

DISSERTATION

THE MOLECULAR ECOLOGY AND EVOLUTION OF PUMA LENTIVIRUS IN BOBCATS
AND MOUNTAIN LIONS IN NORTH AMERICA

Submitted by

Justin S. Lee

Department of Microbiology, Immunology, and Pathology

In partial fulfillment of the requirements

For the Degree of Doctor of Philosophy

Colorado State University

Fort Collins, Colorado

Summer 2013

Doctoral Committee:

Advisor: Sue VandeWoude

Co-Advisor: Kevin R. Crooks

W. Chris Funk

Jennifer L. Troyer

ABSTRACT

THE MOLECULAR ECOLOGY AND EVOLUTION OF PUMA LENTIVIRUS IN BOBCATS AND MOUNTAIN LIONS IN NORTH AMERICA

Host-pathogen dynamics are influenced by ecological and evolutionary processes at all levels of biological organization. Within individuals, viruses that cause chronic infection must either avoid or escape the pressures of the host immune system. Furthermore, viruses adapted to one host environment may have low fitness when transmitted to different individuals, populations, and species. At the landscape level, the movement and distribution of directly transmitted obligate pathogens are inextricably associated with their hosts. We used molecular analyses to investigate the ecology and evolution of feline immunodeficiency virus (FIV) in bobcats and mountain lions within individuals, among populations, and between species of hosts. In Chapter One we investigated the effects of urban development on the movement of bobcats and feline immunodeficiency virus (FIV) among a fragmented landscape in southern California. Our results demonstrate that bobcat movement and gene flow are restricted across a major freeway that bisects the study area, resulting in two genetically and physically distinct subpopulations connected by a low level of migration. However, the FIV population is not similarly structured, suggesting that movements and contacts sufficient for disease transmission continue despite the low level of host migration observed. Chapter Two investigates the causes and effects of FIV evolution among bobcats and mountain lions across North America. Our results illustrate a dynamic host-pathogen relationship characterized by host-immune pressures and a rapidly evolving virus with a highly plastic genome. Finally, in Chapter Three we describe

a pilot project aimed at improving the efficiency with which pathogen genetic data can be collected by combining the use of two modern technologies – targeted genome capture and next-generation sequencing. The results suggest this is a promising approach to detecting and sequencing multiple pathogens from biological samples. Collectively, the work described in this dissertation combines new and existing methodologies to generate, analyze, and interpret molecular data to answer complex questions about the ecological and evolutionary determinants of host-pathogen dynamics.

ACKNOWLEDGEMENTS

The work presented in this dissertation represents a large collaborative effort. Many people contributed throughout the research process either directly by contributing samples and assisting with data analyses, or indirectly through discussion, review, and constructive criticism of my scientific approach. Below is a list of the collaborators contributing to the work described in each dissertation chapter.

Chapter 1 - Emily W. Ruell, Erin E. Boydston, Lisa M. Lyren, Robert S. Alonso, Jennifer L. Troyer, Chris Funk, Kathryn Huyvaert, Kevin R. Crooks, Sue VandeWoude

Chapter 2 - Mat Aldredge, Sarah N. Bevins, Walter M. Boyce, Erin E. Boydston, Kevin R. Crooks, Ken A. Logan, Lisa M. Lyren, Roy McBride, Jill Pecon-Slattery, Seth P. Riley, Melody Roelke-Parker, Laurel E. K. Serieys, Winston Vickers, Sue VandeWoude

Chapter 3 – Anne C. Avery, Christina Boucher, Benjamin Chan, Tyler Eike, Ken A. Logan, Paul Tanger, Sue VandeWoude

I would like to thank my family, friends, and labmates for their never ending support and encouragement throughout this process. I am truly fortunate to have so many wonderful people in my life.

TABLE OF CONTENTS

ABSTRACT.....	ii
ACKNOWLEDGEMENTS.....	iv
TABLE OF CONTENTS.....	v
INTRODUCTION	1
Literature Cited.....	5
CHAPTER ONE: Gene Flow and Pathogen Transmission among Bobcats (<i>Lynx rufus</i>) in a Fragmented Urban Landscape	7
Introduction	7
Methods	10
Results	20
Discussion	25
Figures	33
Literature Cited.....	48
CHAPTER TWO: The Evolution of Puma Lentivirus in Bobcats (<i>Lynx rufus</i>) and Mountain Lions (<i>Puma concolor</i>) in North America.....	52
Introduction	52
Methods	58
Results	65
Discussion	72
Conclusion.....	82
Figures	84
Literature Cited.....	106
CHAPTER THREE: Using Targeted Genome Capture and Next-Generation Sequencing to Detect and Sequence Multiple Feline Pathogens.....	112
Introduction	112
Methods	114
Results	118
Discussion	119
Conclusion.....	122
Figures	124
Literature Cited.....	131
CONCLUSION.....	133

INTRODUCTION

Chapter One – Urbanization can result in the fragmentation of once contiguous natural landscapes into a patchy habitat interspersed within a growing urban matrix. Animals living in fragmented landscapes often have reduced movement among habitat patches due to avoidance of intervening human development, which potentially leads to both reduced gene flow and pathogen transmission between patches (Crooks & Sanjayan 2006; McDonald *et al.* 2008; McKinney 2002). Mammalian carnivores with large home ranges, such as bobcats (*Lynx rufus*), may be particularly sensitive to habitat fragmentation (Crooks 2002).

In Chapter One, we investigated the effects of urbanization on bobcat movement using genetic analyses of bobcats and their directly transmitted viral pathogen, feline immunodeficiency virus (FIV – a lentivirus in the family *Retroviridae*). We hypothesized that urban development, including major freeways, would limit bobcat movement and result in genetically structured host and pathogen populations. We analyzed molecular markers from 106 bobcats and 19 FIV isolates from seropositive animals in urban southern California. Our findings indicate that reduced gene flow between two primary habitat patches has resulted in genetically distinct bobcat subpopulations separated by urban development. However, the distribution of genetic diversity among FIV isolates determined through phylogenetic analyses indicates that pathogen genotypes are less spatially structured – exhibiting a more even distribution between habitat fragments.

We conclude that this bobcat population is structured due to low levels of effective migration resulting in gene flow. However, movements and contacts sufficient for disease transmission occur with enough frequency to preclude structuring among the viral population.

We illustrate the utility in using multiple molecular markers that differentially detect movement and gene flow between subpopulations when assessing connectivity.

Chapter Two – Understanding the evolutionary mechanisms of naturally occurring viral pathogens in the context of the host species they infect is one of the fundamental goals of molecular virology. Retroviruses, with high rates of mutation and recombination, have the capacity to evolve rapidly. At least nine feline species are infected with species-specific lentiviruses and inter-specific transmission events are rare (Troyer *et al.* 2005; Troyer *et al.* 2008). This reciprocal monophyletic relationship between pathogen and host is likely the result of viral-host coevolution and the geographic/social isolation of different host species (Pecon-Slattery *et al.* 2008). Throughout North and South America, mountain lions (*Puma concolor*) are infected with puma lentivirus B (PLVB) (Carpenter *et al.* 1996; Olmsted *et al.* 1992). A second, highly divergent lentiviral strain, PLVA, also infects mountain lions in southern California and Florida (Langley *et al.* 1994; Olmsted *et al.* 1992). Bobcats in these two geographic regions are also infected with PLVA, and to date, this is the only strain of lentivirus sequenced from bobcats (Franklin *et al.* 2007).

In Chapter Two, we characterize the forces and constraints that shape the molecular evolution of PLV in bobcats and mountain lions. We hypothesized that PLVA has coevolved with bobcats, and that mountain lions have become infected with PLVA relatively recently as the result of cross-species transmission events. We therefore predicted that PLVA isolates in mountain lions would exhibit molecular changes associated with viral adaption to a new host species.

We constructed phylogenetic trees, estimated selection pressures (positive vs. purifying), and identified recombination break points using full-length and partial genome sequence data

from PLVA and PLVB isolates across North America. We detected positive selection in three of six mountain lion PLVA isolates but none of the 20 bobcat isolates, consistent with our prediction that the virus would be subject to different selection pressures in native vs. non-native hosts. Positive selection and recombination breakpoints were detected across both PLV genomes, though the patterns of these genetic signatures differed between clades. We conclude that PLV evolution is a dynamic process, characterized by structured virus populations, high genetic diversity, and frequent recombination.

Chapter Three – The development of new strategies for the detection of existing and novel pathogens would greatly benefit infectious disease diagnoses and research (Mardis 2009). Targeted genome capture is a newly developed technology whereby rare nucleic acids within complex samples (i.e. blood) can be enriched using oligonucleotide capture probes (Summerer 2009). The result is amplification of target DNA, originally present as a minute fraction of the total nucleic acids from a sample (Kent *et al.* 2011). One of the most informative applications of targeted genome capture is the subsequent use of enriched DNA for next-generation sequencing, which can provide robust genetic sequence data for targeted nucleic acids while minimizing non-target sequences (Depledge *et al.* 2011). These technologies have many potential uses, but their application in animal health research and clinical settings remains relatively unexplored.

Domestic cats commonly exhibit clinical diseases with symptoms highly reminiscent of infectious etiologies; yet confirmatory diagnostic tests often fail to identify specific pathogen involvement. Many infectious agents cause chronic, latent infections with low replication rates, rendering them undetectable by current assays despite remaining pathogenic or transmissible (Evermann *et al.* 2012). By enriching and sequencing pathogen nucleic acids within a sample, targeted genome capture and next-generation sequencing could potentially improve upon current

diagnostic assays, and provide valuable genetic sequence data for infectious disease research. Furthermore, these technologies may facilitate the discovery of novel pathogens by allowing for the capture and sequencing of pathogens homologous to known microorganisms.

In Chapter Three, we discuss a proof-of-concept study to test the effectiveness of targeted genome capture and next-generation sequencing for gathering pathogen genetic data from biological samples. We also attempt to discover novel feline pathogens, using samples with clinical symptoms and serological results suggestive of pathogen involvement, but for which no infectious etiology has yet been identified. We hypothesized that the combined use of these technologies would enable the detection and characterization of multiple feline viral and bacterial pathogens. Sixteen samples representing domestic cats, bobcats, and pumas with known and unknown pathogen status were screened using a capture probe library designed to enrich diverse pathogen sequences. All 16 post-enrichment samples were pooled and sequenced on a MiSeq platform (Illumina Technologies, Inc., Santa Clara, CA).

We detected most of the targeted pathogens known to be present in the samples, as well as several previously undetected pathogens. The on-target genetic sequence data was sufficient to definitively characterize all pathogens detected. However, the capture assay had low specificity, and therefore, non-target genomic DNA sequences were abundant in most samples. Increasing the enrichment efficiency and utilizing deep sequencing during future applications could improve assay performance. We conclude that this methodology provides a valuable tool for clinicians and researchers to detect and characterize pathogens present in biological samples.

LITERATURE CITED

- Carpenter MA, Brown EW, Culver M, *et al.* (1996) Genetic and phylogenetic divergence of feline immunodeficiency virus in the puma (*Puma concolor*). *Journal of Virology* **70**, 6682-6693.
- Crooks KR (2002) Relative sensitivities of mammalian carnivores to habitat fragmentation. *Conservation Biology* **16**, 488-502.
- Crooks KR, Sanjayan M (2006) *Connectivity Conservation*. Cambridge University Press, Cambridge, UK.
- Depledge DP, Palser AL, Watson SJ, *et al.* (2011) Specific Capture and Whole-Genome Sequencing of Viruses from Clinical Samples. *Plos One* **6**.
- Evermann J, Sellon R, Sykes J (2012) Laboratory Diagnosis of Viral and Rickettsial Infections and Clinical Epidemiology of Infectious Diseases. In: *Infectious diseases of the dog and cat* (ed. Greene CE). Elsevier Saunders, St. Louis, MO.
- Franklin SP, Troyer JL, Terwee JA, *et al.* (2007) Frequent transmission of immunodeficiency viruses among bobcats and pumas. *Journal of Virology* **81**, 10961-10969.
- Kent BN, Salichos L, Gibbons JG, *et al.* (2011) Complete Bacteriophage Transfer in a Bacterial Endosymbiont (*Wolbachia*) Determined by Targeted Genome Capture. *Genome Biology and Evolution* **3**, 209-218.
- Langley RJ, Hirsch VM, O'Brien SJ, *et al.* (1994) Nucleotide sequence analysis of puma lentivirus (PLV-14): genomic organization and relationship to other lentiviruses. *Virology* **202**, 853-864.
- Mardis ER (2009) New strategies and emerging technologies for massively parallel sequencing: applications in medical research. *Genome Medicine* **1**, 41-44.
- McDonald RI, Kareiva P, Forman RTT (2008) The implications of current and future urbanization for global protected areas and biodiversity conservation. *Biological Conservation* **141**, 1695-1703.
- McKinney ML (2002) Urbanization, biodiversity, and conservation. *Bioscience* **52**, 883-890.
- Olmsted RA, Langley R, Roelke ME, *et al.* (1992) Worldwide Prevalence of Lentivirus Infection in Wild Feline Species - Epidemiologic and Phylogenetic Aspects. *Journal of Virology* **66**, 6008-6018.
- Pecon-Slattery J, Troyer JL, Johnson WE, O'Brien SJ (2008) Evolution of feline immunodeficiency virus in Felidae: Implications for human health and wildlife ecology. *Veterinary Immunology and Immunopathology* **123**, 32-44.
- Summerer D (2009) Enabling technologies of genomic-scale sequence enrichment for targeted high-throughput sequencing. *Genomics* **94**, 363-368.
- Troyer JL, Pecon-Slattery J, Roelke ME, *et al.* (2005) Seroprevalence and genomic divergence of circulating strains of Feline immunodeficiency virus among Felidae and Hyaenidae species. *Journal of Virology* **79**, 8282-8294.
- Troyer JL, VandeWoude S, Pecon-Slattery J, *et al.* (2008) FIV cross-species transmission: An evolutionary prospective. *Veterinary Immunology and Immunopathology* **123**, 159-166.

CHAPTER ONE

Gene Flow and Pathogen Transmission among Bobcats (*Lynx rufus*)

in a Fragmented Urban Landscape

INTRODUCTION

Habitat loss and degradation are the leading causes of species declines around the world (<http://www.iucnredlist.org/>). Urbanization, an extreme form of habitat degradation, results in immediate displacement of wildlife from developed areas, followed by the increasing isolation of groups of animals confined to shrinking natural areas (McDonald *et al.* 2008; McKinney 2002). A common effect of urbanization is the fragmentation of once contiguous landscapes into smaller patches of non-contiguous habitat.

Maintaining functional connectivity, the extent to which organisms and genetic material move between habitat patches, can be essential for population persistence in fragmented landscapes (Crooks & Sanjayan 2006). Reduced functional connectivity between habitat patches can result in physically and genetically isolated subpopulations prone to inbreeding and to the loss of genetic diversity through genetic drift (Frankham 2006). However, measuring functional connectivity can be difficult, especially for cryptic solitary species such as large carnivores (Crooks 2002). One commonly used method of evaluating functional connectivity involves characterizing patterns of gene flow using molecular markers to evaluate the distribution of genetic diversity within and among groups of individuals (Balkenhol & Waits 2009; Frankham 2006; Ruell *et al.* in press). Populations with high connectivity should exhibit homogenous distributions of genetic diversity. Populations with low connectivity will be genetically

structured, exhibiting localized variations in genetic diversity as a consequence of reduced gene flow among isolated groups of individuals.

Microsatellites are neutral heritable molecular markers commonly used to evaluate genetic structure, and hence connectivity, among natural populations (Awise 2004; Hedrick 2005a). These polymorphic, codominant markers provide a powerful means for assessing gene flow. However, microsatellite markers only reflect individual movement within structured populations if a migrant is sampled, or if a migrant successfully reproduces and at least one offspring is sampled. Transient movements between subpopulations, or migrants that do not reproduce, may not be detected by analyzing microsatellites (Riley *et al.* 2006). Therefore, while powerful, host genetic markers may not accurately characterize connectivity when individual movements do not result in gene flow.

Because directly transmitted obligate pathogens are inextricably linked to their hosts, pathogens can serve as alternative or additional markers for studies of wildlife population dynamics (Liu *et al.* 2008; Nieberding & Olivieri 2007). Feline immunodeficiency virus (FIV), a retrovirus that naturally infects many felid species (Troyer *et al.* 2005; VandeWoude & Apetrei 2006), has many characteristics that render it potentially useful as a marker of population dynamics in wild cats. First, the mutation rate of FIV is significantly faster than that of host genetic markers (approximately $\mu = 1$ to 3% every 10 years in mountain lions) (Biek *et al.* 2003). Also, infection with FIV is life-long since an obligatory step in viral replication involves the permanent insertion of a copy of the viral genome into the host's chromosomal DNA. Therefore, FIV genotypes have the capacity to serve as life-long molecular markers for each infected individual. Furthermore, the virus cannot be transmitted by insect vectors, nor is it stable in the environment, and thus transmission events are indicators of direct contact between individuals.

Finally, the strains of FIV, which infect domestic cats, are genetically distinct from the strains isolated from wild felids. The domestic cat strains have never been documented to infect non-domestic felids in the wild, and therefore, the genetic diversity of FIV among non-domestic felids is only influenced by the distribution, movement, and contact rates among conspecifics across the landscape.

Indeed, previous studies have demonstrated that patterns of FIV relatedness closely reflect the geographic distribution of bobcats (*Lynx rufus*), mountain lions (*Puma concolor*), and African lions (*Panthera leo*) at various geographic scales (Antunes *et al.* 2008; Biek *et al.* 2006; Franklin *et al.* 2007a). Specifically, Franklin *et al.* (2007a) demonstrated that the FIV isolates infecting bobcats north and south of Los Angeles, CA are genetically distinct, having diverged since the isolation of the two host populations. Because of these characteristics, viral genetic analyses may provide novel and powerful techniques for assessing connectivity and population structure with improved resolution, supplementing that which is currently possible using host genetic markers.

Bobcats, with large home ranges and high resource requirements, are susceptible to the effects of habitat loss and fragmentation in urbanizing systems (Crooks 2002; Riley *et al.* 2006; Riley *et al.* 2003; Riley 2010). We investigated patterns of genetic diversity among bobcats and FIV isolates from a fragmented landscape in southern California to evaluate how urbanization affects connectivity among bobcats in this region. We specifically evaluated the extent to which several large freeways, and the developed areas surrounding them, are barriers to gene flow and pathogen transmission between habitat patches. We predicted that decreases in connectivity would lead to significant genetic structure among both host and pathogen populations. We performed standard population genetics analyses using 16 unlinked microsatellite loci from 106

bobcats. We also constructed phylogenetic trees to assess patterns of relatedness among FIV isolates from 19 of these individuals infected by the virus.

Our findings, presented below, demonstrate that two spatially structured, genetically distinct bobcat subpopulations exist on either side of a large freeway. However, the pathogen phylogeny revealed no association between FIV relatedness and bobcat population structure, suggesting movements, contacts, and disease transmission between subpopulations continue despite very low levels of host gene flow. This finding was inconsistent with our hypothesis that decreased gene flow among bobcats would lead to similar genetic structure among the virus population. However, these results are consistent with other studies that have found major roads and urban development to be more permeable to transient bobcat movements than to effective migration, allowing for the potential movement of pathogens in the absence of gene flow (Riley *et al.* 2006; Ruell *et al.* in press). We thus illustrate the utility in using multiple molecular markers, each with different determinants of movement throughout populations, to assess complex questions of connectivity.

METHODS

Location and Field Sampling

This study was located south and west of Los Angeles, CA and included four habitat patches divided by three large freeways – Interstate 5 (I-5), the Riverside Freeway (SR-91), and the San Joaquin Hills Transportation Corridor (SR-73) (Figure 1.1). I-5 through this region was constructed from 1944 to 1958 and has an Average Annual Daily Traffic volume (AADT) of approximately 262,000 cars per day (California 2009). SR-91 was originally completed in 1971 and underwent a major expansion in 1995. The AADT of this freeway is approximately 264,000 cars per day (California 2009). SR-73 was constructed in 1996 and has an AADT of

approximately 73,000 cars per day (California 2009). Aside from these roads and adjacent urban development, the only potential natural barrier to bobcat movement is the Santa Ana River, which flows about 100 meters to the north of SR-91. No other natural barriers to gene flow (i.e., major mountain ranges) exist between these habitat patches, which are primarily characterized by chaparral, coastal scrub, and grassland vegetative communities.

A total of 106 bobcats were included in this study. The majority of bobcats ($n = 75$) were live-captured between December 2002 and March 2009 using wire cage traps baited with visual and odor attractants (Lyren *et al.* 2008a; Lyren *et al.* 2008b; Lyren *et al.* 2006). Animals were anesthetized and blood samples were collected. Animals were captured, sampled, and released with permission of cooperating agencies after approval by all appropriate animal care and use committees. The remaining individuals ($n = 31$) were opportunistically sampled post mortem; when possible, heart blood clots, thoracic fluid, ear punch, and hair samples were collected from these bobcats. Blood and tissues were stored at USGS facilities in Irvine, CA and aliquots were sent to Colorado State University for analysis as described below.

Putative Subpopulation Assignments

Bobcats were assigned to one of four putative subpopulations based on the GPS coordinates of the capture or road kill location relative to freeways I-5, SR-91, and SR-73 (Fig. 1.1). These three freeways were the focus of our investigation because they completely traverse the study area, (i.e., animals cannot move between habitat fragments without crossing one of these freeways) and these roads represent a gradient of permeability to bobcat movements based on previous studies (see discussion and also Lyren *et al.* 2008a & Lyren *et al.* 2008b). The San Joaquin Hills west subpopulation [SJH-west ($n = 29$)] was located between the coast and SR-73, which merges with I-405 to the north and I-5 to the south. The San Joaquin Hills east

subpopulation [SJH-east (n = 20)] included individuals sampled between SR-73 and I-5. The North Irvine Ranch [NIR (n = 44)] subpopulation comprised individuals captured east of I-5 and south of SR-91, whereas the Chino Puente Hills [CPH (n = 12)] subpopulation consisted of individuals captured east of I-5 and north of SR-91. One individual did not have a recorded capture location and thus it was not included in analyses that required the above subpopulation assignments. The following population genetics analyses were, except where noted, performed using this *a priori* assignment of individuals to putative subpopulations and the multilocus microsatellite genotype data for each individual.

Genotyping Microsatellites

Bobcat genomic DNA was extracted from whole blood, peripheral blood mononuclear cells (PBMCs), or tissue using the QIAamp® DNeasy blood and tissue kit (QIAGEN Inc., Valencia, CA). Seventeen microsatellite loci (Table 1) were amplified using primer pairs for polymerase chain reaction (PCR) developed by (Menotti-Raymond *et al.* 1999): FCA008, FCA023, FCA026, FCA031, FCA043, FCA045, FCA077, FCA090, FCA096, FCA132, FCA149, FCA559; (Menotti-Raymond *et al.* 2005): FCA740, FCA742; and (Faircloth *et al.* 2005): BCE5T, BCD8T, BCG8T. Primers BCE5T and BCG8T were modified to contain the M-13 sequence instead of the CAG sequence as published. We selected these primer pairs from the above publications based on the following criteria: longer repeat units, efficiency of amplification, and maximal heterozygosity. The 5' end of the forward primer of each primer pair was modified with a 16 base pair tail comprising the M-13 sequence (5' – GTA AAA CGA CGG CCA G – 3'). Reverse primers were not modified. All microsatellite PCR products were fluorescently labeled using a second forward primer consisting of the above M-13 sequence with 6-FAM on the 5' end.

PCR methodologies were adapted from (Boutin-Ganache *et al.* 2001) and (Riley *et al.* 2006). PCR reaction conditions included 94°C for 3 min followed by 22 cycles of [94°C for 30 sec; 59°C for 30 sec; 72°C for 45 sec], followed by 10 cycles of [94°C for 30 sec; 53°C for 30 sec; 72°C for 45 sec], and a final cycle of 72°C for 10 min. Randomly selected PCR products, as well as the negative control for each reaction, were visualized under UV light using gel electrophoresis with ethidium bromide in 2% agarose gel to confirm the presence of amplicons of appropriate length. Precise PCR product fragment lengths were determined using an ABI 3730xl DNA Analyzer, and Peak Scanner 1.0 software (Applied Biosystems, Foster City, CA). PCR and genotyping were repeated in ten percent of randomly chosen bobcat samples for each microsatellite locus to confirm genotypes and prevent scoring errors. All duplicated genotypes were consistent with the primary analysis, demonstrating a high degree of assay reproducibility and reducing the likelihood of genotyping errors due to false alleles or allelic dropout.

Validating and Characterizing Microsatellite Data

Microsatellite data were screened for genotyping errors due to stuttering, null alleles, and large allele dropout at all loci in Microchecker 2.2.3 (Van Oosterhout *et al.* 2004). There was no evidence of errors due to stuttering or large allele dropout at any loci. The following three loci showed evidence of null alleles when testing across all individuals with no subpopulation information: FCA045, FCA090, and FCA132. The null alleles at these loci correspond to NIR (FCA045 and FCA090) and SJH-west (FCA132) when the same analysis was run with *a priori* subpopulation assignments.

Tests for linkage disequilibrium among loci were performed in GenePop 4.0 (Raymond & Rousset 1995; Rousset 2008). The results for linkage disequilibrium varied greatly between subpopulations with seven significant tests in SJH-west (120 total tests; $\alpha = 0.009313$), four

significant tests in SJH-east (120 total tests; $\alpha = 0.009313$), two significant tests in NIR (136 total tests; $\alpha = 0.009102$), and one significant test in CPH (104 total tests; $\alpha = 0.009567$). Loci FCA077 and FCA043 were in linkage disequilibrium in all four putative subpopulations and therefore FCA043 was not included in further analyses.

Hardy-Weinberg (HW) equilibrium probabilities were calculated using GenePop 4.0 (Raymond & Rousset 1995; Rousset 2008). Three of four subpopulations significantly deviated from HW equilibrium at one unique locus: FCA023 in SJH-west (15 tests; $\alpha = 0.015068$), FCA045 in NIR (16 tests; $\alpha = 0.014790$), and BCE5T in CPH (16 tests; $\alpha = 0.014790$). The remaining loci in each subpopulation did not deviate from HW equilibrium and therefore all subpopulations were assumed to be in HW equilibrium. Observed and expected heterozygosity and the polymorphic information content (PIC) for each locus were determined using the program Cervus 3.0 (Table 1) (Kalinowski *et al.* 2007).

Assessment of Population Structure

Population differentiation based on allele frequencies was calculated for each pair of putative subpopulations using GenePop 4.0 (96 tests; $\alpha = 0.0097$) (Raymond & Rousset 1995; Rousset 2008). Allelic richness, estimated using rarefaction to avoid bias due to differences in sample size (Leberg 2002), was calculated for each putative subpopulation using Fstat 2.9.3.2 (Goudet 1995). Allelic richness results were confirmed to be normally distributed using a Ryan Joiner Test in Minitab Student Version 14.11.1 (Ryan Joiner test; $p > 0.1$). Analysis of variance (ANOVA) was used to determine whether allelic richness differed significantly between subpopulations ($\alpha = 0.05$). Estimates of subpopulation differentiation (D_{est}) were calculated using the online program Software for Measurement of Genetic Diversity (Crawford 2010; Jost 2008). F_{st} values were calculated in Fstat (Goudet 1995). The use of F_{st} values as measures of

population differentiation has recently been criticized (Hedrick 2005b; Jost 2008). Therefore, we include them here as supplemental information only, to allow a general comparison among similar, previously published studies (Table 1.2).

Bayesian clustering in program Structure 2.3.3 was used to infer the number of genetically distinct subpopulations (K), and to assign each individual to the subpopulation with which they share the highest genetic similarity. Parameters were set to include 50,000 burn-in and 500,000 Markov Chain Monte Carlo (MCMC) iterations (Pritchard *et al.* 2000). Data were first analyzed without *a priori* source population information for individuals. Independent allele frequencies among subpopulations and genetic admixture were included as parameters so as not to introduce an upward bias in the estimation of K (Pritchard *et al.* 2000). This analysis was repeated five times for each K to verify the consistency of likelihood values between runs. K was varied from K = 1 to K = 5, representing a range of greater than expected K values to ensure our analysis included all ecologically plausible values of K.

This was followed by additional analyses with the data set divided into two groups: coastal animals (SJH-west and SJH-east) and inland animals (NIR and CPH), to more closely evaluate possible substructure within each of these two groups (Pritchard *et al.* 2010). The parameters of this model were the same as above, with each analysis repeated five times for each K from K = 1 to K = 3 for each group. For all of the above analyses, posterior probability values were computed for each K according to Pritchard *et al.* (2010). Additionally, ΔK values, which have been shown to accurately reflect the actual number of genetic clusters, were calculated according to (Evanno *et al.* 2005).

Structure was also used to identify individuals that were captured in one subpopulation but genetically assigned to another, and thus, represent migrants. The parameters for this

analysis were the same as described above except subpopulation assignments were included in the analysis with the migration prior set to 0.05. Individuals with a probability of assignment to their source population ≤ 0.01 were considered migrants. Individuals with ambiguous assignment probabilities were considered hybrids.

An individual pair-wise relatedness test was performed after correction for null alleles using the program ML-Relate (Kalinowski *et al.* 2006). The average pair-wise relatedness of each subpopulation was compared using a t-test.

Detection of FIV Infection

All bobcats for which serum samples were available (n = 91) were screened for antibodies to FIV by Western blot as previously described (Franklin *et al.* 2007b). Of these, 24 (26.4%) were scored as ‘weak positive’ or ‘positive’ for FIV antibodies. PCR was used to confirm FIV infection in these samples (two to four PCR attempts per seropositive individual), using DNA extracted from whole blood or PBMCs. We used a set of degenerate nested primers, which was previously shown to amplify a region of the *RT-pol* gene from a diverse set of FIV isolates (Troyer *et al.* 2005). All bobcats that were not screened by Western blot (n=15) were screened for FIV infection by the PCR method only. In total, 19 individual bobcat FIV isolates were amplified by PCR and included in the FIV genetic analyses.

FIV *pol* and *env* PCR Amplification and Sequencing

Two gene segments were analyzed in order to evaluate viral phylogeny using both a highly conserved region (*RT-pol*, encoding the essential viral polymerase) and a region that is less evolutionarily constrained (*env*, encoding the surface envelope protein) (Pecon-Slattery *et al.* 2008b). PCR amplification of a region of the *RT-pol* gene was performed using degenerate primers as previously described (Troyer *et al.* 2005). Primers to amplify a region of the *env* gene

were designed by first performing an alignment of two previously published FIV sequences: PLV-14 (GeneBank accession #U03982) isolated from a Florida Panther (*Puma concolor coryi*), and PLV-1695 (GeneBank accession #DQ192583) isolated from a puma (*Puma concolor cougar*) in British Columbia. Degenerate nested primer pairs were designed from regions of homology including first round primers mJLenvF1 (5' – GTG CAI GTC ATI AGA TGT AGA G – 3'), and mPLVenvR7 (5' – GGG GTG TCA TTA TAA IIA GTA AAA TT – 3'), amplifying a fragment of approximately 700 base pairs, and second round primers mPLVenvF8 (5' – GGG TGC ATT IGT IAA AGA ICC ATT TTT AG – 3') and mPLVenvR6 (5' – GGT GCI TTG AAI GGA CAC ATT CC – 3'), which amplified a 570 base pair product. Underlined bases indicate 5' tail sequences added to the primers to lengthen primers and increase strength of primer binding to the template DNA.

Fifty microliter PCR reactions contained 25 µL iQSuperMix (Qiagen, Valencia CA), 400 nM of each primer, and 10 µL DNA. DNA concentrations varied among samples resulting in a range of approximately 100 ng to 500 ng template per reaction. PCR reaction conditions for both rounds included a hot start at 94°C followed by 20 cycles of melting at 94°C for 30 sec, touchdown annealing temperatures ranging from 55°C to 46°C decreasing by 1°C every 2 cycles for 30 sec, extension at 72°C for 30 sec, followed by 25 cycles of melting at 94°C for 30 sec, annealing at 52°C for 30 sec, and extension at 72°C for 30 sec with a final extension at 72°C for 3 min. This protocol successfully amplified proviral *env* fragments from three FIV-positive bobcats.

The resulting sequences were aligned and the regions of highest homology were used to develop the following nested primer pairs that successfully amplified *env* fragments from all remaining bobcats with amplified *pol* sequences (n = 16). First round primers were envfw201

(5' – TTT CTC ATG TTC CTT GAA TGG TAC – 3') and envrv202 (5' – CAC ATT CCA CTT AAT TGG TAT TG – 3'), resulting in approximately a 450 base pair amplicon. Second round primers were envfw202 (5' – TGG TAC ATT CTG GGT GTT TAA ATC – 3') and envrv201 (5' – CTA TTT TGG TCA CTC TCT GAT GC – 3'), resulting in approximately a 400 base pair product. PCR reagents and reaction conditions were the same as above with the exception that touchdown annealing temperatures ranged from 58°C to 49°C and the annealing temperature for the last 25 cycles was 54°C. PCR products were visualized under UV light using gel electrophoresis with ethidium bromide in 2% agarose gel to confirm the presence of product bands.

PCR products were purified using the QIAquick® PCR Purification Kit (QIAGEN Inc., Valencia, CA) prior to sequencing. Forward and reverse sequences were aligned using BLAST (National Center for Biotechnological Information, Bethesda, MD) and a single consensus FIV sequence was produced for each infected bobcat. All sequences were verified manually. All sequences are available in the NCBI GenBank under accession numbers JN383436 –JN383465.

Genetic Alignments and Phylogenetic Analyses

Sequences were trimmed at the 5' and 3' ends resulting in all sequences having the same length (*pol* = 427 bp, *env* = 347 bp). Trimmed consensus sequences for each gene fragment (n = 19 *pol* & *env*) were converted to coding frame using an online DNA translator tool (Swiss Institute of Bioinformatics; <http://www.isb-sib.ch/>), prior to alignment in Clustal X2 (Larkin *et al.* 2007). Alignments were input into jModeltest (Posada 2008) to estimate the best-fit model of nucleotide substitution, which was the TPM2uf model with among-site rate variation for both gene segments (Kimura 1981). The estimated model parameters used for *pol* were [Lset base = (0.4030 0.1312 0.1592) nst = 6 rmat = (7.3755 50.7381 7.3755 1.0000 50.7381) rates = gamma

shape = 0.2280 ncat = 4 pinvar = 0]. The model parameters for *env* included [Lset base = (0.3707 0.1869 0.1840) nst = 6 rmat = (4.2249 11.3706 4.2249 1.0000 11.3706) rates = gamma shape = 0.2250 ncat = 4 pinvar = 0].

Maximum-likelihood (ML) phylogenetic analyses were conducted in Phylogenetic Analysis Using Parsimony (PAUP) (Sinauer Associates, Sunderland, MA) (Swofford 2003). The corresponding *pol* and *env* gene regions from an FIV isolate sequenced from a Florida panther in 1994 were included to provide a root for each tree (Langley *et al.* 1994). ML Trees were constructed using a neighbor-joining starting tree, followed by a heuristic search using the tree-bisection-reconnection branch-swapping algorithm. The *pol* and *env* trees were found to be congruent, and therefore a single *pol-env* concatenated sequence was used to construct the final phylogenetic tree for analysis. Bootstrap analyses were performed with 100 iterations for all trees. Viral isolates were divided into four 'FIV Groups' based on clusters of related isolates arising from a basal node supported with an ML bootstrap value of 70 or greater.

We estimated the number of FIV migration events between coastal and inland bobcat subpopulations by calculating the *s* statistic for the observed ML phylogenetic tree in Mesquite 2.75 (Maddison & Maddison 2011). The *s* statistic reflects the minimum number of parsimony steps that explain the discord between subpopulations as monophyletic groups on the tree (Slatkin & Maddison 1989).

To estimate the timeline of past virus transmissions within and between subpopulations, the concatenated viral sequence data was analyzed in the coalescent framework as implemented by the program BEAST 1.6.2 (Drummond & Rambaut 2007). The SDR06 substitution model was used with a relaxed uncorrelated lognormal molecular clock (Drummond *et al.* 2006). The tree model included a piece-wise linear Bayesian Skyline prior with five groups and a randomly

generated starting tree (Drummond *et al.* 2005). An initial run of 10,000,000 MCMC iterations, sampled every 1000 runs, was performed to estimate model parameter values. The first ten percent of logged values were discarded as burn-in. From this analysis, the following model parameter priors were changed from default settings: (1) the relative rate parameters were set to vary from zero to ten; (2) the Bayesian Skyline population size was set to vary from zero to 500; (3) the mean rate for the uncorrelated relaxed molecular clock was set to vary from zero to 10.

The final analysis included the above settings with 50,000,000 MCMC iterations sampled every 1000 runs. The first ten percent of logged values were again discarded as burn-in. The estimated values and associated effective sample size (ESS) for each model parameter were viewed in Tracer 1.5 (Rambaut & Drummond, 2007). ESS values for all parameters were greater than 500. The maximum clade credibility tree was produced in TreeAnnotator 1.6.2 (Rambaut 2002). The resulting tree was viewed in FigTree 1.3.1 (Rambaut 2006), and the mean posterior probability heights with 95% highest posterior density (HPD) intervals were labeled on internal nodes.

RESULTS

Bobcat Population Structure

(1) Distribution of Alleles

The analysis of population differentiation indicated that the distribution of alleles among bobcats from the two coastal subpopulations, SJH-west and SJH-east, differed significantly ($p = 0.0064$) at only the FCA008 microsatellite locus. Therefore the coastal bobcats, regardless of whether they were captured east or west of SR-73, had a similar distribution of alleles at 15 of the 16 microsatellite loci examined. Similarly, the two inland subpopulations, NIR and CPH, significantly differed in allelic distribution at only FCA026 ($p = 0.0025$) and FCA077 ($p =$

0.0048). This finding demonstrates that bobcats captured north and south of SR-91 had a similar distribution of alleles at 14 of the 16 microsatellite loci we analyzed. However, the distribution of alleles differed greatly between the coastal and inland subpopulations. SJH-west bobcats significantly differed from NIR and CPH bobcats at 13 and 10 microsatellite loci respectively, and SJH-east bobcats differed from both NIR and CPH bobcats at 13 loci ($p < 0.01$ all significant pair-wise tests). These results indicate a high degree of genetic differentiation exists between bobcats separated by I-5 and its associated urban development.

(2) Genetic Diversity

No pair-wise difference existed in allelic richness (Table 1.3) when comparing the two coastal subpopulations ($F_{1,29} = 2.59$, $p = 0.118$) or the two inland subpopulations ($p = 0.982$, $F_{1,30} = 0.00$). However, the coastal bobcats (combined SJH-W and SJH-E) had significantly lower allelic richness than the inland bobcats (combined NIR and CPH) ($p < 0.001$, $F_{1,62} = 12.62$). This further suggests that coastal and inland bobcats are genetically distinct, and indicates that coastal bobcats have lower genetic diversity than inland bobcats. Interestingly, the BCD8T locus appears to have drifted to fixation in coastal bobcats as only one allele was sampled at this locus from all 49 coastal bobcats. Although the frequency of this allele was also high in NIR (0.82) and CPH (0.67), four and five alleles respectively were present at this locus in these subpopulations.

(3) Departures from Random Mating

The overall estimate of genetic differentiation among the four putative subpopulations, $D_{est} = 0.11$, indicates a moderate amount of genetic structure exists within this population. Pair-wise D_{est} values were lowest when comparing the two coastal subpopulations ($D_{est} = 0.008$, SJH-west:SJH-east) or the two inland subpopulations ($D_{est} = 0.014$, NIR:CPH) and highest between

the coastal and inland subpopulations ($D_{\text{est}} = 0.11$, SJH-West:NIR; $D_{\text{est}} = 0.13$, SJH-West:CPH; $D_{\text{est}} = 0.14$, SJH-East:NIR; $D_{\text{est}} = 0.19$, SJH-East:CPH). The estimated differentiation between the coastal bobcats (combined SJH-W and SJH-E) and the inland bobcats (combined NIR and CPH) was $D_{\text{est}} = 0.14$. Locus-specific D_{est} values are reported in Table 1.3.

We found no evidence of inbreeding within the overall population ($F_{\text{is}} = 0.013$; 95% CI: $-0.29 - 0.065$), nor in any of the putative subpopulations [$F_{\text{is}} = (0.011, \text{SJH-W}; -0.045, \text{SJH-E}; 0.029, \text{NIR}; 0.034, \text{CPH})$].

(4) Individual Assignment Tests

Bayesian clustering indicated the assumption of two genetically distinct subpopulations ($K=2$) best explained the variation in our microsatellite data (Figure 1.2). This result was the same regardless of whether or not source population information was used as a prior in the analysis. All bobcats caught east of I-5 (NIR and CPH bobcats $n = 56$) were assigned to one ‘inland’ subpopulation, while 46 of 49 bobcats caught west of I-5 (SJH-West and SJH-East bobcats) were assigned to a ‘coastal’ subpopulation (Figure 1.3). Three individuals captured west of I-5 were genetically assigned to the inland subpopulation and thus represent possible migrants. However, no bobcats captured east of I-5 were genetically assigned to the coastal subpopulation. If any additional human development and/or freeways in this region (i.e., SR-91, SR-73, SR-241, SR-261) were causing genetic structure, a population model with $K > 2$ should have had the highest support. Therefore, the I-5 corridor is the only human development in this region implicated as a cause of genetic structure among bobcats during our period of sample collection.

Three individuals were identified as first generation migrants as they were captured in the coastal area but had assignment probabilities of 1.00 to the inland population when capture

locations were included in the Structure analysis (denoted by * in Figure 1.3b). Three other bobcats captured in the coastal area had the genetic profile of hybrids with partial assignment to both subpopulations (denoted by # in Figure 1.3b). No individuals sampled from the inland area were identified as migrants or hybrids from the coastal area.

(5) Relatedness of Individuals

The average relatedness of bobcats in the coastal population ($R = 0.096$; $SE = 0.0037$) was significantly higher ($t_{5634} = 7.23$; $p < .001$) than in the inland population ($R = 0.064$; $SE = 0.0026$).

FIV Phylogenetic Analyses

We amplified a 427 base-pair region of *pol* and 347 base-pair region of *env* by PCR from 19 bobcats (4 in SJH-west, 4 in SJH-east, and 11 in NIR). These included 17 of the 24 (70.8%) bobcats putatively seropositive by Western blot, one bobcat that was negative by Western blot, and one bobcat that was screened by PCR only. The inability to amplify FIV sequences from a subset of putatively seropositive bobcats is similar to the findings of previous studies (Franklin *et al.* 2007b; Troyer *et al.* 2005). This is likely the result of a difference in the sensitivity and specificity of these two assays and/or a relatively low FIV proviral load present in a subset of infected animals.

The demographic information for 18 of the 19 FIV-positive bobcats was known: 13 were adult males, one was a yearling male, two were adult females, and two were yearling females. The age-specific prevalence of FIV was 14.2% (3/21) for yearlings, and 23.8% (15/63) among adults.

The ML phylogenetic tree built from *pol-env* concatenated gene sequences (Figure 1.4) shows no evidence of association between capture location (coastal vs. inland) and FIV

relatedness (see also Figure 1.5). The basal nodes have low bootstrap support, suggesting little overall genetic structuring within the virus population. The number of parsimony steps that best explains the discord of FIV relatedness between subpopulations indicates a minimum of three FIV transmission events have occurred between the coastal and inland subpopulations (Slatkin & Maddison 1989).

We estimated the past timeline of virus movement within and between bobcat subpopulations using a coalescent Bayesian model (Drummond & Rambaut 2007; Drummond *et al.* 2005). Figure 1.6 illustrates the maximum clade credibility tree with labels representing the estimated year of coalescence (mean posterior node ages) for each internal node. The ages of the basal nodes are difficult to interpret due to large 95% highest posterior density intervals (HPD - analogous to 95% confidence intervals). Therefore, it is impossible to reconstruct a precise timeline of ancestral divergence into the observed FIV groups. However, the more terminal nodes are insightful as they depict the estimated year of recent FIV coalescent events. FIV Group 4, the largest supported group of isolates, comprises six isolates from inland bobcats (x23, x24, x27, x31, x37, x55) and two isolates from coastal bobcats (x240, x46). The two coastal isolates in this group share a common ancestor with two of the inland isolates (x23 and x27) in approximately 1990 (Figure 1.6; 95% HPD: 1967 – 2003). Similarly, FIV Group 2, with 99% bootstrap support, contains nearly identical sequences from three inland bobcats (x22, x34, x35) and two coastal bobcats (x39, x44), which share a common ancestor in approximately the year 2000 (Figure 1.6; 95% HPD: 1995 – 2003).

FIV group 3 contains four closely related viral isolates (Figure 1.4; x75, x76, x78, x49) from related coastal bobcats ($r > 0.25$). To our knowledge this is the first evidence of FIV

familial transmission among bobcats. Familial transmission of FIV has been previously documented in mountain lions (Poss *et al.* 2008).

DISCUSSION

Bobcat Population Structure

Understanding the degree of connectivity among populations in heterogeneous landscapes is an important goal of ecology, population genetics, and conservation biology (Crooks & Sanjayan 2006; Fischer & Lindenmayer 2007; Taylor *et al.* 1993). We evaluated host and pathogen genetic markers to investigate connectivity among bobcats throughout a fragmented urban landscape in southern California. Our results indicate that two genetically distinct groups of bobcats existed in our study area, defined as coastal and inland subpopulations, separated by urban development including I-5 (Figures 1.1 and 1.3). This finding is in agreement with other analyses performed previously with a small subset of these bobcats and four of the 16 microsatellite loci utilized in this report (Ruell *et al.* in press).

Our results indicate that the coastal and inland bobcat subpopulations had a different distribution of alleles at most of the microsatellite loci examined and a low, unidirectional pattern of migration from the inland to the coastal area. We also observed reduced genetic diversity and increased relatedness among individuals in the coastal population. These findings suggest the observed genetic differentiation is due to decreased migration through the urban matrix between the inland and coastal habitat patches. We therefore conclude that urban development, including I-5, has been a physical barrier that has reduced bobcat movement and gene flow between isolated groups of individuals.

We did not detect substructure among bobcats separated by the two other freeways that we specifically evaluated, SR-91 and SR-73. Nor did we find evidence of genetic structure due

to any other freeways (i.e., SR-241, SR-261) or human development in the region. The maintenance of gene flow across SR-91 and SR-73 is likely explained by the fact that these roadways are perforated by more functional wildlife underpasses, and are bordered by more natural habitat than I-5. Therefore, the distance between habitat patches is shorter across SR-91 and SR-73 than across I-5, a factor that has been shown to be an important determinant in carnivore movement among habitat fragments in this region (Crooks 2002). It is possible that because SR-91 and SR-73 are newer roads, any isolation these may be causing has not yet resulted in detectable genetic structure.

Our findings, revealed by analyses of empirical genetic data, are supported by observations from remotely triggered cameras placed near underpasses of all three focal roadways. Cameras placed near the only potential wildlife corridor under I-5 that directly links SJH-E to NIR did not document any movement of bobcats between these habitat patches during 204 consecutive days of observation (Lyren *et al.* 2008a). Cameras, however, did not monitor another potential path across I-5 connecting NIR to SJH-W to the south; road kill carcasses and models of connectivity both suggest this path may be utilized by bobcats (Lyren *et al.* 2008a). In contrast to the lack of remote camera observations of bobcat movement across I-5, a similar duration of camera monitoring of potential wildlife corridors under SR-91 documented many successful movements between NIR and CPH (unpublished data). Likewise, multiple bobcat movements between SJH-W and SJH-E were also recorded during 358 days of camera observations near SR-73 (Lyren *et al.* 2008b).

Interstate-5 and the surrounding urban matrix have greatly increased in size over time since the original construction was completed in 1958. While it is difficult to know when I-5 became a 'barrier' to bobcat movement, Crooks (2002) estimated that the coastal San Joaquin

Hills might have become effectively isolated from inland natural areas around 1980. Given the generation time of bobcats (~ 2 years) (Knick *et al.* 1985), between 12 (~1980) and 25 (~1958) generations have passed since the inland and coastal bobcat subpopulations became physically isolated.

FIV Phylogenetic Analyses

Directly transmitted pathogens are inextricably linked to their hosts, and therefore the geographic distribution of pathogens reflects the movement and contacts of their hosts throughout the landscape. We analyzed two gene sequences from feline immunodeficiency virus (FIV), a retroviral pathogen of bobcats, to evaluate if the gene flow (transmission) of the virus revealed information about bobcat movement not discernable from traditional host genetic analyses. We hypothesized that urban development, which limits bobcat gene flow, would also limit viral transmission between bobcat subpopulations. Given that only one in five bobcats are infected with FIV, the expected virus migration rate is five times lower than the host migration rate. Thus we expected each of the two bobcat subpopulations would be infected with a genetically distinct viral strain, divergent from one another due to years of isolation, low viral migration rates, and the high mutation rate of FIV.

This prediction is consistent with previous literature demonstrating that geographic or social barriers, which prevent mixing between neighboring host subpopulations, result in genetically structured retrovirus populations (Franklin *et al.* 2007a; Liu *et al.* 2008; see also Figure 1.7- demonstrating clear divergence between FIVs isolated from bobcats north or south of Los Angeles). However, the phylogenetic trees constructed in this study from FIV gene segments, demonstrated a pattern that differs from these previously published findings. The relatedness of FIV isolates is mixed among coastal and inland bobcat subpopulations, indicating

that there is no association between FIV relatedness and the geographic/genetic structure of its host (Figures 1.4 and 1.5). While this finding differs from our prediction based on the ecology of FIV, it is consistent with previous studies of urban bobcats in Southern California, which have found that individual movements (and thus opportunities for disease transmission) between fragmented habitat patches occur more frequently than predicted by observed levels of gene flow (Riley *et al.* 2006; Ruell *et al.* in press).

The topology of the ML phylogenetic tree (Figure 1.4) illustrates that neither the coastal nor the inland bobcats are infected with a monophyletic cluster of viruses. Instead, we identified four groups of related FIV isolates but found no evidence to suggest these groups developed due to the population structure of bobcats. While two of the four groups contained only coastal or inland isolates (FIV Group 1 and 3), these were the smallest groups sampled and the other two FIV groups did not follow this pattern. Instead, FIV Groups 2 and 4 contained closely related viruses arising from both bobcat subpopulations. The coastal isolates within each of these two groups shared recent common ancestry with inland viruses (Figure 1.6) and thus we conclude that FIV infected bobcat migration events across I-5 are responsible for the observed mixing of isolates. Half of the viruses infecting coastal individuals (4/8) recently originated from, or were transmitted to, inland bobcats. The former is likely to have occurred in FIV Group 2 as suggested by the presence of multiple inland isolates basal to the two coastal isolates, and evidence of a long Group 2 residence time within the inland subpopulation.

Utilizing Host and Pathogen Genetics

The presence of related FIV strains on both sides of I-5 suggests that the intervening urban development is somewhat permeable to bobcat movement and disease transmission, despite the presence of distinct genetic structure among the host population. There may be

several explanations for the discrepancy between the population structure of FIV and its host. One hypothesis is that a recent increase in bobcat migration across I-5 has led to the observed mixing of viral genotypes, but not enough time has passed for this recent increase in migration to counteract previously established genetic structure among the bobcats. While possible, we consider this unlikely as human development along the I-5 corridor has increased over time, and no notable changes have been made in the area (e.g., underpasses, culverts, etc.) that would account for a recent increase in connectivity between the two subpopulations.

An alternative hypothesis is that FIV exchange between the two subpopulations is not necessarily linked to bobcat gene flow. Transient movements of individuals across a semi-permeable barrier such as I-5 may not result in the exchange of genetic material, but may involve sufficient contact between individuals to allow for disease transmission. Under this hypothesis, the contrasting patterns of population structure may reflect differences in the underlying ecology of the two molecular markers.

Microsatellite markers are useful for detecting host gene flow; the specific process involving animal movement which results in the exchange of genetic material from one group of individuals to another (Endler 1977). According to this definition, gene flow is dependent upon successful mating after migration. FIV transmission, however, can occur both vertically *and* horizontally, allowing FIV isolates to move between individuals and subpopulations in the absence of gene flow. For example, Biek et al., (2003) reported that for one population of mountain lions, horizontal transmission among adults resulted in the majority of new FIV infections and accounted for the observed increasing prevalence of FIV with age. Vertical transmission, resulting in a cohort of young individuals infected prior to adolescence, was equally important in explaining the dynamics of FIV in the population.

Given that bobcats and mountain lions share many life-history characteristics, it is likely that similar FIV transmission dynamics occur in bobcats. Indeed, the relative prevalence of FIV in yearlings (14%) versus adults (24%) in this study is similar to the age-prevalence relationship described in Biek et al., (2003). Young infected individuals may therefore play an important role in the maintenance and spread of FIV in this population. At adolescence, juvenile (usually male) bobcats often make transient movements during dispersal from their natal range while attempting to establish a new home range (Hansen 2007; Kitchings & Story 1984; Knick 1990). Our observation of familial FIV transmission in this study is relevant anecdotal support of this hypothesis. Young bobcats infected prior to dispersal therefore represent a potentially important mode of virus movement within and between subpopulations.

In fragmented landscapes such as southern California, where urban development and freeways act as boundaries limiting animal movement, bobcat home ranges may shrink, and the amount of overlap between neighboring home ranges may increase (Riley 2006; Riley *et al.* 2006). This pattern of ‘home-range pile-up’ has been described in other bobcat populations in California (Riley *et al.* 2006). This phenomenon decreases the probability that juveniles dispersing to a neighboring subpopulation will successfully mate. Under these conditions, the actual rate of bobcat movements, contacts, and opportunities for disease transmission between subpopulations would be higher than expected based on migration rates estimated from gene flow.

The low level of gene flow we detected across I-5 occurred in a unidirectional pattern from the inland area toward the coast. The coastal population, while reduced in overall genetic diversity, contains both migrants and hybrids from the inland population. This suggests that, while rare, inland bobcats can successfully migrate to, and breed in, the coastal population. The

reverse was not seen. However, inland FIV isolates do not form a monophyletic group; two of the three FIV groups infecting coastal bobcats also infect inland bobcats. There are two possible explanations to this pattern. One is that movement of FIV has occurred repeatedly one-way from the inland subpopulation to the coast, causing the diversity among coastal isolates to closely mirror that observed inland. This hypothesis is consistent with the one-way pattern of bobcat movement we detected from the microsatellite analysis.

The other possibility is that a coastal virus population diverged from inland viruses after the two groups were isolated by urban development and has since been transmitted back into the inland subpopulation multiple times via transient movements and contacts resulting in disease transmission but not gene flow. Tracking the movements of individual bobcats, evaluating FIV diversity in other neighboring bobcat populations, and utilizing spatially explicit phylogeographic analyses may help to distinguish between these alternatives.

In conclusion, our findings indicate that bobcats inhabiting this fragmented landscape in southern California are physically isolated and genetically structured. This pattern is consistent with decreased connectivity across urban development, resulting in low levels of migration and/or a low probability that migrants establish a home range and successfully mate. However, movements are apparently made, allowing for disease transmission between these habitat fragments. This conclusion carries with it conservation implications as populations with these characteristics are susceptible to decline due to a continued loss of genetic diversity from genetic drift and decreased individual fitness due to inbreeding depression (Frankham 2006). In addition, both subpopulations should be managed as a unit when considering treatment and prevention interventions during future disease outbreaks. Habitat conservation and restoration, as well as connectivity enhancements such as functional underpasses to better facilitate

movement of bobcats under roadways may help safeguard their persistence in the face of substantial ongoing threats posed by humans in this region.

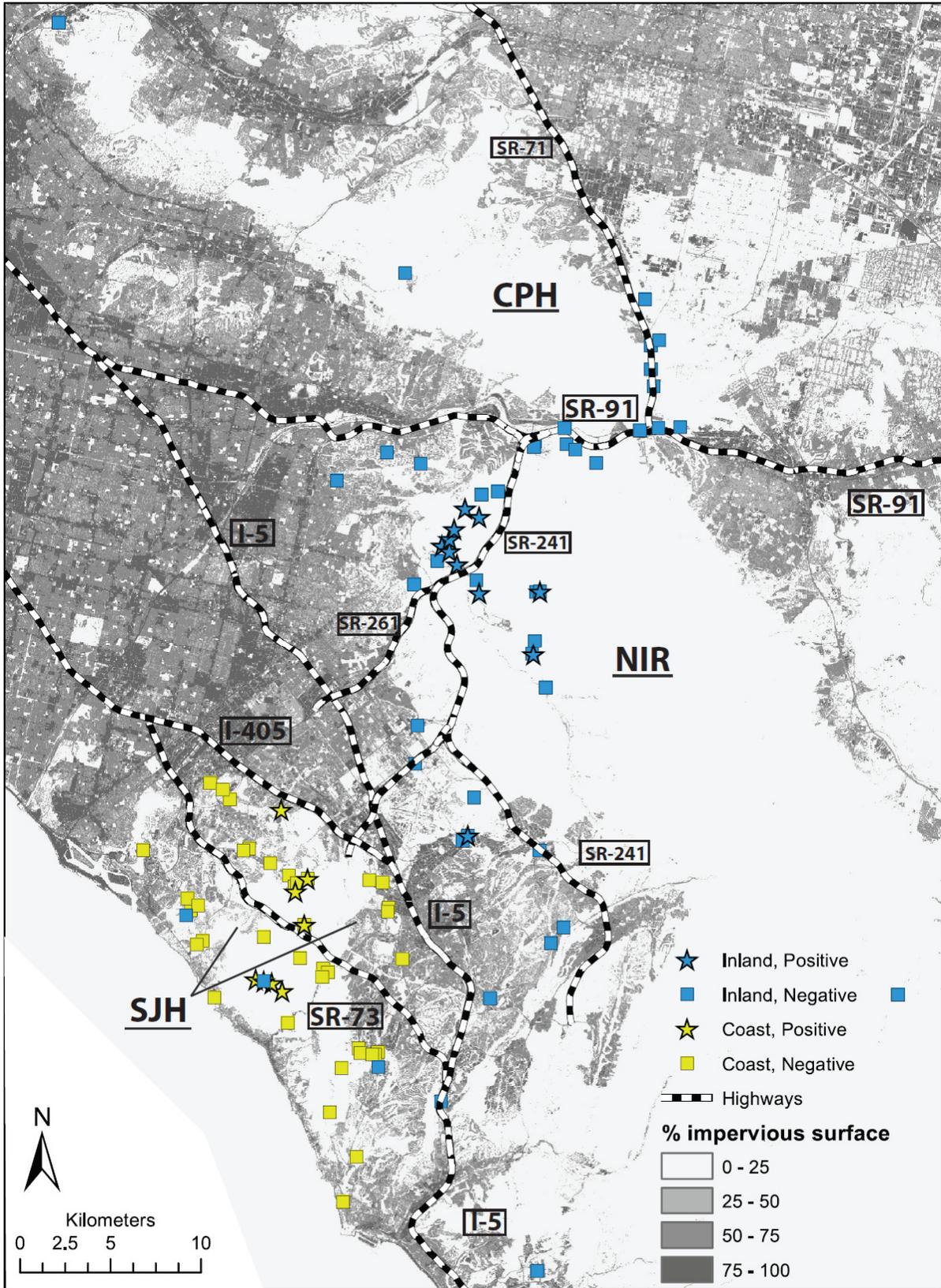


Figure 1.1: Map of study locations south-east of Los Angeles, California. GPS capture locations of 105 bobcats indicated by squares (FIV negative) or stars (FIV positive). Blue or yellow fill of each symbol represents the genetic assignment of an individual to an inland (east of I-5), or coastal (west of I-5) subpopulation, respectively (see Figure 1.3). The *a priori* assignment of individuals to putative subpopulations (SJH-W: San Joaquin Hills West; SJH-E: San Joaquin Hills East; NIR: North Irvine Ranch; CPH: Chino Puente Hills) was based on the capture location of bobcats relative to the focal roads SR-73, I-5, and SR-91, which were investigated as potential barriers to gene flow. Degree of urban development (% impervious surface) is represented by gray shading.

Table 1.1: Characterization of 17 microsatellite loci in 106 bobcats.

Locus	Size Range	# alleles	Repeat	Chromosome	Ho	He	PIC	Reference
FCA008	140-156	8	di	A1	0.71	0.77	0.73	1
FCA023	144-158	6	di	B1	0.67	0.72	0.67	1
FCA026**	138-166	13	di	D3	0.79	0.83	0.81	1
FCA031	237-255	8	di	E3	0.78	0.80	0.77	1
FCA043*	131-139	5	di	C2	0.78	0.73	0.68	1
FCA045**	147-173	7	di	A1	0.63	0.83	0.81	1
FCA077*	130-140	6	di	C2	0.77	0.74	0.70	1
FCA090**	108-126	7	di	A1	0.41	0.52	0.48	1
FCA096	189-209	8	di	A2	0.80	0.77	0.75	1
FCA132**	182-194	7	di	D3	0.66	0.78	0.74	1
FCA149	133-149	9	di	B1	0.76	0.78	0.75	1
FCA559	115-135	6	tetra	B1	0.64	0.67	0.60	1
FCA740	333-353	6	tetra	C1	0.84	0.79	0.76	2
FCA742	104-134	7	tetra	D4	0.65	0.67	0.61	2
BCD8T	156-180	5	tetra	unknown	0.21	0.21	0.20	3
BCE5T	256-280	7	tetra	unknown	0.70	0.75	0.71	3
BCG8T	275-299	11	di	unknown	0.73	0.78	0.74	3

Ho: Observed Heterozygosity; He: Expected Heterozygosity; PIC: Polymorphic Information Content. * FCA077 and FCA 043 were found to be in linkage disequilibrium; FCA043 was not used in population genetics analyses. ** Null alleles may exist in one subpopulation.

References: (1) Menotti-Raymond et al. 1999; (2) Menotti-Raymond et al. 2005; (3) Faircloth et al. 2005

Table 1.2: Measures of genetic structure among bobcats calculated from 16 microsatellite loci. (A) - Allelic richness values for each locus reported for each study location. Allelic richness is a measure of genetic diversity standardized to compare across differences in sample size (Leberg 2002). The mean allelic richness of the coastal bobcats (SJH-W and SJH-E) is significantly lower than the inland bobcats (NIR and CPH). (B) - Estimates of subpopulation differentiation (D_{est}) for each locus with 95% confidence intervals. All D_{est} values are significantly different from zero.

Locus	(A)				(B)	
	SJH-W	SJH-E	NIR	CPH	D_{est}	95% CI of D_{est}
FCA008	6.5	4.1	5.9	5.0	0.35	0.26-0.49
FCA023	3.9	2.8	4.7	5.0	0.22	0.16-0.35
FCA026	7.0	5.3	8.4	10.0	0.37	0.29-0.54
FCA031	5.9	5.3	6.9	6.0	0.26	0.19-0.41
FCA045	6.6	6.6	6.2	6.0	0.03	0.03-0.26
FCA077	4.8	3.8	4.3	4.0	0.23	0.16-0.39
FCA090	3.0	3.2	4.8	4.0	0.03	0.02-0.09
FCA096	6.7	5.5	6.6	7.0	0.04	0.04-0.18
FCA132	4.3	3.8	6.4	7.0	0.15	0.10-0.31
FCA149	5.3	3.9	7.5	7.0	0.09	0.07-0.29
FCA559	3.0	2.6	4.2	4.0	0.16	0.12-0.27
FCA740	4.4	4.9	5.7	5.0	0.12	0.07-0.26
FCA742	4.3	3.9	4.4	4.0	0.14	0.07-0.29
BCD8T	1.0	1.0	3.2	5.0	0.04	0.01-0.12
BCE5T	4.9	3.7	5.5	5.0	0.17	0.12-0.33
BCG8T	4.8	4.2	7.6	8.0	0.26	0.17-0.41
Mean	4.8	4.0	5.8	5.8	Overall D_{est} = 0.11	
Standard Deviation	1.6	1.3	1.4	1.7	D_{est} for coastal vs. inland bobcats = 0.14	

Table 1.3: Fst values calculated from 16 bobcat microsatellite loci (Goudet 1995). p-values for pair-wise tests were obtained using 120 permutations. Significant pair-wise Fst values (indicated by *) were determined after a Bonferroni correction for multiple tests ($\alpha = 0.0083$).

Subpopulations	SJH-E	NIR	CPH
SJH-W	0.015	0.055*	0.082*
SJH-E		0.098*	0.14*
NIR			0.015
Overall Fst	0.062 (95% CI: 0.045 - 0.079)		
Fst between coastal and inland bobcats: 0.074 (95% CI: 0.054 - 0.095)			

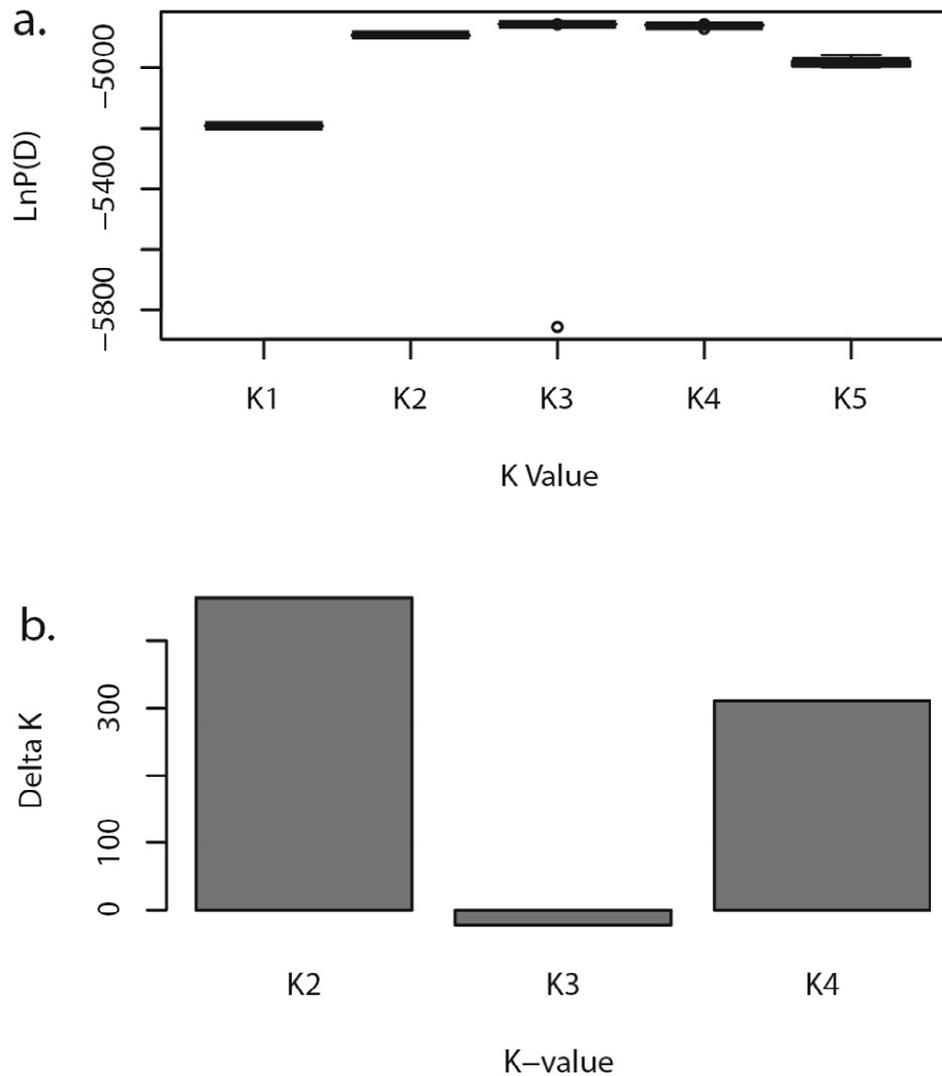


Figure 1.2: Estimating the number of genetic subpopulations (K) of bobcats. K was estimated without *a priori* information about capture location. 2a - Boxplot of likelihood values associated with each estimated number of subpopulations (K) after five simulations for each K . Horizontal lines represent mean log-likelihood values with bars (present but not always visible) showing \pm one standard deviation. 2b - ΔK values used to estimate the number of distinct genetic subpopulations. According to criteria described by Pritchard et al., 2010 (2a) and Evanno et al., 2005 (2b), $K = 2$ most accurately characterizes the structure of this population.

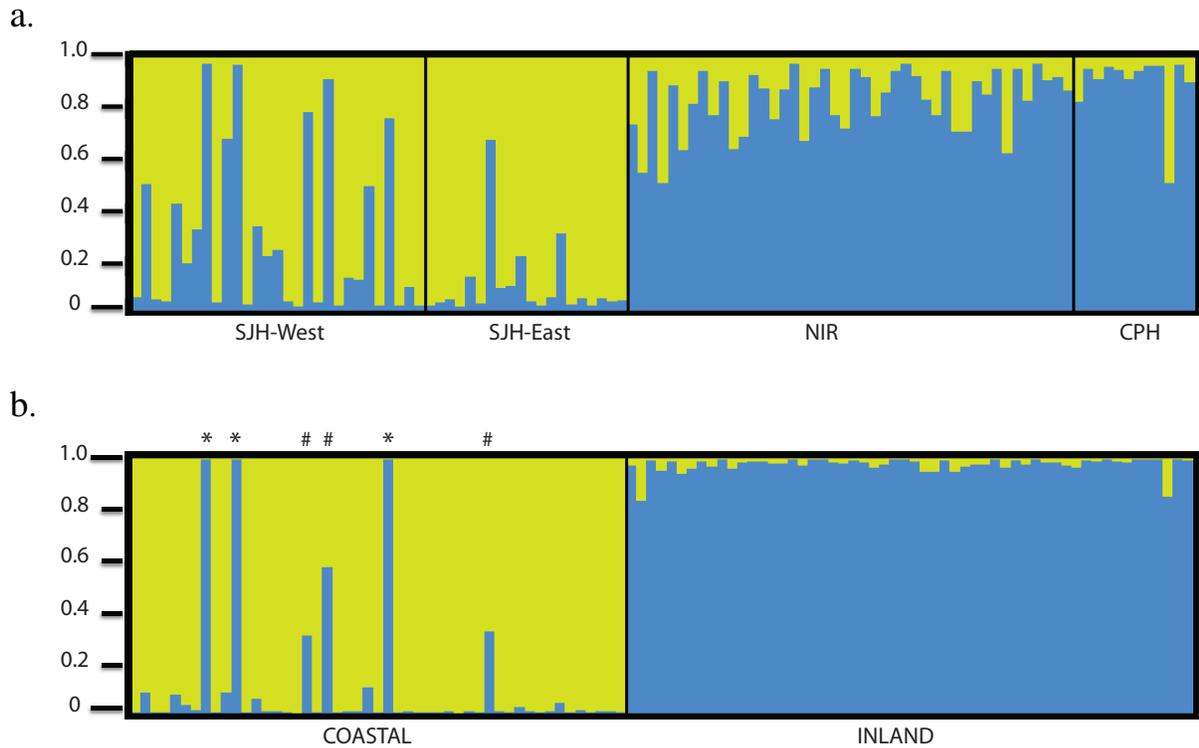


Figure 1.3: Individual bobcat genetic assignments to each of two distinct subpopulations. Each vertical bar represents one individual. Values on the y-axis are the probability of assignment of each individual to one of the two genetic groups identified. The shading of each bar corresponds to the probability of genetic assignment to either the coastal group (yellow) or the inland group (blue) of bobcats. 3a – Results for simulation of $K = 2$ without including *a priori* capture locations. The majority of individuals captured west of I-5 had strong assignment to one subpopulation (coastal - yellow), while all of the individuals captured east of I-5 had a high probability of assignment to a second subpopulation (inland - blue). 3b – Three migrants (*) and three hybrids (#) were identified in Structure using *a priori* assignments to coast or inland subpopulations based on capture locations. All gene flow was detected from the inland to the coastal subpopulation.

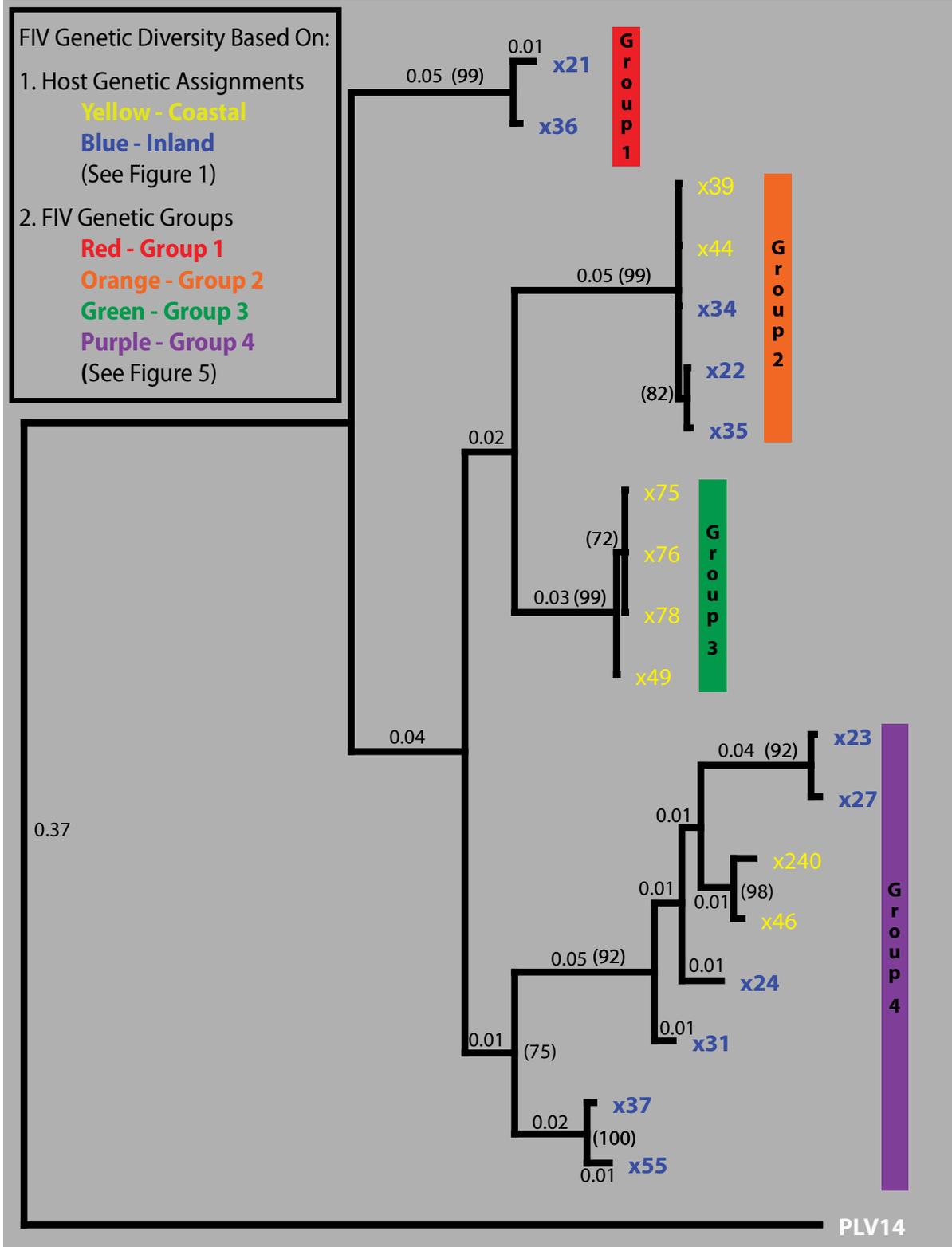


Figure 1.4: Maximum-likelihood phylogenetic tree of FIV concatenated *pol-env* sequences from 19 infected bobcats. The tree was constructed using a single 774 bp sequence (427 bp from *pol*, 347 bp from *env*) of proviral DNA from each individual. All sequences are in coding frame. Bootstrap values greater than 70 are indicated in parentheses. Branch lengths greater than or equal to 0.01 are also indicated. Individuals have been colored to represent their genetic assignment based on the results of the *Structure* analysis (see Figures 1.1 and 1.3). Because no migrants were infected with FIV, the color of an individual also represents whether an individual was captured west (yellow) or east (blue) of I-5. FIV isolates were assigned to groups based on the four most basal supported nodes, to illustrate the distribution of FIV genetic diversity across the landscape (See Figure 1.5). The tree is rooted with PLV-14, a viral sequence from the same FIV clade, which was sequenced from a Florida panther in 1994.

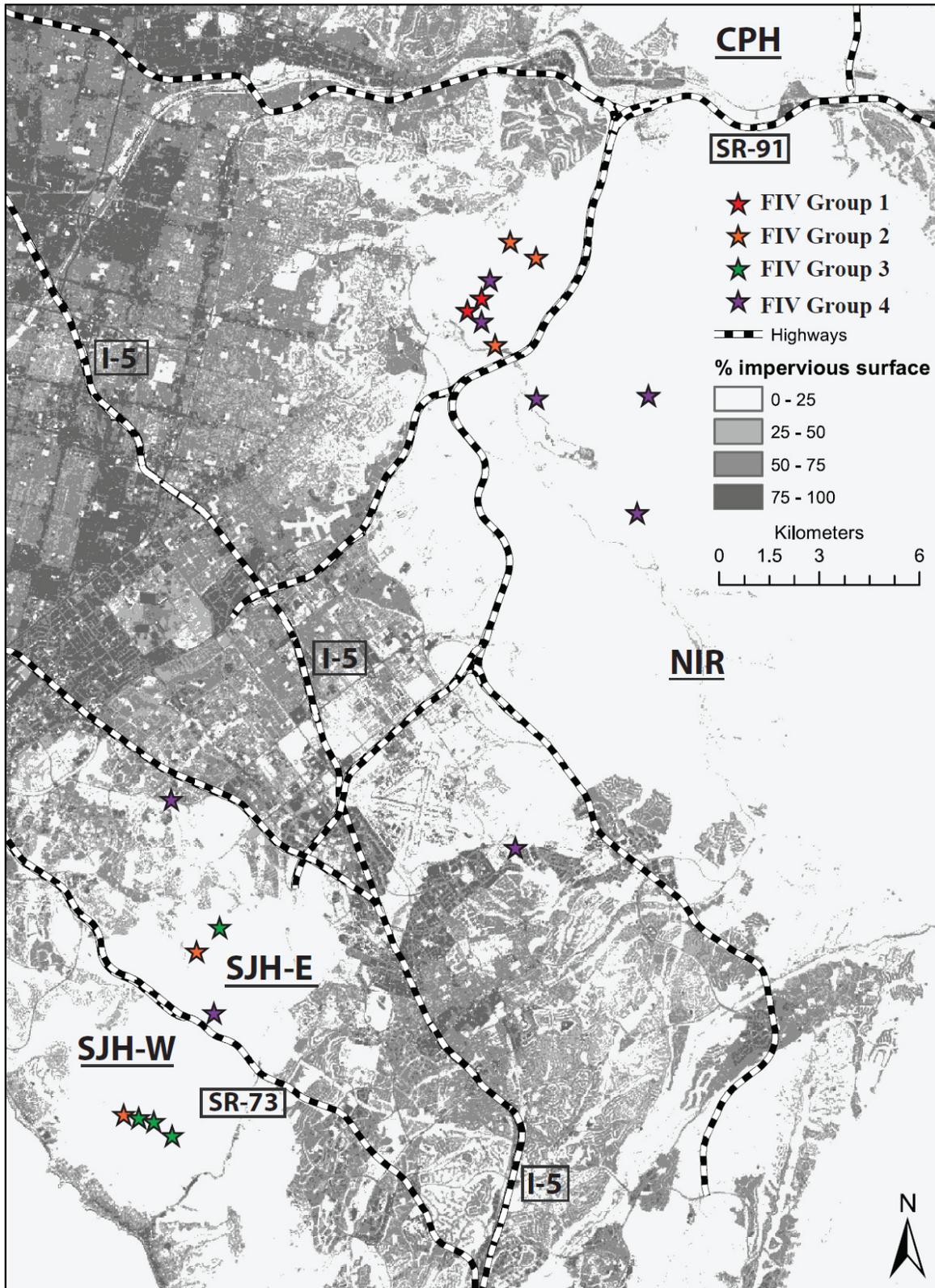


Figure 1.5: Geographic distribution of FIV genetic diversity among inland and coastal bobcats. Each star indicates the capture location of one FIV-positive bobcat. The stars are colored corresponding to which of the four groups of related FIV isolates was sampled from each bobcat (See Figure 1.4). FIV Groups 1 (red) and 3 (green) contain isolates sampled only from inland or coastal bobcats respectively. FIV Groups 2 (orange) and 4 (purple), containing both inland and coastal bobcats, resulted from the movement of FIV across Interstate-5. Degree of urban development (% impervious surface) is represented by gray shading.

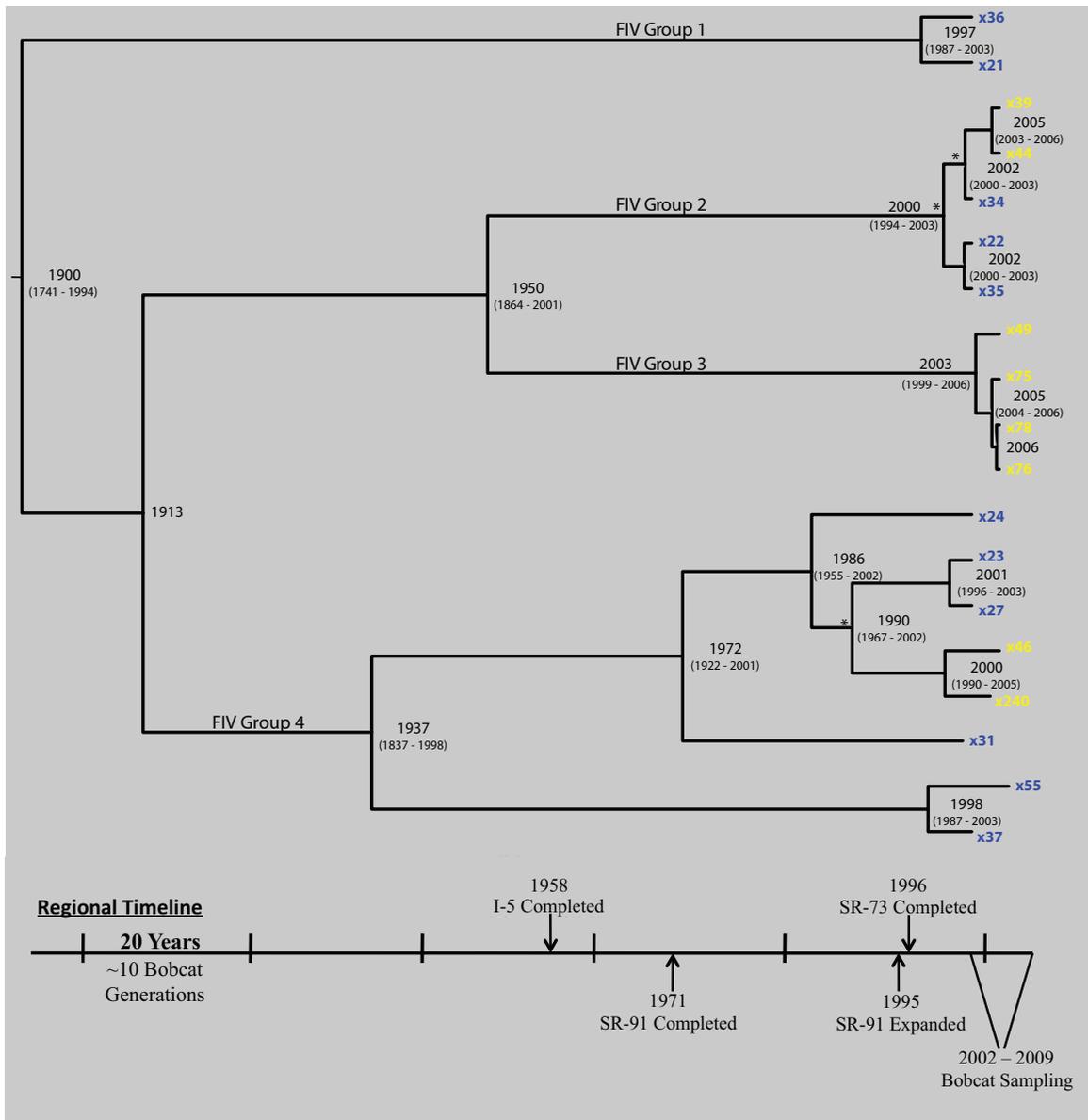


Figure 1.6: The estimated timeline of viral coalescent events. Maximum clade credibility Bayesian phylogenetic tree of *pol-env* concatenated sequences constructed using dated tips. Internal node labels correspond to the highest mean posterior probability estimate (with 95% HPD interval) for the year of coalescence. The estimated dates on basal nodes lack precision; therefore, it is impossible to accurately reconstruct the timeline of ancestral diversification into the current four-group structure of FIV. Estimates of recent coalescent events are more precise, and depict contemporary FIV transmission within and between bobcat subpopulations.

* Indicates common ancestor to contemporary isolates sequenced from both inland and coastal bobcat isolates. Viral movement across I-5 has likely occurred since the estimated date of these coalescent events, resulting in the presence of related viral strains sampled in both subpopulations (See Figure 1.5).

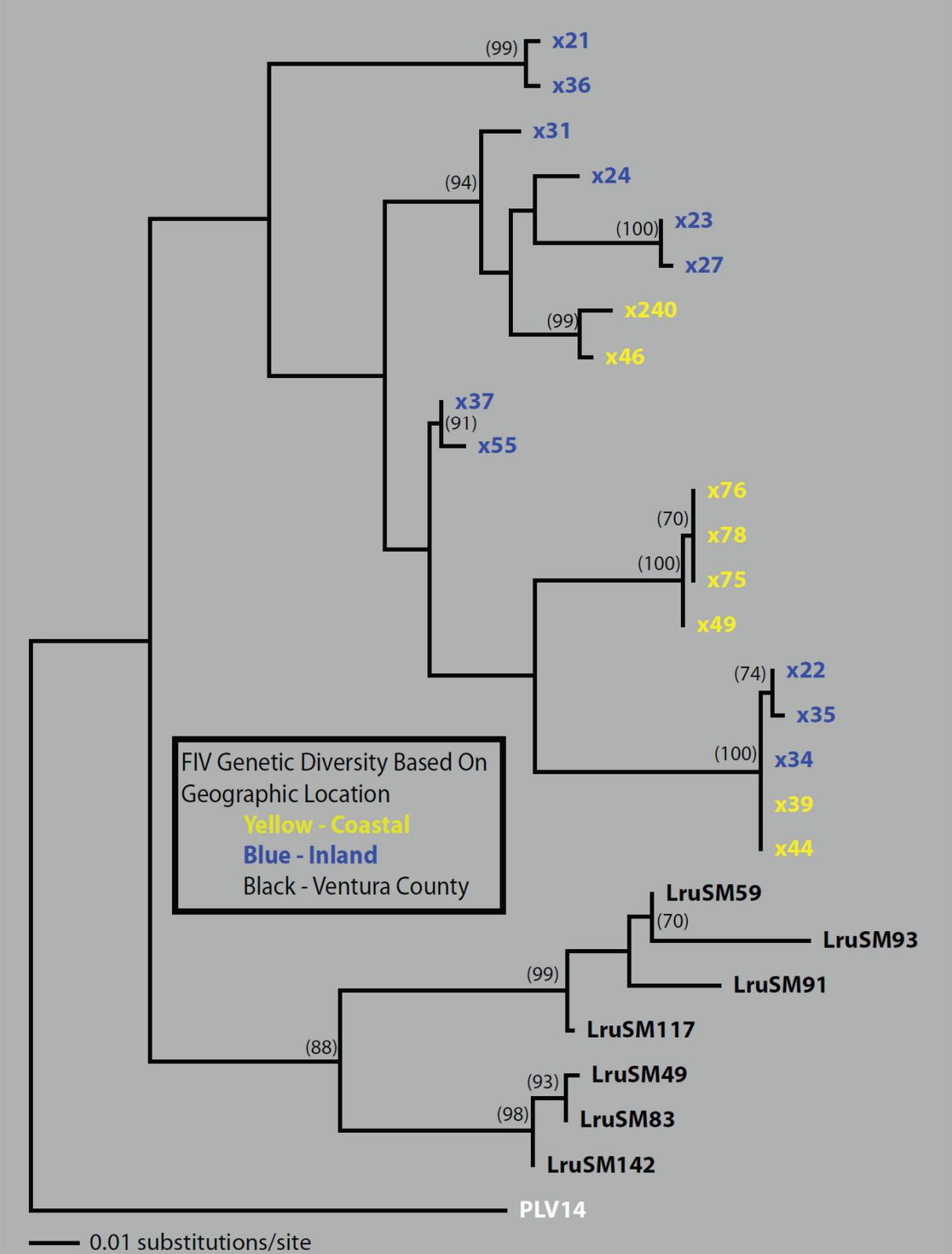


Figure 1.7: Maximum-likelihood phylogenetic tree built from *pol* sequences using the same model parameters as those described to produce the ML tree in Figure 4 (see METHODS). Bootstrap values greater than 70 are indicated in parentheses. This tree includes seven viral isolates from bobcats captured north of Los Angeles, CA, which are paraphyletic to coastal and inland isolates, illustrating a structured viral population based on geographic location of the hosts. Our initial hypothesis, based on the low level of bobcat gene flow detected across I-5, predicted that viral isolates from coastal and inland bobcats would form well-supported paraphyletic groups.

LITERATURE CITED

- Antunes A, Troyer JL, Roelke ME, *et al.* (2008) The Evolutionary Dynamics of the Lion *Panthera leo* Revealed by Host and Viral Population Genomics. *Plos Genetics* **4** (11). e1000251.
- Avise JC (2004) *Molecular Markers, Natural History, and Evolution*, Second Edition edn. Sinauer Associates, Inc., Sunderland, MA.
- Balkenhol N, Waits LP (2009) Molecular road ecology: exploring the potential of genetics for investigating transportation impacts on wildlife. *Molecular Ecology* **18**, 4151-4164.
- Biek R, Drummond AJ, Poss M (2006) A virus reveals population structure and recent demographic history of its carnivore host. *Science* **311**, 538-541.
- Biek R, Rodrigo AG, Holley D, *et al.* (2003) Epidemiology, genetic diversity, and evolution of endemic feline immunodeficiency virus in a population of wild cougars. *Journal of Virology* **77**, 9578-9589.
- Boutin-Ganache I, Raposo M, Raymond M, Deschepper CF (2001) M13-tailed primers improve the readability and usability of microsatellite analyses performed with two different allelizing methods. *Biotechniques* **31**, 24-28.
- California DOT (2009) All Traffic Volumes on California State Highway System. <http://traffic-counts.dot.ca.gov/2008all.htm>. Traffic Operations Division. Traffic and Vehicle Data Systems Unit.
- Crawford NG (2010) SMOGD: software for the measurement of genetic diversity. *Molecular Ecology Resources* **10**, 556-557.
- Crooks KR (2002) Relative sensitivities of mammalian carnivores to habitat fragmentation. *Conservation Biology* **16**, 488-502.
- Crooks KR, Sanjayan M eds. (2006) *Connectivity Conservation*. Cambridge University Press, Cambridge, UK.
- Drummond AJ, Ho SYW, Phillips MJ, Rambaut A (2006) Relaxed phylogenetics and dating with confidence. *Plos Biology* **4**, 699-710.
- Drummond AJ, Rambaut A (2007) BEAST: Bayesian evolutionary analysis by sampling trees. *Bmc Evolutionary Biology* **7** (214), doi:10.1186/1471-2148-7-214.
- Drummond AJ, Rambaut A, Shapiro B, Pybus OG (2005) Bayesian coalescent inference of past population dynamics from molecular sequences. *Molecular Biology and Evolution* **22**, 1185-1192.
- Endler JA (1977) *Geographic variation, speciation, and clines* Princeton University Press, Princeton, N.J.
- Evanno G, Regnaut S, Goudet J (2005) Detecting the number of clusters of individuals using the software STRUCTURE: a simulation study. *Molecular Ecology* **14**, 2611-2620.
- Faircloth BC, Reid A, Valentine T, *et al.* (2005) Tetranucleotide, trinucleotide, and dinucleotide loci from the bobcat (*Lynx rufus*). *Molecular Ecology Notes* **5**, 387-389.
- Fischer J, Lindenmayer DB (2007) Landscape modification and habitat fragmentation: a synthesis. *Global Ecology and Biogeography* **16**, 265-280.
- Frankham R (2006) Genetics and landscape connectivity. In: *Connectivity conservation* (ed. Crooks KR, Sanjayan M.), pp. 73 - 96. Cambridge University Press.

- Franklin SP, Troyer JL, Terwee JA, *et al.* (2007a) Frequent transmission of immunodeficiency viruses among bobcats and pumas. *Journal of Virology* **81**, 10961-10969.
- Franklin SP, Troyer JL, TerWee JA, *et al.* (2007b) Variability in assays used for detection of lentiviral infection in bobcats (*Lynx rufus*), pumas (*Puma concolor*) and ocelots (*Leopardus pardalis*). *Journal of Wildlife Diseases* **43**, 700-710.
- Goudet J (1995) FSTAT (Version 1.2): A computer program to calculate F-statistics. *Journal of Heredity* **86**, 485-486.
- Hansen K (2007) *Bobcat: master of survival* Oxford University Press, USA.
- Hedrick PW (2005a) *Genetics of Populations*, Third Edition edn. Jones and Bartlett, Sudbury, MA.
- Hedrick PW (2005b) A standardized genetic differentiation measure. *Evolution* **59**, 1633-1638.
- Jost L (2008) *G_{ST}* and its relatives do not measure differentiation. *Molecular Ecology* **17**, 4015-4026.
- Kalinowski ST, Taper ML, Marshall TC (2007) Revising how the computer program CERVUS accommodates genotyping error increases success in paternity assignment. *Molecular Ecology* **16**, 1099-1106.
- Kalinowski ST, Wagner AP, Taper ML (2006) ML-RELATE: a computer program for maximum likelihood estimation of relatedness and relationship. *Molecular Ecology Notes* **6**, 576-579.
- Kimura M (1981) Estimation of evolutionary distances between homologous nucleotide sequences. *Proceedings of the National Academy of Sciences of the United States of America* **78**, 454.
- Kitchings JT, Story JD (1984) Movements and dispersal of bobcats in east Tennessee. *The Journal of wildlife management* **48**, 957-961.
- Knick ST (1990) Ecology of bobcats relative to exploitation and a prey decline in southeastern Idaho. *Wildlife Monographs* **108**, 3-42.
- Knick ST, Brittell JD, Sweeney SJ (1985) Population Characteristics of Bobcats in Washington State. *Journal of Wildlife Management* **49**, 721-728.
- Langley RJ, Hirsch VM, O'Brien SJ, *et al.* (1994) Nucleotide sequence analysis of puma lentivirus (PLV-14): genomic organization and relationship to other lentiviruses. *Virology* **202**, 853-864.
- Larkin MA, Blackshields G, Brown NP, *et al.* (2007) Clustal W and clustal X version 2.0. *Bioinformatics* **23**, 2947-2948.
- Leberg PL (2002) Estimating allelic richness: Effects of sample size and bottlenecks. *Molecular Ecology* **11**, 2445-2449.
- Liu WM, Worobey M, Li YY, *et al.* (2008) Molecular ecology and natural history of simian foamy virus infection in wild-living chimpanzees. *Plos Pathogens* **4** (7): e1000097.
- Lyren LM, Alonso RS, Crooks KR, Boydston EE (2008a) Evaluation of Functional Connectivity for Bobcats and Coyotes across the Former El Toro Marine Base, Orange County, California. *Administrative Report*, 179.
- Lyren LM, R. S. Alonso, K. R. Crooks, Boydston EE (2008b) GPS Telemetry, Camera Trap, and Mortality Surveys of Bobcats in the San Joaquin Hills, Orange County, California, U.S. *Geological Survey Administrative Report*, 132.
- Lyren LM, Turschak GM, Ambat ES, *et al.* (2006) Carnivore activity and movement in a southern California protected area, the North/Central Irvine Ranch. In: *Technical Report*.

- Department of the Interior, U.S. Geological Survey, Western Ecological Research Center, Sacramento, CA.
- Maddison WP, Maddison DR (2011) Mesquite: a modular system for evolutionary analysis.
- McDonald RI, Kareiva P, Forman RTT (2008) The implications of current and future urbanization for global protected areas and biodiversity conservation. *Biological Conservation* **141**, 1695-1703.
- McKinney ML (2002) Urbanization, biodiversity, and conservation. *Bioscience* **52**, 883-890.
- Menotti-Raymond M, David VA, Lyons LA, *et al.* (1999) A genetic linkage map of microsatellites in the domestic cat (*Felis catus*). *Genomics* **57**, 9-23.
- Menotti-Raymond MA, David VA, Wachter LL, Butler JM, O'Brien SJ (2005) An STR forensic typing system for genetic individualization of domestic cat (*Felis catus*) samples. *Journal of Forensic Science* **50**, 1061-1070.
- Nieberding CM, Olivieri I (2007) Parasites: proxies for host genealogy and ecology? *Trends in Ecology & Evolution* **22**, 156-165.
- Pecon-Slatery J, Troyer JL, Johnson WE, O'Brien SJ (2008) Evolution of feline immunodeficiency virus in Felidae: Implications for human health and wildlife ecology. *Veterinary Immunology and Immunopathology* **123**, 32-44.
- Posada D (2008) jModelTest: Phylogenetic model averaging. *Molecular Biology and Evolution* **25**, 1253-1256.
- Poss M, Ross H, Rodrigo A, *et al.* (2008) The molecular biology and evolution of feline immunodeficiency viruses of cougars. *Veterinary Immunology and Immunopathology* **123**, 154-158.
- Pritchard JK, Stephens M, Donnelly P (2000) Inference of population structure using multilocus genotype data. *Genetics* **155**, 945-959.
- Pritchard JK, Wen X, Falush D (2010) Documentation for *structure* software: Version 2.3.
- Rambaut A, Drummond AJ (2002) TreeAnnotator: MCMC output analysis.
- Rambaut A, Drummond AJ (2006) FigTree: Tree figure drawing tool.
- Raymond M, Rousset F (1995) Genepop (Version-1.2) - Population-Genetics Software for Exact Tests and Ecumenicism. *Journal of Heredity* **86**, 248-249.
- Riley SPD (2006) Spatial ecology of bobcats and gray foxes in urban and rural zones of a national park. *Journal of Wildlife Management* **70**, 1425-1435.
- Riley SPD, Pollinger JP, Sauvajot RM, *et al.* (2006) A southern California freeway is a physical and social barrier to gene flow in carnivores. *Molecular Ecology* **15**, 1733-1741.
- Riley SPD, Sauvajot RM, Fuller TK, *et al.* (2003) Effects of urbanization and habitat fragmentation on bobcats and coyotes in southern California. *Conservation Biology* **17**, 566-576.
- Riley SPDB, E. E.; Crooks, K. R.; Lyren, L. M.; in Gehrt, Stanley D., Riley, Seth P. D., Cypher, Brian L. editors (2010) Bobcats (*Lynx rufus*). Urban Carnivores - Ecology, Conflict, and Conservation. The Johns Hopkins University Press.
- Rousset F (2008) GENEPOP'007: a complete re-implementation of the GENEPOP software for Windows and Linux. *Molecular Ecology Resources* **8**, 103-106.
- Ruell EW, Riley SD, Douglas MR, *et al.* (In Press) Urban Habitat Fragmentation and Genetic Population Structure of Bobcats in Coastal Southern California. *American Midland Naturalist*.

- Slatkin M, Maddison WP (1989) A Cladistic Measure of Gene Flow Inferred from the Phylogenies of Alleles. *Genetics* **123**, 603-613.
- Swofford DL (2003) PAUP*. Phylogenetic analysis using parsimony (*and other methods), Version 4. Sinauer Associates, Sunderland.
- Taylor PD, Fahrig L, Henein K, Merriam G (1993) Connectivity Is a Vital Element of Landscape Structure. *Oikos* **68**, 571-573.
- Troyer JL, Pecon-Slattery J, Roelke ME, *et al.* (2005) Seroprevalence and genomic divergence of circulating strains of Feline immunodeficiency virus among Felidae and Hyaenidae species. *Journal of Virology* **79**, 8282-8294.
- Van Oosterhout C, Hutchinson WF, Wills DPM, Shipley P (2004) MICRO-CHECKER: software for identifying and correcting genotyping errors in microsatellite data. *Molecular Ecology Notes* **4**, 535-538.
- VandeWoude S, Apetrei C (2006) Going wild: Lessons from naturally occurring T-lymphotropic lentiviruses. *Clinical Microbiology Reviews* **19**, 728-762.

CHAPTER TWO

The Evolution of Puma Lentivirus in Bobcats (*Lynx rufus*) and Mountain Lions (*Puma concolor*) in North America

INTRODUCTION

RNA Virus Evolution

Genetic sequences can provide valuable information about the evolutionary interplay between pathogens and their hosts. RNA viruses mutate more rapidly, and therefore attain more genetic diversity, than any other organisms on the planet (Drake 1993). Characteristics of RNA viruses that contribute to this diversity are the high error-rate of the RNA polymerase enzyme, rapid replication rates, large population sizes, and the ability to recombine or reassort genomes between two or more distinct viral lineages. Because genetic diversity is the medium through which natural selection acts, RNA viruses can adapt and evolve rapidly. The sources of natural selection driving this evolution include intra-host factors, such as the innate and adaptive immune responses, inter-host pressures including transmission efficiency and differences in host biology, and environmental conditions outside the host species (Domingo *et al.* 1996).

Although high mutation rates are a mechanism that enables viral adaptation, mutations can also be costly given that the majority of non-synonymous genetic mutations negatively impact fitness. As a result, the rapid error rate that generates high levels of genetic diversity among RNA viruses also creates many deleterious mutations (Holmes 2003). The number of potentially deleterious mutations is proportional to genome length, and is likely the factor limiting the length of RNA viral genomes, which are among the shortest of all known organisms

(Eigen 1971, 1996). All necessary viral genes must be efficiently encoded in compact genomes, and the potential for genetic diversity is restricted to a confined number of nucleotides.

Consequently, RNA viral genomes often encode proteins in overlapping open reading frames and/or complex mRNA splice patterns. Mutations in these regions will often alter multiple proteins, which may result in fitness trade-offs between the various functions of the proteins affected (Baranowski *et al.* 2001). Therefore, the high genetic diversity generated by RNA viruses creates an interesting dynamic, which both increases and constrains the plasticity of viral genomes and their capacity to evolve.

Retroviruses are unique among RNA viruses, in that they utilize their reverse transcriptase enzyme to generate DNA copies of their genome (provirus). The provirus is subsequently inserted into the chromosome of infected host cells. This allows retroviruses to replicate their genome via RNA transcription and through mitotic division of infected cells. As a result, retroviruses can exploit the benefits of high viral mutation rates while simultaneously maintaining stable genomes that are subject to the low error rates of host polymerase enzymes. Proviruses therefore provide a 'safe haven' for highly fit viral genomes to persist and replicate without incurring decreases in fitness through deleterious mutations (Cavrois *et al.* 1996; Overbaugh & Bangham 2001).

Another aspect of retroviral replication that affects pathogen evolution is the common occurrence of recombination between the two genomic copies of single-stranded RNA packaged into each mature viral particle (Bruen & Poss 2007; Jung *et al.* 2002; Poss *et al.* 2007).

Recombination occurs during reverse transcription in newly infected cells, and potentially yields progeny viruses that are genetically distinct from either parental strain. Similar to high mutation rates, recombination can have large impacts on viral fitness, as it is a mechanism that can quickly

purge deleterious alleles, acquire advantageous alleles, or result in defective viruses by disrupting open reading frames (Boerlijst *et al.* 1996). The existence of many circulating recombinant forms amongst various retrovirus species is evidence that recombination is common and can generate highly fit viral lineages (Bachmann *et al.* 1997; Bailes *et al.* 2003; Bruen & Poss 2007; Moutouh *et al.* 1996; Robertson *et al.* 1995).

Virus-Host Interactions

Given the ability of retroviruses to rapidly mutate and adapt, it is interesting that many retroviruses have a narrow host range. For example, species in the genus *Lentivirus* typically infect hosts in a species-specific manner. Distinct viral subtypes or clades are generally associated with a single host species (Olmsted *et al.* 1992; Troyer *et al.* 2005, VandeWoude & Apetrei 2006). Transmission of these host-adapted viruses to new species appears to be rare (Troyer *et al.* 2008). Host restriction is likely due to many factors, including lower viral fitness in the novel host, intrinsic anti-viral defense mechanisms, and/or limited contact sufficient for transmission between different host species (Munk *et al.* 2010; VandeWoude *et al.* 2010).

At least nine species of the carnivore family *Felidae* harbor unique subtypes of feline immunodeficiency virus (FIV) (Troyer *et al.* 2005). FIV is a lentivirus that is transmitted by the exchange of infected fluids during direct contact (i.e. fighting, mating, parturition, grooming, etc.) and results in life-long infection via integration of provirus as described above. These host-specific viruses result in productive infection with little documented morbidity or mortality – except the FIV of domestic cats, which causes immune depletion analogous to human immunodeficiency virus (HIV). This host-specificity and apparent lack of significant pathology are thought to be the result of long virus-host coevolution and geographic isolation among different cat species (O'Brien *et al.* 2006; Pecon-Slattery *et al.* 2008b). Cross-species

transmissions of FIVs have been documented, but these events appear to be infrequent and propagation of the virus within the new host species is rare (Carpenter *et al.* 1996; Nishimura *et al.* 1999).

The only known exception to this reciprocal monophyletic pattern among felids and FIVs comprises two divergent FIV subtypes infecting the wild cats of North America (Franklin *et al.* 2007a). The mountain lion (*Puma concolor* – also referred to as puma, cougar, and panther), one of the world's most successful terrestrial apex predators, inhabits a geographic range from western Canada to southern Chile. Bobcats (*Lynx rufus*) are sympatric mesopredators throughout much of North America (Figure 2.1). Both species are habitat generalists but are sensitive to anthropogenic influences and have experienced regional extinctions and population subdivision due to overhunting and habitat degradation (Crooks 2002; Lee *et al.* 2012; Riley *et al.* 2006; Riley 2010).

The species of FIV that infects mountain lions and bobcats is currently referred to as puma lentivirus (PLV) or FIV_{Pco} (Pco designates *Puma concolor* – the first host species from which this virus was isolated). Puma lentivirus clade B (PLVB) infects mountain lions throughout their geographic range (Figure 2.1) (Carpenter *et al.* 1996). Therefore, PLVB may have infected mountain lions prior to their recolonization of North America after the last Ice Age (10,000 – 15,000 years ago) (Johnson *et al.* 2006). Puma lentivirus clade A (PLVA), first identified in a Florida panther (*Puma concolor coryi*), has subsequently been identified in other Florida panthers and several mountain lions in southern California (Carpenter *et al.* 1996; Langley *et al.* 1994; Olmsted *et al.* 1992). Interestingly, Franklin *et al.* (2007) identified PLVA isolates from sympatric bobcats in Florida and California. These studies demonstrate that mountain lions are infected by two highly divergent FIV subtypes, and PLVA is the only FIV

subtype documented to infect two different species in the wild (Franklin *et al.* 2007a). Franklin and Troyer *et al.* (2007) hypothesized that PLVA has coevolved with bobcats and that mountain lions are infected in these two geographic regions as a result of cross-species transmission events. This hypothesis is supported by empirical findings, which demonstrate differences between PLV virus-host relationships and those observed among other FIV subtypes. First, the genetic distance between PLVA and PLVB is similar to that separating the other species-specific strains of FIV, and is suggestive of divergent evolution in separate host species (Troyer *et al.* 2005). Second, PLVA is the only FIV that has been isolated from bobcats, and its prevalence in the populations studied suggests that it is an endemic virus (Franklin *et al.* 2007a; Lee *et al.* 2012). In comparison, PLVA is rare in mountain lions – it is not present throughout most of this species' geographic range and it is much less common than PLVB in regions where the two viruses co-circulate (Carpenter *et al.* 1996; Franklin *et al.* 2007a; Lee *et al.* 2012).

These characteristics of PLVA infection in pumas are reminiscent of the emergence of HIVs via cross-species transmissions of simian immunodeficiency viruses (SIVs) from African primates (reviewed in Sharp & Hahn 2011). Similar to the host-virus relationship of FIVs, many old world primates are infected with species-specific lentiviruses. Cross-species transmissions have been documented, but most transmissions occur intra-specifically. Viral genome organization, phylogenetic relatedness, and epidemiological linkages support the hypothesis that SIVs from sooty mangabeys (*Cercocebus atys*) and chimpanzees (*Pan troglodytes troglodytes*) were transmitted to humans and became HIV-2 and HIV-1 group M respectively (Gao *et al.* 1999; Hirsch *et al.* 1989; Sharp *et al.* 1994). The HIVs diversified into many subtypes after becoming established in humans – high levels of genetic diversity, frequent recombination, and rapid spread through human populations all contributed to the adaptation of HIV in the human

host (Li *et al.* 1988; Sharp *et al.* 1995; Wain *et al.* 2007). Together, these genetic analyses have contributed substantially to our understanding of the ecological and evolutionary aspects of HIV emergence, and provide a context for the analyses described herein.

While several studies have described the phylogenetic relationships among FIVs (including PLV), most of these have been based solely on nucleotide sequence alignments from a short, conserved region of the *pol* gene. Little sequence data has been generated from other regions of the FIV genome, with the exception of domestic cat FIVs that are typically compared using a short region of the *env* gene. Recent publications describing phylogenetic relationships among full-length sequences of lion lentivirus (LLV) isolates across Africa, and PLVB isolates from central North America, demonstrate that complex evolutionary histories cannot be resolved through single gene analyses (Bruen & Poss 2007; Burkala & Poss 2007; Poss & Ross 2008; Rigby *et al.* 1993; Slattery *et al.*, 2008). To date, only one full-length PLVA sequence has been published so no genomic evolutionary studies have been conducted on this clade (Langley *et al.* 1994).

Here we describe the genetic diversity and molecular evolution of PLVA and PLVB using new and existing full-length viral sequence data. We describe in detail: (1) phylogenetic relationships among PLVs in bobcats and mountain lions; (2) selection forces and constraints that shape PLV evolution; (3) patterns of recombination and its effect on genetic diversity, and; (4) positive selection and viral adaptation in different hosts. This is the first robust analysis of lentiviral genome evolution in the context of natural cross-species transmission outside of studies on HIV and SIV. Therefore, this is an important contribution toward furthering our understanding of virus-host dynamics, natural selection, and adaptation of retroviruses in natural host populations.

METHODS

Sample Collection and DNA Extraction

Mountain lion and bobcat samples were collected from natural populations in three US states (Figure 2.1). Sampling locations included: (1) two sites in coastal southern California (CA-south and CA-north), adjacent to one another but divided by the metropolitan area of Los Angeles; (2) two sites in Colorado (CO-west, and CO-east), separated by the Rocky Mountains; and (3) one site in Florida (FL). Published PLV sequences from mountain lions in Wyoming (WY), Montana (MT), British Columbia (BC), and FL were also included in our analyses (Table 2.1).

All samples were collected from live, free-ranging animals captured using baited cage traps or scent-trained tracking hounds, as previously described (Bevins *et al.* 2012). Animals were chemically sedated for blood collection. All capture and animal handling protocols followed approved Animal Care and Use Committee guidelines, and where applicable, local government regulations. Aliquots of blood samples were sent to Colorado State University for PLV diagnosis by serology and sequence characterization as described below.

DNA was extracted from whole blood or peripheral blood mononuclear cells (PBMCs) using the standard DNeasy® Blood and Tissue protocol (Qiagen Inc., Valencia, CA). Some PBMC samples were cultured *in vitro* (up to 14 days) to propagate the virus prior to DNA extraction, as previously described (Pecon-Slattey *et al.* 2008a). Two of the viral sequences derived from cultured cells were compared to sequences derived from primary PBMCs in order to identify genetic changes arising as an artifact *ex vivo*. No significant changes were observed in either case. Table 2.1 provides the sex, age, location, and collection date for the samples included in this study.

PCR Amplification, Sequencing, and Assembly of Viral Genomes

We designed a standardized set of polymerase chain reaction (PCR) primers and protocols to amplify and sequence diverse viral isolates from each PLV clade. PLVB primers were designed to target conserved regions from an alignment of published PLVB genomes (Bruen & Poss 2007). Because only one PLVA genome has been published (PLV14), the initial PLVA primers were designed to anneal to conserved areas identified from an alignment of this sequence with several PLVB sequences (Langley *et al.* 1994). The remaining PLVA primers in this study were designed using the following iterative process: (1) align existing sequences in Clustal-W using default parameters (Larkin *et al.* 2007); (2) visually identify conserved regions in the alignment; (3) design primers that have a high probability of binding to diverse viral isolates while minimizing primer self-complementarity; (4) amplify and sequence new viral isolates. These steps were repeated until enough sequences were compiled such that primers could be designed to consistently amplify highly divergent PLVA genomes. The majority of primers designed by this method contained degenerate bases. All potential primer sequences were input into the online program OligoCalc v.3.26 to identify and discard sequences likely to form problematic secondary structures, such as self-dimers and hairpins (Kibbe 2007). The genome coordinates and sequences for all PLVA and PLVB primers are listed in Table 2.2 and a map of primer locations is depicted in Figure 2.2.

PLVA and PLVB genomes were amplified using three subtype-specific overlapping nested PCR reactions (Table 2.3). For the samples that failed to amplify by one of these protocols, alternative primer combinations were tested until the complete protein coding-region of the viral genome was successfully amplified (i.e. from the start of *gag* to the end of *env*). The 5'- or 3'-end of some PLVA isolates from mountain lions in California and Florida failed to

amplify by any primer pairs – only partial genomes were sequenced for these individuals (Table 2.1).

All PCR reactions were performed with Platinum *Taq*® DNA Polymerase High-Fidelity (Invitrogen Inc., Carlsbad, CA), using the manufacturer-suggested protocol modified to include twice the recommended units of enzyme. Five microliters of DNA (100 ng to 250 ng) were used as the template for each first round PCR reaction. Second round PCR reactions used 2 ul of first-round product as the template. All PCR reactions began with touchdown cycles, in which the annealing temperature (T_a) decreased from the optimal upper primer T_a to the optimal lower primer T_a in 0.5°C increments each cycle. These conditions were followed by additional cycles of amplification at the optimal lower T_a for a total of 40 cycles per reaction. Optimal primer T_a values were defined as three to five degrees below the melting temperature (T_m) determined by the Nearest-Neighbor calculation in OligoCalc (Kibbe 2007). The extension time in each PCR was adjusted for product length at a rate of one minute/kilobase, as recommended by the manufacturer (Invitrogen Inc., Carlsbad, CA).

PCR products were separated by size with agarose gel electrophoresis and visualized with ethidium bromide. Specific PCR products were purified using the QIAquick® PCR Purification Kit (Qiagen Inc., Valencia, CA). If products contained multiple bands, the band at the correct fragment length was extracted and purified using the QIAquick® Gel Extraction Kit (Qiagen Inc., Valencia, CA). PCR products were sequenced on an ABI 3130xL Genetic Analyzer (Applied Biosystems Inc., Foster City, CA) using internal forward and reverse primers spanning approximately every 500-700 base pairs (Table 2.2 and Figure 2.2). To minimize sequencing errors, we sought to achieve at least 2x-sequencing coverage at all positions in each viral genome. However, small regions in some viral genomes with high quality chromatograms had

single coverage. All viral sequences will be deposited in GenBank, the open-access genetic sequence database (Benson *et al.* 2011).

All chromatogram files were manually screened to ensure bases were scored correctly. Individual sequence files were trimmed and assembled into longer, contiguous sequences using default settings in Geneious Pro 5.6.4 . A single consensus sequence was generated for each sample using default settings in Geneious. Intra-host variation was observed in many individuals by the consistent presence of two or more chromatogram peaks at a single position across multiple sequencing reads. These positions were scored as a single base if the ‘highest quality’ option for handling ambiguous bases in Geneious resolved the ambiguity. Otherwise, the position was scored as ambiguous, using the appropriate degenerate character to indicate which bases co-occur at that site in a given sample.

Sequence Analysis

The open reading frame (ORF) predictor in Geneious, and previously annotated sequences, were used to identify the location of each ORF (Langley *et al.* 1994; Poss *et al.* 2006). Only the *gag*, *pol*, *env*, and *vif* ORFs could be identified with confidence in all PLVA and PLVB genomes (see RESULTS), and therefore, the following analyses were performed on only these four genes. Putative cleavage sites within precursor proteins (*gag*, *pol*, and *env*) were determined from conserved amino acid motifs identified in protein alignments with annotated FIV sequences (Burkala & Poss 2007; Elder *et al.* 1993).

Each consensus sequence was trimmed into the four ORFs encoding *gag*, *pol*, *vif*, and *env*. Only completely sequenced ORF’s (i.e. start codon to stop codon) were analyzed from partially sequenced viral isolates. Codon alignments and translated amino acid sequences were constructed for each ORF using Muscle (Edgar 2004) with default parameters. All translated

sequences contained intact ORFs without premature stop codons. To characterize the levels of intra- and inter-clade genetic diversity at the protein level, pair-wise identity and the proportion of invariant sites were estimated from amino acid alignments . Throughout the remaining text, proteins will be referred to in plain text (i.e. Gag), and ORFs will be written in italics (i.e. *gag*).

Recombination Breakpoints

Codon alignments for each viral ORF were uploaded into Data Monkey, the online server of the HyPhy package (Delpont *et al.* 2010b). Recombination breakpoints within each ORF were identified using the genetic algorithm for recombination detection (GARD) with a general discrete distribution of sites into two rate classes (Pond *et al.* 2006). Breakpoints identified by improved Akaike information criterion (AIC) values were only considered significant if subsequently supported by a Kishino-Hasegawa test result indicating significant topological incongruence between the trees constructed from the regions right and left of the breakpoint ($\alpha = 0.05$). This is a conservative approach since recombination does not always change phylogenetic relationships (Wiuf *et al.* 2001). An initial run of GARD with each full-length ORF was performed. Due to time limitations on the open access server, most analyses were stopped before the full search for all potential breakpoints was complete. Therefore, each alignment with at least one significant breakpoint was divided at a breakpoint and rerun in GARD as two shorter alignments. This process was repeated until all significant breakpoints were identified and/or GARD could run to completion without finding additional breakpoints.

We performed two separate analyses to identify recombinant viruses. First, the PLVA and PLVB codon alignments were separated into segments between the recombination breakpoints identified in the GARD analysis. The evolutionary history of each region of the genome was estimated by constructing a consensus neighbor-joining (NJ) tree from 1000

bootstrap replicates in Geneious. The reference sequence for PLVA (PLV14) and PLVB (PLV1695) were included in the other clade's alignment, and used to root each NJ tree. NJ input parameters included the Hasegawa, Kishino, and Yano model of nucleotide substitution and 50% minimum consensus threshold for tree clusters.

Second, we analyzed the entire viral protein-coding region in the Recombination Detection Program v.3 (RDP3) (Martin *et al.* 2010). This software package applies multiple algorithms to genetic alignments in order to locate breakpoints and identify recombinant sequences. Putative recombination signals identified by RDP3 were considered significant if at least three statistical tests were in agreement (Bonferroni correction $p \leq 0.05$), and at least one of these tests were based on detecting topological incongruence (i.e. Bootscan).

Phylogenetic Analyses

A Bayesian phylogenetic tree of PLVA and PLVB sequences was constructed from a 1439bp alignment of a non-recombinant region of *pol* using the software Bayesian Evolutionary Analysis Sampling Trees v1.6 (BEAST) (Drummond & Rambaut 2007). The input file was compiled using the Bayesian Evolutionary Analysis Utility v1.6 (BEAUti). Model parameters included dated tips, the SDR06 model of nucleotide substitution, and a piecewise-linear coalescent Bayesian skyline tree-model prior (Drummond *et al.* 2005). An initial run of BEAST with 10M Markov chain Monte Carlo (MCMC) iterations sampled every 1000 iterations was performed using default parameter settings in BEAUti. The first 10% of samples were discarded as burn-in. Estimated model parameters were viewed in Tracer v1.5 (Rambaut & Drummond 2007), and the results were used to modify the following model priors for the final analysis: relative rate parameters (0-10); relaxed-clock mean rate (0-10); skyline population size (0-10,000); gamma shape parameters (0-10); transition-transversion initial value (codon positions 1

& 2 = 4; codon position 3 = 14). The final analysis ran for 20M MCMC iterations, which provided sufficient sampling for model convergence as all posterior parameter estimates were supported by Effective Sample Size (ESS) values greater than 200. A maximum clade credibility (MCC) tree with median node heights was constructed from the sampled trees using TreeAnnotator v1.6. Time to the most recent common ancestor (TMRCA) estimates with 95% highest posterior density (HPD) intervals, and posterior probability support values were visualized in FigTree v 1.3. The reported TMRCA estimates are likely biased due to the time-dependency inherent in substitution rate estimates based solely on samples collected in shallow time (Ho *et al.* 2005). Therefore, while we present median node age estimates from the BEAST analysis, the actual TMRCA values are likely between the median and the upper HPD estimates.

Evolutionary Hypothesis Testing

Specific hypotheses about the evolutionary history of PLV in bobcats and mountain lions were tested using analyses available in Data Monkey (Delpont *et al.* 2010a). The mixed effects model of evolution (MEME) was used to screen all ORFs for amino acid residues under positive selection ($dN > dS$; $\alpha = 0.05$) (Kosakovsky *et al.* 2011). MEME has been shown to be more sensitive at detecting positive selection than fixed effects methods, yet has similar Type I error rates (Kosakovsky *et al.* 2011). The fixed effects likelihood analysis (FEL) was used to identify amino acid residues under negative selection ($dN < dS$; $\alpha = 0.05$) (Pond & Frost 2005). Neighbor-joining trees for all of the non-recombinant genome fragments identified in GARD were used as input in these analyses. The branch-site random effects likelihood analysis (BSR) was used to identify specific viral lineages diversifying under positive selection (i.e. those containing a proportion of sites with $dN > dS$ ($\alpha = 0.05$)) (Kosakovsky *et al.* 2011).

RESULTS

Viral Sequences

Twenty-five full-length PLVA isolates (20 from bobcats; 5 from mountain lions), and 20 PLVB isolates (all from mountain lions) were sequenced (Figure 2.1; Table 2.1). All PLVA mountain lion isolates with full-length sequences were sampled in CA. Seven additional PLVA mountain lion isolates were partially sequenced (four from CA and three from FL). Thirteen bobcat isolates from CA and seven from FL were sequenced. Of the PLVB isolates sequenced, 11 were from CA and nine were from CO. PLVA was not amplified from 80 bobcats and 30 mountain lions from CO screened for PLV using primers that amplify both viral clades (Lagana *et al.* in review; unpublished data).

Although we planned to analyze the phylogenetic and evolutionary characteristics of all viral proteins, we were not able to analyze OrfA and Rev. We did not identify an ORF in the PLVA sequences that was consistent with the location and length of the *orfA* gene of other FIVs (Chatterji *et al.* 2002; Poss & Ross 2008). Several possibilities for the length, location, and organization of *rev* in PLV have been posited, but not empirically proven (Poss & Ross 2008). The potential variation in *rev* length due to multiple possible splice sites is large and would render analyses speculative at best.

Phylogenetic Relationships

Most of the genetic distance within the isolates evaluated in this study exists between PLV clades A and B, suggesting a long, separate evolutionary history since these lineages diverged from a common ancestor (Figure 2.3A). Each clade has subsequently diversified into several distinct viral lineages, which vary in geographic distribution.

Within PLVA, most of the viruses sampled belong to one of three strongly supported clusters based on the geographic region from which they were collected (FL, CA-south, CA-north) (Figure 2.3B). These three contemporary lineages coalesce approximately 879 years ago (95% HPD: 348 - 2089 years). Most of the viruses sampled in CA belong to one of two clusters that were predicted to coalesce 471 years ago (95% HPD: 192 -1129 years). One virus (Pco7), isolated from a mountain lion in CA-south, clusters generally with the other CA isolates, but is paraphyletic to both CA-south and CA-north groups. This isolate is estimated to coalesce with the other CA viruses 625 years ago (95% HPD: 249 – 1488).

The PLVA CA-south cluster contains viral isolates from three mountain lions and eight bobcats. Within CA-south, there are distinct viral lineages that have diversified from a common ancestor approximately 323 years ago (95% HPD: 123 - 764 years). This cluster contains more genetic diversity than the other PLVA groups, which may reflect a sampling bias as it comprises a larger sample size than the other clusters. However, the CA-south sampling area is also the largest from which PLVA isolates were collected, and thus, this level of diversity may represent viral population structure within this region. The clustering of isolates into sub-groups representing the northern and southern parts of the CA-south supports this possibility. Only one isolate (Pco6) differs from this trend as it was sampled in the southern part of CA-south, but clusters with viruses from the northern part of the sampling area.

The CA-north isolates are estimated to coalesce 195 years ago (95% HPD: 74 - 466). Within this group, most of the genetic diversity is represented by a single divergent isolate (Lru2). The other five viruses, which include four bobcat and one mountain lion isolates, exhibit little diversity. Mountain lion isolates cluster with bobcat isolates in both CA sites,

recapitulating earlier phylogenetic comparisons based on a shorter, partially overlapping fragment in *pol* (Franklin *et al.* 2007a).

The FL cluster is highly divergent from the viruses sampled from animals in CA. Six of the seven newly sequenced FL isolates were sampled within relatively close proximity to one another. These viruses exhibit very little divergence, suggesting a relatively homogeneous PLVA population circulates in this area. Lru15 was sampled from a captive bobcat approximately 150 kilometers away from the other FL isolates (Figure 2.1), and thus, the genetic distance between this virus and the rest of the Florida cluster may reflect geographically distinct viral subpopulations that circulate in Florida. These two lineages were estimated to coalesce approximately 217 years ago (95% HPD: 82 – 517). PLV14 was adapted to a domestic cat cell line *in vitro* prior to sequencing (Langley *et al.* 1994), and thus the genetic distance and TMRCA estimated from this isolate to the rest of the FL viruses cannot be interpreted directly. It is included only as a reference for comparison with previously published PLV studies.

The PLVB viruses from CA-south and MT form monophyletic groups. Most of the other PLVB isolates exhibit less-strict geographic clustering. The single PLVB isolate from CA-north clusters with the reference sequence from BC. The Colorado cluster comprises seven isolates from CO (n = 3 CO-west; n = 4 CO-east) and two isolates from WY-west. Within this cluster, isolates sampled east and west of the Rocky Mountains are intermixed, suggesting that a relatively unstructured virus population circulates on both sides of the Continental Divide. The Wyoming cluster comprises seven isolates in three distinct subgroups sampled in four different areas: WY-west (n = 4), WY-east (n = 1), CO-west (n = 1), and CO-east (n = 1).

Two divergent PLVB clusters (Colorado and Wyoming) co-circulate in the central US mountain lion populations (Figure 2.3C), as do recombinant viruses that have arisen from these

lineages (Figure 2.4). The Colorado cluster shares ancestry with the viruses circulating in MT and CA-south. Among this group of clusters the CA-south and Colorado clusters coalesce prior to the coalescence of the Montana cluster, and thus, the Montana cluster appears to be ancestral to both the CA-south and Colorado clusters. However, this relationship changes when phylogenetic relationships are estimated from other genomic regions (Figure 2.5). The Wyoming cluster shares no significant ancestry with the Colorado cluster in this gene region, and the TMRCA estimate for the basal node suggests they diverged 587 years ago (95% HPD: 279 – 1167). The basal node that estimates the divergence of the PLVB lineage comprising the CA-south, Colorado, and Montana clusters from the CA-north/BC cluster has low posterior support. Hence, the pattern and timing of divergence into these lineages cannot be estimated from this gene region.

Genetic Diversity and Natural Selection

Two parameters of protein genetic diversity (pair-wise identity and invariant amino acids) were evaluated from amino acid alignments of all PLVA and PLVB sequences. The pair-wise identity of each alignment reports the percentage of pair-wise sites with identical amino acid residues. This value accounts for insertions and deletions as gap vs. non-gap comparisons are included. Pair-wise identity values vary inversely with protein diversity. Invariant amino acids represent monomorphic sites in the alignment and are considered to be the result of strict purifying selection. The percentage of these fixed residues is generally highest in regions critical for protein function, such as receptor-ligand and catalytic binding domains. The genetic diversity among PLV proteins, as characterized by these two parameters, is briefly described below (detailed results are presented in Table 2.4 and Table 2.5).

Gag

Among both PLVA and PLVB translation products, Gag has the highest pair-wise identity and largest proportion of invariant sites (Table 2.4). Consistent with this, 53% of the amino acids in Gag are under purifying selection, while less than 1% are under positive selection. Among the products of gag proteolytic cleavage, the capsid, matrix, and nucleocapsid proteins have increasing levels of genetic diversity within both PLV clades (Table 2.5 and Figure 2.4). Capsid is the most conserved protein for both PLVA and PLVB, with the highest percentage of both invariant residues and sites under negative selection. No sites in capsid are under positive selection in either PLV clade – this is the only PLVA protein analyzed without at least one site under positive selection. The genetic diversity in matrix is intermediate to capsid and nucleocapsid, as are the proportion of sites under positive and negative selection. Nucleocapsid is the most variable protein in gag, with approximately 30% more variant-sites than capsid. Nucleocapsid is also the region of gag with the highest percentage of sites under positive selection, and the most length variation due to insertions and deletions (indels). These differences among the gag proteins illustrate that amino acid homology and selection pressures can vary greatly among proteins translated from a single transcript.

Pol

The translated Pol polyprotein is cleaved by the viral protease enzyme into five enzymes – protease (PR), reverse transcriptase (RT), RNaseH, dUTPase (DU), and integrase (IN) (Figure 2.4). Additionally, a small peptide (pre-PR) is produced from the 5' end of the protein upon the N-terminal cleavage of PR. Pol exhibits intermediate levels of genetic diversity compared to the other PLV translation products (Table 2.4). Most of the Pol proteolytic cleavage products vary only slightly in pair-wise identity and the percent of invariant sites (Table 2.5). The exception to this is the pre-PR peptide, which has the lowest pair-wise identity and percent of invariant sites

of any PLV gene-region analyzed. Pol exhibits relatively little variation in length considering it is the longest open reading frame in the genome. Most of the observed length variation in both PLVA and PLVB Pol results from truncated IN proteins due to premature stop codons in some isolates. More sites in pol are under positive selection in PLVA than PLVB. Of the six Pol protein domains, three in PLVA (pre-PR, PR, RNaseH) and two in PLVB (Pre-PR, RT) have more than one percent of amino acids under positive selection.

Env

The viral protease enzyme cleaves the Env polyprotein into the leader (L), surface (SU), and transmembrane (TM) proteins. Env exhibits high levels of genetic diversity, with less than 60% of residues being invariant within each PLV clade (Table 2.4). Similarly, a relatively large proportion of Env residues are under positive selection and a small proportion of residues are under negative selection (Table 2.5). All env protein domains contain indels, many of them being unique to individual viral isolates (alignments not shown). The L region is the least conserved of the Env cleavage products for both PLV clades, having only slightly more than 50% invariant residues. This domain has dual functions; it targets the *env* mRNA to the rough endoplasmic reticulum for translation and acts as the first exon of *rev* (Knipe 2007). SU has a high proportion of sites under purifying selection and intermediate levels of positive selection compared to L or TM. Between viral clades, the PLVA SU and TM have similar or slightly higher levels of genetic diversity than PLVB. In light of this, it is interesting that SU and TM in PLVB have a higher proportion of residues under positive selection and lower proportion of residues under negative selection than in PLVA. This suggests that the marginally higher PLVA genetic diversity is the result of neutrally evolving sites rather than higher positive selection in this clade.

Vif

The diversity and selection characteristics of Vif vary between PLVA and PLVB more than any other protein (Table 2.4). Among PLVA viruses, Vif has no indels and is similar to pol in both pair-wise identity and proportion of invariant sites. However PLVB vif has higher levels of genetic diversity (similar to that observed in PLVB Env), and varies in length more than any other protein analyzed. The length variations in PLVB Vif are due to additional residues at the 3' end of the protein in four isolates previously described by (Poss & Ross 2008).

Diversifying Viral Lineages

To investigate whether PLVA is under different selection pressures in bobcats than in mountain lions, we examined all non-recombinant genomic regions for evidence of positive selection in any of the PLVA isolates ($dN > dS$). In contrast to identifying specific residues evolving under positive selection (above), this analysis provides insight into which viruses are diversifying under positive selection. Evidence of positive selection was found in at least one viral protein in each of three mountain lion isolates (Table 2.6). No bobcat isolates were estimated to be diversifying under positive selection. Many internal nodes (theoretical ancestral viruses) also showed evidence of positive selection in one or more gene regions (results not shown). Of these, the Florida lineage was under positive selection in at least one region of all four genes (*gag*, *pol*, *vif*, and *env*).

Recombination

Recombination is common within both PLVA and PLVB clades, though the distribution of recombination breakpoints differs greatly between the two clades. The PLVA isolates contain ten breakpoints distributed relatively uniformly throughout the viral genome (Table 2.5 and Figure 2.4). In contrast, the PLVB breakpoints occur almost exclusively in the 3'-half of the

genome, with only one of nine breakpoints located in the 5'-half. This finding is in agreement with a previous study of PLVB that analyzed the sequences from WY and MT (Bruen & Poss 2007).

We were able to identify recombinant viruses with significantly incongruent phylogenetic signals among 11 PLVA and 10 PLVB trees, which were constructed from each non-recombinant genome region. Many putative recombination events identified by this method and through analyses in RDP3 could not be confirmed due to a lack of resolution amongst trees constructed from many of the short non-recombinant regions. One tree from each PLV clade is depicted in Figure 2.5 to illustrate the influence of recombination on genetic diversity (full tree set is not shown), but additional analyses will be required to definitively identify all of the recombinant isolates sampled.

DISCUSSION

We analyzed genetic diversity, natural selection, and recombination in four viral proteins (Gag, Pol, Vif, and Env) of puma lentivirus clade A and clade B isolates. The purpose of this study was two-fold: first, we sought to characterize the evolutionary history of each PLV clade; second, we investigated the hypothesis that PLVA has co-evolved with bobcats and infects mountain lions in two geographic regions as the result of cross-species transmission events see (Franklin *et al.* 2007a). Our analyses included bobcat and mountain lion PLV isolates from three geographic areas across North America – southern California, Colorado, and Florida. Additionally, we included all previously published PLVB sequences, which were sampled from mountain lions in Wyoming, Montana, and British Columbia. Thus, the sequences included in our analyses span the known range of each PLV clade in North America - PLVB has been isolated from mountain lions throughout western North America, and PLVA has been isolated

from bobcats and mountain lions in southern California and Florida (Biek *et al.* 2003; Carpenter *et al.* 1996; Olmsted *et al.* 1992; Troyer *et al.* 2005).

Previous to this work, only one full-length PLVA sequence had been published (Langley *et al.* 1994), and therefore, this is the first detailed characterization of PLVA genome diversity. It is interesting that PLVA appears to lack a single intact gene region encoding the OrfA protein. All lentiviruses encode a transactivator of proviral translation, which is considered essential for viral replication (Knipe 2007). OrfA, the transactivator protein of feline lentiviruses, is thought to be translated in PLVB from a single unspliced open reading frame (ORF) overlapping the *vif* and *env* genes (Figure 2.2) (Poss & Ross 2008). It is possible that the *orfA* gene is present in PLVA, but that it is translated from a spliced RNA or an internal ribosome entry site as occurs in other lentiviruses (i.e. HIV/EIAV) (Knipe 2007). In either case, this difference in genomic organization between PLVA and PLVB may have resulted from, or contributed to, the divergence of these clades. If indeed *orfA* is absent and is not necessary for productive PLVA infection, investigating proviral transcription in the absence of this protein may reveal unique mechanisms of lentivirus replication.

Because the FIVs of non-domestic cats cause little to no known pathogenicity, it has been posited that co-evolution has resulted in ‘stable’ virus-host relationships (O'Brien *et al.* 2006). Supporting this idea, previously published studies of PLVB concluded that *pol*, *env*, and *vif* evolve under purifying or neutral selection (Biek *et al.* 2003; Franklin *et al.* 2007a). Only the nucleocapsid domain of *gag* has been found to be under positive selection (Burkala & Poss 2007). However, our results reveal that positive selection is an important force driving the evolution of PLVA and PLVB proteins. Some amino acid residues were found to be evolving under positive selection in all four genes analyzed for both clades. The discrepancy between our

results and previously published findings is likely due to differences in the samples analyzed and algorithms used for each analysis. Our analysis included genetically diverse isolates collected over a broad geographic range, enabling the detection of diversifying or directional evolution in multiple viral populations. We also used a recently developed algorithm (MEME), which has been demonstrated to have more power to detect residues under positive selection than previously employed methods (Kosakovsky *et al.* 2011). The level of positive selection we detected in PLV is consistent with the evolutionary dynamics described in other lentiviruses (Evans *et al.* 1999; Ross & Rodrigo 2002), and illustrates that despite apparently low pathogenicity, these viruses evolve and adapt in response to selection pressures from their hosts.

The PLV structural genes *gag*, *pol*, and *env* vary in levels of genetic diversity and selection, reflecting the different functions and host pressures on each protein. *Gag* is the most conserved ORF, and the capsid protein is the most conserved region of the Gag polyprotein with about 90% of capsid amino acid sites being invariant within each clade. Intra-population diversity in *gag* is even lower (data not shown), corresponding to the pervasive negative selection detected in this gene. Together these results suggest that nearly all non-synonymous mutations in *gag* decrease viral fitness, and thus, circulating viruses are highly homogeneous at the protein level. This strong purifying selection likely maintains secondary and tertiary structural conformations vital for the packaging, budding, and maturation of viral particles (Luttge & Freed 2010). However, some of the diversification in the PLVA and PLVB matrix and nucleocapsid protein domains resulted from positive selection. Therefore, the fitness landscape for *gag* amino acid diversity appears to be characterized by steep peaks, with divergent virus populations evolutionarily constrained on different fitness peaks.

The *pol* gene, which encodes enzymes necessary for viral replication, exhibits levels of genetic diversity similar to *gag*. Changes in amino acid composition that alter the structure, charge, and/or hydrophobicity of protein domains could disrupt enzyme function, explaining the high levels of homology and negative selection in *pol*. However, the protein domain with the highest percentage of sites under positive selection in both clades is the short pre-PR peptide transcribed from within *pol*. Because the *pol* polyprotein is translated via ribosomal frame shift within the *gag* ORF, *pre-PR* entirely overlaps the 3'-end of *gag*. Therefore, mutations occurring in this region will affect both *gag* and *pol* polyproteins. Interestingly, while the PLVA and PLVB pre-PR products have approximately 10% of sites under positive selection, no *gag* residues encoded from the same region are under positive selection in either PLV clade. Negative selection among the overlapping *gag* and pre-PR residues is the lowest of any section of the genome analyzed. This overlapping coding region is characterized by high levels of mutation, which evolve under neutral or purifying selection in the *gag* ORF, but mostly neutral and positive selection in the *pol* ORF. The biological significance of this level of selection on pre-PR is unknown, as no function has been ascribed to this peptide. However, our results suggest pre-PR may have important implications for viral fitness.

The *env* polyprotein and its subunits were the least conserved proteins analyzed. Interestingly, the highest percentage of residues under positive selection among the Env protein domains of both clades is the L peptide. This region of *env* has two putative functions, acting both as a signal to target the *env* transcript to the rough endoplasmic reticulum for translation, and as the first exon of the accessory protein Rev (Phillips *et al.* 1992; Poss & Ross 2008). Rev is encoded by all lentiviruses, and is required for the export of unspliced and partially spliced mRNAs from the nucleus (Knipe 2007). This is accomplished by altering the host-cell nuclear-

export mechanism, which in the absence of rev only allows for the export of fully spliced mRNAs from the nucleus (Malim *et al.* 1989). While both functions of the env L domain rely on specific residue motifs, diversity of Rev form and function may be evolutionarily advantageous. For example, highly efficient Rev proteins in equine infectious anemia virus rapidly produces virions during acute infection, but Rev proteins with low efficiency limit virus production and allow the virus to evade the host immune response during chronic infection (Belshan *et al.* 2001). Unfortunately, no empirical studies on *rev* transcription, translation, or protein function have been performed with puma lentiviruses. Such work could elucidate the mechanisms underlying the selective forces acting on the highly variable L peptide domain.

The SU and TM domains of env were more conserved than the L peptide, but exhibited higher levels of genetic diversity and positive selection than most other protein domains analyzed. The *env* ORF also has the highest number of indels and recombination breakpoints, illustrating that the genetic diversity observed in Env results from several different aspects of viral polymerase activity. This high level of genetic diversity, together with the low levels of negative selection, is consistent with the immune driven antigenic variation observed in other lentiviruses (Braun *et al.* 1987; Holmes *et al.* 1992; Payne *et al.* 1987). The SU and TM domains of lentiviral Env proteins are targets of the adaptive immune response, and this pressure often selects for viral diversity both within and between hosts (Braun *et al.* 1987; Holmes *et al.* 1992; Payne *et al.* 1987). Our analysis suggests PLV Env evolution may be similarly influenced by host immune pressures, resulting in high genetic diversity within and among viral populations.

The *vif* ORF encodes a lentiviral accessory protein that functions in primate and feline lentiviruses to overcome an anti-viral intrinsic-immune restriction factor known as APOBEC (apolipoprotein B mRNA-editing enzyme-catalytic polypeptide-like) (Mariani *et al.* 2003;

Zielonka *et al.* 2010). In the absence of a functional Vif, APOBEC produces high levels of G-to-A mutations during reverse transcription of the viral genome (Bishop *et al.* 2004). The inability of host-adapted viral Vif to counteract APOBEC activity in alternative hosts is thought to be an important factor limiting cross-species transmission events of lentiviruses (VandeWoude *et al.* 2010; Zielonka *et al.* 2010). It is therefore interesting that Vif diversity is higher among PLVB viruses than PLVA, given that the former infects one species and the latter infects two. The mountain lion and bobcat APOBEC proteins are highly homologous (Zielonka *et al.* 2010), and our findings suggest that the PLVA Vif may successfully counteract both species anti-viral proteins. This is supported by the similar levels of positive selection detected in PLVA and PLVB *vif*, which are comparable to the positive selection observed in *env*. This could be one viral mechanism contributing to the cross-species transmission of PLVA to pumas.

PLV Geographic Distribution and Phylogenetic Relationships

We did not isolate PLVA from any animals in CO, despite testing 110 samples using PCR primers that readily amplify PLVA from bobcats and mountain lions (Lagana *et al.* in press). We predicted mountain lions in CO would not be infected with PLVA since hundreds of animals in central North America have been previously screened, and only PLVB has been detected (Biek *et al.* 2006; Biek *et al.* 2003; Carpenter *et al.* 1996). However, because PLVA is the only lentivirus known to infect bobcats, we predicted PLVA would likely circulate in bobcats across their geographic range. Ten of 217 bobcats from CO and Iowa showed a low level of seroreactivity to PLV antigens via Western blot, but PLV was not amplified by PCR from any bobcats in these regions (Lagana *et al.* in press; unpublished data). One possible explanation is that bobcats in CO are infected with a divergent subtype of PLVA that we were unable to amplify with our PCR primers. We consider this unlikely given the high success rate of our

protocols in amplifying other divergent PLVA isolates. It is more likely that the low level of seroreactivity we observed in a small number of individuals from this region was non-specific (Franklin *et al.* 2007b). We conclude that bobcats in this portion of their geographic range may not be infected with the lentiviruses that circulate in FL and southern CA. Bobcats from areas surrounding FL and southern CA could be screened for PLVA to further delineate the geographic range of this clade.

This bi-coastal pattern of PLVA distribution may have resulted from historical species-level fluctuations in bobcats. A recent study found that bobcats in the Midwestern US, including CO, have a genetic signature reflecting secondary contact between distinct east and west bobcat lineages (Reding *et al.* 2012). Significant genetic structure further divides the two lineages into distinct populations. Together, these patterns are consistent with founder effects resulting from long-distance dispersers that recolonized the central and northern US from a small number of Pleistocene refugia on each coast (Reding *et al.* 2012). Haplotype maps suggest these refugia were most likely located in FL and the Pacific Northwest, as bobcats in these two locations have similar mitochondrial DNA (mtDNA) haplotypes that were predicted to be ancestral to haplotypes from bobcats in the continental interior. The same founder effects may also explain the lack of PLVA in the Midwestern US, if early colonizers to the region were PLV negative. This could have occurred due to stochastic effects given that this region was likely recolonized by a small number of individuals, and that gene flow among populations in this region has been relatively low since colonization (Reding *et al.* 2012). Alternatively, even small fitness differences between infected and uninfected individuals could have increased the likelihood that long-distance dispersers were primarily PLV negative.

Similarly, the current distribution of PLVB lineages across North America is in accordance with the ancestry of mountain lions. Mountain lions span from Canada to southern Chile, but North American mountain lions are genetically distinct and relatively homogeneous compared to populations in Central and South America (Culver *et al.* 2000). Culver *et al.*, (2000) proposed that mountain lions were completely displaced from North America during the late Pleistocene. They hypothesized that contemporary North American mountain lions have descended from a relatively small number of individuals that migrated north from Central America after the glaciers receded. However, one population of mountain lions in British Columbia carries a unique mtDNA haplotype, similar to the haplotype of mountain lions in Brazil, which was predicted by haplotype mapping to be the most ancestral population of the species (Culver *et al.* 2000). Interestingly, PLVB isolates from BC and Brazil form the basal node of the PLVB clade in phylogenetic trees containing North, South, and Central American viruses (Figure 2.6) (Franklin *et al.* 2007a). Therefore, ancestral genotypes among mountain lions, and among PLVB isolates, reside in British Columbia and Brazil. This observation is intriguing as it suggests that, similar to bobcats, a population of mountain lions infected with PLVB may have survived the most recent Ice Age in a Pacific Northwest refugium that has since recolonized parts of the west coast of North America.

While historic geological conditions may have similarly affected the distribution of the two North American felids and their lentiviruses, the contemporary geographic distribution and genetic diversity of PLVA and PLVB differ in several ways. The PLVA viruses sampled are highly spatially structured at both small and large geographic scales, reflecting the population dynamics of their primary host - the bobcat (Reding *et al.* 2012). Bobcats inhabit smaller home-ranges, live at higher densities, and disperse shorter distances from their natal area than mountain

lions (Logan 2001). Bobcat population structure has been detected at very fine scales where urban development decreases gene flow among nearby populations (Lee *et al.* 2012; Riley *et al.* 2006). While some PLVB viruses exhibit local population structure (i.e. southern CA), diverse PLVB isolates co-circulate in some areas and genetically related viruses circulate in populations separated by thousands of kilometers (Figures 2.3, 2.5, and 2.6). These findings are consistent with the dynamics of mountain lions, which inhabit large home ranges and can disperse more than 1000km from their natal range (Thompson & Jenks 2005).

The PLVB viruses circulating in WY and CO arose from at least two genetically distinct lineages that diverged from one another early in the ancestry of this clade in North America (Figures 2.3 and 2.6). The Colorado cluster shares ancestry with viruses circulating in CA-south and MT. Viruses related to this lineage also circulate in BC and Mexico (Figure 2.6) (Carpenter *et al.* 1996). Therefore, viruses arising from this lineage circulate in mountain lion populations spanning all of western North America. Other viruses arising from the same lineage as the Wyoming cluster have been sampled in Idaho, Arizona, and Texas (Figure 2.6) (Carpenter *et al.* 1996). Given the distribution of this lineage across the current eastern limit of the mountain lion's geographic distribution in western North America, it is possible that this divergent lineage also circulated in mountain lion populations now absent from the eastern part of the continent.

Pco2 (CA-north) and PLV1695 (BC) are distinct from the viruses circulating in CA-south and the continental interior (Figure 2.3C), and the evolutionary history shared by these two isolates varies across the genome (Figure 2.5). Two other viruses from CA-north, which were partially sequenced (Table 2.1), also cluster with Pco2 and PLV1695 in a phylogenetic tree constructed from a non-recombinant region of *env* (tree not shown). Therefore, viruses circulating north of Los Angeles share a more recent common ancestor with viruses circulating

approximately 2000 kilometers north in British Columbia, than with viruses circulating south of Los Angeles. This finding appears to be consistent with the genetic structure of mountain lion populations in southern California. Ernest et al., (2003) detected reduced gene flow between the lion populations separated by Los Angeles, but observed that animals from north of Los Angeles to northern California form a continuous, albeit dispersed, population (Ernest *et al.* 2003). Our results suggest related viruses may also circulate in northern California, Oregon, and Washington.

PLVA in Two Host Species

PLVA is the only feline lentivirus known to naturally infect two host species. Although originally discovered in a mountain lion, PLVA is rare in this species, but appears to be endemic in the bobcat populations studied here. Franklin and Troyer et al. (2007) hypothesized that PLVA has co-evolved with bobcats, and may only infect mountain lions opportunistically in areas where habitat degradation has altered population dynamics leading to increased contact and opportunity for disease transmission. According to this hypothesis, we predicted that PLVA would be under pressure to adapt in mountain lions because host-adapted pathogens are expected to experience different selection pressures in a non-adapted host species. Our findings support this hypothesis – three of six mountain lion PLVA isolates, but none of the 20 bobcat isolates, showed evidence of positive selection. This implies that host selection pressures differ between the two species, and that PLVA viruses are under pressure to adapt in mountain lions.

Furthermore, all three mountain lion isolates were found to be diversifying under positive selection in *env*, indicating that the viral proteins that interact with the host cellular receptors and immune system are an important component of viral adaptation. Based on these findings, we would predict that PLVA viruses have lower fitness in mountain lions than bobcats. This

hypothesis could be tested by quantifying the viral loads in these two species (Blake *et al.* 2006). Additional studies investigating the interaction between PLVA-vif and each host's APOBEC proteins could provide a molecular basis for the observation that cross-species transmission occurs with unusually high frequency in these species.

CONCLUSION

PLVA and PLVB are highly divergent, and though historically classified as two clades of species-specific viruses infecting mountain lions, they are likely distinct viral species with different primary feline hosts. The geographic distribution and prevalence of PLVA are consistent with the ecology and evolutionary history of bobcats. Highly structured PLVA lineages circulate in distinct bobcat populations, and PLVA is likely absent from bobcats in central North America. PLVA does not appear to be fully adapted to mountain lions, as we detected positive selection acting on Env from half of the isolates sampled from this species. PLVB is common in mountain lions throughout their geographic range and, in contrast to PLVA, distinct viral lineages co-circulate in some areas corresponding to the ability of this virus to move long distances in its vagile host.

Analyzing full-genome sequences from divergent virus populations provided new information about the organization of the PLVA genome, levels of positive and negative selection on different viral proteins, and different patterns of recombination break points in PLVA versus PLVB. We have shown that high levels of genetic diversity, frequent recombination, and positive selection on viral proteins characterize PLV evolution in bobcats and mountain lions. These findings contrast previously published reports (Biek *et al.* 2003; Burkala & Poss 2007; Franklin *et al.* 2007a), which were primarily based on studies of short gene fragments and/or isolates from a single population. These virus-host relationships appear to

be dynamic, illustrating that the co-evolution of pathogens and hosts does not necessarily result in a stable equilibrium, as has been proposed for some FIVs (O'Brien *et al.* 2006; Pecon-Slattery *et al.* 2008b). Analyzing intra-host diversity and levels of circulating provirus in diverse isolates could provide a better understanding of how genetic diversity and selection affect viral fitness.

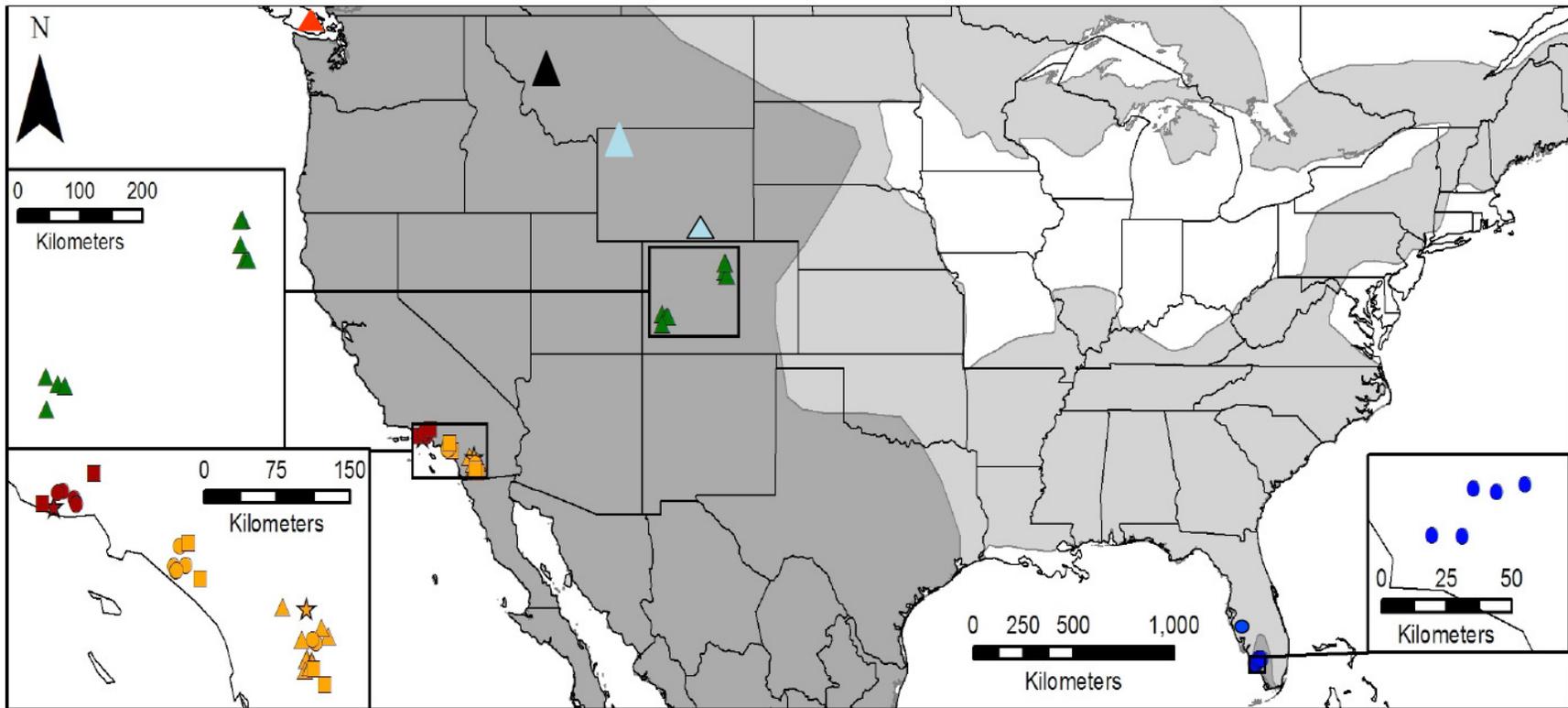


Figure 2.1: Host geographic ranges in North America and sample locations. Bobcat and mountain lion geographic ranges are depicted by light and dark grey shading respectively (<http://www.iucnredlist.org/>). The dark shading also represents areas of sympatry because the geographic range of bobcats completely overlaps that of mountain lions within the map boundaries. Animal capture locations are indicated by: circles – bobcats (PLVA); squares – mountain lions (PLVA), small triangles – mountain lions (PLVB), and; stars – PLVA/PLVB co-infected mountain lions. Large triangles represent approximate capture locations for published PLVB sequences from mountain lions included in this study (Bruen & Poss 2007; Langley *et al.* 1994). Symbol colors represent geographic regions: CA-south – orange; CA-north – dark red; CO-west and CO-east – green; FL – blue; WY-west light blue; WY-east – light blue with black outline; MT – black; BC – bright red.

Table 2.1: Demographic information for all bobcat (Lru) and puma (Pco) samples included in this study. GenBank accession numbers for previously published sequences are shown in parentheses.

PLVA Isolates					
Isolate Identifier	Location	Year Collected	Sex	Age	
Lru1	CA-north	1996	M	A	
Lru2	CA-north	2001	M	A	
Lru3	CA-north	2001	M	A	
Lru4	CA-north	2002	M	A	
Lru5	CA-north	2002	F	Y	
Lru6	CA-south	2003	M	A	
Lru7	CA-south	2005	M	A	
Lru8	CA-south	2006	M	A	
Lru9	CA-south	2007	M	A	
Lru10	CA-south	2006	M	A	
Lru11	CA-south	2009	M	A	
Lru12	CA-south	2009	F	A	
Lru13	CA-south	2009	M	A	
Lru14	FL	1984	-	-	
Lru15	FL	2007	F	A	
Lru16	FL	2010	M	A	
Lru17	FL	2010	M	A	
Lru18	FL	2010	M	A	
Lru19	FL	2010	M	A	
Lru20	FL	2010	M	A	
Pco1	CA-north	2004	F	A	
Pco2*§	CA-north	2002	M	A	
Pco3*	CA-north	2009	M	A	
Pco4*§	CA-south	2006	M	A	
Pco5	CA-south	2004	F	A	
Pco6	CA-south	2003	M	A	
Pco7	CA-south	2004	F	A	
Pco8	CA-south	2002	F	A	
Pco9*	CA-south	2010	M	A	
Pco10*	FL	1991	-	-	
Pco11*	FL	1987	-	-	
Pco12*	FL	1991	-	-	
PLV14 (U03982)	FL	1991	-	-	
* - only a portion of the PLVA genome was sequenced					
§ - co-infected with both PLVA and PLVB viruses					

Table 2.1 continued

PLVB Isolates					
Isolate Identifier	Location	Year Collected	Sex	Age	
Pco2§	CA-north	2002	M	A	
Pco4§	CA-south	2006	M	A	
Pco13	CA-south	2005	F	A	
Pco14	CA-south	2002	F	A	
Pco15	CA-south	2002	M	A	
Pco16	CA-south	2011	M	-	
Pco17	CA-south	2003	M	A	
Pco18	CA-south	2002	F	A	
Pco19	CA-south	2003	F	A	
Pco20	CA-south	2003	F	A	
Pco21	CA-south	2003	F	A	
Pco22	CO-east	2009	M	A	
Pco23	CO-west	2010	F	A	
Pco24	CO-east	2010	F	Y	
Pco25	CO-east	2010	F	A	
Pco26	CO-east	2010	F	A	
Pco27	CO-east	2008	F	A	
Pco28	CO-west	2009	M	A	
Pco29	CO-west	2009	F	A	
Pco30	CO-west	2009	M	A	
PLV1695 (DQ192583)	BC	1995	-	-	
Pco603 (EF455603)	MT	2001	-	-	
Pco604 (EF455604)	MT	2002	-	-	
Pco605 (EF455605)	MT	2001	-	-	
Pco606 (EF455606)	MT	2001	-	-	
Pco607 (EF455607)	WY-west	1992	-	-	
Pco608 (EF455608)	WY-west	1992	-	-	
Pco609 (EF455609)	WY-west	2004	-	-	
Pco610 (EF455610)	WY-west	2003	-	-	
Pco611 (EF455611)	WY-west	2001	-	-	
Pco612 (EF455612)	WY-west	2001	-	-	
Pco613 (EF455613)	WY-east	2001	-	-	
§ - co-infected with both PLVA and PLVB viruses					

Table 2.2: PCR primers used to amplify PLVA and PLVB isolates. Primer sequences are listed in the 5'- to 3'-direction. The location of the 5'-base (forward primers) or 3'-base (reverse primers) is listed relative to the reference sequence for each clade (PLV14 – PLVA; PLV1695 - PLVB). Primers denoted with 1 and 2 were originally published in Lee *et al.* (2012) and Troyer *et al.* (2005) respectively. All other primers were designed in this study.

PLVA		
Primer	Sequence	Location
PLVA1F	AAC TAG CTT AAC CGY AAA CCG CA	100
PLVA2F	CCA CAT CCT ATA GAA ATG ACA AWA AGA	131
PLVA3F	AGA TCC TCA GGT ATG CTT TAA TAA AGA G	208
PLVA4F	GTG AAC CTT GGT GGC TGC CTG	241
PLVA5F	GGC CAG AAA CYC TGC AGT TGG	296
PLVA6F	GCT TGG TTG AAG AGA TAC AAA CTG	482
PLVA6R	CAG TTT GTA TCT CTT CAA CCA AGC	
PLVA7F	TAG GTT CTC AAG CGG GAC ACC A	554
PLVA7R	TGG TGT CCC GCT TGA GAA CCT A	
PLVA8F	GAT GGA AAA RGC TAG AGG AGG	1089
PLVA8R	CCT CCT CTA GCY TTT TCC ATC	
PLVA9F	GAR GAA GCA GTN TTA TGG TTT ACT G	1123
PLVA9R	AGT AAA CCA TAA GAC TGC TTC	
PLVA10F	AAT GCW AAT GGR GAG TGT AGA ARR GC	1573
PLVA10R	GCY YTT CTA CAC TCY CCA TTW GCA TT	
PLVA11F	GGG GAA TAG TAT ATT TGA TGG ATA TCA	2172
PLVA11R	GGG GTG ATA TCC ATC AAA TAT ACT ATT	
PLVA12F	GGC CAT TGT TTT ATT TGA GGN CC	2439
PLVA12R	GGN CCT CAA ATA AAA CAA TGG CC	
PLVA13F	TTC WGT AAR AGC TTC TAT YTT TTC ATT WGT TA	2464
PLVA13R	CAA ATG AAA AAA TAG AAG CTT TAA C	
PLVA14F	GAA CCY CCY TAT AAR TGG ATG GGA TAT	3054
PLVA14R	ATA TCC CAT CCA TTT ATA GGG RGG TTC	
PLVA15F	GGN TRG GAA GRA TGA ATA GRC AAA AGA ARA	3553
PLVA15R	TYT TCT TTT GYC TAT TCA TYC TTC CYA NCC	
PLVA16F	TTA GGA GGW GTA ATT GAT CAA GGA T	4362
PLVA16R	ATC CTT GAT CAA TTA CWC CTC CTA A	
PLVA17F	TAA ART TAG GWG AAG GDA TAT GGC ARA T	4744
PLVA17R	ATY TGC CAT ATH CCT TCW CCT AAY TTT A	

Table 2.2 continued

PLVA (continued)		
<u>Primer</u>	<u>Sequence</u>	<u>Location</u>
PLVA18F	GAA AGG GTG GAT TAG GGG GTA TYA CAC	5143
PLVA18R	GTG TRA TAC CCC CTA ATC CAC CCT TTC	
PLVA19F	CWT GGG AWT ATT GTG GAG ATT GTT GG	6919
PLVA19R	CCA ACA ATC TCC ACA ATA WTC CCA WG	
PLVA20F	GGA TYG GAC CTG AAG AAG GAG AAA TG	6281
PLVA20R	CAT TTC TCC TTC TTC AGG TCC RAT CC	
PLVA21F	GCT ATA GAT CCT CCT TGG GTK ATT CC	6766
PLVA21R	GGA ATA AYC CAA GGD GGA TCT AWA GC	
PLVA22F	TGT TCA GGR ATA CCG GGA GTA GAT	7450
PLVA22R	ATC TAC TCC CGG TAT YCC TGA ACA	
envfw201 ¹	TTT CTC ATG TTC CTT GAA TGG TAC	7761
envfw202 ¹	TGG TAC ATT CTG GGT GTT TAA ATC	7779
PLVA23F	GTC ACT GCT GGR ATG ATW GGG	8047
PLVA23R	CCC WAT CAT YCC AGC AGT GAC	
PLVA24F	GCA CTG CAG CCC TGR CGG TAT C	8078
PLVA24R	GAT ACC GYC AGG GCT GCA GTG C	
PLVA25F	GCG ACT CAA WGA GWT AAT GCT CC	8103
PLVA25R	GGA GCA TTA WCT CWT TGA GTC GC	
PLVA26F	GGA TTG TTC AGA GGA GAC TGC	8165
PLVA26R	GCA GTC TCC TCT GAA CAA TCC	
PLVA27F	CAG AGG AGA CTG CAA GAT GCA G	8173
PLVA27R	CTG CAT CTT GCA GTC TCC TCT G	
envrv201 ¹	GCA TCA GAG AGT GAC CAA AAT AG	8214
envrv202 ¹	CAA TAC CAA TTA AGT GGA ATG TG	8254
PLVA28R	CTT CCC AGT CCA CCC TTT CTT CTT	8773
PLVA29R	TGC GGT TTR CGG TTA AGC TAG TT	8889
PLVA30R	TCT TAT TGT CAT TTC TAT AGG ATG TGG	8920
PLVA31R	CTC TTT ATT AAA GCA TAC CTG AGG ATC T	8997
PLVA32R	CAG GCA GCC ACC AAG GTT CAC	9033

Table 2.2 Continued

PLVB		
<u>Primer</u>	<u>Sequence</u>	<u>Location</u>
PLVB1F	CTA GCT TTR RCC RYA AAC CGC AAR T	51
PLVB2F	CAG RCT GYC CCT CAG GTA GAA TAA A	201
PLVB3F	AAC CCT GAC WTC TGC CTG AGA	240
PLVB4F	TCT TAT GTG GGT CTA AGG RAT CCG	268
PLVB5F	CCC AGT CAA GAG TAA GGC TTG GTA G	367
PLVB5R	CTA CCA AGC CTT ACT CTT GAC TGG G	
PLVB6F	CTT GGT TGA AGA GAC ACT GRC TG	447
PLVB6R	CAG YCA GTG TCT CTT CAA CCA AG	
PLVB7F	CTT CAC GGA TCW TCA AGC CAG	565
PLVB7R	CTG GCT TGA WGA TCC GTG AAG	
PLVB8F	CTG TCT GTC ATG GGG AAT GAG T	630
PLVB8R	ACT CAT TCC CCA TGA CAG ACA G	
PLVB9F	TTG CCT TGG TAG CTA CAG GAC	760
PLVB9R	GTC CTG TAG CTA CCA AGG CAA	
PLVB10F	ACA GCT TTC AAY CCT AGA ACA GTA GC	1074
PLVB11F	AGA RGG AAT ACA TAG TGA AGA RGC CAT	1119
PLVB11R	ATG GCY TCT TCA CTA TGT ATT CCY TCT	
PLVB12F	ATA CAT AGT GAA GAA GCC ATT CTG	1128
PLVB12R	CAG AAT GGC TTC TTC ACT ATG TAT	
PLVB13F	TGT CAG CTC CAG GAT GTG CT	1207
PLVB13R	AGC ACA TCC TGG AGC TGA CA	
PLVB14F	CCA GAG AAA TCA AAT GCA GGT	1943
PLVB14R	ACC TGC ATT TGA TTT CTC TGG	
PLVB15F	CGT GTT GAG GCC TGG ATA AAT G	2054
PLVB15R	CAT TWA TCC AGG CCT CAA CWC G	
PLVB16F	ACA GAA AAY GAA GGA ARA TGT TGT AA	2240
PLVB16R	TTA CAA CAT YTT CCT TCR TTT TCT GT	
PLVB17F	TTC TGC TGA TAA TGG CCA TTG TT	2427
PLVB17R	AAC AAT GGC CAT TAT CAG CAG AA	

Table 2.2 continued

PLVB (continued)		
<u>Primer</u>	<u>Sequence</u>	<u>Location</u>
PLVB18F	GAT TTG AAA CWC CAG ARG ATA AGC T	3000
PLVB18R	AGC TTA TCY TCT GGW GTT TCA AAT C	
PLVB19F	AAT TGG GCM ACT CAA ATA ATA GG	3155
PLVB19R	CCT ATT ATT TGA GTK GCC CAA TT	
PLVB20F	GCA AAT WAT GGA RAT AGA AGG ATC TAA TC	3772
PLVB20R	GAT TAG ATC CTT CTA TYT CCA TWA TTT GC	
PLVB21F	CCT TAT GGA ARG AGA TTA TAG AAG A	3921
PLVB21R	TCT TCT ATA ATC TCY TTC CAT AAG G	
PLVB22F	CCW GGA AAT AAR GAA ATW GAT GA	4001
PLVB22R	TCA TCW ATT TCY TTA TTT CCW GG	
PLVB23F	AAA GGW TTA GAT GTV YTA GGA GGA GT	4220
PLVB23R	ACT CCT CCT ARB ACA TCT AAW CCT TT	
PLVB24F	ACA AGG GTT TGG RAG YAC AGG	4411
PLVB24R	CCT GTR CTY CCA AAC CCT TGT	
PLVB25F	ATA GAT TGY ACA CAT WTA GAA GGA CA	4646
PLVB25R	TGT CCT TCT AWA TGT GTR CAA TCT AT	
PLVB26F	AGT RGA AAC TTT ACA RGC RGC AGT AG	4961
PLVB26R	CT ACT GCT GCY TGT AAW GTT TCY ACT TC	
PLVB27F	GAA TTA GTA GCA GGR ACA GGR C	5493
PLVB27R	GGY CCT GTY CCT GCT ACT AAT TC	
PLVB28F	TCY TGG TAT TGT AAA CCT YCT TAC AGG	5880
PLVB28R	YCT GTA AGG RGG TTT ACA ATA CCA	
PLVB29F	TTA GAA TTT GAR GAR GCA ATA GA	6122
PLVB29R	TCT ATT GCY TCY TCA AAT TCT AA	
PLVB30F	GAG AAG ATC ARA GRA TCC CWT CAG GAA	6362
PLVB30R	TTC CTG AWG GGA TYC TYT GAT CTT CTC	
PLVB31F	ACA AAT RCT GGA AGG AGG TGA GTT	6474
PLVB31R	AAC TCA CCT CCT TCC AGY ATT TGT	

Table 2.2 continued

PLVB (continued)		
<u>Primer</u>	<u>Sequence</u>	<u>Location</u>
PLVB32F	TGG AGT RTD GCT TGG TGG ACA TG	7012
PLVB32R	CAT GTC CAC CAA GCH AYA CTC CA	
PLVB33F	GAG AAA TAT GCA CTC AGC CRA CT	7562
PLVB33R	AGT YGG CTG AGT GCA TAT TTC TC	
PLVB34F	ATG AAT AAT GCA TCT TGG AAT TGG	7855
PLVB34R	CCA ATT CCA AGA TGC ATT ATT CAT	
PLVB35F	ACC TCG GTG GCA GGG CTG ATA	8080
PLVB35R	TAT CAG CCC TGC CAC CGA GGT	
PLVB36F	TGG CRG GGC TGA TAG GAG CA	8087
PLVB36R	TGC TCC TAT CAG CCC YGC CA	
PLVB37F	ACC ACT GGC ACC ACG GCC TT	8107
PLVB37R	AAG GCC GTG GTG CCA GTG GT	
PLVB38F	CTC AGA ACY TAA GAG GGA TTA TGC T	8135
PLVB38R	AGC ATA ATC CCT CTT ARG TTC TGA G	
PLVB39F	CAA ACA GAR ATA GAT GAG CAA ACC TT	8164
PLVB39R	AAG GTT TGC TCA TCT ATY TCT GTT TG	
PLVB40F	ATG GCM AAT CTG ATA GAA ATA AAA GAA G	8641
PLVB40R	CTT CTT TTA TTT CTA TCA GAT TKG CCA T	
PLVB41R	CAT YCC TCC CAG TCY ACC CTT	8777
PLVB42R	CGC AYT TGC GGT TTR YGG YYA AGC TAG	8837
PLVB43R	TTT ATT CTA CCT GAG GGR CAG	8991
PLVB44R	TCT CAG GCA GAW GTC AGG GTT	9026
PLVB45R	CGG ATY CCT TAG ACC CAC ATA AG	9054

PLVA AND PLVB		
<u>Name</u>	<u>Sequence</u>	<u>Location</u>
PLV330F	GCA GTT GGC GCC CGA ACA G	309
PLV330R	CCT GTT CGG GCG CCA ACT G	
FIV P1F ²	TGG CCW YTA WCW AAT GAA AAR ATW GAA	2457
FIV P2F ²	TGA AAA RAT WGA AGC HTT AAC AGA MAT AG	2471
FIV P1R ²	GTA ATT TRT CTT CHG GNG TYT CAA ATC CCC	3019
FIV P2R ²	GTA TTY TCT GCY TTT TTC TTY TGT CTA	3569

Table 2.3: PCR primer combinations for amplifying the full-protein coding region in PLVA and PLVB viral isolates. Table 2.2 provides the sequence and genome location for each primer.

PLVA primers		Forward	Reverse	Product Length	PLVB primers		Forward	Reverse	Product Length
PCR #1	Round 1	A3F	A16R	4.2kb	PCR #1	Round 1	B3F	P2R	3.3kb
	Round 2a	A4F	A12R	2.2kb		Round 2a	B4F	B9R	500bp
	Round2b	A9F	A15R	2.4kb		Round 2b	B7F	P1R	2.5kb
PCR #2	Round 1	P1F	A26R	5.7kb	PCR #2	Round 1	B14F	B39R	6.2kb
	Round 2a	A14F	A21R	3.7kb		Round 2a	B15F	B25R	2.6kb
	Round 2b	A19F	A24R	2.1kb		Round 2b	B24F	B30R	2.0kb
Round 2c						B29F	B38R	2.0kb	
PCR #3	Round 1	A21F	A32R	2.3kb	PCR #3	Round 1	B30F	B44R	2.7kb
	Round 2	A22F	A31R	1.6kb		Round 2	B34F	B43R	1.1kb

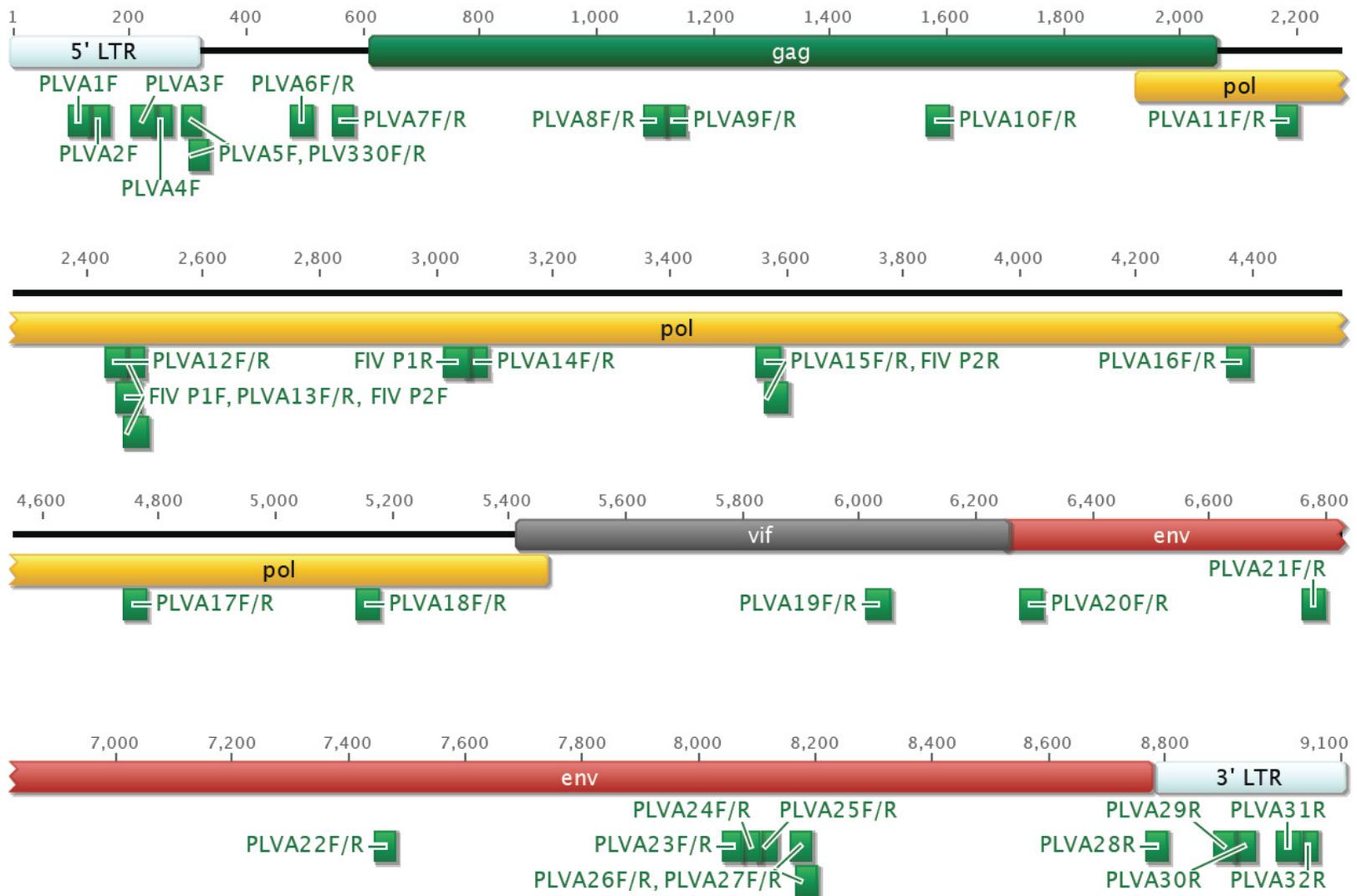


Figure 2.2: PLVA and PLVB primer maps showing the location of the primers (small green boxes) utilized in this study. Each figure is annotated with the location of the 5'- and 3'- long terminal repeats (LTR - light blue) and open reading frames (*gag*, *pol*, *vif*, *OrfA* (PLVB only) and *env*). Primer sequences and genome coordinates are listed in Table 2.2.

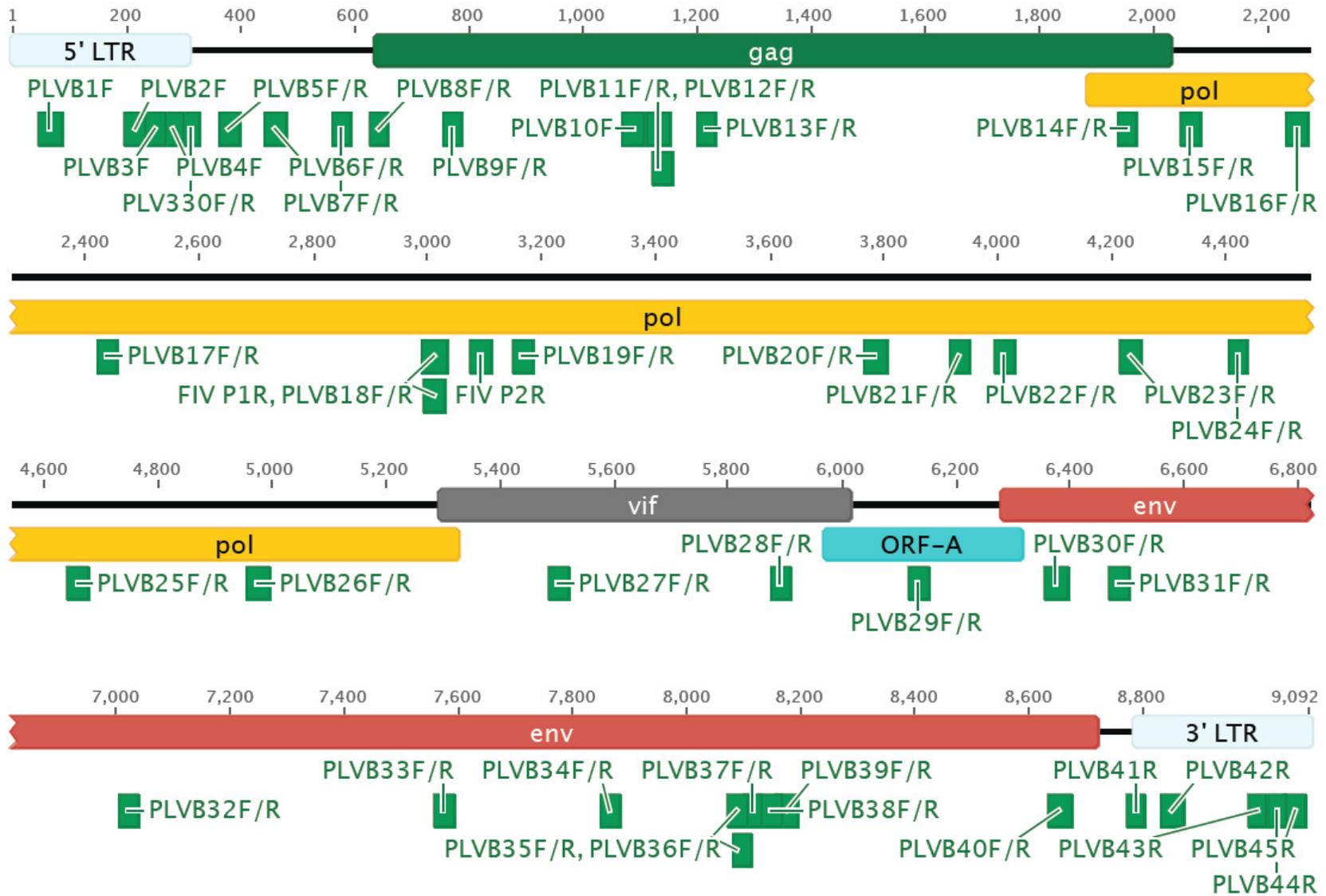


Figure 2.2 continued

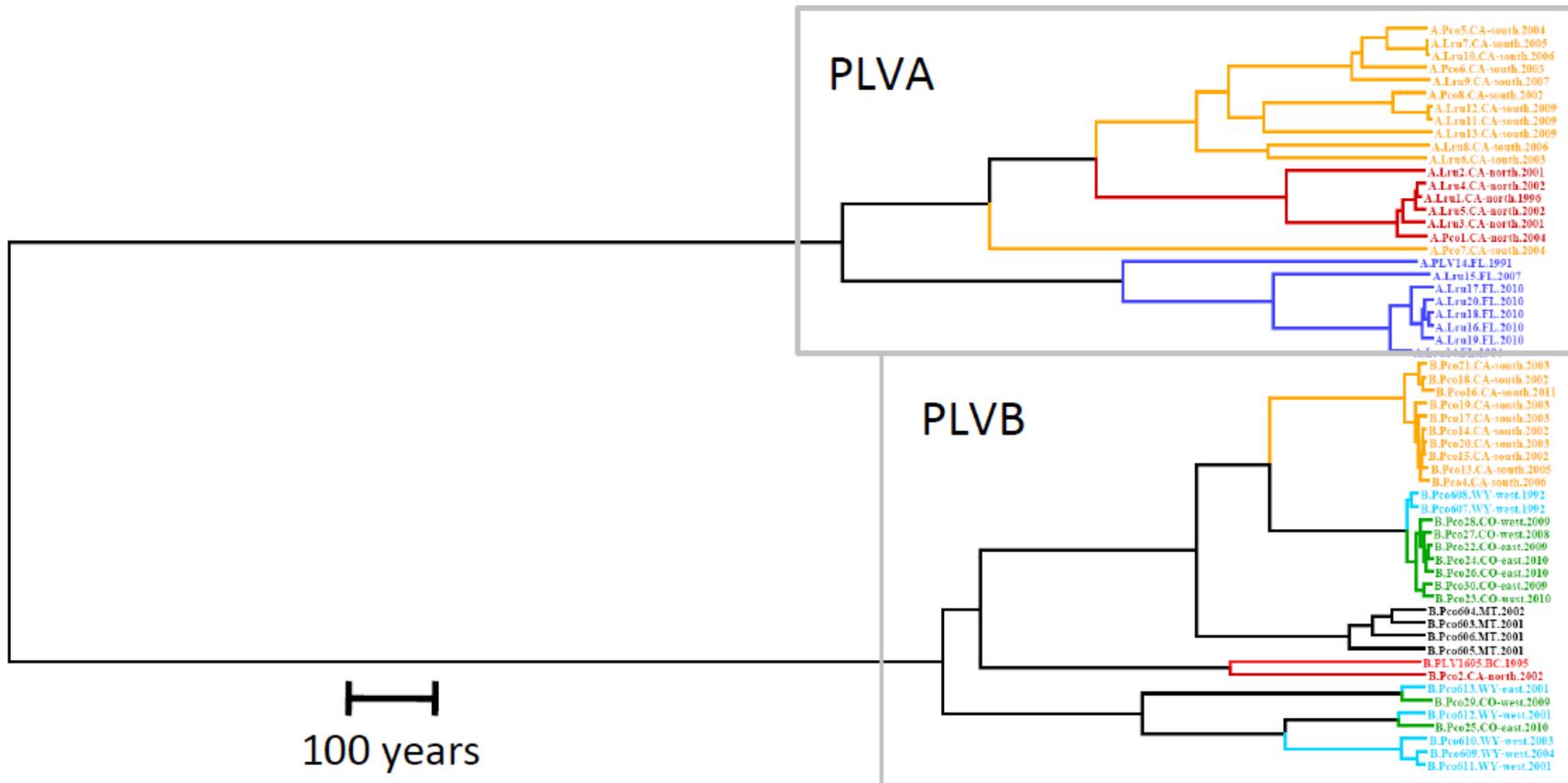


Figure 2.3A: The phylogenetic relationships among 26 PLVA and 32 PLVB viral isolates. This maximum clade credibility Bayesian phylogenetic tree of a 1439bp non-recombinant region of *pol* was constructed using dated tips. Isolates are color coded according to the geographic region from which they were sampled (Figure 2.1). Clusters are labeled based on the geographic location from which the majority of isolates were sampled. Internal nodes are labeled with the estimated year of coalescence (95% highest posterior density intervals listed in RESULTS). Support values (posterior probability greater than 80) are also labeled on each node (** > 90; * > 80). Isolate names provide the following information: (1) PLV clade (A or B); (2) host species (Lru – Bobcat; Pco - Puma); (3) animal ID number (Table 2.1); (4) sampling location, and; (5) sample year (1987 – 2011).

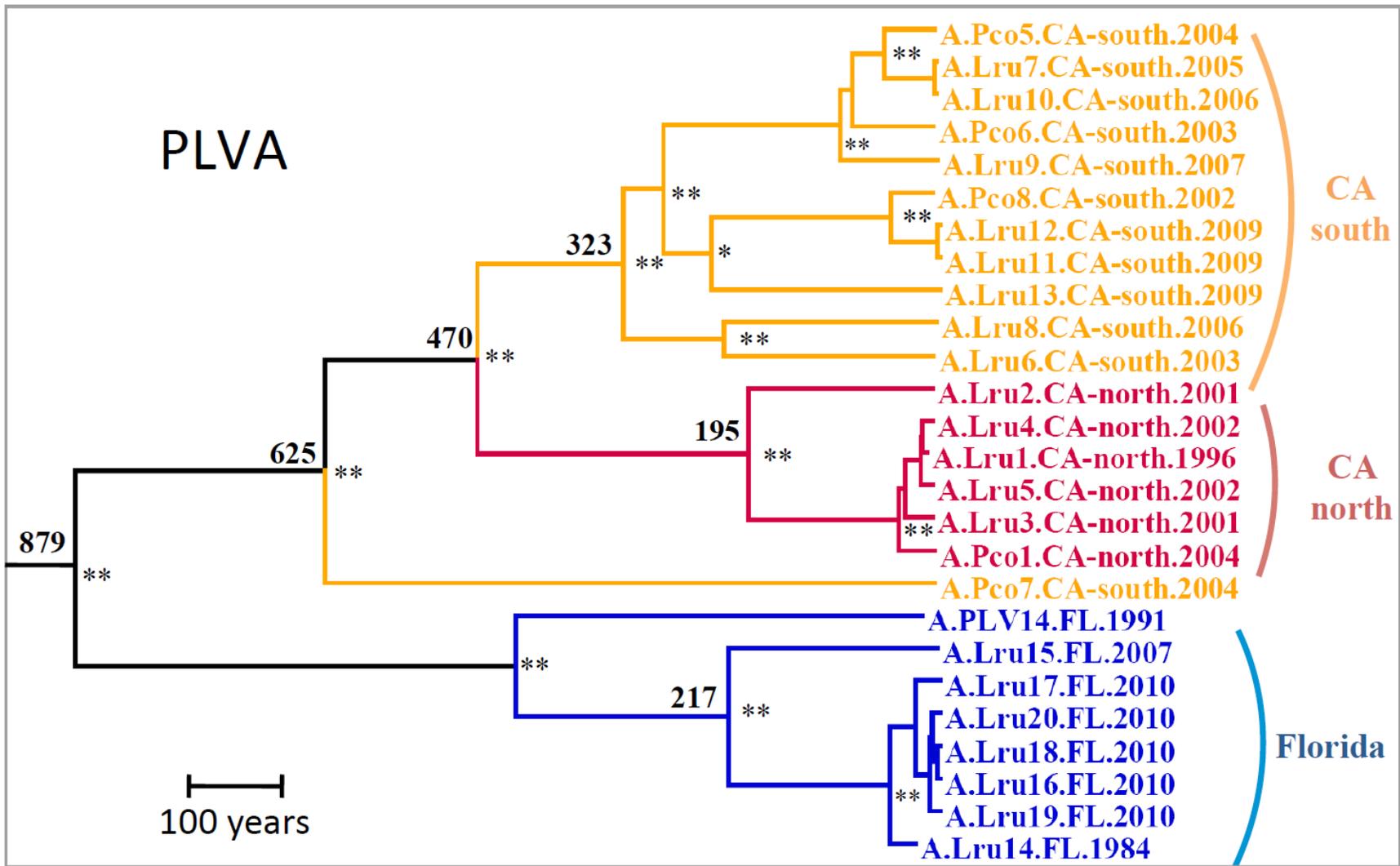


Figure 2.3B

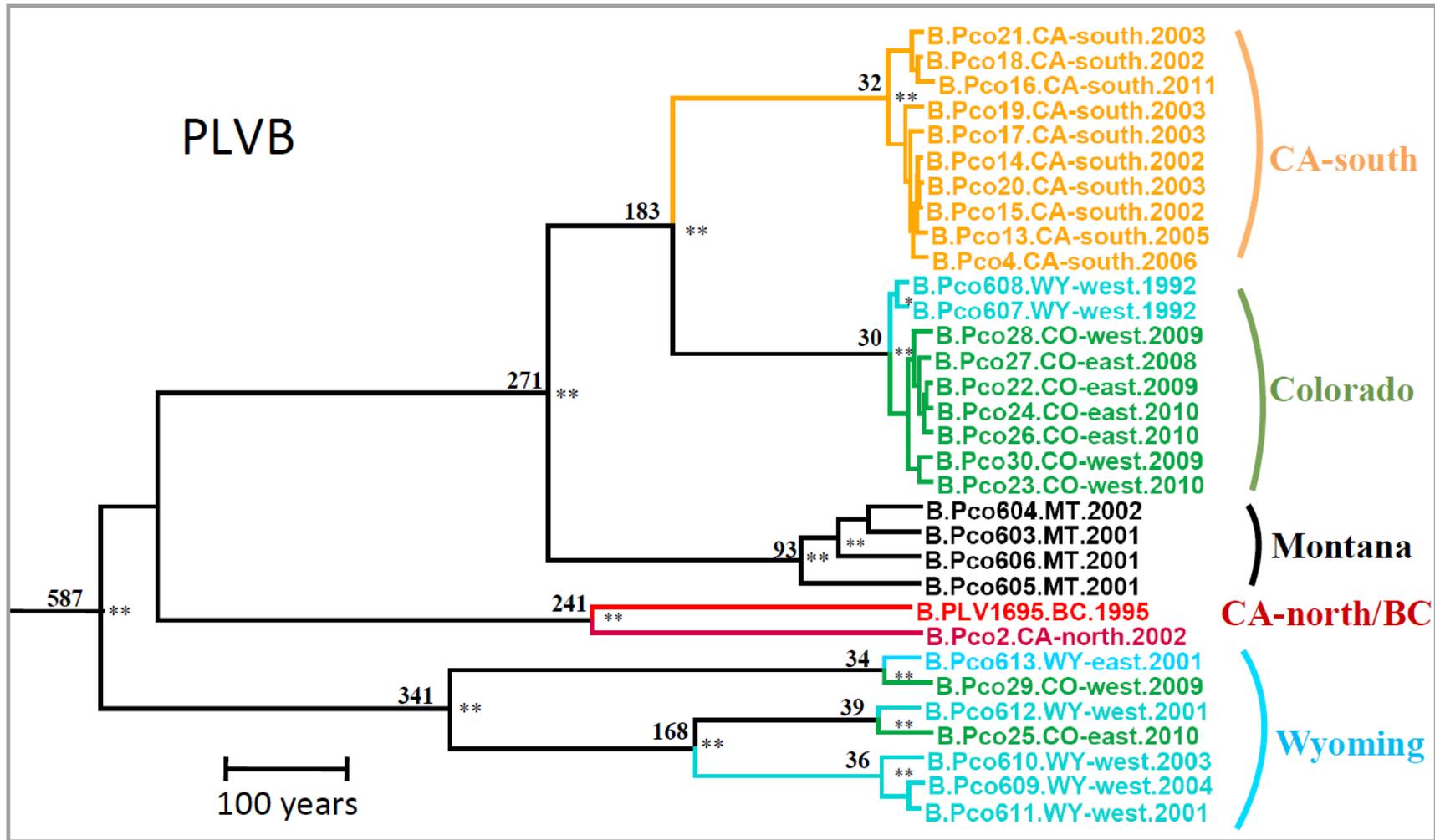


Figure 2.3C

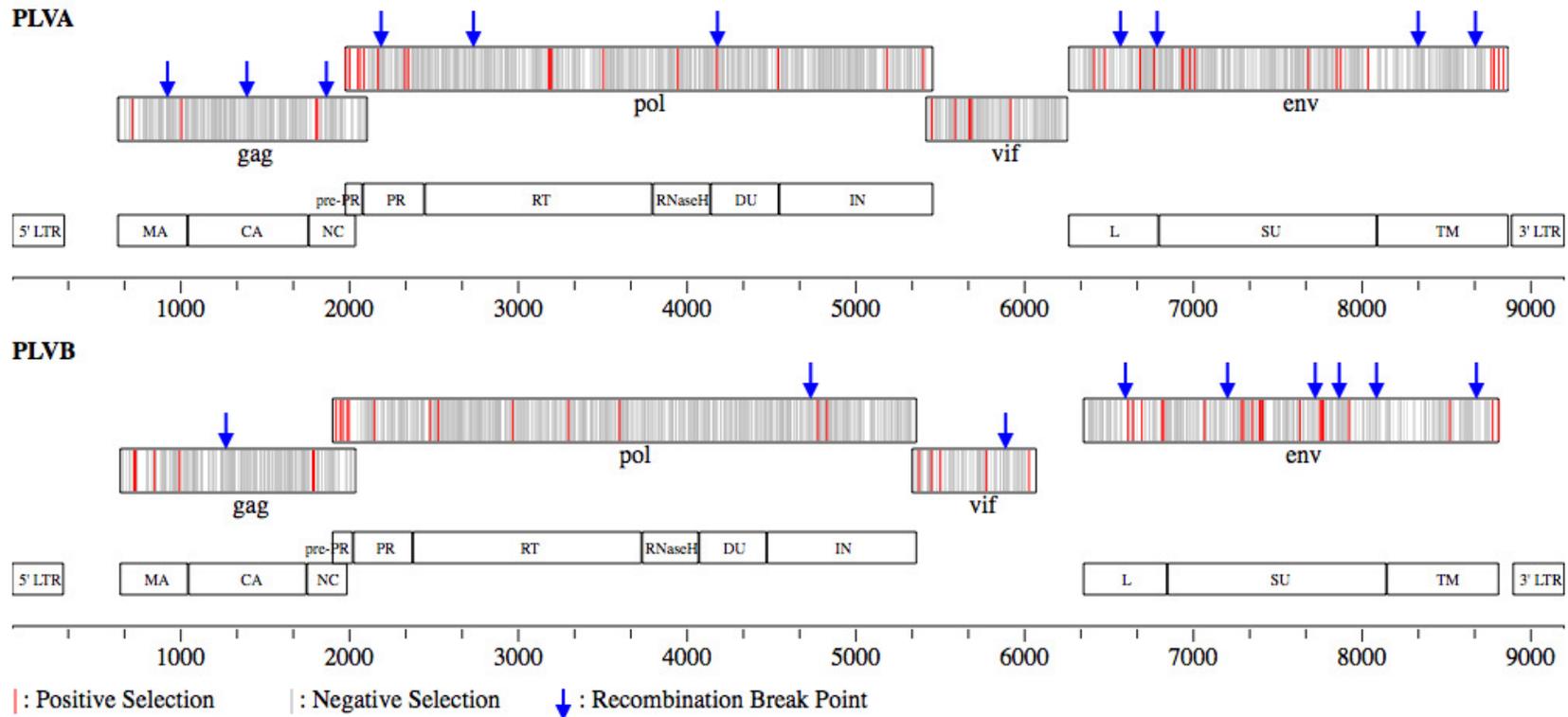


Figure 2.4: Location of recombination breakpoints and residues under selection (see Table 2.5) within the PLVA and PLVB genomes. PLV14 and PLV1695 were used as the reference genomes for PLVA and PLVB respectively (Bruen & Poss 2007; Langley *et al.* 1994). Each genome is annotated with the four viral open reading frames studied (above - *gag*, *pol*, *vif*, and *env*) as well as the protease cleavage products for the polyproteins (below). The scale below each genome represents base pairs starting at base #1 in the 5'-LTR.

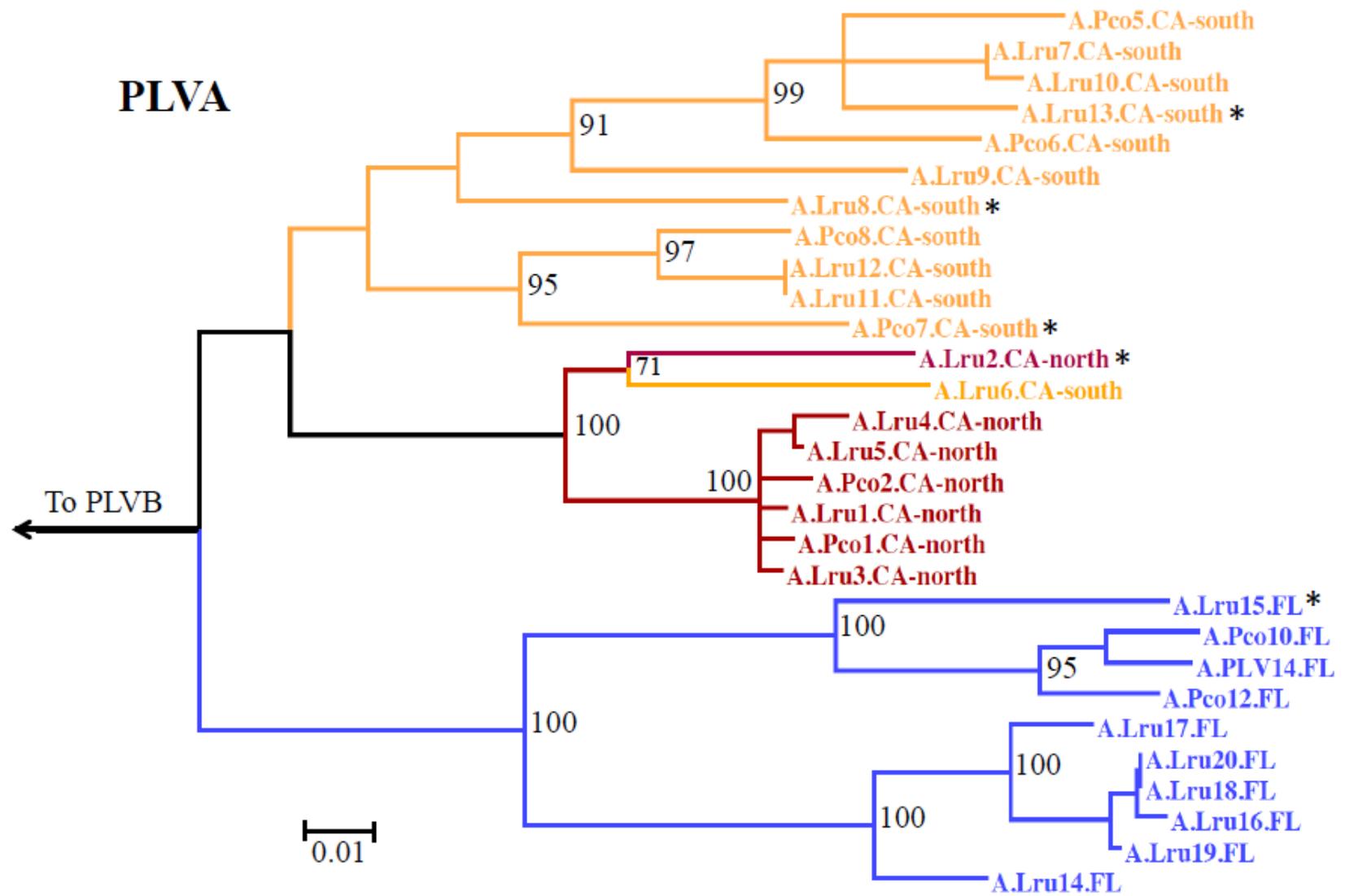


Figure 2.5A: Phylogenetic trees depicting alternative evolutionary histories of genomic regions resulting from recombination between isolates from different viral lineages. Neighbor-joining trees were constructed from all non-recombinant genome regions (Figure 2.4). One tree from each clade is presented to illustrate the effect of recombination on viral genetic diversity. The trees represent the following genome regions: PLVA – *gag* 905-1368 (520bp); PLVB – *env* 6529 – 7122 (620bp). Bootstrap values estimated from 1000 replicates are labeled on nodes with greater than 70% cluster support. Isolates labeled with * are likely recombinant viruses because their position on this tree is incongruent (supported by bootstrap values > 70) with their position in Figure 2.3

Table 2.4: Summary of genetic diversity, natural selection, and recombination results for four viral proteins. Pair-wise identity is the percent of identical pair-wise amino acids. Invariant sites are those columns in the alignment with a single amino acid shared across all sequences. Recombination breakpoints were identified using GARD (Pond *et al.* 2006). Positively selected sites were identified using MEME (Kosakovsky *et al.* 2011). Negatively selected sites were identified using FEL (Pond & Frost 2005). Table 2.5 provides additional details as well as results for all viral polyprotein cleavage products. The location of recombination breakpoints and residues under selection can be viewed in Figure 2.4.

Viral Protein	Protein Length (amino acids)		Pairwise Identity			% Invariant Sites			# Recombination Breakpoints		# Sites Under + Selection		# Sites Under - Selection	
	PLVA	PLVB	PLVA	PLVB	PLVA/PLVB	PLVA	PLVB	PLVA/PLVB	PLVA	PLVB	PLVA	PLVB	PLVA	PLVB
<i>gag</i>	479-485	461-462	91.5	94.9	72.3	77.0	80.8	43.4	3	1	4	8	257	222
<i>pol</i>	1135-1140*	1142-1146	88.2	89.7	74.4	67.2	64.4	40.1	3	1	17	14	593	519
<i>vif</i>	277	232-245	88.8	85.7	62.7	65.3	57.4	31.7	0	1	5	5	142	93
<i>env</i>	834-841	809-813	83.2	87.4	63.2	55.2	57.8	25.2	4	6	16	23	400	325

* The PLVA results in this table exclude a unique 41 aa insert present in the PLV-14 pol sequence, which may have arisen during adaptation to cell culture.

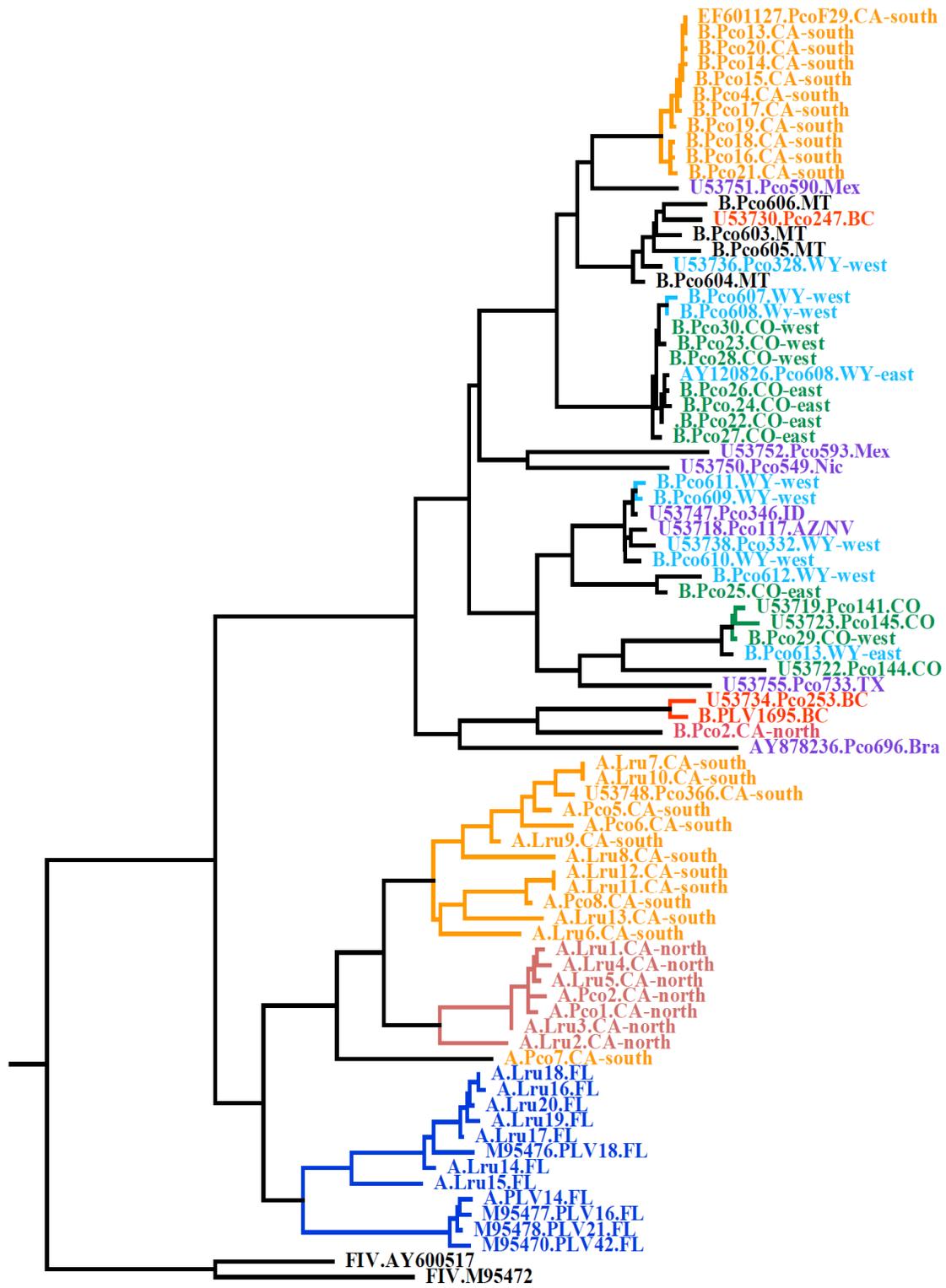
Table 2.5: Detailed genetic diversity, natural selection, and recombination results for all viral proteins studied.

Viral protein	Protein Length (amino acids)		Pairwise Identity		% Invariant Sites		Recombination Breakpoints ¹		Sites Under + Selection ²		Sites Under - Selection ²	
	PLVA	PLVB	PLVA	PLVB	PLVA	PLVB	PLVA	PLVB	PLVA	PLVB	PLVA	PLVB
<i>gag</i>	479-485	461-462	91.5	94.9	77.0	80.8			4 (0.8)	8 (1.7)	257 (53.0)	222 (48.1)
MA	136	134	89.2	94.9	72.1	77.6			2 (1.5)	4 (3.0)	66 (48.5)	52 (38.8)
CA	234	232	96.3	98.9	88.5	90.9	905		0	0	144 (61.5)	132 (56.9)
NC	83-90	78-79	83.3	84.5	58.1	58.8	1368	1262*	2 (2.2)	4 (5.1)	38 (41.3)	31 (39.2)
post-NC	22-25	17	90.4	87.4	72.0	70.6	1820		0	0	9 (39.1)	7 (41.2)
<i>pol</i>	1135-1140³	1142-1146	88.2	89.7	67.2	64.4			17 (1.4)	14 (1.2)	593 (50.2)	591 (51.6)
pre-PR	32-34	40-41	76.2	76.0	44.1	31.7			4 (11.8)	5 (12.2)	2 (5.9)	2 (4.9)
PR	119-120	117	91.4	89.3	74.2	65.0	2128		4 (3.3)	1 (0.9)	64 (53.3)	54 (46.2)
RT	442	449	88.6	90.9	67.6	67.7	2665		4 (0.9)	5 (1.1)	230 (52.0)	250 (55.7)
RNaseH	112	112	86.0	91.4	63.4	67.0			2 (1.8)	0	66 (58.9)	59 (52.7)
dUTPase	133	133	87.9	87.8	67.7	63.2	4207		1 (0.8)	1 (0.8)	74 (55.6)	70 (52.6)
IN	297-299	290-294	88.4	90.0	67.6	63.3		4696	2 (0.7)	2 (0.7)	157 (52.5)	156 (53.1)
<i>vif</i>	277	232-245	88.8	85.7	65.3	57.4		5849	5 (1.8)	5 (2.0)	142 (51.4)	93 (38.0)
<i>env</i>	834-841	809-813	83.2	87.4	55.2	57.8			16 (1.9)	23 (2.8)	400 (46.9)	325 (39.9)
leader	172-175	162-165	78.3	84.5	51.4	50.3	6551, 6758	6529	4 (2.3)	5 (3.0)	80 (45.7)	61 (37.0)
SU	410-416	425-429	86.0	88.5	57.4	59.3		7122, 7633, 7779, 7995	8 (1.9)	12 (2.8)	219 (51.8)	191 (44.5)
TM	249-254	219-220	81.8	87.3	53.3	60.5	8337*, 8643*	8575	4 (1.6)	6 (2.7)	101 (39.6)	73 (33.2)
Overall	2724-2739	2645-2664	87.2	89.4	65.1	64.2	10 breakpoints	9 breakpoints	42 (1.5)	50 (1.9)	1392 (50.8)	1231 (46.2)

1 - Numbers represent the nucleotide position in either the PLVA (PLV14) or PLVB (PLV1695) reference sequences. All breakpoints significant at alpha = 0.01 except those labeled with (*) which are significant at alpha = 0.05.
2 - The number (and %) of sites under selection within each specified genome region.
3 - Numbers do not include a unique 41aa insert in PLV14 RT, which may be an artifact of prolonged cell culture (Langley *et al.*, 1994).
Abbreviations: matrix (MA); capsid (CA), nucleocapsid (NC); protease (PR); reverse transcriptase (RT); deoxy-uridine transferase (dUTPase); integrase (IN); surface (SU); transmembrane (TM)

Table 2.6: PLVA isolates evolving under positive selection in at least one non-recombinant region of the genome (Kosakovsky *et al.* 2011). P-values shown were corrected for multiple tests by the Holm-Bonferroni method.

Isolate	Genome Region (gene)	Proportion of Sites with dN > dS	p-value
A.Pco7.CA-south	5419 - 6249 (<i>vif</i>)	0.03	0.020
	6551 - 6758 (<i>env</i>)	0.04	0.019
	8643 - 8772 (<i>env</i>)	0.24	0.014
A.Pco4.CA-sorth	6250 - 6551 (<i>env</i>)	0.02	0.048
A.Pco6.CA-south	6759 - 8337 (<i>env</i>)	0.03	0.035



0.2

Figure 2.6: Maximum-likelihood phylogenetic tree constructed from a 471 bp region of *pol* illustrating the relationship among the isolates analyzed in this study and previously published viral sequences. Two FIV sequences from domestic cats form the root of the tree. Isolates are color coded according to the geographic region from which they were sampled as in Figures 2.3 and 2.5. Sequences from geographic regions not represented in these figures are colored purple. Isolate names provide the following information: (1) PLV clade (A or B); (2) host species (Lru – bobcat; Pco - puma); (3) animal ID number or GenBank accession number, and; (4) sampling location (BC – British Columbia; CA – California, CO – Colorado, FL – Florida, MT – Montana, WY – Wyoming; AZ/NV – Arizona/Nevada; ID – Idaho; Mex – Mexico; Nic – Nicaragua; Bra – Brazil; TX – Texas).

LITERATURE CITED

- Bachmann MH, Mathiason Dubard C, Learn GH, *et al.* (1997) Genetic diversity of feline immunodeficiency virus: Dual infection, recombination, and distinct evolutionary rates among envelope sequence clades. *Journal of Virology* **71**, 4241-4253.
- Bailes E, Gao F, Bibollet-Ruche F, *et al.* (2003) Hybrid origin of SIV in chimpanzees. *Science* **300**, 1713-1713.
- Baranowski E, Ruiz-Jarabo CM, Domingo E (2001) Evolution of cell recognition by viruses. *Science* **292**, 1102-1105.
- Belshan M, Baccam P, Oaks JL, *et al.* (2001) Genetic and biological variation in equine infectious anemia virus rev correlates with variable stages of clinical disease in an experimentally infected pony. *Virology* **279**, 185-200.
- Benson D, Karsch-Mizrachi I, Lipman D, Sayers E (2011) GenBank. *Nucleic Acids Research* **39**, D32-37.
- Bevins SN, Carver S, Boydston EE, *et al.* (2012) Three Pathogens in Sympatric Populations of Pumas, Bobcats, and Domestic Cats: Implications for Infectious Disease Transmission. *Plos One* **7**.
- Biek R, Drummond AJ, Poss M (2006) A virus reveals population structure and recent demographic history of its carnivore host. *Science* **311**, 538-541.
- Biek R, Rodrigo AG, Holley D, *et al.* (2003) Epidemiology, genetic diversity, and evolution of endemic feline immunodeficiency virus in a population of wild cougars. *Journal of Virology* **77**, 9578-9589.
- Bishop KN, Holmes RK, Sheehy AM, *et al.* (2004) Cytidine deamination of retroviral DNA by diverse APOBEC proteins. *Current Biology* **14**, 1392-1396.
- Blake DJ, Graham J, Poss M (2006) Quantification of Feline immunodeficiency virus (FIV_{pco}) in peripheral blood mononuclear cells, lymph nodes and plasma of naturally infected cougars. *Journal of General Virology* **87**, 967-975.
- Boerlijst MC, Bonhoeffer S, Nowak MA (1996) Viral quasi-species and recombination. *Proceedings of the Royal Society B-Biological Sciences* **263**, 1577-1584.
- Braun MJ, Clements JE, Gonda MA (1987) The Visna Virus Genome - Evidence for a Hypervariable Site in the Env-Gene and Sequence Homology among Lentivirus Envelope Proteins. *Journal of Virology* **61**, 4046-4054.
- Bruen TC, Poss M (2007) Recombination in feline immunodeficiency virus genomes from naturally infected cougars. *Virology* **364**, 362-370.
- Burkala E, Poss M (2007) Evolution of feline immunodeficiency virus Gag proteins. *Virus Genes* **35**, 251-264.
- Carpenter MA, Brown EW, Culver M, *et al.* (1996) Genetic and phylogenetic divergence of feline immunodeficiency virus in the puma (*Puma concolor*). *Journal of Virology* **70**, 6682-6693.
- Cavrois M, Gessain A, Wain-Hobson S, Wattel E (1996) Proliferation of HTLV-1 infected circulating cells in vivo in all asymptomatic carriers and patients with TSP/HAM. *Oncogene* **12**, 2419-2423.
- Chatterji U, de Parseval A, Elder JH (2002) Feline immunodeficiency virus OrfA is distinct from other lentivirus transactivators. *Journal of Virology* **76**, 9624-9634.

- Crooks KR (2002) Relative sensitivities of mammalian carnivores to habitat fragmentation. *Conservation Biology* **16**, 488-502.
- Culver M, Johnson WE, Pecon-Slaterry J, O'Brien SJ (2000) Genomic ancestry of the American puma (*Puma concolor*). *Journal of Heredity* **91**, 186-197.
- Martin D, Lemey P, Lott M, *et al.* (2010) RDP3: a flexible and fast computer program for analyzing recombination. *Bioinformatics* **26**, 2462-2463.
- Delport W, Poon AFY, Frost S, Kosakovsky Pond S (2010a) Datamonkey 2010: a suite of phylogenetic analysis tools for evolutionary biology. *Bioinformatics* **26**, 2455-2457.
- Delport W, Poon AFY, Frost SDW, Pond SLK (2010b) Datamonkey 2010: a suite of phylogenetic analysis tools for evolutionary biology. *Bioinformatics* **26**, 2455-2457.
- Domingo E, Escarmis C, Sevilla N, *et al.* (1996) Basic concepts in RNA virus evolution. *Faseb Journal* **10**, 859-864.
- Drake JW (1993) Rates of Spontaneous Mutation among Rna Viruses. *Proceedings of the National Academy of Sciences of the United States of America* **90**, 4171-4175.
- Drummond AJ, Rambaut A (2007) BEAST: Bayesian evolutionary analysis by sampling trees. *Bmc Evolutionary Biology* **7**.
- Drummond AJ, Rambaut A, Shapiro B, Pybus OG (2005) Bayesian coalescent inference of past population dynamics from molecular sequences. *Molecular Biology and Evolution* **22**, 1185-1192.
- Edgar RC (2004) MUSCLE: multiple sequence alignment with high accuracy and high throughput. *Nucleic Acids Research* **32**, 1792-1797.
- Eigen M (1971) Self-organization of matter and the evolution of biological macromolecules. *Naturwissenschaften* **58**, 465-523.
- Eigen M (1996) *Steps Towards Life* Oxford University Press.
- Elder JH, Schnolzer M, Hasselkuslight CS, *et al.* (1993) Identification of Proteolytic Processing Sites within the Gag and Pol Polyproteins of Feline Immunodeficiency Virus. *Journal of Virology* **67**, 1869-1876.
- Ernest HB, Boyce WM, Bleich VC, *et al.* (2003) Genetic structure of mountain lion (*Puma concolor*) populations in California. *Conservation Genetics* **4**, 353-366.
- Evans DT, O'Connor DH, Jing PC, *et al.* (1999) Virus-specific cytotoxic T-lymphocyte responses select for amino-acid variation in simian immunodeficiency virus Env and Nef. *Nature Medicine* **5**, 1270-1276.
- Franklin SP, Troyer JL, Terwee JA, *et al.* (2007a) Frequent transmission of immunodeficiency viruses among bobcats and pumas. *Journal of Virology* **81**, 10961-10969.
- Franklin SP, Troyer JL, TerWee JA, *et al.* (2007b) Variability in assays used for detection of lentiviral infection in bobcats (*Lynx rufus*), pumas (*Puma concolor*) and ocelots (*Leopardus pardalis*). *Journal of Wildlife Diseases* **43**, 700-710.
- Gao F, Bailes E, Robertson DL, *et al.* (1999) Origin of HIV-1 in the chimpanzee *Pan troglodytes* troglodytes. *Nature* **397**, 436-441.
- Geneious version 5.6.4 created by Biomatters. Available from <http://geneious.com/>
- Hirsch VM, Olmsted RA, Murphey-Corb M, Purcell RH, Johnson PR (1989) An African primate lentivirus (SIVsm) closely related to HIV-2. *Nature* **339**, 389-391.
- Ho SYW, Phillips MJ, Cooper A, Drummond AJ (2005) Time dependency of molecular rate estimates and systematic overestimation of recent divergence times. *Molecular Biology and Evolution* **22**, 1561-1568.

- Holmes EC (2003) Error thresholds and the constraints to RNA virus evolution. *Trends in Microbiology* **11**, 543-546.
- Holmes EC, Zhang LQ, Simmonds P, Ludlam CA, Brown AJL (1992) Convergent and Divergent Sequence Evolution in the Surface Envelope Glycoprotein of Human-Immunodeficiency-Virus Type-1 within a Single Infected Patient. *Proceedings of the National Academy of Sciences of the United States of America* **89**, 4835-4839.
- International Union for the Conservation of Nature (<http://www.iucnredlist.org/>).
- Johnson WE, Eizirik E, Pecon-Slattery J, *et al.* (2006) The Late Miocene radiation of modern Felidae: A genetic assessment. *Science* **311**, 73-77.
- Jung A, Maier R, Vartanian JP, *et al.* (2002) Recombination: Multiply infected spleen cells in HIV patients. *Nature* **418**, 144.
- Kibbe WA (2007) OligoCalc: an online oligonucleotide properties calculator. *Nucleic Acids Research* **35**, W43-W46.
- Knipe DM, Howley, Peter M (2007) *Fields Virology*. Lippincott Williams & Wilkins.
- Kosakovsky P, Murrell B, Fourment M, *et al.* (2011) A random effects branch-site model for detecting episodic diversifying selection. *Molecular Biology and Evolution* **28**, 3033-3043.
- Lagana DM, Lee JS, Lewis JS, *et al.* (In Press) Characterization of feline immunodeficiency virus sequences in bobcats (*Lynx rufus*). *Journal of Wildlife Diseases*.
- Langley RJ, Hirsch VM, O'Brien SJ, *et al.* (1994) Nucleotide sequence analysis of puma lentivirus (PLV-14): genomic organization and relationship to other lentiviruses. *Virology* **202**, 853-864.
- Larkin MA, Blackshields G, Brown NP, *et al.* (2007) Clustal W and clustal X version 2.0. *Bioinformatics* **23**, 2947-2948.
- Lee JS, Ruell EW, Boydston EE, *et al.* (2012) Gene flow and pathogen transmission among bobcats (*Lynx rufus*) in a fragmented urban landscape. *Molecular Ecology* **21**, 1617-1631.
- Li WH, Tanimura M, Sharp PM (1988) Rates and dates of divergence between AIDS virus nucleotide sequences. *Molecular Biology and Evolution* **5**, 313-330.
- Logan KaS, L (2001) *Desert Puma: Evolutionary Ecology and Conservation of an Enduring Carnivore* Island Press, Washington D.C.
- Lutge BG, Freed EO (2010) FIV Gag: Virus assembly and host-cell interactions. *Veterinary Immunology and Immunopathology* **134**, 3-13.
- Malim MH, Hauber J, Le SY, Maizel JV, Cullen BR (1989) The Hiv-1 Rev Trans-Activator Acts through a Structured Target Sequence to Activate Nuclear Export of Unspliced Viral Messenger-Rna. *Nature* **338**, 254-257.
- Mariani R, Chen D, Schrofelbauer B, *et al.* (2003) Species-specific exclusion of APOBEC3G from HIV-1 virions by Vif. *Cell* **114**, 21-31.
- Moutouh L, Corbeil J, Richman DD (1996) Recombination leads to the rapid emergence of HIV-1 dually resistant mutants under selective drug pressure. *Proceedings of the National Academy of Sciences of the United States of America* **93**, 6106-6111.
- Munk C, Hechler T, Chareza S, Lochelt M (2010) Restriction of feline retroviruses: Lessons from cat APOBEC3 cytidine deaminases and TRIM5 alpha proteins. *Veterinary Immunology and Immunopathology* **134**, 14-24.
- Nishimura Y, Goto Y, Yoneda K, *et al.* (1999) Interspecies transmission of feline immunodeficiency virus from the domestic cat to the Tsushima cat (*Felis bengalensis euphilura*) in the wild. *Journal of Virology* **73**, 7916-7921.

- O'Brien SJ, Troyer JL, Roelke M, Marker L, Pecon-Slattery J (2006) Plagues and adaptation: Lessons from the Felidae models for SARS and AIDS. *Biological Conservation* **131**, 255-267.
- Olmsted RA, Langley R, Roelke ME, *et al.* (1992) Worldwide Prevalence of Lentivirus Infection in Wild Feline Species - Epidemiologic and Phylogenetic Aspects. *Journal of Virology* **66**, 6008-6018.
- Overbaugh J, Bangham CRM (2001) Selection forces and constraints on retroviral sequence variation. *Science* **292**, 1106-1109.
- Payne SL, Fang FD, Liu CP, *et al.* (1987) Antigenic Variation and Lentivirus Persistence - Variations in Envelope Gene-Sequences during EIAV Infection Resemble Changes Reported for Sequential Isolates of Hiv. *Virology* **161**, 321-331.
- Pecon-Slattery J, McCracken CL, Troyer JL, *et al.* (2008a) Genomic organization, sequence divergence, and recombination of feline immunodeficiency virus from lions in the wild. *Bmc Genomics* **9**(66), doi:10.1186/1471-2164-9-66.
- Pecon-Slattery J, Troyer JL, Johnson WE, O'Brien SJ (2008b) Evolution of feline immunodeficiency virus in Felidae: Implications for human health and wildlife ecology. *Veterinary Immunology and Immunopathology* **123**, 32-44.
- Phillips TR, Lamont C, Konings DA, *et al.* (1992) Identification of the Rev transactivation and Rev-responsive elements of feline immunodeficiency virus. *Journal of Virology* **66**, 5464-5471.
- Pond SLK, Frost SDW (2005) Not so different after all: A comparison of methods for detecting amino acid sites under selection. *Molecular Biology and Evolution* **22**, 1208-1222.
- Pond SLK, Posada D, Gravenor MB, Woelk CH, Frost SDW (2006) Automated phylogenetic detection of recombination using a genetic algorithm. *Molecular Biology and Evolution* **23**, 1891-1901.
- Poss M, Idoine A, Ross HA, *et al.* (2007) Recombination in feline lentiviral genomes during experimental cross-species infection. *Virology* **359**, 146-151.
- Poss M, Ross H (2008) Evolution of the long terminal repeat and accessory genes of feline immunodeficiency virus genomes from naturally infected cougars. *Virology* **370**, 55-62.
- Poss M, Ross HA, Painter SL, *et al.* (2006) Feline lentivirus evolution in cross-species infection reveals extensive G-to-A mutation and selection on key residues in the viral polymerase. *Journal of Virology* **80**, 2728-2737.
- Rambaut A, Drummond A (2007) Tracer v1.4, Available from <http://beast.bio.ed.ac.uk/Tracer>.
- Reding DM, Bronikowski AM, Johnson WE, Clark WR (2012) Pleistocene and ecological effects on continental-scale genetic differentiation in the bobcat (*Lynx rufus*). *Molecular Ecology* **21**, 3078-3093.
- Rigby MA, Holmes EC, Pistello M, *et al.* (1993) Evolution of Structural Proteins of Feline Immunodeficiency Virus - Molecular Epidemiology and Evidence of Selection for Change. *Journal of General Virology* **74**, 425-436.
- Riley SPD, Pollinger JP, Sauvajot RM, *et al.* (2006) A southern California freeway is a physical and social barrier to gene flow in carnivores. *Molecular Ecology* **15**, 1733-1741.
- Riley SPDB, E. E.; Crooks, K. R.; Lyren, L. M.; in Gehrt, Stanley D., Riley, Seth P. D., Cypher, Brian L. editors (2010) Bobcats (*Lynx rufus*). Urban Carnivores - Ecology, Conflict, and Conservation. The Johns Hopkins University Press.

- Robertson DL, Sharp PM, Mccutchan FE, Hahn BH (1995) Recombination in Hiv-1. *Nature* **374**, 124-126.
- Ross HA, Rodrigo AG (2002) Immune-mediated positive selection drives human immunodeficiency virus type 1 molecular variation and predicts disease duration. *Journal of Virology* **76**, 11715-11720.
- Sharp PM, Hahn BH (2011) Origins of HIV and the AIDS pandemic. *Cold Springs Harbor Perspective Medicine*, 1-22.
- Sharp PM, Robertson DL, Gao F, Hahn BH (1994) Origins and diversity of human immunodeficiency viruses. *AIDS* **8**, S27-S42.
- Sharp PM, Robertson DL, Hahn BH (1995) Cross-species transmission and recombination of 'AIDS' viruses. *Philosophical Transactions of the Royal Society of London Series B-Biological Sciences* **349**, 41-47.
- Pecon-Slattery J, McCracken CL, Troyer JL, VandeWoude S, Roelke M, Sondgeroth K, Winterbach C, Winterbach H, O'Brien S (2008). Genomic organization, sequence divergence, and recombination of feline immunodeficiency virus from lions in the wild. *BMC genomics* **9.1**, 66.
- Thompson DJ, Jenks JA (2005) Long-distance dispersal by a subadult male cougar from the Black Hills, South Dakota. *Journal of Wildlife Management* **69**, 818-820.
- Troyer JL, Pecon-Slattery J, Roelke ME, *et al.* (2005) Seroprevalence and genomic divergence of circulating strains of Feline immunodeficiency virus among Felidae and Hyaenidae species. *Journal of Virology* **79**, 8282-8294.
- Troyer JL, VandeWoude S, Pecon-Slattery J, *et al.* (2008) FIV cross-species transmission: An evolutionary prospective. *Veterinary Immunology and Immunopathology* **123**, 159-166.
- VandeWoude S, Troyer J, Poss M (2010) Restrictions to cross-species transmission of lentiviral infection gleaned from studies of FIV. *Veterinary Immunology and Immunopathology* **134**, 25-32.
- Wain LV, Bailes E, Bibollet-Ruche F, *et al.* (2007) Adaptation of HIV-1 to its human host. *Molecular Biology and Evolution* **24**, 1853-1860.
- Wiuf C, Christensen T, Hein J (2001) A simulation study of the reliability of recombination detection methods. *Molecular Biology and Evolution* **18**, 1929-1939.
- Zielonka J, Marino D, Hofmann H, *et al.* (2010) Vif of Feline Immunodeficiency Virus from Domestic Cats Protects against APOBEC3 Restriction Factors from Many Felids. *Journal of Virology* **84**, 7312-7324.

CHAPTER THREE

Using Targeted Genome Capture and Next-Generation Sequencing to Detect and Sequence Multiple Feline Pathogens

INTRODUCTION

Infectious agents pose the single largest threat to domestic feline health, and can cause significant morbidity and mortality among wild cats (Brown *et al.* 2008; Kapil & Lamm 2008; Meli *et al.* 2009; Roelke-Parker *et al.* 1996). Hence, the ability to definitively identify and characterize pathogens is of the utmost importance for clinical veterinarians and wildlife health researchers (Evermann *et al.* 2012). Pathogen genetic data provide important information about prevention strategies, pathogenesis, treatment options, and population management decisions (Gillespie 2002; Kapil & Lamm 2008; Rambaut *et al.* 2004). Furthermore, genetic data can provide insights into the ecology, evolution, and emergence pathogens through phylogenetic analyses, the development of transmission networks, and the discovery of natural reservoirs (McCarthy *et al.* 2007; Parrish *et al.* 2008; Woolhouse *et al.* 2005).

Despite considerable advancements in the ability to generate and utilize genetic sequence data from many common pathogens, significant limitations remain. For example, many infectious agents cause chronic, latent infections with low replication rates, rendering them undetectable by current assays despite remaining pathogenic and/or transmissible. Additionally, pathogens are extremely successful at evading host immune mechanisms by undergoing genetic mutations. High mutation rates can also complicate detection. For example, polymerase chain reaction (PCR) assays can fail to detect pathogens with mutations that disrupt primer annealing.

Therefore, improving upon current assays to detect and sequence feline pathogens would have broad applications for clinicians, researchers, and wildlife managers.

Targeted genome capture (TGC) is a new methodology that enriches specific genetic sequences within a heterogeneous mixture of DNA (Gnirke *et al.* 2009). This process involves mixing whole DNA extracts with pre-designed oligonucleotide ‘probes’, which will hybridize to genetically similar sequences in the extract ($\geq 80\%$ homology). All other (non-target) DNA in the sample is separated and discarded, and the DNA bound to the probes is eluted. The process can result in dramatic enrichment of the DNA sequences of interest – effectively concentrating target pathogen DNA from nearly undetectable pre-enrichment levels to $> 80\%$ post-enrichment concentrations. Notable examples of this include: (1) concentrating *Wolbachia* DNA (an intracellular bacteria) from whole insect DNA extracts (Kent *et al.* 2011); (2) herpesvirus genome enrichment from human saliva and CSF samples (Depledge *et al.* 2011); and, (3) capture of *Yersinia pestis* DNA from Black Death victims (Schuenemann *et al.* 2011). In all three of these studies, the enriched DNA was suitable for downstream deep-coverage sequencing of the target pathogen.

The uses of TGC in veterinary medicine remain unexplored, yet there are many potential applications of this technology in the field of animal health. For example, many feline clinical presentations have long been thought to be the result of an infectious etiology (i.e. fever of unknown origin, chronic lymphocytosis, feline poliomyelitis, CD4 T-cell leukemia) (Avery & Avery 2007). Yet despite many efforts, no pathogens have been associated with these syndromes. By targeting evolutionarily conserved genomic regions of pathogen taxa that cause similar disease states in other species, TGC may assist in the discovery of related feline pathogens, which have to date evaded detection.

Next-generation sequencing (NGS) produces non-specific sequence data from input DNA. One of the most informative applications of TGC is the subsequent use of enriched DNA for NGS (Mardis 2009; Summerer 2009). Due to the low pathogen:host genome ratio in extracted DNA (often 1:1x10⁶ or less), NGS of whole extracted DNA is an ineffective means of sequencing infectious agents from blood, tissue, or *in vitro* cell culture samples. However, with the ability of TGC to specifically enrich pathogen DNA to 80% or greater of a sample (up to 98% in one study), NGS can be used to effectively sequence entire pathogen genomes post-enrichment (Bos *et al.* 2011; Depledge *et al.* 2011; Hu *et al.* 2011).

This study was designed as a proof-of-concept to utilize TGC-NGS to detect and sequence diverse feline pathogens from *in vitro* and *in vivo* samples (Table 3.1). Our results illustrate strengths and weaknesses of the assay design, which will inform the optimization of the assay for future uses. We conclude that this novel approach to detecting and characterizing diverse pathogens from biological samples is a powerful and effective methodology with practical applications in clinical and research settings.

METHODS

Design of Capture Probes

Probes were designed for the capture of nucleic acids from the viral and bacterial feline pathogens listed in Table 3.1. Target pathogens were chosen because they fit one or more of the following characteristics: (1) relatively small genome size allowed redundancy in the capture design, providing enhanced sensitivity to divergent strains; (2) the pathogen is highly relevant to feline health, yet current assays often fail to detect or completely characterize infection due to latency and/or high mutation rate; and, (3) the virus family is involved in diverse pathologic processes (i.e. neoplasia) in one or more host species. Probes targeting RNA viruses were

included in the library design for future uses of this assay (only DNA was enriched and sequenced for this project).

Full-length and partial genetic sequences representing the known diversity of each target pathogen were downloaded from GenBank (www.ncbi.nlm.nih.gov/genbank/) (Table 3.2). Nucleotide alignments were constructed using default parameters in Muscle (Edgar 2004). Maximum-likelihood (ML) trees were built in MEGA v.5, using the Hasegawa-Kishino-Yano model of nucleotide substitution (Tamura *et al.* 2011) (trees not shown). Rate variation among sites was estimated with a two-category gamma distribution, allowing for invariant sites. Cluster support was estimated with 100 bootstrap replicates. One or more sequences were selected randomly for the probe sequence library from each polytomy and each monophyletic cluster of terminal nodes (Table 3.3).

All sequences in the probe library were cross-referenced to the domestic cat genome using the BLAST-like Alignment Tool (BLAT) (Kent 2002). Sequences with greater than 80% homology to the cat genome (2011-Felis_catus 6.2 (GenBank ID GCA_000181335.2)), were removed from the probe library to prevent non-specific capture of host chromosomal DNA. This was especially important for the exclusion of probes complementary to the endogenous FeLV sequences that exist in the cat genome.

The library sequences were uploaded into the eArrayXD capture probe design software (Agilent Technologies Inc., Santa Clara, CA). A custom library of 120 base-pair capture probes complementary to the selected pathogen sequences was designed at 4X coverage (probes tiled every 30 bases). This resulted in approximately 54,605 capture probes that are complementary to the sequences in the probe library (Table 3.2).

DNA Library Preparation, Target Enrichment, and Sequencing

Sixteen samples were selected to represent three categories of pathogen status: (1) positive controls - samples known to contain one or more target pathogen(s) based on previous diagnostic tests; (2) negative controls - samples known to be free of target pathogens (i.e. specific-pathogen free cats, naïve cell-cultures), and; (3) samples from animals with clinical presentation or serological results suggestive of pathogen involvement, but for which no infectious etiology has been identified (Table 3.4). DNA was extracted from blood, tissue, or cell culture samples from domestic cats, bobcats, and mountain lions using the DNeasy® Blood and Tissue Kit (Qiagen Inc., Valencia, CA). For each sample, 3µg of single or pooled DNA extracts were used to construct the DNA library according to the SureSelectXT Automated Target Enrichment for Illumina Paired-End Multiplexed Sequencing Protocol (Agilent Technologies Inc., Santa Clara, CA). Briefly, each DNA extract was sheared to produce short DNA fragments (~150 to 200bp). Adaptors were ligated to the ends of each fragment, and adaptor-specific primers were used to randomly amplify the DNA libraries prior to enrichment (4 to 6 PCR cycles).

The prepared DNA library was hybridized to the custom-designed capture probes for 24 hours at 65°C and unbound (non-target) DNA was washed away. Target DNA was separated from the capture probes and isolated. Unique barcodes (6-mer oligonucleotides) were ligated to the DNA fragments from each enriched sample for multiplex sequencing. All 16 samples were pooled and a non-indexed positive control was added to the samples (1% bacteriophage phiX174 DNA). The samples were sequenced together in a 16 sample multiplex as paired-end reads on a MiSeq instrument (Illumina, Inc., San Diego, CA).

Bioinformatics

Sequencing reads were assessed for quality, and separated by barcode and read direction by the MiSeq Reporter software (Illumina, Inc., San Diego, CA). All reads were trimmed on the 5'- and 3'-ends to remove low quality bases with a 5% error threshold (Drummond *et al.* 2012). Reads were aligned to reference sequences using the Burrows-Wheel Aligner (BWA) with default settings (Li & Durbin 2009). Alignment files created by BWA were sorted and merged using SAMtools (Li *et al.* 2009). All reads were aligned to the domestic cat genome to quantify and remove non-target sequences. It is unlikely that this process filtered all host-chromosomal reads because the currently available cat genome may not be completely sequenced, and regions of low complexity (i.e. tandem repeats) may not be efficiently aligned by BWA. The remaining 'non-cat reads' were then aligned to sequences of the target pathogens (i.e., sequences included in the custom capture probe library design) (Table 3.4).

All reads were also screened against a database of all viral sequences from GenBank. From this 'pan-viral' search, hit-rates were calculated by dividing the number of reads that matched or 'hit' a pathogen taxon in the database by the total number of reads in the sample. Non-specific or spurious hits in the database were accounted for by subtracting the corresponding hit-rates from the negative control samples (K. Fischer, manuscript in preparation).

RESULTS

Paired-end Sequencing

The 16 multiplexed samples produced 16,572,796 total reads of 151base pairs in length (Table 3.5). Eighty nine percent of bases had quality scores of Q30 or greater (error probability ≤ 0.001). After trimming low quality bases from the 5'- and 3'-ends of the reads, the mean length was 146 base pairs (standard deviation = 17.1 base-pairs). The number of reads per

sample ranged from 639,092 to 3,098,822 (median = 790,189). The distribution of reads across samples was non-uniform – two samples (Samples #8 and #9) accounted for 35% of the total reads. These samples had the lowest percentage of reads map to the cat genome (3% and 1% respectively) (Table 3.5). The other samples had 51% to 82% of reads map to the cat genome, indicating that the majority of post-enrichment DNA was non-target genomic DNA in most samples (Table 3.5). 20% to 46% of the total reads for each sample did not map to the cat genome or the target pathogen reference sequences (Table 3.5).

Detection and Characterization of Pathogens

Ten (of 16) samples were known to have one or more target pathogen(s) based on previous diagnostic assays (16 total pathogens) (Table 3.4). At least one target pathogen was detected in nine of these samples. All target pathogens were detected, except PLVA, which was not detected by BWA alignment or by the pan-viral screen in any of the three samples in which it was present. Nine pathogens were detected that were not known to be present prior to this study. The percentage of each targeted genome region that was covered by sequencing reads varied from zero to 100%, with an average coverage of 84% among the 22 pathogens detected (Table 3.4).

No reads from any samples were aligned to the gamma herpesvirus or T-lymphotropic virus sequences included for discovery of related feline pathogens. These viral taxa had low hit-rates for some samples in the pan-viral screen but the number and pattern of hits suggests these were non-specific (Figure 3.1). Similarly, reads homologous to many non-target viral taxa were identified in the pan-viral screen. None of these associations appear to be specific either, as attempts to align the hit-reads to reference sequences from the associated viral taxa failed.

Therefore, no results from either analysis were consistent with the identification of novel feline pathogens.

All samples (including the negative controls) had reads that aligned to the FHV-1 and phiX174 reference sequences. It is likely that the ubiquitous FHV-1 reads originated from Samples #8 and #9 but were incorrectly assigned to the other samples due to an error that occurred during the de-multiplexing of reads in the MiSeq software (A. Becker personal communication). This would also explain the presence of the phiX174 reads as the phiX174 DNA was not barcoded and should not have appeared in any set of sample reads. The error likely resulted in the incorrect assignment of a low percentage of reads from each sample. Because pathogen reads represented a small proportion of the total reads for most samples, zero or very few incorrectly assigned reads were detected for most pathogens (i.e., PLVB reads in domestic cat samples). However, because FHV-1 reads from Samples #8 and #9 represented nearly 25% of the total sequencing reads, the de-multiplexing error resulted in a detectable number of FHV-1 reads in all samples (including all negative controls).

DISCUSSION

This proof-of-concept study was designed to detect and characterize multiple feline pathogens using targeted genome capture and next-generation sequencing. To our knowledge this is the first application of these technologies as a large multiplex pathogen screening and discovery tool. The custom capture probe library included divergent DNA sequences selected to represent the known diversity of each target pathogen (Table 3.3). Our results reflect both strengths and weaknesses of the assay as currently implemented, and suggest that if optimized, this approach could greatly expand our ability to generate pathogen genetic data in clinical and research settings.

Six of the seven target pathogen taxa known to be present in at least one sample were detected by this assay. For each of these, the majority of the targeted genome region was sequenced (Table 3.2). This included single and pooled DNA extracts, with up to three pathogens per sample originating from *in vitro* or *in vivo* sources. The total number and percent of on-target reads varied widely between samples (Table 3.5). The two samples over-represented during sequencing (Samples #8 and #9) both contained DNA extracted from FHV-1 cell cultures, and together these two samples had almost four million reads mapped to FHV-1. The high ratio of viral to genomic DNA in these cultures probably contributed to this outcome.

At least one pathogen was detected in all *in vivo* samples except for samples #3 and #11. Some of the detected pathogens were known to be present prior to this study from previous diagnostic assays (Table 3.2). Other pathogens were detected by the TGC-NGS assay through the mapping of reads to reference sequences. In total, 17 pathogens were detected in the 11 test samples (negative controls and *in vitro* samples not included). Puma lentivirus clade A was not detected in any of the three samples in which it was known to be present. This cannot be explained by a lack of homology to the capture probes because full-genome PLVA sequences generated from two of the three samples were included in the probe design. It is more likely that the amount of PLVA DNA in these samples was too low to be detected by the assay. The proviral copy number of PLVA in mountain lions is low (approximately one copy per 1×10^5 cells), rendering approximately 50% of samples undetectable after 40 cycles of PCR (S. Templin, unpublished data). At this copy number, the ratio of proviral DNA to cat chromosomal DNA is approximately $1:3 \times 10^{11}$ ($\sim 3 \times 10^9$ bases/cat genome and two genomes/cell), and as few as 40 proviral copies may have been present in the DNA input for enrichment. It is possible that either the capture probes failed to hybridize and enrich this low level of target DNA, or that

enrichment occurred but was not sufficient to be sequenced by the ‘shallow’ sequencing technique performed. In either case, our results suggest that a typical PLVA infection in mountain lions is below the level of detection for this assay as currently implemented. Future uses of this assay could use deeper sequencing and/or an optimized enrichment protocol to improve the sensitivity of the assay for PLVA and other low copy-number pathogens.

Previous serological results suggested that bobcats in Colorado may be infected with PLV-like and FFV-like viruses known to circulate in other bobcat populations or in sympatric felids (Bleiholder et al, unpublished data). Previous attempts to identify these viruses have been unsuccessful, and therefore two samples were included to determine if the TGC-NGS assay could detect divergent or low-copy pathogens undetected by conventional PCR. However, despite including diverse PLV and FFV sequences in the capture probe library, no reads from these bobcat samples mapped to any of the reference sequences included for these pathogens.

Two other samples, each containing pooled DNA from two domestic cats with either lymphoma or T-cell leukemia, were also included to search for infectious etiologies for these common feline disease states. Viral pathogens produce clinically similar pathologies in other species – T-cell leukemia can result from HTLV-1 infection in humans and many species harbor gamma herpes viruses that initiate neoplastic conditions such as lymphomas. We included HTLV-1, STLV, and gamma herpes virus sequences in the capture probe library, attempting to discover homologous feline viruses. We did not assemble reads to reference sequences consistent with these hypothesized viral etiologies. Therefore, our initial attempt to identify novel pathogens with this assay was not successful.

CONCLUSION

We have shown that TGC and NGS can detect multiple pathogens from individual and pooled samples. In addition, sequencing coverage allowed for accurate characterization of previously undetected pathogens in several samples. While successful as a proof-of-concept, the assay could be improved by optimizing the capture probe design and utilizing deeper sequencing to increase the sensitivity for detecting divergent or low copy number pathogens.

The results of this project suggest that this assay design could be modified for specific research or clinical applications. For example, the number of probes per target pathogen could be increased if the goal is to investigate the genetic diversity among a smaller set of target pathogens. This may improve the capture efficiency and sequencing depth for each pathogen, enabling analyses into, for example, molecular evolution and intra-host diversity. Alternatively, by targeting only conserved regions of pathogens, instead of full-length genomes, more target pathogens could be included in a single capture probe library. This could allow researchers to ‘cast a big net’ when screening for pathogens within individuals or populations. The sensitivity of any probe-design modifications would be maximized by utilizing a deep-sequencing technique, which can produce ten-fold more reads per run than the method used here. This would theoretically enable the sequencing of even extremely rare nucleic acids captured by probe-hybridization.

Finally, while we were not successful in identifying novel feline pathogens, we consider this approach to pathogen discovery valid and deserving of further investigation. While we were limited to testing a small number of samples in this study, we will continue to utilize this approach to identify unknown pathogens in wildlife and to search for infectious etiologies to common diseases of unknown origin in domestic cats.

Table 3.1: Summary of the pathogens for targeted DNA enrichment and sequencing.

Pathogen family	Infectious agents	Genome	Rationale for inclusion
<i>Caliciviridae</i>	Feline calicivirus ¹	7.7 kb RNA	Persistent infection; rapidly mutates; pathogenicity varies greatly by strain
<i>Coronaviridae</i>	Feline enteric coronavirus ¹	29.1 kb RNA	Suspected genetic basis for switch from apathogenic to pathogenic phenotype
<i>Herpesviridae</i>	Feline herpesvirus* Gamma herpesvirus ²	134kb DNA	Potential co-factor in lymphoma; highly prevalent; latent infection; improved detection methods needed
<i>Mycoplasmataceae</i>	<i>M. haemominutum</i> * <i>M. haemofelis</i> <i>M. turicensis</i>	1.1Mb DNA ³	Chronic transmissible infections possible; high morbidity and mortality if untreated; common in wild cats
<i>Retroviridae</i> ⁴	Feline leukemia virus*	9.6 kb RNA	Often latent; associated with tumors; exogenous infection difficult to characterize
	Feline immunodeficiency virus*	9.1kb RNA	Latent infection; mutates rapidly; role in a variety of disease states
	Puma lentivirus*	9.1kb RNA	Difficult to diagnose; emerging pathogen in endangered Florida panther; cross-species transmission and host range expansion documented
	Feline foamy virus*	11kb RNA ³	Suspected co-factor in many feline diseases; high prevalence in wild cats
	Human and Simian T-lymphotrophic viruses ²	9.0 kb RNA	Infection in humans results in CD4 T-cell leukemia
* - The six target pathogens known to be present in at least one sample included in this study. 1 - Pathogens included in the bait design for future applications; 2 - Infectious agents in other host species that result in disease states similar to clinical presentations in cats for which no feline pathogens have been identified; 3- Only a portion of the genome was targeted; 4 - The DNA copy (provirus) of the RNA genome was targeted.			

Table 3.2: Summary of the custom capture probe library design.

Pathogen	# Sequences Analyzed	Base Pairs Analyzed	# Probes included
Feline calicivirus	481	493,030	2,594
Feline coronavirus	430	431,290	10,161
Feline foamy virus	30	142,620	3,566
Feline herpesvirus	1	135,800	9,054
Feline immunodeficiency virus	1,531	1,273,790	3,888
Puma lentivirus A & B	63	531,910	17,731
Feline leukemia virus	25	50,400	143
Bovine immunodeficiency virus	3	24,000	283
Equine infectious anemia virus	1	8,300	275
Lion lentivirus	70	57,050	1,155
Simian immunodeficiency virus	2	19,140	639
Gamma herpesviruses	40	42,900	1430
Human T-lymphotropic virus	12	80,800	1,164
Simian T-lymphotropic virus	13	101,080	1,802
<i>Mycoplasma haemofelis</i>	26	35,180	274
<i>Mycoplasma haememinitum</i>	41	55,020	227
<i>Mycoplasma turicensis</i>	25	32,130	219
TOTAL	2,794	3,515,160	54,605

Table 3.3: Pathogen sequences included in the custom capture probe library. Each published sequence is identified by the GenBank accession number. Unpublished PLV sequences are listed by the isolate identifier from Chapter 2. Unpublished FFV sequences from four mountain lions provided by M. Lochelt.

Calicivirus	Herpesvirus	Mycoplasma	Gammaretrovirus
FCV	FHV-1	<i>M. haememinitum</i>	FeLV
AF479590	FJ478159	AY150978	AB060732
AY560115	Gammaherpesvirus	AY297712	AY374199
AY560117	AAC55644	DQ825446	AY374202
D31836	AAC55645	DQ825452	AY374206
DQ910788	AAC55648	FJ004275	AY374207
DQ910791	AAC57974	<i>M. haemofelis</i>	AY374210
DQ910794	AAC57976	DQ157159	AY374217
DQ910795	AAC59452	DQ825447	AY374219
GU214989	AAC59454	DQ825453	AY374220
JN210884	AAD30141	EU145745	AY374222
JN210890	AAD30142	U88563	AY706368
M86379	AAD56945	U95297	AY706369
	AAF23082	<i>M. turicensis</i>	FLU03213
Coronavirus	AAF23083	DQ157150	GU731413
FCoV	AAG10783	DQ157153	M18247
AY994055	AAG21351	DQ464424	NC_001940
EU663798	AAG21352	EU789559	
EU663833	AAG23140	JQ689949	Spumavirus
EU663842	AAG23142		FFV
EU663872	AAG39060	Deltaretrovirus	AB052796
EU663887	AAG39061	HTLV-1	AB052797
EU663915	AF81664	AY563953	AB052798
EU663949	AF141888	J02029	AJ564746
EU663950	AF141889	L02534	NC_001871
EU663958	AF204166	M86840	U21247
EU663990	AF250884	STLV	U78765
EU663994	AF250886	AF074966	U78766
EU663995	AF282937	AF391797	x94bet
EU664000	AF282943	AY217650	x94envSU
EU664018	AF290601	AY222339	x103bel
EU664031	AJ25157	AY590142	x301pol
EU664093	AY270026	U90557	x431envSU
EU664102	CAB61753		x431gag
EU664170	CAB61754		Y08851
EU664177	DQ093191		
FJ938055	DQ120516_1		
FJ938058	DQ120516_2		
FJ938060	F346488		
HQ012367	GQ169129_1		
HQ012368	GQ169129_2		
HQ012370	NC_001826_1		
HQ012371	NC_001826_2		
HQ012372			
JN183882			

Table 3.3 continued

Lentivirus			
FIV		PLVA	PLVB
AB027301	DQ250716	A.Lru1.CA-north	B.Pco2.CA-north
AB027302	E03581	A.Lru2.CA-north	B.Pco4.CA-south
AB514980	EF413006	A.Lru3.CA-north	B.Pco13.CA-south
AB515047	EF413015	A.Lru4.CA-north	B.Pco14.CA-south
AB515092	EF413017	A.Lru5.CA-north	B.Pco15.CA-south
AB515116	EF667010	A.Lru6.CA-south	B.Pco16.CA-south
AB515133	EU375593	A.Lru7.CA-south	B.Pco17.CA-south
AB515155	FJ654797	A.Lru8.CA-south	B.Pco18.CA-south
AB515224	FJ654807	A.Lru9.CA-south	B.Pco19.CA-south
AB515247	FJ654824	A.Lru10.CA-south	B.Pco20.CA-south
AF047703	FJ654904	A.Lru11.CA-south	B.Pco21.CA-south
AF474246	FJ654965	A.Lru12.CA-south	B.Pco22.CO-east
AF531054	GQ339807	A.Lru13.CA-south	B.Pco23.CO-west
AF531058	GQ339831	A.Lru16.FL	B.Pco24.CO-east
AF531061	HQ456792	A.Lru17.FL	B.Pco25.CO-east
AF531066	HQ456793	A.Lru18.FL	B.Pco26.CO-east
AJ304951	M25381	A.Lru19.FL	B.Pco27.CO-east
AJ304972	M36968	A.Lru20.FL	B.Pco28.CO-west
AY196342	M59418	A.Pco1.CA-north	B.Pco29.CO-west
AY196343	U02399	A.Pco2.CA-north	B.Pco30.CO-west
AY220050	U03982	A.Pco5.CA-south	B.Pco603.MT
AY220063	U11820	A.Pco6.CA-south	B.Pco604.MT
AY220074	U53757	A.Pco7.CA-south	B.Pco605.MT
AY369378	X57000	A.Pco8.CA-south	B.Pco606.MT
AY600517	X57002	LLV	B.Pco607.WY-west
D31929	X68019	AY552696	B.Pco608.WY-west
D31938	Z96126	AY552697	B.Pco609.WY-west
DQ083386	Z96133	AY552702	B.Pco610.WY-west
	BIV, ELAV, SIV	AY552722	B.Pco611.WY-west
	M32690	AY552739	B.Pco612.WY-west
	AF247394	AY552740	B.Pco613.WY-east
	NC_001549	EF106736	B.P901.FL
	NC_004455	EF106737	B.P910.FL
		EU117991	B.P1033.FL
		EU117992	B.P1094.FL
		U56928	B.Pco1123.FL
			B.Pco1124.FL
			B.PLV1695.BC

Table 3.4: Samples, target pathogens, and alignment results. Outlined cells in the table indicate pathogens known to be present prior to this study. The percent of each targeted genome region sequenced is indicated for all detected pathogens. Mean read depth for samples with > 90% sequencing coverage indicated in parentheses. Dark shading indicates targets for pathogen discovery predicted to be present in some samples based on previous serological assays or clinical presentations (none of the predicted pathogens were identified by the TGC/NGS assay).

Sample	Host Species	Source of DNA	Alignment of Reads to Target Pathogens										
			FFV	FHV-1	FIVA	FIVC	PLVB	PLVA	FeLV ¹	M. haemo ²	HTLV-like	Gamma	
1	Puma (CA)	whole blood	97% (5.9)								23%		
2	Puma (CA)	whole blood	99% (13.6)					56%					
3	Puma (FL)	whole blood											
4	Puma (CO)	whole blood	71%					72%			86%		
5	Domestic cat	cell culture			100% (672.3)								
6	Domestic cat	cell culture				100% (3872.6)							
7	Domestic cat	whole blood				99% (16.4)							
8	Domestic cat	cell culture		99% (2212.0)	74%		93% (9.5)						
9	Domestic cat	cell culture		99% (2170.5)									
10	Bobcat (CO)	whole blood		97% (18.2)							100% (724.2)		
11	Bobcat (CO)	whole blood											
12	Domestic cat	whole blood	60%	97% (7.3)									
13	Domestic cat	lymphoma	54%		98% (72.0)								
14	Puma (CA)	spleen						92% (5.3)		100% (15.9)			
15	Domestic cat	PBMC	Negative Control										
16	Domestic cat	cell culture	Negative Control										

1 - only the long-terminal repeat region of the genome was targeted; 2 - only the 16s ribosomal RNA region of the genome was targeted; Abbreviations: FFV - Feline foamy virus; FHV-1 - Feline herpesvirus-1; FIVA - Feline immunodeficiency virus clade A; FIVC - Feline immunodeficiency virus clade C; PLVB - Puma lentivirus clade B; PLVA - Puma lentivirus clade A; FeLV - Feline leukemia virus; M. haemo - *Mycoplasma haemofelis*; HTLV - Human T-lymphotropic virus; Gamma - Gamma herpesvirus; CA - California; FL - Florida; CO - Colorado; PBMC - peripheral blood mononuclear cells

Table 3.5: The number (and percent) of reads per sample that mapped to the cat genome, pathogen reference sequences, or were unmapped.

	Sample 1	Sample 2	Sample 3	Sample 4	Sample 5	Sample 6	Sample 7	Sample 8
	Puma	Puma	Puma	Puma	In vitro	In vitro	Domestic cat	In vitro
Total reads	710956	865136	622356	754766	704662	815416	704810	2713940
Cat genomic reads	461083 (64.9)	549047 (63.5)	375688 (60.4)	488779 (64.8)	521589 (74.0)	411768 (50.5)	562273 (79.8)	84319 (3.1)
Target pathogen reads	2234 (0.3)	2311 (0.3)	1688 (0.3)	3814 (0.5)	47094 (6.7)	239758 (29.4)	4695 (0.7)	2040620 (75.2)
Unmapped reads	247639 (34.8)	313778 (36.3)	244980 (39.4)	262173 (34.7)	135979 (19.3)	163890 (20.1)	137842 (19.6)	589001 (21.7)
	Sample 9	Sample 10	Sample 11	Sample 12	Sample 13	Sample 14	Sample 15	Sample 16
	In vitro	Bobcat	Bobcat	Domestic cat	Domestic cat	Puma	Domestic cat	Domestic cat
Total reads	3098822	905292	639092	784480	874376	745644	795898	837150
Cat genomic reads	29957 (1.0)	573426 (63.3)	411822 (64.4)	641557 (81.8)	605693 (69.3)	400540 (53.7)	630115 (79.2)	673740 (80.5)
Target pathogen reads	1878706 (60.6)	26507 (2.9)	10982 (1.7)	8860 (1.1)	7496 (0.9)	5271 (0.7)	2610 (0.3)	3002 (0.36)
Unmapped reads	1190159 (38.4)	305359 (33.7)	216288 (33.8)	134063 (17.1)	261187 (29.9)	339833 (45.6)	163173 (20.5)	160408 (19.2)

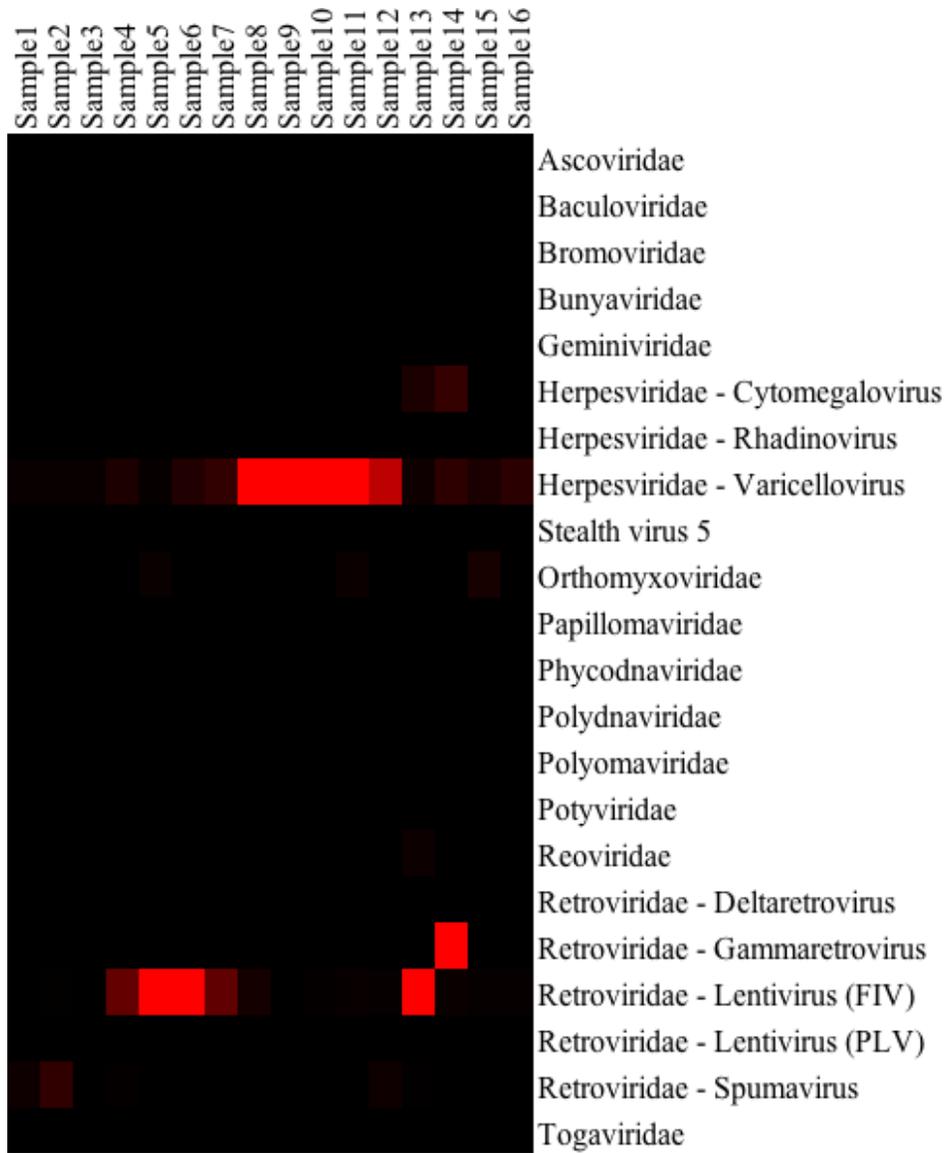


Figure 3.1: Heat map depicting the results of the pan-viral database screen. Samples are listed on the x-axis and viral taxa with at least one 'hit' are listed on the y-axis. The intensity of color in each cell reflects the proportion of reads in each sample that hit a particular viral taxa ranging from low (black) to high (red). Figure produced by B. Chan and K. Fisher.

LITERATURE CITED

- GenBank Sequence Database. National Center for Biotechnology Information. <http://www.ncbi.nlm.nih.gov/genbank/>.
- Avery AC, Avery PR (2007) Determining the significance of persistent lymphocytosis. *Veterinary Clinics of North America: Small Animal Practice* **37**, 267-282.
- Bos KI, Schuenemann VJ, Golding GB, *et al.* (2011) A draft genome of *Yersinia pestis* from victims of the Black Death. *Nature* **478**, 506-510.
- Brown MA, Cunningham MW, Roca AL, *et al.* (2008) Genetic characterization of feline leukemia virus from Florida panthers. *Emerging Infectious Diseases* **14**, 252-259.
- Depledge DP, Palser AL, Watson SJ, *et al.* (2011) Specific Capture and Whole-Genome Sequencing of Viruses from Clinical Samples. *Plos One* **6**.
- Drummond A, Ashton B, Buxton S, *et al.* (2012) Geneious Pro 5.6 Available from <http://www.geneious.com/>.
- Edgar RC (2004) MUSCLE: multiple sequence alignment with high accuracy and high throughput. *Nucleic Acids Research* **32**, 1792-1797.
- Evermann J, Sellon R, Sykes J (2012) Laboratory Diagnosis of Viral and Rickettsial Infections and Clinical Epidemiology of Infectious Diseases. In: *Infectious diseases of the dog and cat* (ed. Greene CE). Elsevier Saunders, St. Louis, MO.
- Gillespie SH (2002) Evolution of drug resistance in *Mycobacterium tuberculosis*: Clinical and molecular perspective. *Antimicrobial Agents and Chemotherapy* **46**, 267-274.
- Gnirke A, Melnikov A, Maguire J, *et al.* (2009) Solution hybrid selection with ultra-long oligonucleotides for massively parallel targeted sequencing. *Nature Biotechnology* **27**, 182-189.
- Hu B, Xie G, Lo CC, Starckenburg SR, Chain PSG (2011) Pathogen comparative genomics in the next-generation sequencing era: genome alignments, pangenomics and metagenomics. *Briefings in Functional Genomics* **10**, 322-333.
- Kapil S, Lamm C (2008) Emerging and Reemerging Viruses of Dogs and Cats. In: *Veterinary Clinics of North America*. Elsevier, Inc., Philadelphia, PA.
- Kent BN, Salichos L, Gibbons JG, *et al.* (2011) Complete Bacteriophage Transfer in a Bacterial Endosymbiont (*Wolbachia*) Determined by Targeted Genome Capture. *Genome Biology and Evolution* **3**, 209-218.
- Kent W (2002) BLAT - the BLAST-like alignment tool. *Genome Research* **12**, 656-664.
- Li H, Durbin R (2009) Fast and accurate short read alignment with Burrows-Wheeler Transform. *Bioinformatics* **25**, 1754-1760.
- Li H, Handsaker B, Wysoker A, *et al.* (2009) The sequence alignment/map (SAM) format and SAMtools. *Bioinformatics* **25**, 2078-2079.
- Mardis ER (2009) New strategies and emerging technologies for massively parallel sequencing: applications in medical research. *Genome Medicine* **1**, 40.
- McCarthy AJ, Shaw MA, Goodman SJ (2007) Pathogen evolution and disease emergence in carnivores. *Proceedings of the Royal Society B-Biological Sciences* **274**, 3165-3174.
- Meli ML, Cattori V, Martinez F, *et al.* (2009) Feline Leukemia Virus and Other Pathogens as Important Threats to the Survival of the Critically Endangered Iberian Lynx (*Lynx pardinus*). *Plos One* **4**.

- Parrish CR, Holmes EC, Morens DM, *et al.* (2008) Cross-species virus transmission and the emergence of new epidemic diseases. *Microbiology and Molecular Biology Reviews* **72**, 457-470.
- Rambaut A, Posada D, Crandall KA, Holmes EC (2004) The causes and consequences of HIV evolution. *Nature Reviews Genetics* **5**, 52-61.
- RoelkeParker ME, Munson L, Packer C, *et al.* (1996) A canine distemper virus epidemic in Serengeti lions (*Panthera leo*). *Nature* **379**, 441-445.
- Schuenemann VJ, Bos K, DeWitte S, *et al.* (2011) Targeted enrichment of ancient pathogens yielding the pPCP1 plasmid of *Yersinia pestis* from victims of the Black Death. *Proceedings of the National Academy of Sciences of the United States of America* **108**, E746-E752.
- Summerer D (2009) Enabling technologies of genomic-scale sequence enrichment for targeted high-throughput sequencing. *Genomics* **94**, 363-368.
- Tamura K, Peterson D, Peterson N, *et al.* (2011) MEGA 5: Molecular Evolutionary Genetics Analysis Software.
- Woolhouse MEJ, Haydon DT, Antia R (2005) Emerging pathogens: the epidemiology and evolution of species jumps. *Trends in Ecology & Evolution* **20**, 238-244.

CONCLUSION

Pathogens and their hosts coexist in complex relationships influenced by ecological and evolutionary forces across all levels of biological organization – from intricate molecular interactions to landscape-level processes. Molecular analyses, such as those described in Chapter One and Chapter Two, provide a mechanism for investigating the causes and effects of pathogen-host dynamics. Though these chapters focused on questions specifically related to feline immunodeficiency virus in non-domestic cats, the approaches, methods, and analyses described can be applied to questions in many other host-pathogen systems.

However, limitations inherent in traditional methods for generating genetic data (i.e. PCR), can render the use of molecular analyses difficult or ineffective in some systems. New technologies may make it possible to circumvent such limitations, increasing the depth and breadth of genetic data that can be collected from complex biological samples. In Chapter Three we described the use of targeted genome capture and next-generation sequencing to detect and sequence diverse feline pathogens from single and pooled biological samples. The strengths and weaknesses in this approach, as identified by this pilot project, will be useful to improve future studies.

The collective work described in this dissertation follows many like it in the past. The goal appears simple: to explore a small part of the natural world in order to further our understanding of why that one small part works the way it does. And yet, what at first appears simple is indeed very complex. This I have now learned. It is probably this very contradiction, and the elusive nature of what is not yet known, that keeps us always wanting, working, and searching for more...