf RH; Andes WA; West M; Montelaro RC. Isolation of an HTLV-III-related retrovirus from macaques with simian AIDS and its possible origin in asymptomatic mangabeys. Nature. 1986;321(22):435-7.


BIOGRAPHICAL SKETCH

Brittany obtained her Bachelor of Science degree in biology at James Madison University, where she carried out her research in virology and neuroscience, respectively, in the labs of Dr. Louise Temple-Rose brook and Dr. Corey Cleland. Her honors thesis, as part of Dr. Cleland’s lab, focused on the development of spatial transformations in the nociceptive withdrawal response in rats, driven by her curiosity in the complexity of neural network programming in higher level organisms. During her thesis research, her fascination quickly turned to neuroinfectious diseases and pathogens, particularly viruses, which led her to Dr. Temple's lab, wherein she took part in bacteriophage characterization, or “viral discovery,” in a large effort to gain a better understanding of the genetic and mechanistic diversity of viruses. She is still captivated by this biodiversity and strives to learn more about its behavior within differing biological systems.

Brittany obtained her doctoral degree in medical science, with a concentration in biochemistry and molecular biology, in the summer of 2017. During the pursuit of her doctoral degree, she worked as a researcher in the lab of Dr. Marco Salemi in the Emerging Pathogens Institute. In addition to her work in the Salemi lab on HIV and SIV phyloanatomy in the context of neuroAIDS, she worked in collaboration with the Center for HIV RNA studies (CRNA), using a phylogenetic approach to evaluate the appositeness of biochemical methods in HIV/SIV de novo RNA secondary structure prediction. She plans to continue her work as a post-doctoral researcher in the lab of Dr. Sergei Kosakovsky Pond at Temple University in the area of molecular evolution, particularly at the intra-host level, aiming to better appreciate the immunological and biochemical selective pressures that drive the emergence of virulence.