Distribution Agreement

In presenting this thesis as a partial fulfillment of the requirements for a degree from Emory University, I hereby grant to Emory University and its agents the non-exclusive license to archive, make accessible, and display my thesis in whole or in part in all forms of media, now or hereafter known, including display on the world wide web. I understand that I may select some access restrictions as part of the online submission of this thesis. I retain all ownership rights to the copyright of the thesis. I also retain the right to use in future works (such as articles or books) all or part of this thesis.

Signature:

Elizabeth Rogawski April 1, 2010

Tracking and Jumping:
A Cophylogenetic Analysis of Host Switching in the Lyssaviruses

by

Elizabeth Tacket Rogawski

Adviser Dr. Leslie A. Real

Department of Biology

_____
Dr. Leslie A. Real
Adviser

_____
Dr. Lance A. Waller
Committee Member

_____
Dr. Jacobus de Roode
Committee Member

April 1, 2010
Date

Tracking and Jumping:
A Cophylogenetic Analysis of Host Switching in the Lyssaviruses

By

Elizabeth Tacket Rogawski

Adviser Dr. Leslie A. Real

An abstract of
A thesis submitted to the Faculty of Emory College of Arts and Sciences
of Emory University in partial fulfillment
of the requirements of the degree of
Bachelor of Sciences with Honors

Department of Biology

2010

Abstract

Tracking and Jumping:
A Cophylogenetic Analysis of Host Switching in the Lyssaviruses

By Elizabeth Tacket Rogawski

Host switching, in which an infection by a pathogen in a novel host results in sustained transmission, is an infrequent phenomenon yet is responsible for viral zoonoses that have caused many emerging infectious diseases in humans. RNA viruses of the *Lyssavirus* genus, including the Rabies virus, are found in bat host reservoirs, and the frequency and risk factors for host switching among these associations have not yet been characterized. Since the lyssaviruses diverged after their hosts speciated, a history of cospeciation can be rejected, and instead this study distinguishes between *host tracking*, in which host and pathogen divergences are tied to each other, and *host jumping*, which describes switches that are not constrained by host phylogeny. The study aims to identify host jumps in lyssavirus history and to characterize the influence of genetic and geographic distance between hosts as determinants of successful host jumping. Lyssavirus and bat host phylogenies were generated in BEAST v1.5.2, and host jumps were identified in TreeMap v2.02β. Genetic distances between hosts and overlap of geographic bat ranges for identified host jumps were then compared to the same distances for random pairings of hosts. Eight host jumps were identified to explain the current host-virus associations. Genetic similarity between donor and recipient hosts does not appear to constrain successful host jumping. Host jumps occurred between both closely related and more distantly related hosts, and the genetic distances between hosts of identified jumps were not significantly smaller than those for random pairings of hosts. Conversely, host jumps were more common between hosts with greater overlapping ranges, and hosts involved in jumps generally shared similar foraging and roosting habitats. While genetic similarity may also have an impact, these results suggest that geographic proximity to new hosts and the number and intensity of contacts between bat species are the driving factors in host jumping events.

Tracking and Jumping:
A Cophylogenetic Analysis of Host Switching in the Lyssaviruses

By

Elizabeth Tacket Rogawski

Adviser Dr. Leslie A. Real

A thesis submitted to the Faculty of Emory College of Arts and Sciences
of Emory University in partial fulfillment
of the requirements of the degree of
Bachelor of Sciences with Honors

Department of Biology

2010

Acknowledgements

I would like to thank foremost my advisor, Dr. Leslie Real, for welcoming me as his first undergraduate in several years and for guiding me through a project that has not only led to meaningful results, but that has also broadened my interests in biology and the evolution of disease. Without his extensive knowledge, experience, and many collaborations with others at the forefront of the field, this project would not have been possible.

I would also like to thank the members of the Real Lab, including Dr. Vijay Panjeti for providing more sensible alternatives to my brute force techniques and for helping me to interpret my data. I am grateful for Dr. Gavino Puggioni's mastery of R and could not have generated such beautiful figures without him. To Anand Bhardwaj, co-inhabitant of our Office of Phylogenetic Phun, I could not be more thankful for the endless support through the constant trivial roadblocks. Finally, thank you to Nelle Couret for her expertise in ArcGIS. The Real Lab has provided consistent support, guidance, and perhaps most importantly, stimulating conversation over an endless cup of coffee.

I am also grateful to the many others who have contributed ideas and advice to shape my project, most importantly my committee members, Dr. Lance Waller and Dr. Jaap de Roode, as well as other collaborators, Dr. Nicole Gerardo, Quinn McFrederick, and Daniel Streiker.

Table of Contents

Introduction

Emerging Zoonotic Diseases

Infectious diseases have plagued human societies since early history. Disease has consistently been a defining factor of society, influencing not only mortality rates, but also culture, art, and politics. Recently, most notably with the development of penicillin and successful vaccines, some have declared triumph over infectious diseases (Barrett, Kuzawa et al. 1998). Yet as we continue to develop successful prevention and treatment strategies, barriers to the eradication of infectious diseases remain high. From the evolution of drug resistance to inadequate primary health care in developing nations, the challenges we face in controlling infectious diseases continue to warrant our attention and resolve.

A major threat concerns emerging and re-emerging infectious diseases, which are largely unpredictable and can initiate widespread epidemics. Emerging diseases are defined as those which have recently increased in incidence or geographic area, been reported in new regions, or been identified for the first time (Cleaveland, Laurenson et al. 2001). The US Centers for Disease Control and Prevention (CDC) identified 29 novel pathogens that emerged between 1973 and 1995, which add to a much longer list of re-emerging diseases with continued importance (Barrett, Kuzawa et al. 1998). While public health officials have been successful in implementing measures to minimize the risks and impacts of these diseases, the mechanisms driving disease emergence are thought to be roughly divided between ecology and evolution and are not well understood.

Commonly, emerging diseases are the result of zoonotic pathogens, which are maintained in animal reservoirs and can occasionally infect and cause disease in humans. Almost two-thirds (61%) of all human diseases are zoonotic (Cleaveland, Laurenson et al. 2001) and of 175 pathogens associated with emerging diseases, 132 (75%) are zoonotic (Paul-Pierre 2009). These statistics indicate the major impact zoonotic diseases have on both human health and that of animal populations.

The influence of humans on ecosystem processes through globalization, climate change, and new agricultural practices has created novel opportunities for zoonotic diseases to emerge in human populations (Paul-Pierre 2009). Because viral diseases have a high rate of evolution and are generally difficult to control, viral zoonoses are a sustained and growing threat. These pathogens pose a particularly serious risk for disease emergence due to the potential for viruses currently in animal reservoirs to adapt to and cause disease in human hosts. Because many human diseases are zoonotic, infectious disease control efforts must focus not only on human health, but also on the dynamics of the disease in animals. A greater understanding of how and when zoonotic diseases emerge is vital to our current infectious disease control strategies.

Host Switching

For a zoonotic disease to emerge in human populations, the pathogen must first be transmitted to humans from an animal reservoir. To then become established in human populations, the pathogen must experience sustained transmission between humans. This process of transfer from one host reservoir to another that results in sustained transmission is termed host switching and applies to diseases transmitted both to humans

and among animal species. The emergence of Human Immunodeficiency Virus (HIV), which originated in simian hosts, is an example of host switching that has caused devastating disease in humans worldwide. Similarly, recently emerging diseases, including highly-pathogenic strains of H1N1 and H5N1 influenza A, originated in wildlife hosts and have become high-impact diseases in humans through the complicated process of host switching (Parrish, Holmes et al. 2008). While successful host switches are generally considered to be rare, the potential of these events to cause disease and their impact on human populations are demonstrated by many of today's major infectious disease concerns.

Host switching succeeds when spillover infections into a new recipient host result in sustained transmission in the new population, creating a new host reservoir. While spillover infections often cause disease in the recipient, these infections are contained because transmission within the recipient species is rare, and the disease does not persist in the population. On the other hand, sustained transmission of the disease in the new species, where $R_0 > 1$, poses a persistent threat and can severely impact the dynamics of a population. At minimum, this process of transmission requires contact between the old and new host. Most agree that adaptation of the virus to the new host is also essential, although the absolute necessity or degree of adaptation required is unknown (Parrish, Holmes et al. 2008).

Host switching is rare due to a variety of ecological and genetic barriers that impede interactions between hosts and between the host and the pathogen (Kuiken, Holmes et al. 2006). In many cases, minimal cross-species contact between potential hosts may be a sufficient barrier to host switching. The increase in spillover infections

and emerging zoonotic pathogens in humans may be partially explained by the increased success of cross-species transmission as contact patterns are modified by globalization and other human-induced changes (Parrish, Holmes et al. 2008).

After initial transmission of the pathogen to the recipient host, it is generally accepted that due to genetic barriers within the host, multiple and complex adaptive changes in the virus are required before it can be maintained in the recipient population. Genetic differences between hosts can impact viral success in receptor binding, viral fusion and entry, genome replication and expression, and production and shedding of infectious virus. Innate antiviral responses in the new host may also prevent viral establishment (Kuiken, Holmes et al. 2006). Genetic mutations that overcome these barriers and optimize the ability of the virus to infect a new host will likely reduce its fitness in the donor host, and lower-fitness intermediates are prone to extinction (Parrish, Holmes et al. 2008). Therefore, it is improbable that multiple adaptive mutations will occur to allow the virus to cross "low-fitness valleys" between peaks of fitness in the old and new host (Parrish, Holmes et al. 2008). Although the exact criteria for success is unknown and varies among pathogens, this rare process is necessary for host switching.

Cophylogenetics and Cospeciation

While host switches are commonly studied as individual events in conjunction with the emergence of a human disease, studies of host switching patterns in related species can also be instructive in determining the history of associations between pathogens and their hosts. In turn, these patterns can be useful in predicting where future host switches may occur and possibly in directing public health interventions.

Cophylogenetics uses a range of techniques to compare host and parasite phylogenies to determine ancestral associations and the events that have led to current host-pathogen associations. Typically, researchers question if cospeciation has occurred in the history of a specific association, which indicates that adaptation and character change have been correlated between host and pathogen for an extended period of time (Page 2003).

Cospeciation occurs when pathogens diverge with their host population, resulting in the joint speciation of their lineages (Page 2003; Holmes 2004). This phenomenon can be inferred from congruent phylogenies, in which host and pathogen trees show the same branching pattern and mirror each other. Incongruent phylogenies, on the other hand, can be attributed to a variety of processes, including but not limited to host switching (Page 2003; Holmes 2004). Five plausible events are commonly considered to explain current associations: *cospeciation*, or codivergence of host and pathogen species; *duplication*, in which a pathogen speciates without host speciation so two pathogen lineages are associated with the same host; *host switching*, in which a pathogen infects a novel host species; *sorting*, in which a pathogen becomes extinct from a host lineage; and *inertia*, in which a host speciates without pathogen speciation so a single pathogen infects multiple hosts (Paterson and Banks 2001; Ronquist 2003). Likely, a complex combination of these events has caused the current associations between hosts and pathogens (Paterson and Banks 2001). Figure 1, adapted from Paterson and Banks, gives an example of how these events can be depicted in the comparison of host and pathogen phylogenies (2001).

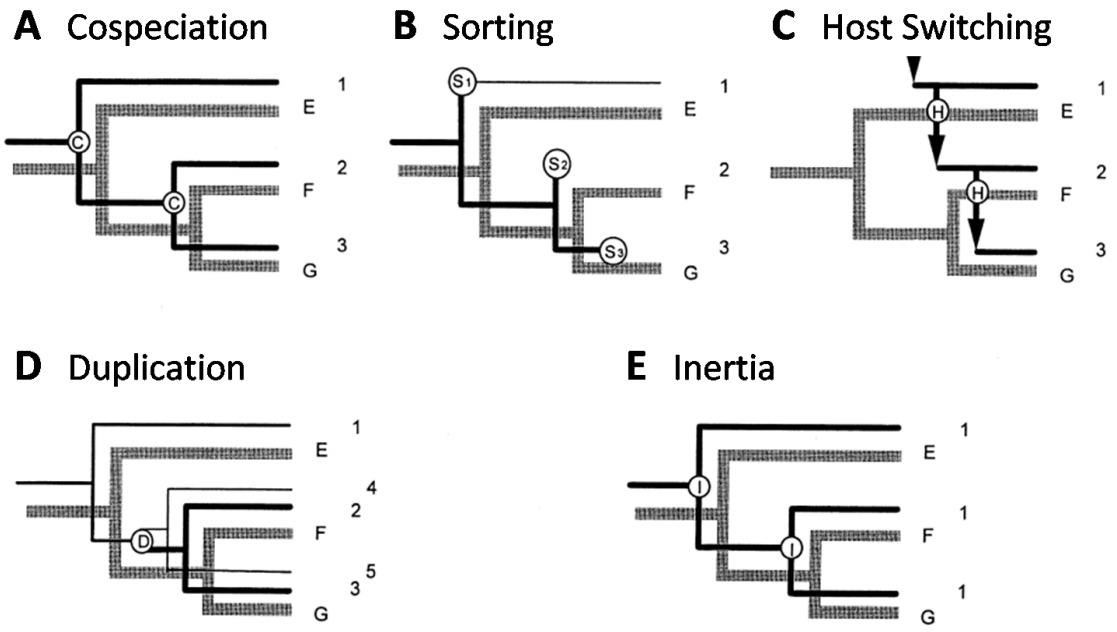

Figure 1. The five cophylogenetic events that can explain incongruent phylogenies. Theoretical host phylogenies with three taxa (E, F, and G) are shown in thick gray lines, and pathogen phylogenies with up to five taxa (1-5) are overlaid in thin black lines. A. cospeciation events ('C'), indicating codivergence of host and pathogen species; B. sorting events ('S'), in which a pathogen becomes extinct from a host lineage; C. host switching ('H'), in which a pathogen infects novel host species; D. duplication ('D'), which indicates pathogen speciation without host speciation; E. inertia ('I'), in which a host speciates without pathogen speciation (Paterson and Banks 2001).

Computational Methods

A variety of computational methods are currently available to study cophylogenetics, and the debate over the value and limits of each method continues (Stevens 2004). The first major division in methods separates pattern-based and event-based methods. Without the assumption of a model, pattern-based methods, such as that implemented by Brooks Parsimony Analysis (BPA), transform pathogen information as character states into additive binary code to be mapped onto host character-state trees (Paterson and Banks 2001). However, these methods cannot interpret all evolutionary

events, most importantly host switches. Conversely, event-based methods employ parsimony analysis with a model that specifies a cost for each cophylogenetic event. Observed associations are explained by the set of events resulting in the least cost. Several of these more recent techniques consider host switches as plausible events and provide a proposed sequence of events throughout the species' histories to explain their current associations (Ronquist 2003).

In this study, tree reconciliation is considered, which is a model-based method that implements the Jungles event-cost algorithm in TreeMap v2.02β (Charleston 1998). This program allows the user to define event-costs for cospeciation, duplication, lineage sorting, and host switching events and to set criteria limiting the maximum number of each event (Stevens 2004). The result is in the form of multiple potentially optimal (POpt) reconstructions, which minimize the total cost or the total number of specific events. Reconstructions can then be compared to determine which sequence of events and evolutionary explanations receive the most support.

Though often overlooked, the relative ages of host and parasite lineages are an important consideration in the study of cospeciation. Many computational techniques to study cophylogenetics do not compare the relative times that the host and pathogen species diverged. For example, the input file for Jungles analysis in TreeMap v2.02β describes the topology of the host and pathogen trees without branch length or divergence time information. Therefore, it is possible for these techniques to suggest that cospeciation has occurred even if the host and pathogen did not diverge at the same time. The pathogen must be present in the host at the time it speciates or the two could not possibly speciate together. Such incorrect matches are termed *pseudocospeciation*, and

strategies that compare divergence times better discriminate these events from true cospeciations. Alternatively, solutions may also predict host switches between hosts that did not exist at the same time, which are inherently impossible (Page 2003). Therefore, matching divergence times in the host and pathogen phylogenies are an important criterion for both cospeciation and host switching events (Holmes 2004). While cophylogenetic methods that address this issue are an active area of research, there have been no significant advancements in software for techniques that consider divergence times. Huelsenbeck developed a Bayesian Maximum Likelihood approach to quantitatively determine the degree of congruence between phylogenies using genetic sequence data and rates of evolution. However, this method has since become unavailable (Huelsenbeck, Rannala et al. 2000).

Lyssaviruses

RNA viruses are responsible for the majority of emerging diseases in humans (Badrane and Tordo 2001). In addition to their relevance for human disease, they are commonly used as models in studying disease dynamics due to their rapid evolution. RNA viruses evolve quickly relative to DNA viruses due to high mutation rates, short generation times, and large population sizes (Holmes 2004). In comparison to other pathogens, these viruses are most likely to develop beneficial mutations necessary for adaptation to new hosts (Parrish, Holmes et al. 2008). Therefore, RNA viruses are reasonable candidates for successful host switching and warrant analysis of their cophylogenetic history.

Viruses of the *Lyssavirus* genus have a single-strand, negative-sense RNA genome approximately 12 kb in length. The genus is most commonly known for the type species Rabies virus (RABV), which causes fatal encephalitis in both humans and other mammals. First documented over 4,000 years ago (Nel and Markotter 2007), rabies is commonly referred to as the most important and deadly viral zoonotic disease worldwide, causing over 50,000 deaths per year (Delmas, Holmes et al. 2008). Six other recognized lyssavirus genotypes and four recently identified molecular variants cause similar diseases in mammals that are clinically indistinguishable from rabies. Recognized genotypes include Rabies virus (RABV), Lagos Bat virus (LBV), Mokola virus (MOKV), Duvenhage virus (DUVV), European Bat Lyssavirus 1 (EBLV-1), European Bat Lyssavirus 2 (EBLV-2), and Australian Bat Lyssavirus (ABLV). The four unclassified lyssaviruses are designated Aravan virus (ARAV), Khujand virus (KHUV), Irkut virus (IRKV), and West-Caucasian Bat virus (WCBV). While RABV is distributed nearly worldwide, LBV, MOKV, and DUVV have been isolated exclusively from Africa, EBLV-1 and EBLV-2 from Europe, and ABLV from Australia. The new molecular variants have been isolated mainly from Asia (Nel and Markotter 2007).

Bat Hosts

While lyssavirus infections commonly spillover into humans and a range of mammals, these infections do not result in sustained transmission and are considered dead-end infections. Conversely, the natural wildlife reservoirs for each of the lyssavirus genotypes are distinct and relatively consistent for each species, largely confined to bats and terrestrial carnivores. RABV is established in both carnivores and bats, while the

other *Lyssavirus* species are found only in bat reservoirs. MOKV is the only genotype that has not been isolated from bats, and its effective reservoir is still unknown (Nel and Markotter 2007). This study focuses on the major bat reservoirs of the lyssaviruses. Species of both suborders of the Chiroptera (bats), the Megachiroptera and Microchiroptera, serve as hosts for these pathogens. Although RABV has been isolated from nearly all bat species studied in North America, a majority of isolates are found in *Eptesicus fuscus* (47.1%) and *Tadarida brasiliensis* (28.3%) (Hughes, Orciari et al. 2005). These two species are considered as representative of the hosts for RABV in this study. Known hosts of the other ten genotypes are shown in Table 1, adapted from Calisher, Childs et al. 2006.

Table 1. Virus genotypes and their known hosts.

Virus Genotype	Bat Species	Common Name
RABV	*Eptesicus fuscus*	Big brown bat
	Tadarida brasiliensis	Mexican free-tailed bat
LBV	*Eidolon helvum*	African straw-colored fruit bat
	Micropteropus pusillus	Peters' lesser epauletted fruit bat
	Epomophorus wahlbergi	Wahlberg's epauletted fruit bat
	Epomops dobsonii	Dobson's epauletted fruit bat
	Nycteris gambiensis	Gambian slit-faced bat
DUVV	*Miniopterus* sp.	
	Nyctalus noctula	Noctule
	Vespertilio murinus	Particolored bat
	Nyteris thebaica	Egyptian slit-faced bat
EBLV-1	*Eptesicus serotinus*	Common serotine
	Vespertilio murinus	Particolored bat
	Rousettus aegyptiacus	Egyptian rousette
EBLV-2	*Myotis* sp.	
	Miniopterus schreibersii	Schreibers' long-fingered bat
	Rhinolophus ferrumequinum	Greater horseshoe bat
ABLV	*Pteropus* sp.	

	Saccolaimus flaviventris	Yellow-bellied pouched bat
ARAV	*Myotis blythii*	Lesser mouse-eared myotis
KHUV	*Myotis daubentonii*	Daubenton's myotis
	Myotis mystacinus	Whiskered bat
IRKV	*Murina leucogaster*	Greater tube-nosed bat
WCBV	*Miniopterus schreibersii*	Schreibers' long-fingered bat

The total number of bat species (925) comprises about one fifth of all mammalian species, indicating their potential by sheer number to be host to a variety of pathogens. Notably, 66 viruses have been found in bat tissue (Calisher, Childs et al. 2006). Many bat species live sympatrically, with 110 sympatric species in the Neotropics for example (Simmons and Conway 2003). Overlap of species' ranges introduces many opportunities for host switching between bat species. With the ability to fly and distinct migratory behavior, these mammals are unique in their ability to exchange novel variants relatively easily between populations of the same species and with other bat species. In addition, bats have a longer life span than expected when compared to that of other mammals of similar size. Coupled with persistent infections, extended longevity increases the infectious period and thereby increases the probability of viral transmission to new populations. Finally, bat populations are widely distributed and occur commonly in high density. Their crowded roosting behavior may also contribute to an increased potential for viral transfer and successful host switching (Calisher, Childs et al. 2006).

Objectives

Despite numerous barriers that render host switching events uncommon, this phenomenon has been documented in the history of lyssaviruses, most notably from the *Chiroptera* to the *Carnivora* orders about 1,000 years ago (Badrane and Tordo 2001). Basing their analysis on the glycoprotein (G) gene, Badrane and Tordo constructed a phylogeny for the seven recognized genotypes of the lyssaviruses and estimated both their rate of evolution and time since most recent common ancestor. However, while speculating when rabies arose in terrestrial mammals from the chiropteran lyssaviruses, they do not explore host switching events among the bat hosts of the lyssaviruses (Badrane and Tordo 2001). In 2004, Jackson and Charleston applied the study of cophylogenetics directly to the lyssaviruses in examining the relationships among rabies viruses and their canid hosts with RABV phosphoprotein (P) gene sequences. They found no significant congruence between host and virus phylogenies, which can be attributed to the recent origin of rabies in terrestrial mammals and frequent viral transfer among species (Jackson and Charleston 2004). However, these workers focus on classic rabies (RABV) and their carnivoran hosts. Cophylogenetic processes have not been characterized quantitatively or comprehensively for the entire *Lyssavirus* genus, and the importance of cospeciation and host switching in the evolutionary history of the lyssaviruses and their bat hosts has not been well described.

The necessity for matching divergence times suggests that a history of cospeciation is impossible for the lyssaviruses because their bat and terrestrial mammal hosts speciated before the viruses were introduced into these populations. The two major lineages of bats, Megachiroptera and Microchiroptera, are estimated to have diverged

over 53 million years ago (Simmons and Conway 2003) and the four major Microchiroptera lineages (Rhinolophoidea, Emballonuroidea, Noctilionoidea, and Vespertilionoidea) are estimated to have originated 52 to 50 million years ago (Teeling, Springer et al. 2005). EBLV-1 has been isolated from bat hosts in both major suborders (*E. serotinus* and *R. aegyptiacus*), yet the most recent common ancestor of EBLV-1 is estimated to have existed only 500 to 750 years ago (Davis, Holmes et al. 2005). Therefore, these viruses could not have exclusively diverged with their host species. Cospeciation considered in its strictest sense does not appear to operate in the lyssaviruses. Every association between lyssaviruses and their hosts can then be considered the result of a host switching event.

However, even without matching divergence times, speciation of the lyssaviruses may still be tied to the divergence of their hosts. Jackson and Charleston suggest that phylogenetic congruence may still occur for viruses and hosts that did not diverge at the same time if the pathogen preferentially infects related hosts (Jackson and Charleston 2004). A broader definition of cospeciation requires only that the host phylogeny impose some constraint on the speciation of the viruses. To avoid confusion in coevolutionary terms, the term *host tracking* will be used to define divergences that are tied to each other but did not occur at the same time. These events may be identified as cospeciation or codivergence events by computational programs that do not consider divergence times. The term *host jumping* will be used to distinguish classical host switching events, which occur when the pathogen jumps from one host to another, unconstrained by the hosts' phylogenetic relationships. Therefore, while the evolutionary timescales of the lyssaviruses and their hosts demonstrate that cospeciation likely did not occur, the host

switching events in their histories can be further explored by distinguishing host tracking from host jumping events as defined above.

Through the generation and analysis of lyssavirus and host phylogenies, this study aims:

1. to identify the cophylogenetic events that have shaped current associations between the lyssaviruses and their hosts and to specifically determine the host jumping events that have occurred in their evolution
2. to determine how genetic distances among hosts constrain host jumping by examining the patterns of genetic distance between hosts and viruses of identified host jumps
3. to determine how geographic distance between hosts affects host jumping by comparing the overlaps in range and ecological habitat of hosts involved in jumps

Hypotheses

Aim 1: Reconstruction Analysis

Host jumps are hypothesized to have been common in the evolution of the lyssaviruses and likely can explain a majority of the current host-virus associations. Incongruence between virus and host phylogenies and the identification of host jumping events are expected in support of this hypothesis. As RNA viruses, the lyssaviruses evolve quickly and are likely to develop beneficial mutations necessary for adaptation to new hosts (Parrish, Holmes et al. 2008). Similarly, lyssaviruses cause acute infections in which the virus remains in the host for only a short period of time. Without causing persistent infection, the virus has limited opportunities to follow host divergence necessary for cospeciation (Holmes 2004).

Behavioral patterns induced by lyssaviruses also support host switching by increasing the transmissibility of the virus. Rabies-like diseases reduce behavioral separation between species by causing aggressive behavior in the host. Specifically, bats infected with rabies show increased aggression towards bats of other species. Since many bat species live sympatrically, the frequency of cross-species contact increases, which is necessary for viral introduction in a new host (Calisher, Childs et al. 2006). Because spillover infections are common, the probability that some of these spillovers may result in the establishment of the virus in a new species is relatively high. Finally, nearly all lyssavirus genotypes have been associated with multiple host reservoirs. This qualifies them as "generalist" viruses, which have adapted to use the cell mechanisms of a variety of hosts to infect and replicate. Without restriction to any one host's cellular

machinery, fewer mutations may be necessary to adapt to a new host (Parrish, Holmes et al. 2008).

Aim 2: Genetic Distance Analysis

Genetic distance between hosts is hypothesized to constrain host jumps in addition to host tracks, which are defined as switches between closely related hosts. Genetic distances between hosts of identified jumps are expected to be smaller than genetic distances between random pairings of hosts. This would indicate that genetic distance is a limiting factor, and a certain degree of genetic relatedness is necessary for jumps to be successful. This hypothesis is supported by the evidence that high genetic homology between hosts decreases the number of beneficial mutations that are necessary for the virus to adapt to a new host. The low-fitness valley that must be crossed is shallower between closely related species, and a lesser genetic requirement increases the probability that host jumps will occur. Therefore, the frequency of host jumping is predicted to be correlated with genetic similarity of hosts.

Aim 3: Geographic Distance Analysis

Host jumping is hypothesized to also be correlated with geographic proximity to new host populations. Therefore, jumps are predicted to be more common among hosts with greater overlap in range and habitat than among random pairings of hosts. Because a higher rate and intensity of contact between host species increases the rate of host switching, host jumps should occur between species that are similarly located geographically. Greater overlap in range and habitat increases both the contacts between

species that lead to spillover infections and the opportunities for adaptation to new hosts. Similarly, host populations in geographic areas with greater species diversity and intermixing of species should experience host switching more frequently than isolated host populations. The hypotheses for the impacts of both genetic distance and geographic distance on host jumping reinforce each other, as generally species that are closely related genetically are also found in similar areas geographically. However, these factors are not necessarily always correlated. Host jumps that occur between distantly related species may be explained by the hosts' similar geographic ranges and vice versa, jumps between geographically distant hosts may be explained by their close genetic relatedness.

Methods

Virus and host phylogenies were generated, compared, and analyzed to determine the cophylogenetic history of the lyssaviruses and their hosts. Cophylogenetic analysis requires robust phylogenetic trees for both the viruses and their hosts, which were generated from alignments of public sequence data. Lyssavirus phylogeny was based on the nucleoprotein (N) gene from sequences available in GenBank (Benson, Karsch-Mizrachi et al. 2008). The sequences selected included all geo-tagged and time-stamped full length gene segments that indicated the host source of the isolate. Sequences were aligned using Geneious Basic 4.7.6 (Biomatters, Auckland, New Zealand, www.geneious.com). Cytochrome *b* (cyt *b*) gene sequences for the bat species that the viral sequences were isolated from were selected for host phylogenies.

MrModeltest 2.3 (Nylander 2004) was used to determine the best nucleotide substitution model for the aligned sequences by comparing 24 models by hierarchical likelihood ratio tests (hLRTs), Akaike information criterion (AIC), and Bayesian information criterion (BIC) (Posada and Crandall 1998). For both virus and host alignments, MrModeltest 2.3 selected the general time-reversible GTR + I + Γ substitution model, which was used in all subsequent analyses. Through a Bayesian Markov Chain Monte Carlo (MCMC) technique implemented by BEAST v1.5.2 (Drummond and Rambaut 2007), multiple phylogenetic trees were generated under a variety of priors until the effective sample size (ESS) for all parameters was greater than 200 and nodes had high posterior support. Final phylogenetic trees were generated with sites partitioned into three codon positions. At least two analyses were run for both virus and host sequences with a chain length of 20,000,000 steps and parameter logging every

5000 steps. Tree information generated by BEAST v1.5.2 was summarized in TreeAnnotator1.5.2 (Drummond and Rambaut 2007) to produce a single target tree, and log files were combined in LogCombiner v1.5.2 (Drummond and Rambaut 2007).

To determine the cophylogenetic events that occurred in *Lyssavirus* history, the final host and virus trees generated in BEAST were simplified into a format suitable for TreeMap v2.02β (Charleston 1998), which required removing branch length information and representing replicate isolates from the same host and virus species with single sequences. The trees were imported including information concerning host-virus associations, which were determined from annotations in GenBank sequence records and confirmed by published results that based their analyses on the same sequences (Davis, Holmes et al. 2005; Kuzmin, Hughes et al. 2005; Delmas, Holmes et al. 2008). For each independent Jungles run in TreeMap v2.02β, a unique set of costs was assigned for cospeciation, duplication, lineage sorting, and host switching events as shown in Table 2.

Table 2. Costs assigned to cophylogenetic events TreeMap v2.02β.

Run	Codivergence	Duplication	Sorting	Host switching
1	0	1	1	1
2	0	2	2	1
3	0	3	3	1
4	0	1	1	2

Due to computational limitations, bounds were set on the maximum number of events allowed in POpt solutions: 40 total events, 25 sorting events, and 8 host switches. These constraints allowed for a greater number of events than were predicted in final reconstructions. Variants of ABLV that infect different species of *Pteropus* were

considered separately, also due to limitations in computing power. Subset trees were created for these three host-virus associations, and the same cost assignments and bounds were applied to the subset tanglegram. Results from this separate Jungles run were added to the results for the full set of lyssavirus associations.

Multiple POpt reconstructions determined by Jungles analysis were manually condensed into a single reconstruction to determine consensus events, defined as those events that occurred in a majority of reconstructions. While some reconstructions maximized the number of codivergence events, only POpt reconstructions that minimized total cost were considered in this analysis. The consensus reconstruction was then compared to those created under alternate cost assignments. Since there were no discrepancies between the events that were predicted, these events were considered in later analyses as identified consensus codivergence, sorting, duplication, and host switching events. According to the criteria outlined above, codivergence events determined in TreeMap v2.02β were interpreted as host tracking events, and host switches were considered as host jumps.

To determine the genetic distances between sequences, pairwise distance matrices were generated from lyssavirus and host sequence alignments in PAUP 4.0b10 (Swofford 2003). Consistent with the model used to generate trees, a general time-reversible (GTR) model with unequal base frequencies and three substitution types was used to measure these distances. Genetic distances between sequences of identical host-virus associations were averaged to remove duplicate associations from the matrices. A Mantel test, performed in R version 2.3.1 (R Development Core Team 2008), was performed to determine the overall correlation between virus and host genetic distances. Genetic

distances between hosts and their viruses were plotted against each other to show the correlation between distances for potential host jumps. All potential host jumps were considered by evaluating every pairwise combination of hosts. Among all potential host jumps, points corresponding to identified consensus host jumps were isolated, and genetic distances between hosts of identified jumps were compared to those of all potential jumps.

The overlaps of host ranges were analyzed using Digital Distribution Maps of The IUCN Red List of Threatened Species (IUCN 2009), which contain information on the distribution of all host species considered in the analysis. In ArcGIS 9.3.1 (ESRI, Redlands, CA, www.esri.com), maps of host ranges were projected from a geographic coordinate system into the Lambert conformal conic projection specific for the continents on which the bats were found. Eurasian species' ranges were projected with the Asia Lambert conformal conic projection, African species with the same projection for Africa, Australian species with the same projection for Asia South, and the new world species with the same projection for North America. Overlapping areas of species' ranges were isolated in ArcMap, and the total area of overlap was measured in square kilometers for each pairwise combination of hosts. The sizes of overlapping ranges for identified host jumps were then compared to those for all potential host jumps. Specifically, geographic overlaps were divided into three equal groups based on the size of overlapping area, and the number of corresponding identified jumps in each group was determined.

Information concerning the ecological habitats of the bat species that are host to the lyssaviruses was also gathered from the IUCN Red List of Threatened Species (IUCN 2009). Characteristics of population size, group size, foraging habitat, and roosting sites

were noted for each of the species. These characteristics were compared for each pair of hosts involved in consensus host jumps identified by TreeMap v2.02β.

Results

Phylogenetic Analysis

A phylogenetic tree for the lyssavirus sequence alignment based on the 1.4 kb N gene is shown in Figure 1. The effective sample sizes (ESS) for all statistics were between 298 and 3513, indicating high support for the tree. Posterior values for nodes were greater than 0.95 for the majority, although several were less than 0.60. Because the isolation dates for the sequences were known, this information was used in BEAST to estimate the rate of substitution. The estimated clock rate was 2.237×10^{-4} (95% confidence intervals (CI) 2.066×10^{-7}, 5.023×10^{-4}) substitutions per site per year, and the most recent common ancestor was estimated to have existed 11,452 (95% CI 785, 30,273) years ago. Sequences from distinct viral genotypes cluster into monophyletic clades, validating their assignment to their respective genotypes.

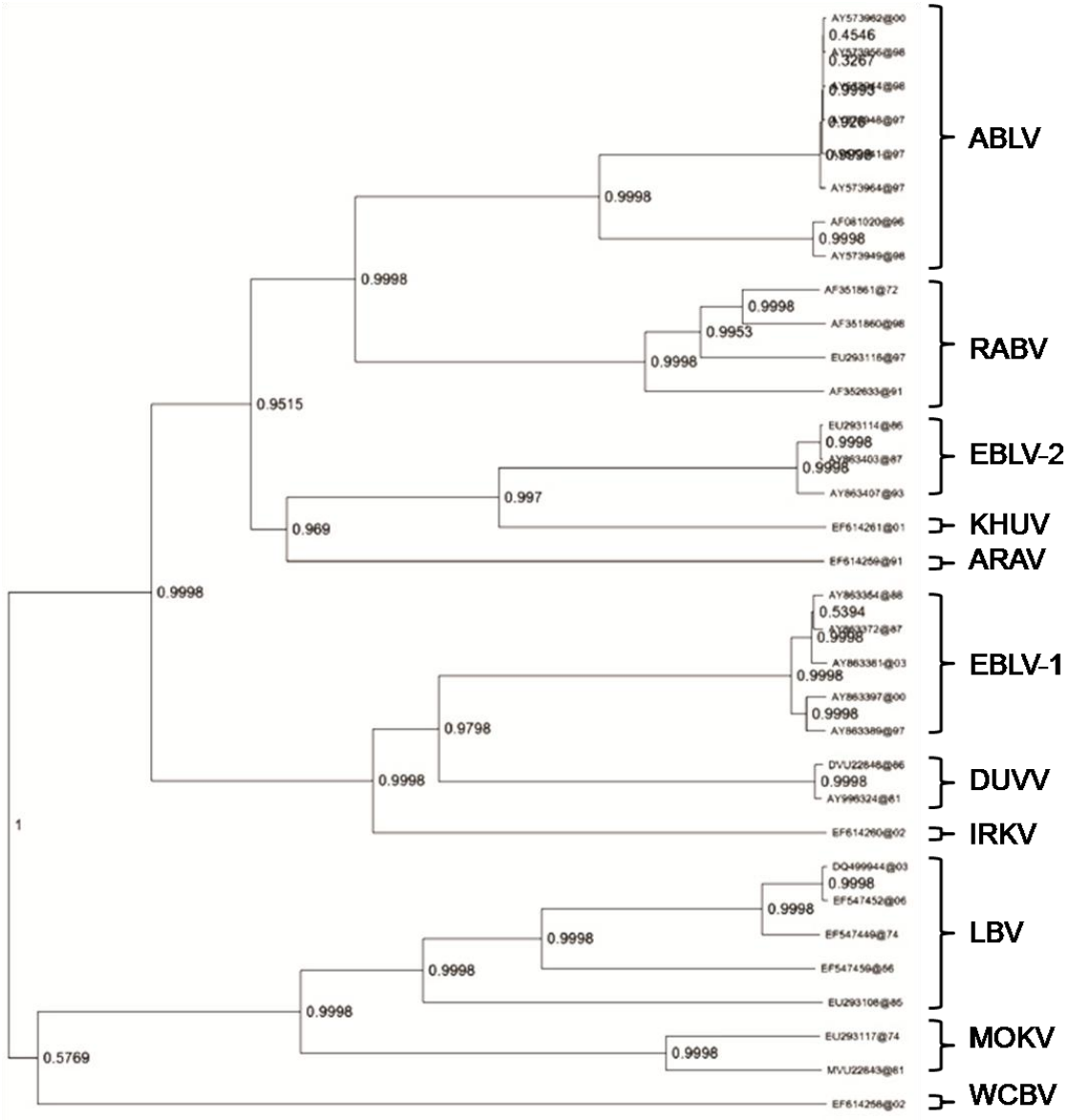

Figure 2. Bayesian MCMC phylogenetic tree based on the N gene of the *Lyssavirus* genus. Nodes are labeled with posterior values.

A phylogenetic tree for the bat sequence alignment based on the cyt *b* gene is shown in Figure 3. The host phylogeny had ESS values between 776 and 6001 with a majority of posterior values greater than 0.95. Because isolation dates were unknown for the host sequences, substitution rate was not estimated. Sequences from distinct bat species cluster in monophyletic clades, and sequences from the same genus, such as the

Myotis species, also cluster. Species of the Megachiroptera form a monophyletic clade, while the Microchiroptera are paraphyletic.

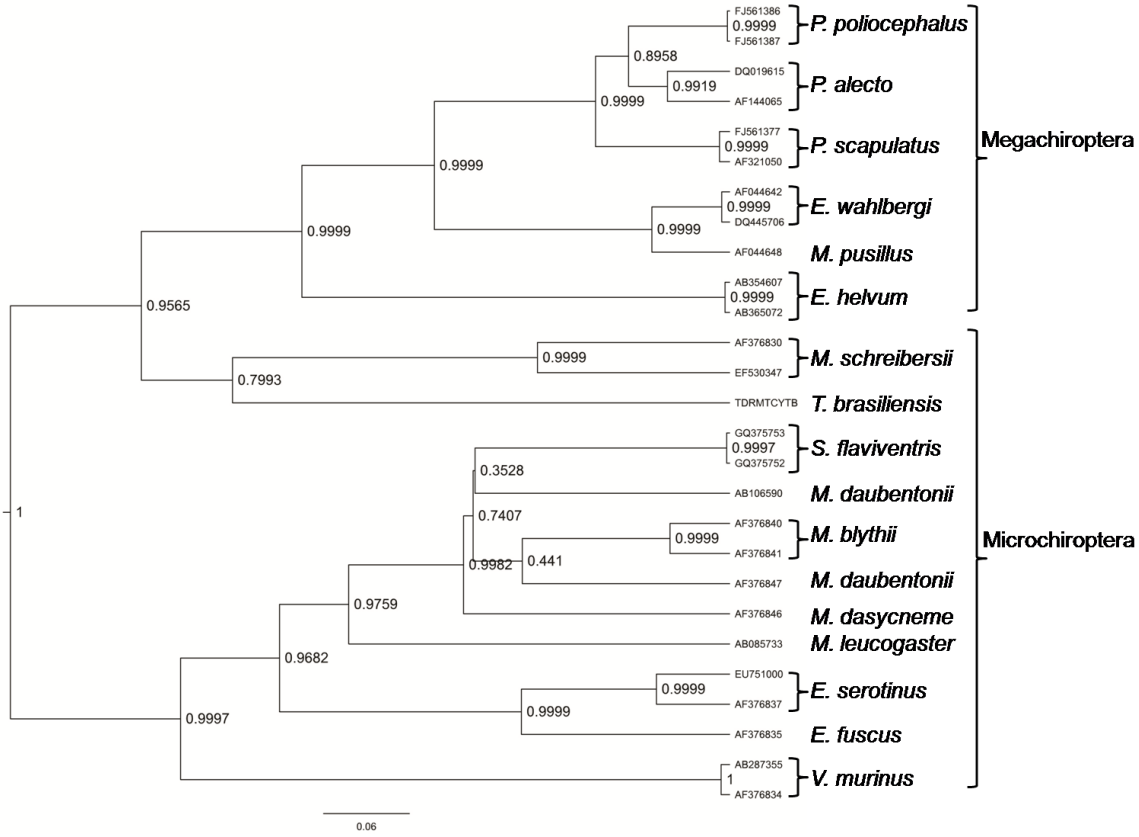

Figure 3. Bayesian MCMC phylogenetic tree based on the cyt *b* gene of bat species that serve as hosts for the lyssaviruses. Nodes are labeled with posterior values.

Reconstruction Analysis

A tanglegram was generated in TreeMap v2.02β to visually represent host-virus associations as shown in Figure 4. Host and virus trees are associated by lines in red to indicate which bat lineages were host to each viral type. Host and pathogen phylogenies clearly do not mirror each other, and several distantly related lyssaviruses are associated with closely related, or even the same, bat species.

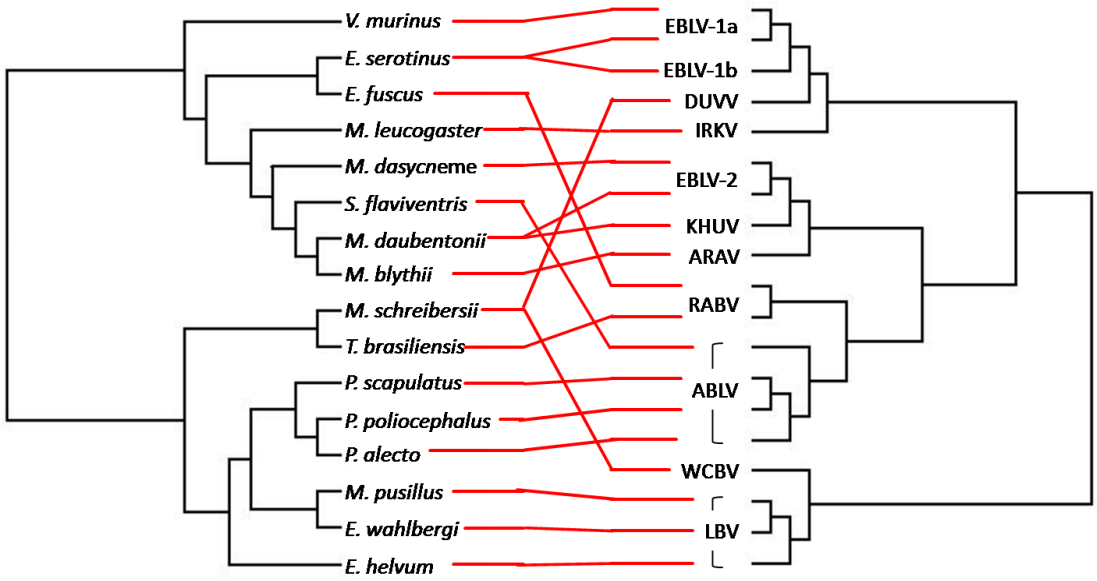

Figure 4. Tanglegram showing the bat (left) and lyssavirus (right) phylogenies. Host-virus associations are depicted by red lines connecting host and viral lineages.

Reconstruction analyses under the four sets of cost assignments in TreeMap v2.02β found 62 POpt solutions in the second Jungles run and 54 POpt solutions in each of the other three runs. Only four solutions from each analysis minimized total cost and were considered for further analysis. These reconstructions were identical under all alternate cost assignments considered. Discrepancies among the four solutions concerned the direction and order of two host jumps, but did not impact which species experienced host switching. A consensus reconstruction, shown in Figure 5, was determined by accepting the sequence of host jumps which resulted in the least genetic distance traversed by the virus in the switches.

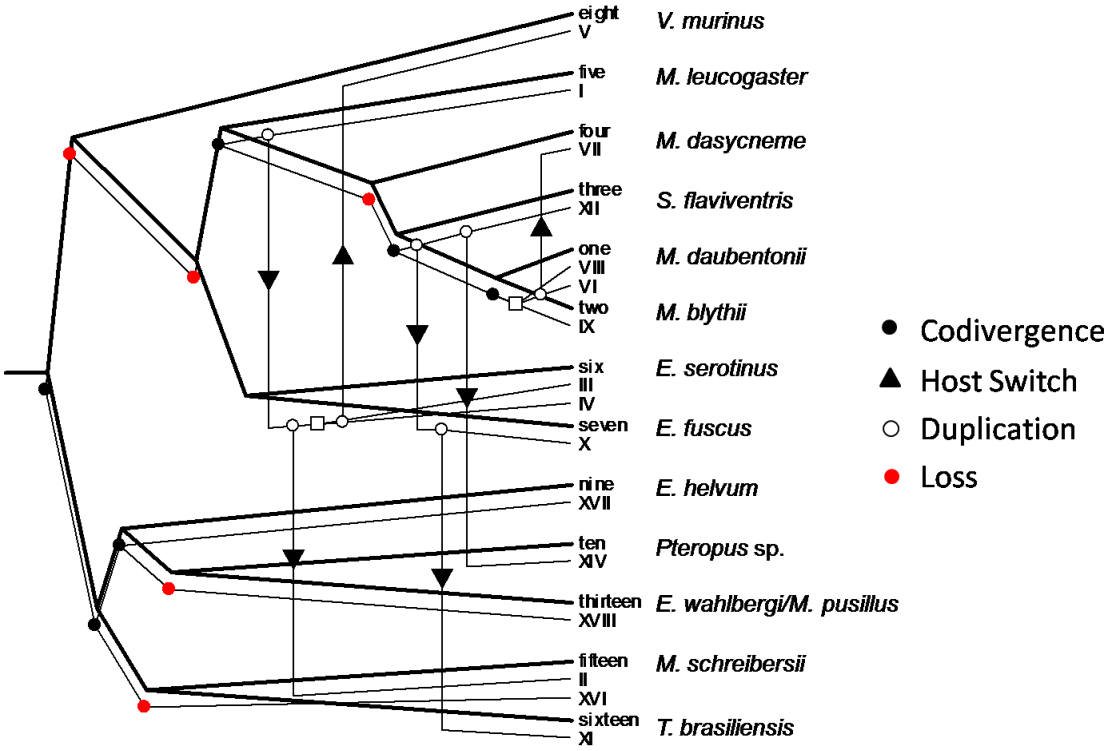

Figure 5. Optimal reconstruction determined by TreeMap v2.02β for current associations between lyssaviruses and their hosts.

The optimal reconstruction identified 12 codivergence or host tracking events, 7 host switches, considered here as host jumps, 18 duplications, and 5 losses, for a total of 42 events. Analyzed separately, the associations for ABLV sequences infecting *Pteropus* hosts produced a single POpt reconstruction of minimal cost shown in Figure 6, which identified an additional host jumping event between *P. poliocephalus* and *P. scapulatus*.

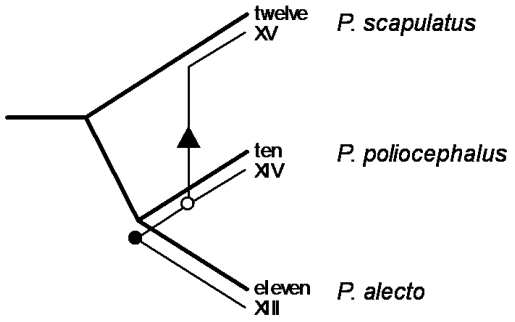

Figure 6. Optimal reconstruction determined by TreeMap v2.02β for current associations between ABLV and *Pteropus* sp.

In sum, complete analysis of the *Lyssavirus* genus identified 8 consensus host jumps between the species shown in Table 3.

Table 3. Consensus host jumps

	Donor Host	Recipient Host
1	*M. leucogaster*	*E. serotinus*
2*	*E. serotinus*	*M. schreibersii*
3	*E. serotinus*	*V. murinus*
4	*S. flaviventris*	*E. fuscus*
5*	*E. fuscus*	*T. brasiliensis*
6	*S. flaviventris*	*P. alecto*
7	*M. daubentonii*	*M. dasycneme*
8	*P. poliocephalus*	*P. scapulatus*

* Direction of host jump was inconsistent for all POpt solutions; donor and recipient hosts chosen to minimize total distance travelled by the viruses.

Genetic Distance Analysis

A Mantel test performed on the pairwise distance matrices for lyssavirus and host sequence alignments revealed a positive association between matrices. The result was significant with a moderate correlation coefficient of r = 0.418 and a simulated p-value of 0.0001.

A plot of genetic distances between viruses against genetic distances between hosts is shown in Figure 7a. This plot considers the genetic distance between all pairwise combinations of host-virus associations and therefore represents the genetic distances between hosts and viruses for all potential host jumps. The average distance between hosts was 0.247 changes per nucleotide (changes/nucleotide) while the average distance between their corresponding viruses was 0.290 changes/nucleotide. Although potential

jumps occurred over a relatively wide range of virus distances, with a variance of 0.00465, the distribution of host distances was smaller, with a variance of 0.00188. Nearly all pairs of hosts had a distance greater than 0.15 changes/nucleotide between their sequences. Points corresponding to more closely related hosts (less than 0.15 changes/nucleotide) represent relationships among the *Pteropus* species and between the closely related species, *E. wahlbergi* and *M. pusillus*. Figure 7b shows the density of genetic distances for all pairwise combinations of hosts and viruses. Peak density of the virus distances occurred at 0.318 changes/nucleotide, while peak density of the host distances occurred at 0.257 changes/nucleotide.

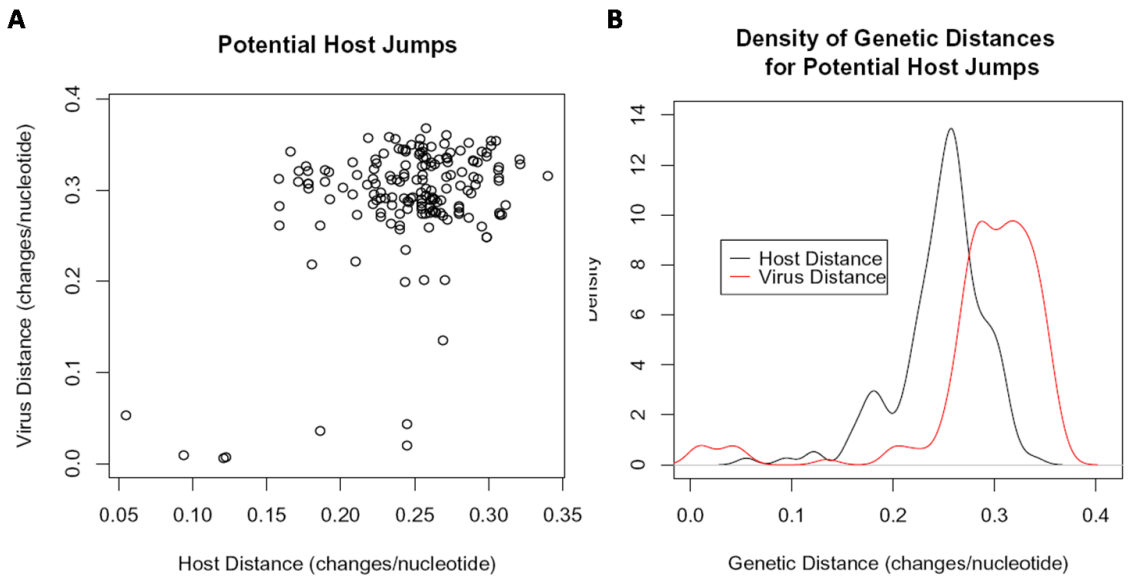

Figure 7. A. Plot of genetic distances between hosts (Host Distance) and genetic distances between viruses (Virus Distance) for all pairwise combinations of host-virus associations; B. Density plots of genetic distances between hosts (black) and viruses (red).

A genetic distance plot that highlights the genetic distances between hosts and viruses of consensus host jumps identified in previous Jungles analysis is shown in Figure 8a.

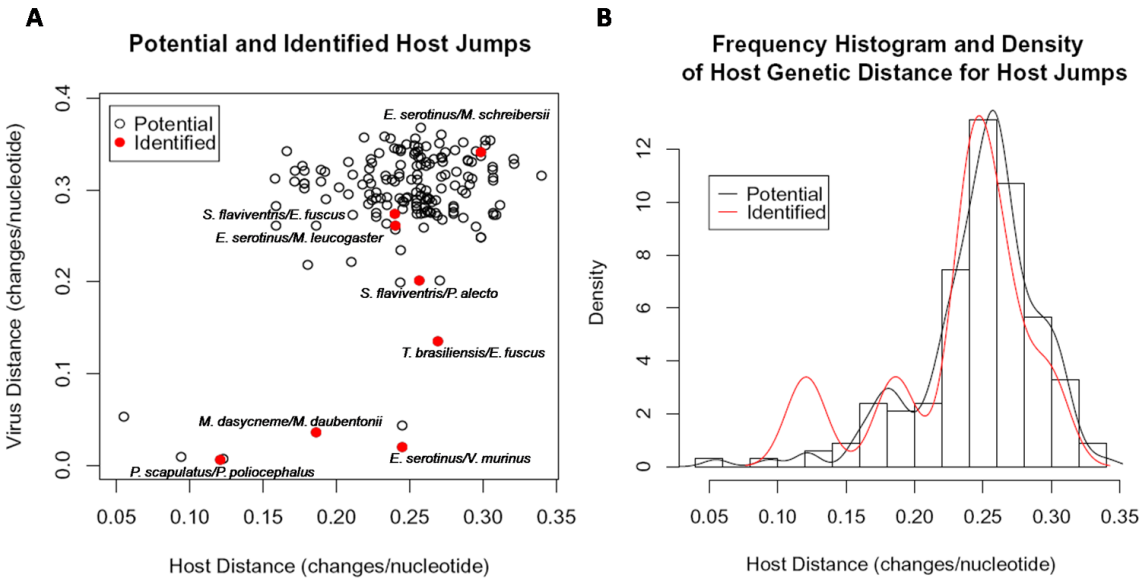

Figure 8. A. Genetic distance plot for all pairwise combinations of host-virus associations, representing genetic distances for potential host jumps (black) and genetic distances for identified consensus host jumps (red). Each identified host jump is labeled with the names of the two bat species involved in the jump; B. Frequency histogram and density plot for distances between hosts of potential host jumps (black) and density plot for distances between hosts of identified jumps (red).

The average distance between hosts for identified jumps was 0.232 changes/nucleotide compared to 0.247 changes/nucleotide for potential jumps. Identified host jumps were spread over the range of host distances, with a variance of 0.00303, compared to a variance of 0.00465 for potential jumps. The density of identified jumps along host genetic distance nearly matched the corresponding density curve for potential jumps, and the distribution of identified jumps does not appear to be significantly different from that of potential jumps. Conversely, Figure 9 demonstrates that the density of virus distances for identified jumps was relatively consistent over the range of virus distances, lacking the peak in density seen for potential jumps at more distantly related viruses. Average virus distance for identified jumps (0.159 changes/nucleotide) was nearly half the average of virus distance for potential jumps (0.290 changes/nucleotide).

In addition, the variance in virus distance for identified jumps, 0.0168, was much larger than that for potential jumps, 0.00465.

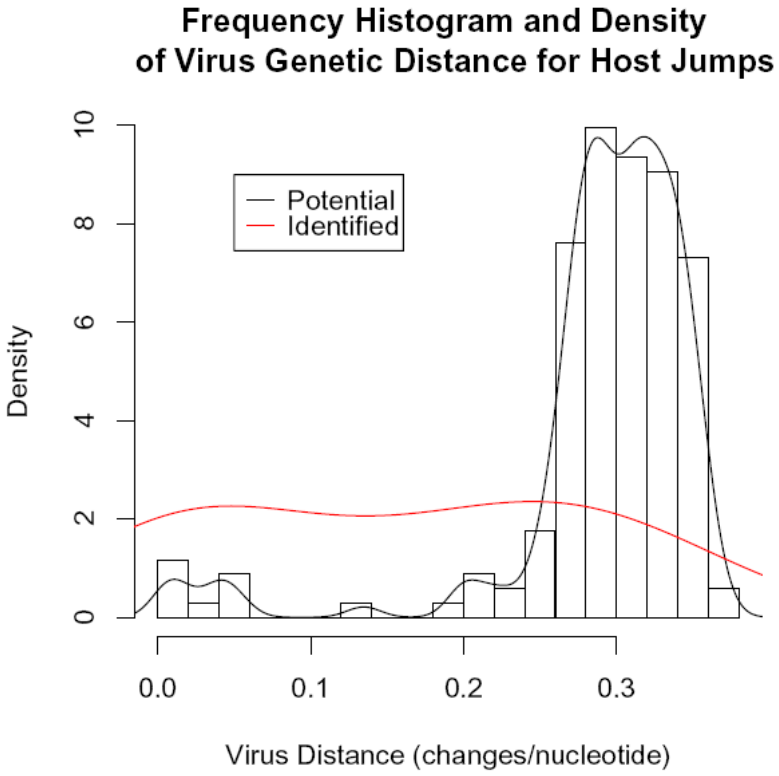

Figure 9. Frequency histogram and density plot for virus distances of potential host jumps (black) and density plot for virus distances of identified host jumps (red).

Geographic Distance Analysis

After the ranges of each of the sixteen bat species that are host to the lyssaviruses were isolated, overlapping area was calculated for all pairwise combinations of hosts that had any overlapping area in their ranges. A total of thirty overlapping areas were identified, representing the overlaps in host ranges for potential host jumps. The ranges of all seven species found in Eurasia, *E. serotinus*, *M. daubentonii*, *M. dasycneme*, *M. blythii*, *M. schreibersii*, *V. murinus*, and *M. leucogaster*, overlapped with the others,

except for that of *M. leucogaster*, which did not overlap with *M. schreibersii*, *M. dasycneme*, or *V. murinus*. Similarly, all four species found in Africa, *E. helvum*, *E. wahlbergi*, *M. pusillus*, and *M. schreibersii*, had overlapping ranges except for *M. schreibersii*, whose range did not overlap with that of *E. wahlbergi*. The two species from the new world, *T. brasiliensis* and *E. fuscus* had overlapping ranges, and all four species from Australia, *P. alecto*, *P. poliocephalus*, *P. scapulatus*, and *S. flaviventris*, had areas of overlap. Overlapping areas for bat species ranged from 8,893 km^2 to 9,604,054 km^2. The mean area of overlap for all pairwise combinations of overlapping hosts was 2,259,386 km^2. Total range area for each of the species is shown in Table 4. The ranges of all sixteen species are shown against a map of the world in Figure 10.

Table 4. Total area of bat species ranges.

Native Continent	**Bat Species**	**Range Area (km^2)**
Eurasia	*Miniopterus schreibersii*	3,559,826
	Eptesicus serotinus	11,587,047
	Vespertilio murinus	14,724,817
	Myotis daubentonii	13,789,398
	Myotis dasycneme	5,145,209
	Myotis blythii	5,857,809
	Murina leucogaster	624,594
Africa	*Eidolon helvum*	10,662,687
	Epomophorus wahlbergi	4,522,579
	Micropteropus pusillus	4,764,330
	Miniopterus schreibersii	5,759,491
Australia	*Pteropus alecto*	1,221,190
	Pteropus scapulatus	2,828,483
	Pteropus poliocephalus	250,094
	Saccolaimus flaviventris	5,675,716
Americas	*Eptesicus fuscus*	12,320,776
	Tadarida brasiliensis	21,633,614

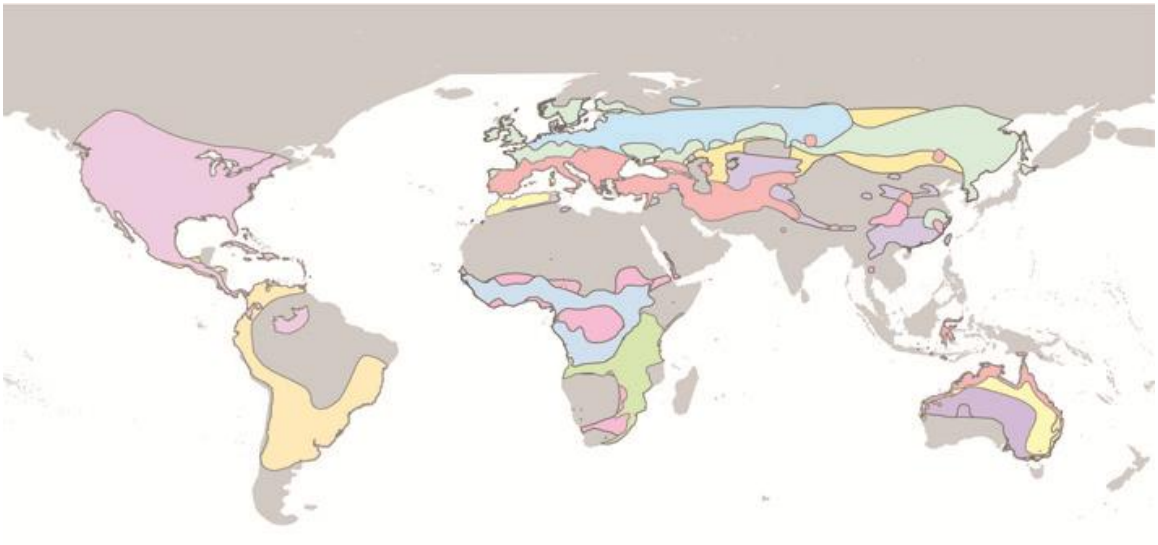

Figure 10. Map of the ranges for the sixteen bat hosts of the lyssaviruses. Ranges are colored arbitrarily to show separation between ranges.

Seven out of the eight host jumps identified by TreeMap 2.02 occurred between hosts with overlapping ranges. Ranges and areas of overlap for these seven identified host switches are shown in Figure 11. The host jump identified between *S. flaviventris* and *E. fuscus*, which are host to ABLV and RABV respectively, occurred without any overlap in host range since *S. flaviventris* is native to Australia and *E. fuscus* is found in the Americas.

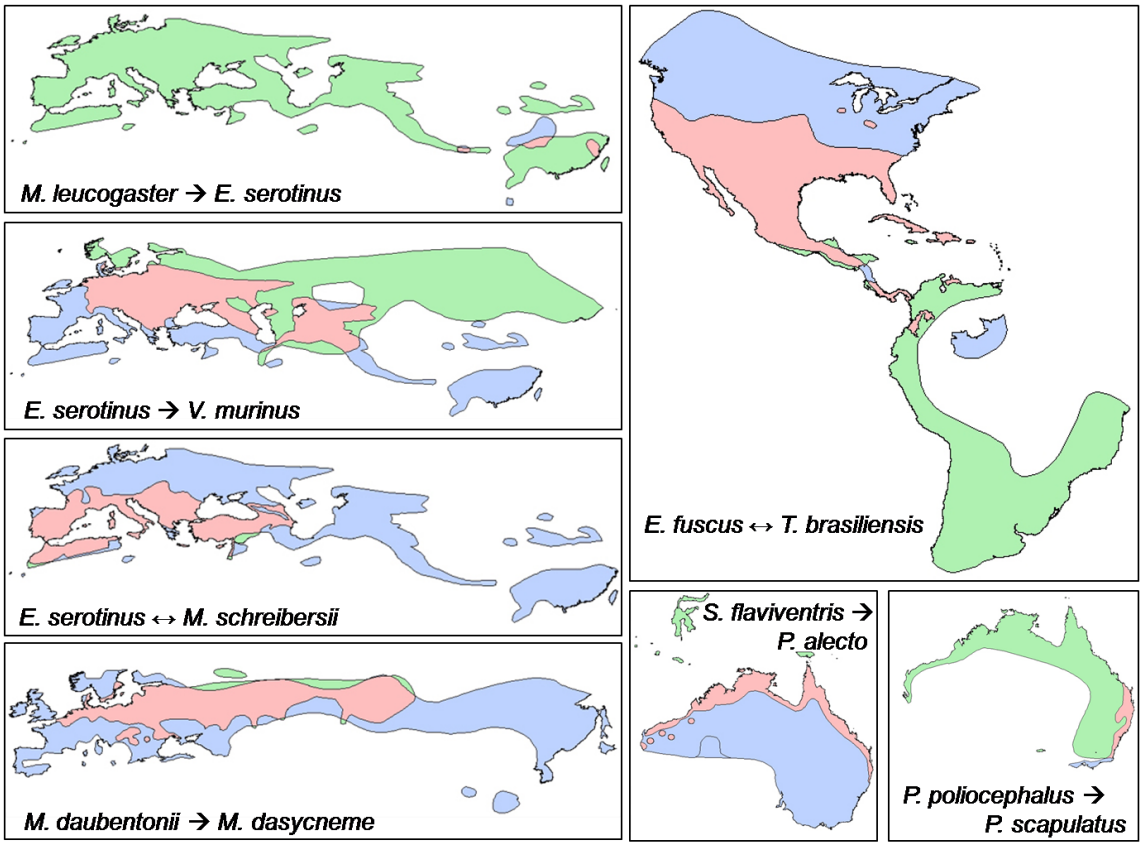

Figure 11. Maps of overlapping ranges for hosts of identified host jumps. Donor host ranges are shown in blue, recipient host ranges in green, and overlapping range areas in pink.

Overlapping areas for bat species of identified host jumps ranged from 195,991 km^2 to 5,536,360 km^2, with an average overlap area of 2,877,076 km^2. The thirty overlapping areas were separated into three equal groups of ten, characterizing small areas of overlap less than 900,000 km^2, medium sized areas between 900,000 and 3 million km^2, and large areas of overlap greater than 3 million km^2. Four identified host jumps occurred between species with greater than 3 million km^2 in range overlap. Conversely, only one jump occurred between hosts with overlap of 900,000-3 million km^2, and only two jumps occurred between hosts with less than 900,000 km^2 of range overlap (Figure 12).

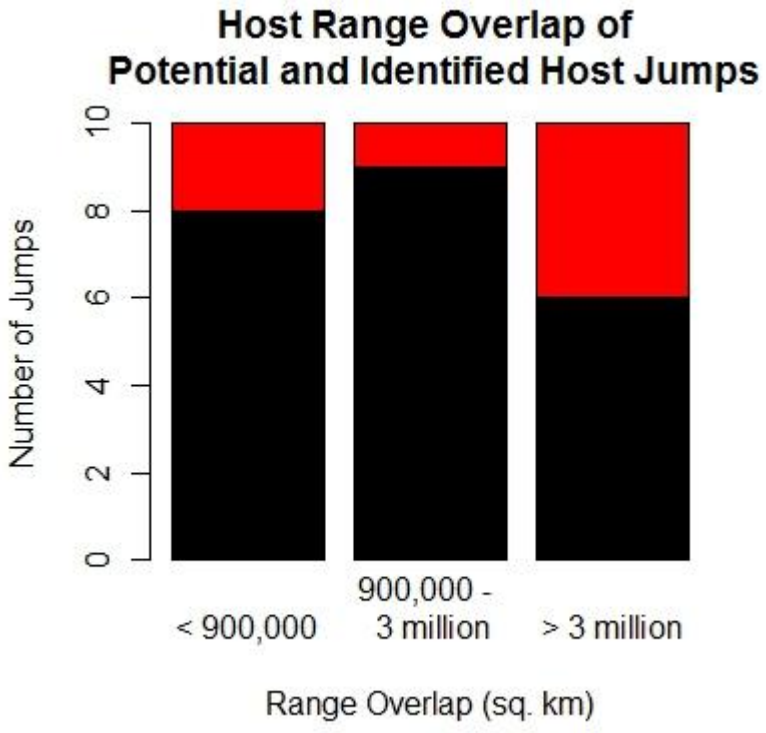

Figure 12. The number of pairs of hosts with areas of overlap less than 900,000 km^2, between 900,000 and 3 million km^2, and greater than 3 million km^2 corresponding to potential host jumps (black) and identified host jumps (red).

The ecologies of bat species that are host to the lyssaviruses varied in population size, foraging habitat, roosting sites, and group size. Nine out of the twelve species involved in host jumps were considered to be generally or extremely abundant in population size. *M. dasycneme*, however, is rare and decreasing in population size, and *M. leucogaster* and *P. poliocephalus* are also considered not abundant and decreasing. *E. serotinus*, *V. murinus*, and *M. schreibersii* use a wide range of habitats and are known to forage in forests, shrub land, and grassland. *M. leucogaster* forages mainly in forests and potentially also in open areas. Conversely, *M. daubentonii* and *M. dasycneme* forage over natural and artificial bodies of water. All three species of *Pteropus* are found and forage in forests. *S. flaviventris* is found in forests and a variety of other habitats

including open savannahs. *E. fuscus* is commonly found near human dwellings, and *T. brasiliensis* occurs in a range of habitats (IUCN 2009).

Nearly all species studied roost in buildings or trees, except for *M. schreibersii*, which is mainly cave dwelling. All species except those found in Australia also roost or spend winters in caves and underground. It is unknown whether *M. leucogaster* roosts in caves. The size of groups in roosting sites varied among bat species from solitary individuals to groups with millions of individuals. *E. serotinus* and *S. flaviventris* are generally solitary or form colonies of small numbers. *V. murinus*, *M. dasycneme*, and *E. fuscus* reside in medium size colonies ranging from fifty to several hundred individuals. *P. alecto* and *M. schreibersii* form groups in the thousands, while P. *scapulatus* and *T. brasiliensis* are found in groups of millions. *M. schreibersii* is commonly found in mixed groups with other cave-dwelling species. Group sizes for the other species are unknown or not mentioned in the IUCN Red List of Threatened Species (IUCN 2009).

Discussion

Published phylogenies of the lyssaviruses do not agree across all tree building techniques and gene sequences used. A Bayesian method in BEAST is used here because of the model's sophistication, its appropriate treatment of uncertain parameters, and the ability to handle large amounts of sequence data in a reasonable amount of time. This technique is also especially useful for its ability to estimate clock rates and divergence times from dated sequences (Aris-Brosou and Xuhua 2008). While most nodes in the final lyssavirus phylogeny had high posterior support, several values fell below 0.60, which indicates the branching patterns at these nodes were present in less than 60% of the steps in the MCMC chain. Poorly supported nodes near extant taxa in the tree can be expected because the sequences are very closely related and alternate branching patters are nearly congruent. Since duplicate virus sequences for the same host are eliminated in further analysis, these instances do not impact identified host jumps. The basal node marking the split between WCBV and the common ancestor of LBV and MOKV was also poorly supported with a posterior value of 0.5769. This low value could not be eliminated by multiple iterations in BEAST and is accepted as a limitation in the data. Only one sequence was available for WCBV which likely contributed to this problem. The split may be better supported in the future once WCBV has been isolated more frequently and more sequences become available.

Kuzmin, Hughes et al. found a similar branching pattern for WCBV, LBV, and MOKV in their neighbor-joining tree of the N gene, and similarly, this node had a less than optimal bootstrap value of 92 in their phylogeny (2005). Interestingly, their neighbor-joining trees based on the G and P genes do not show this same branching

pattern for WCBV, instead having WCBV as an out-group of all the other lyssaviruses. Nonetheless, the N gene was ultimately used for the lyssavirus phylogenies in this study because the N gene was found by Kuzmin, Hughes et al. as producing the least ambiguous division between lyssavirus genotypes. Their phylogenetic tree using the N gene was of higher resolution than those for the G and P genes (Kuzmin, Hughes et al. 2005). In addition, evidence for strong purifying selection in the N gene suggests that this gene has had a relatively consistent rate of substitution and is reliable for evolutionary analysis (Hughes, Orciari et al. 2005). The G gene, conversely, is a common target of the host immune system and shows a higher rate of amino acid substitution, which makes it less suitable for clock-based phylogenetic analyses (Hughes 2008).

The evolutionary rate estimated for the N gene, 2.237×10^{-4} substitutions per site per year, is nearly congruent to the estimate by Hughes et. al. using the same gene from American RABV sequences, 2.32×10^{-4} substitutions/site/year (Hughes, Orciari et al. 2005). In their study of the European Bat Lyssaviruses, Davis, Holmes et al. estimate the rate of evolution for the N gene of a subset of EBLV-1, EBLV-1b, at 2.39×10^{-4} substitutions/site/year, which is also nearly equivalent to the estimate from this study for all lyssaviruses (Davis, Holmes et al. 2005). In addition, the estimate is of the same magnitude as the rate of synonymous substitution found by Badrane and Tordo for the G gene of all seven recognized lyssavirus genotypes, 3.1×10^{-4} to 5.5×10^{-4} d_s/site/year (2001). The clock rate is relatively low when compared to estimates for other negative stranded ssRNA viruses that also cause acute infections, such as arenavirus (5.06×10^{-3} d_s/site/year) and ephemerovirus (2.23×10^{-3} d_s/site/year). However, the lyssaviruses have

a faster rate than that of related vesicular stomatitis virus, which is estimated at 7.20×10^{-5} d_s/site/year (Hanada, Suzuki et al. 2004). The estimate from this study for the time since the most recent common ancestor of the lyssaviruses, 11,452 years ago, is consistent with Badrane's and Tordo's estimated range of 7,080 to 11,631 years ago (Badrane and Tordo 2001).

Sequence data for the bat hosts was less available than that for the viruses, and because dates of sequence isolation were unknown, these values could not be used in BEAST to estimate clock rate. Sequence data for the cyt *b* gene of the bats was available for all species, although several sequences were only 300 bp in length, compared to a full gene sequence of about 1.14 kb. Sequences of short length may have diminished the accuracy of aligning the sequences and comparing them correctly. Nonetheless, the cyt *b* gene sequence was the most available gene for the hosts. In addition, the cyt *b* gene is a well-known housekeeping gene in the mitochondrial genome, which suggests a stable rate of evolution and makes it ideal for use in phylogenetic analyses. Because the hosts of the lyssaviruses are spread throughout the range of bat species, there were many bat taxa—not host to the lyssaviruses—that were not included in the phylogenies.

Due to the complex relationships between bat species and the paucity of sequence data for several species, it is not surprising that the host phylogeny is poorly resolved in some areas. Recent phylogenetic studies of the bats have found conflicting results, and in general, published estimates of chiroptera phylogeny do not agree. In fact, while most studies support the monophyly of bats, the debate continues over whether the microchiroptera form a monophyletic clade (Jones, Andy et al. 2002). Recent molecular analyses strongly support the paraphyly of microbats, and classifications within families

and subfamilies remain uncertain (Teeling, Springer et al. 2005). The bat phylogeny generated in this study likely provides a reasonable estimate for the relationships between bat species that are host to the lyssaviruses. Several of the low posterior support values can be attributed to the short length of cyt *b* sequences that were available for *T. brasiliensis* and *S. flaviventris.* Full length sequences may increase support for these nodes.

A limitation of Jungles in TreeMap v2.02β requires that the input phylogenies are fully resolved, as alternative branching patterns for the trees are not considered in the analysis (Jackson and Charleston 2004). However, despite the inadequacy of available sequence data, the generated phylogeny was accepted as a reasonable estimate of the relationships among bat hosts for the purposes of further analyses.

A variety of available methods to assess cospeciation were considered for cophylogenetic analysis in addition to Jungles in TreeMap v2.02β, including Brooks Parsimony Analysis (BPA) (Brooks 1988) and several statistical tests of congruence, such as an optimized version of ParaFit implemented by CopyCat (Meier-Kolthoff, Auch et al. 2007). However, to compare genetic and geographic constraints on host tracking versus host jumping events, a description of where these host switching events occurred was necessary. Statistical tests that merely conclude if two trees are congruent were not sufficient. TreeMap v2.02β proved the most useful in providing POpt reconstructions of the cophylogenetic events between the viruses and their hosts by specifically delineating where and when cospeciation, duplication, lineage sorting, and host switching events occurred.

The tanglegram of associations between the lyssaviruses and their hosts is consistent with the hypothesis that host jumping has occurred in lyssavirus history. Because host and pathogen phylogenies do not show congruent branching patterns, a history of pure host tracking can be rejected, and a combination of other cophylogenetic events must explain the current associations. While insufficient computational power in TreeMap v2.02β unfortunately required constraining the maximum number of host switches allowed to eight, this limitation was acceptable because under all cost assignments, least cost solutions predicted only seven host switches. This suggests that even with a more relaxed limit on number of host switching events, optimal solutions would continue to predict seven host switches. Predictions of other non-codivergence events similarly did not reach the upper bounds specified in TreeMap v2.02β.

Lowest cost POpt solutions from reconstruction analysis were consistent under different costs assignments, which indicates that their conclusions were robust. The cost assignments for the first Jungles run (Table 2) were the default values implemented by Jungles in TreeMap v2.02β and assume that all events other than codivergence are equally likely. The costs of duplication and sorting in the second and third runs were increased to effectively eliminate these events and focus the analysis on host tracking and jumping. Because a history of true cospeciation with matching divergence times was rejected, duplication and sorting events, which generally accompany cospeciation, were also not expected. In the last run, costs were reversed so that switching events were made more costly than duplication and sorting. Each model returned the same results, indicating that identified consensus host jumps were robust despite alternating the probabilities assigned to non-codivergence events.

The host species involved in host jumps were consistent for all four POpt solutions identified in each Jungles run. Because differences arose only in the direction and intermediate step of two switches, these discrepancies did not affect genetic or geographic distance analyses, which do not discriminate between donor and recipient host. The final consensus reconstruction (Figure 5) was chosen to minimize the total genetic distance traversed by the virus in the jumps under an argument of parsimony. Because the *Pteropus*-ABLV associations were analyzed separately, the specific recipient species of the host switch into the *Pteropus* genus had to be determined. *P. alecto* was selected because the virus associated with this species is the most basal branch, suggesting an extended period of association. In addition, ABLV was first isolated from this species, and *P. alecto* is most frequently associated with ABLV infection (Calisher, Childs et al. 2006).

Results from the Mantel test are consistent with the hypotheses that closely related viruses infect closely related hosts and that host tracking has occurred in the history of the lyssaviruses. The moderate correlation between host and virus distances suggests that in general, the genetic distances between viruses and their branching patterns may be constrained by the genetic distances between their hosts. This result further validates the assessment of codivergence in TreeMap v2.02β, which considers host tracking events likely. However, while the correlation between host and virus distance was significant, it was not perfect, which indicates that variables other than genetic relatedness also influence host-virus associations. Consistent with the hypotheses being studied, the correlation between distances is not strong enough to eliminate the possibility of host jumping events as predicted. The average genetic distance between

hosts was about 0.04 changes/nucleotide smaller than the average distance between viruses. This result supports the evidence that the viruses have a faster mutation rate than their mammalian hosts and therefore have diverged more quickly than their hosts. The high rate of evolution for these viruses has caused greater genetic distance between viruses, even though they diverged more recently than their hosts.

Comparison of identified host jumps to potential host jumps in genetic distance analysis suggests that genetic distance between hosts does not influence the probability of successful host jumping. Although the number of identified host jumps was small, the distribution of host distances for identified jumps was similar to that for all potential jumps, which represents what would be expected if jumps occurred between hosts randomly. This causes the average host distance for identified jumps, 0.232 changes/nucleotide, to be nearly equivalent to that for potential jumps, 0.290 changes/nucleotide. If close genetic similarity between hosts was a requirement for successful host jumping, the average distance between hosts for identified jumps would instead be expected to be smaller, and identified jumps would be expected to cluster at the lower end of the range of potential host jump distances. Figure 13 shows a modified genetic distance plot for identified host jumps to demonstrate that six out of eight identified host jumps occur at distances very close to or higher than the mean distance between hosts, which is marked by the vertical black dashed line. Since a majority of host jumps occur at the higher range of host distances (to the right of the mean), distance between hosts does not appear to constrain host jumps.

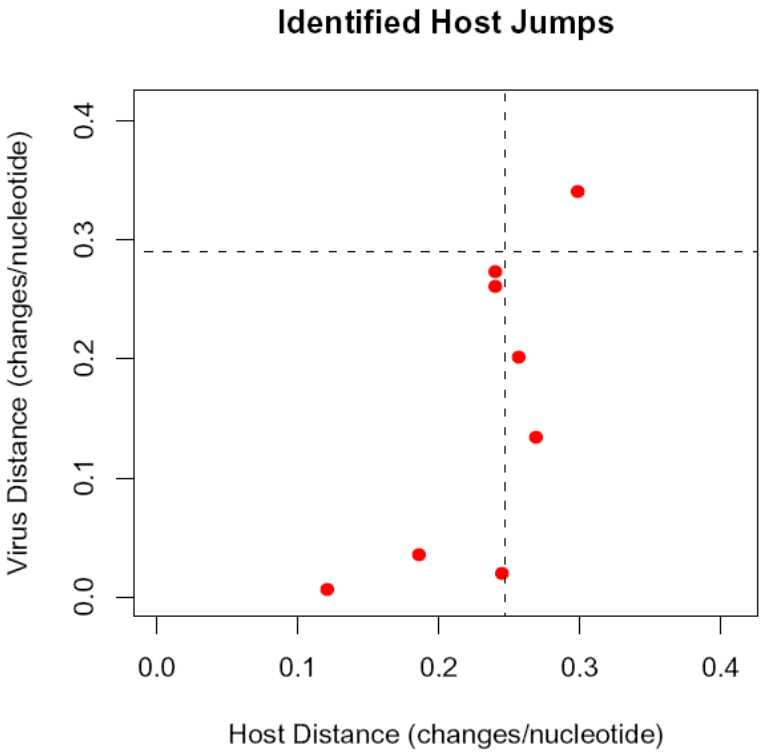

Figure 13. Plot of genetic distances between hosts and viruses for identified host jumps. Dashed black lines indicate the mean value of genetic distance between hosts (vertical) and the mean value of genetic distance between viruses (horizontal) for all potential host jumps.

Similarly, the dashed red curve in Figure 14 demonstrates what the density of distances between hosts of jumps would be expected to look similar to if genetic distance constrained these events. The expected peak of host distances would occur at a smaller distance between hosts than was observed for identified jumps.

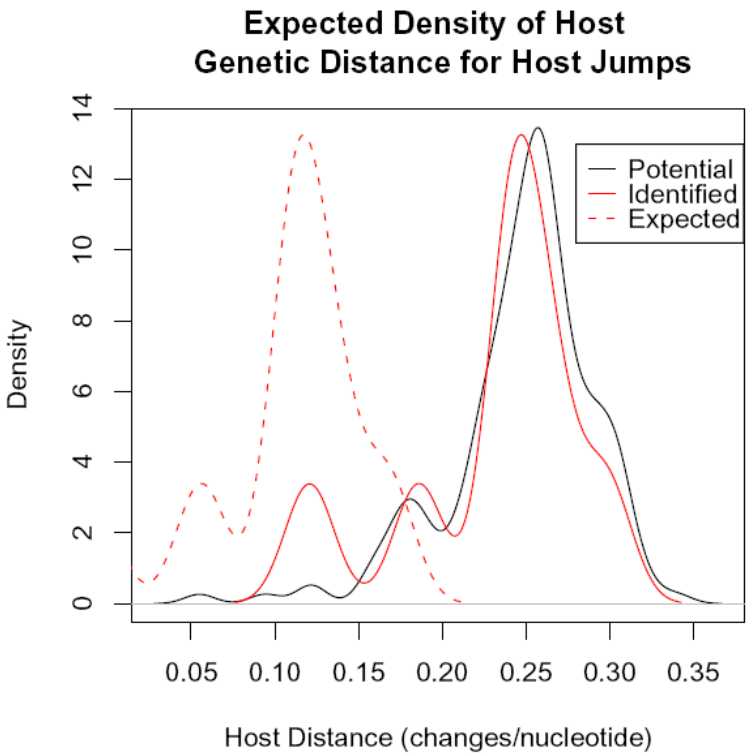

Figure 14. Density plot comparing the expected distances between hosts of jumps if genetic distance constrained jumping (red dashed) to actual distances between hosts of potential jumps (black) and identified host jumps (red).

In terms of virus genetic distance, a greater proportion of identified jumps were found between more closely related viruses than would be expected when compared to the distribution of virus distances for all potential jumps (Figure 9). While a majority of viruses were separated by 0.25 to 0.35 changes/nucleotide, seven out of eight identified host jumps were separated by a distance of near to or less than 0.2 changes/nucleotide. These seven jumps occurred below the mean distance between viruses, which is shown with a horizontal black dashed line in Figure 13. This leads the average virus distance for identified jumps, 0.159 changes/nucleotide, to be nearly half of that for potential jumps, 0.290 changes/nucleotide. Distance between viruses of jumps indicates the timing of the jump. Viruses of more recent jumps have had little time to diverge and are more closely

related, while viruses of more ancient jumps have had more time to diverge and are less closely related. The relatively small distance between viruses of identified jumps suggests that these switches have occurred recently. This trend is expected for host jumps since jumping generally results in current host-virus associations that are not ancient.

Conversely, hosts and viruses of host tracking events would be expected to have a nearly perfect correlation between host distance and virus distance since their speciations are correlated, constrained by the host phylogeny. If genetic distance constrained all host switching events, genetic distances between hosts and viruses of identified jumps would be similar to genetic distances of tracks. In this scenario, genetic distance would be a constant limiting factor on both events, and a certain degree of genetic closeness would be necessary for switches irrespective of whether they are tracking or jumps. However, since identified jumps occurred between hosts more distantly related than expected, the genetic distances between hosts of jumps are greater than those that would be expected for tracks at the same virus distances. This evidence supports the conclusion that host jumping is distinct from host tracking and that genetic distance between hosts does not appear to constrain host jumps. Instead, host jumps can occur between hosts that are more distantly related, and likely some other geological or ecological factor is constraining these events.

Overall, with the limited number of identified host switches, no trend was found to support the hypothesis that host jumps are more common between closely related hosts. However, while two host jumps did occur between especially closely related hosts, with distances of less than 0.2 changes/nucleotide, no jumps occurred between hosts

separated by more than 0.3 changes/nucleotide. Because no jumps occurred near the upper extreme of the range of host distances, it is possible that there is a limit, around 0.3 changes/nucleotide, to the amount of distance between hosts that is acceptable for successful jumps. Therefore, the hypothesis that genetic relatedness plays a role in the probability of successful host jumping cannot be rejected. Genetic distance may contribute to this phenomenon alongside other factors explored in the study such as geographic distance and similarity of ecological habitat.

The geographic analysis of host ranges demonstrates that many of the bat species host to the lyssaviruses live sympatrically. For bat species found on the same continent, only two species did not overlap with the others in some area of their ranges. The many combinations of overlapping species' ranges supports the possibility for host jumping between species, as it indicates the plausibility of cross-species contact between hosts living in the same areas. However, the amount of overlap for pairwise combinations of species' ranges varied by three orders of magnitude, which suggests that some species may come into contact more frequently than others.

The result that seven out of eight host jumps occurred between species with overlapping ranges is consistent with the hypothesis that overlapping range allows for cross-species contacts that are necessary for colonization and adaptation to a new host species. The lack of range overlap for the host jump between *S. flaviventris* and *E. fuscus* is not surprising considering ABLV is found only in Australia. The jumps of the lyssaviruses into both Australia and the New World were likely directly caused or at least influenced by human movements into these areas. This possibility is not considered by TreeMap v2.02β, and the alternative explanation of a jump between *S. flaviventris* and *E.*

fuscus is reasonable considering ABLV and RABV are sister taxa. Therefore, excluding this special case, the analysis of host range overlap suggests that overlapping host ranges are necessary for successful host jumping.

Areas of overlap were separated into three groups, less than 900,000 km^2, between 900,000 and 3 million km^2, and greater than 3 million km^2, to determine if the amount of range overlap can predict host jumping events. While the number of identified host jumps was small, the greatest proportion of jumps was found for species in the group with the largest overlap in range. The trend towards larger areas of overlap is not entirely consistent since a greater number of switches occurred between hosts with less than 900,000 km^2 of overlap than between hosts with 900,000 to 3 million km^2 of overlap. However, the difference between these two groups is only of one host jump and the correlation between the size of range overlap and the number of host jumps may be further supported with data for additional identified jumps. Also, the average area of overlap for hosts of identified jumps, 2,877,076 km^2, was larger than that for potential jumps, 2,259,386 km^2. These results support the hypothesis that hosts of identified jumps have a greater area of range overlap than would be expected by chance. While no identified jumps occurred between hosts at the lower extreme of range overlaps (in the lowest 20% of overlapping ranges), two identified jumps occurred between hosts with overlapping areas in the highest 10%, and three occurred in the top 20%. Although jumps can occur between species with relatively small areas of range overlap, the data suggest that larger areas of overlap increase the probability of successful host jumping.

The relatively small amount of range overlap for the identified host jump between *E. serotinus* and *M. leucogaster* can be explained by the small size of the total range of

M. leucogaster. This species is found over a range of less than one million square kilometers, and therefore the amount of overlap for *M. leucogaster* with other species cannot exceed this value. Any jump involving *M. leucogaster* will automatically fall into the smallest class of range overlaps because the range of the species itself is so small. Similarly, the range of *P. poliocephalus* is even smaller, approximately a quarter of a million square kilometers. The jump between this species and *P. scapulatus* corresponds to a necessarily small amount of range overlap because the entire range of *P. poliocephalus* is small. Therefore, while the areas of overlap for these jumps may appear uncharacteristically small considering the expected trend towards larger areas of overlap, the amount of range overlap is constrained by the total range of these species.

Habitat comparisons for the bat species with overlapping ranges revealed that the hosts of identified host jumps generally use similar ecological niches. Since a majority of the species involved in jumps are abundant, the probability for cross-species contact between hosts of jumps is high. Host jumps between *M. leucogaster, E. serotinus, M. schreibersii,* and *V. murinus* are plausible considering these species all use a wide range of habitats from forests to grassland. Similarly, the switch between *M. daubentonii* and *M. dasycneme* is supported by the fact that they share a common foraging habitat over open water. Jumps among the species native to Australia may have occurred in the forests, where all four species are found. Similarly, the wide range of habitats host to *T. brasiliensis* suggests that this species may come in contact with *E. fuscus* near its home in human-inhabited and urban areas. Roosting sites for hosts of jumps are also similar, mainly found underground and in caves. The mixed colonies of *M. schreibersii* in caves strongly support the possibility for cross-species contacts that may lead to infection of

novel hosts. Finally, in six out of the seven host jumps between hosts that had overlapping ranges, at least one of the species of the jump is found in medium or large group sizes of at least hundreds of individuals. Large roosting groups increase the transmission of pathogens among bats and the potential to transfer them to new species. In sum, the available ecological data for the bat species support the possibility of the host jumps identified between species with overlapping host ranges.

In their study of the factors that influence pathogen sharing among host primate species, Davies and Pedersen found overall that pathogens are most often shared between genetically related hosts. However, specifically for viruses, they found that geographic overlap was a better predictor of viral sharing than genetic distance. Interestingly, while hosts with geographic overlap shared a greater portion of pathogens, the magnitude of overlap did not affect how often hosts were infected with the same viruses (Davies and Pedersen 2008). The results from this current study agree that range overlap is an important requirement for the transmission of pathogens between hosts. However, the data expand on their results by also providing evidence that a larger magnitude of overlap increases the probability of host jumping, and hence pathogen sharing. Davies and Pedersen further suggest that geography plays an important role because host jumps may occur more frequently among viruses than other pathogens such as protozoa (Davies and Pedersen 2008). This interpretation is consistent with the identification of host jumps in the lyssaviruses, and together the results of this study support Davies' and Pedersen's conclusion that range overlap is correlated with pathogen sharing among host species.

Future Directions

The results of this study are contingent on the robustness of the phylogenies generated since TreeMap v2.02β considers input phylogenies to be fully resolved. The conclusions may be strengthened with better sequence data that would more correctly estimate the ancestries of both the viruses and especially their bat hosts. Viral sequences were not available for every documented host-virus association, so several host species were not included in the analysis. In the future, these gaps could be filled by data from new isolates that would potentially identify more host jumps in lyssavirus history. For the bat species analyzed, several cyt *b* sequences were of short length and could be replaced with full length sequences in the future for better alignment and analysis. The resolution of the phylogenies might also be increased by generating and comparing multiple phylogenies based on different gene sequences. Specifically, lyssavirus phylogeny could also be estimated from the G and P genes. Ideally, concatenated sequences of these genes with the N gene would produce the best estimates of the relationships between genotypes. Similarly, the bat phylogeny might be confirmed with trees based on other commonly used housekeeping genes, such as the mitochondrial gene, NADH dehydrogenase (ND1). With the addition of more sequence data, these steps would better resolve the input data for Jungles analysis in TreeMap v2.02β.

To further explore genetic distance as potentially constraining the success of host jumps, sequences from other genes could be used to generate alternate pairwise distance matrices for comparisons between potential and actual host jumps. The N gene may not be the best sequence to correctly estimate genetic similarity between viruses. The use of G or P gene sequences in the same analysis may be informative in corroborating current

results. Similarly, results from the cyt *b* gene for the bat species could be compared to that for other genes such as ND1. In this way, the genetic similarities between viruses and hosts could be better estimated for comparisons of genetic distances of host-virus associations.

In the geographic distance analysis, this study considers the absolute area of host range overlap as representative of the geographic distance between hosts and as a proxy for the number of contacts between host species. However, the amount of geographic overlap may not be perfectly correlated with the number of cross-species contacts because population densities vary across ranges (Davies and Pedersen 2008). To further demonstrate that the number and intensity of cross-species contacts are the best predictors of successful host jumping, data on the actual distribution of bats throughout their ranges is necessary. The frequency of cross-species transmission could be better predicted by considering the densities of hosts specifically in the areas of range overlap. Contact frequencies may be more directly correlated to host jumping than absolute area of range overlap, and these results might strengthen the conclusion that geographic proximity is a determinant of successful host jumping.

While the data allow speculations on the trends for both genetic and geographic distance between hosts as predictors for successful host jumping, the number of identified host jumps (8) was too small to generate statistically significant results to characterize host jumping. However, the analyses performed in this study demonstrate a set of methods to explore host jumps for other host-virus associations as well. Specifically, associations between related viruses and their hosts, including other viruses of the *Rhabdoviridae* family such as vesiculovirus and ephemerovirus, could be studied in the

future to identify more host jumps in collective RNA virus history. These results could be added to the currently identified host jumps to increase the sample size for genetic and geographic distance analyses. By expanding the study into a meta-analysis with data from multiple viral genuses, the conclusions from the results observed in this initial work may be strengthened.

Conclusion

This cophylogenetic study of the lyssaviruses and their hosts identified eight host jumping events in the history of host-virus associations. These host jumps, in conjunction with host tracking, sorting, and duplication events, can explain the current associations between the lyssaviruses and their bat hosts. A history of pure cospeciation was rejected because the lyssaviruses were estimated to have diverged approximately 11,452 years ago with a clock rate of 2.237×10^{-4} substitutions/site/year. Their recent evolution makes the speciation of the viruses and their bat hosts at the same time an impossibility. Genetic similarity between donor and recipient hosts does not appear to constrain the possibility of successful host jumping. Host jumps were as likely to occur between closely related hosts as between more distantly related hosts, and the genetic distances between hosts of identified jumps were not significantly smaller than those for random pairings of hosts. Conversely, host jumps were more common between hosts with larger overlapping ranges, suggesting that a higher number of cross-species contacts increases the probability of successful jumping. For the majority of identified jumps, the hosts involved shared similar foraging and roosting habitats, which is consistent with the necessity for contact between hosts for host jumping.

In sum, these experiments successfully explored the aims of the project—to identify host jumps in lyssavirus history and to analyze the genetic and geographic distances between hosts. Geographic overlap of hosts has a larger impact than genetic distance between hosts in predicting the success of host jumping. While genetic similarity likely also has an impact, the number and intensity of contacts between bat species may be the driving factors in host jumping events. These results suggest that the

lyssaviruses are strongly generalist viruses, similar to universal vectors that can infect a wide variety of mammalian hosts. Since spillover events commonly occur into distant species, it is not surprising that the lyssaviruses can also occasionally cause sustained transmission among fairly distantly related hosts. This ability to infect a wide range of hosts is comparable to that of a related RNA virus, vesicular stomatitis virus (VSV), which enters the cell through universal components of mammalian cell membranes and is able to infect nearly all animal cells (Lichty, Power et al. 2004).

Based on the trends indicated by the data, future host jumps can be predicted to occur between species that have large areas of overlap in their ranges and that share common ecological niches in terms of foraging and roosting behavior. For example, *V. murinus* and *M. daubentonii* have the greatest area of range overlap, 9,604,054 km^2, and both roost in trees or buildings. While more data is needed concerning population densities and frequency of contacts between species, one may speculate that a host jump could occur between these species in the future.

The ability to predict host switching events is an important public health concern considering the potential for virulent pathogens in non-human hosts to adapt and infect human populations. While studied here in the context of the lyssaviruses, the underlying principles that govern host switching apply to a variety of emerging infectious diseases (Kuiken, Holmes et al. 2006). With increasing globalization, in which human populations have expanded to all corners of the globe, our contact with many mammalian species has increased. As our distance geographically from these species is eliminated, the threats from viral zoonoses as potential emerging infectious diseases will only continue. In agreement with the wide body of work on zoonotic pathogens, this study

supports the need for further understanding of host switching dynamics and specifically cross-species contact in order to prevent the future emergence of disease.

Appendix A

Lyssavirus isolates analyzed in this study.

Virus	Host species	Year of isolation	Geographic origin	GenBank accession no.
RABV	*Tadarida brasiliensis*	1991	Mexico	AF352633
RABV	*Tadarida brasiliensis*	1997	Argentina	EU293116
RABV	*Eptesicus fuscus*	1972	Ontario, Canada	AF351861
RABV	*Eptesicus fuscus*	1998	Connecticut, USA	AF351860
LBV	*Eidolon helvum*	1956	Nigeria	EF547459
LBV	*Eidolon helvum*	1985	Senegal	EU293108
LBV	*Micropteropus pusillus*	1974	Central African Republic	EF547449
LBV	*Epomophorus wahlbergi*	2003	South Africa	DQ499944
LBV	*Epomophorus wahlbergi*	2006	South Africa	EF547452
MOKV	?	1974	Cameroun	EU293117
MOKV	?	1981	Zimbabwe	MVU22843
DUVV	*Miniopterus schreibersii*	1986	Republic of South Africa	DVU22848
DUVV	Human	1981	South Africa	AY996324
EBLV-1a	*Eptesicus serotinus*	1988	Germany	AY863354
EBLV-1a	*Eptesicus serotinus*	2003	France	AY863381
EBLV-1a	*Vespertilio murinus*	1987	Ukraine	AY863372
EBLV-1b	*Eptesicus serotinus*	1997	Netherlands	AY863389
EBLV-1b	*Eptesicus serotinus*	2000	France	AY863397
EBLV-2a	*Myotis dasycneme*	1986	Holland	EU293114
EBLV-2a	*Myotis dasycneme*	1987	Netherlands	AY863403
EBLV-2b	*Myotis daubentonii*	1993	Switzerland	AY863407
ABLV	*Pteropus alecto*	1997	Australia	AY573964
ABLV	*Pteropus alecto*	2000	Australia	AY573962
ABLV	*Pteropus scapulatus*	1998	Australia	AY573944
ABLV	*Pteropus scapulatus*	1998	Australia	AY573956
ABLV	*Pteropus poliocephalus*	1997	Australia	AY573941
ABLV	*Pteropus poliocephalus*	1997	Australia	AY573948
ABLV	*Saccolaimus flaviventris*	1996	Australia	AF081020
ABLV	*Saccolaimus flaviventris*	1998	Australia	AY573949
ARAV	*Myotis blythii*	1991	Kyrgyzstan	EF614259
KHUV	*Myotis daubentonii*	2001	Tajikistan	EF614261
IRKV	*Murina leucogaster*	2002	Russia	EF614260
WCBV	*Miniopterus schreibersii*	2002	Russia	EF614258

Appendix B

Bat host isolates analyzed in this study.

Bat species	**GenBank accession no.**
Tadarida brasiliensis	TDRMTCYTB
Eptesicus fuscus	AF376835
Eidolon helvum	AB354607
Eidolon helvum	AB365072
Micropteropus pusillus	AF044648
Epomophorus wahlbergi	DQ445706
Epomophorus wahlbergi	AF044642
Miniopterus schreibersii	EF530347
Miniopterus schreibersii	AF376830
Eptesicus serotinus	AF376837
Eptesicus serotinus	EU751000
Vespertilio murinus	AF376834
Vespertilio murinus	AB287355
Myotis dasycneme	AF376846
Myotis daubentonii	AF376847
Myotis daubentonii	AB106590
Pteropus alecto	AF144065
Pteropus alecto	DQ019615
Pteropus scapulatus	AF321050
Pteropus scapulatus	FJ561377
Pteropus poliocephalus	FJ561387
Pteropus poliocephalus	FJ561386
Saccolaimus flaviventris	GQ375752
Saccolaimus flaviventris	GQ375753
Myotis blythii	AF376840
Myotis blythii	AF376841
Murina leucogaster	AB085733
Miniopterus schreibersii	EF530347
Miniopterus schreibersii	AF376830

References

Aris-Brosou, S. e. and X. Xuhua (2008). "Phylogenetic Analyses: A Toolbox Expanding towards Bayesian Methods." International Journal of Plant Genomics **2008**: 1-16.

Badrane, H. and N. Tordo (2001). "Host switching in Lyssavirus history from the Chiroptera to the Carnivora orders." J Virol **75**(17): 8096-104.

Barrett, R., C. W. Kuzawa, et al. (1998). "Emerging and Re-Emerging Infectious Diseases: The Third Epidemiologic Transition." Annual Review of Anthropology **27**: 247-271.

Benson, D. A., I. Karsch-Mizrachi, et al. (2008). "GenBank." Nucl. Acids Res. **36**(suppl_1): D25-30.

Brooks, D. R. (1988). "Macroevolutionary Comparisons of Host and Parasite Phylogenies." Annual Review of Ecology and Systematics **19**(1): 235-259.

Calisher, C. H., J. E. Childs, et al. (2006). "Bats: Important Reservoir Hosts of Emerging Viruses." Clin. Microbiol. Rev. **19**(3): 531-545.

Charleston, M. A. (1998). "Jungles: a new solution to the host/parasite phylogeny reconciliation problem." Math Biosci **149**(2): 191-223.

Cleaveland, S., M. K. Laurenson, et al. (2001). "Diseases of humans and their domestic mammals: pathogen characteristics, host range and the risk of emergence." Philosophical Transactions of the Royal Society of London. Series B: Biological Sciences **356**(1411): 991-999.

Davies, T. J. and A. B. Pedersen (2008). "Phylogeny and geography predict pathogen community similarity in wild primates and humans." Proceedings of the Royal Society B: Biological Sciences **275**(1643): 1695-1701.

Davis, P. L., E. C. Holmes, et al. (2005). "Phylogeography, Population Dynamics, and Molecular Evolution of European Bat Lyssaviruses." J. Virol. **79**(16): 10487-10497.

Delmas, O., E. C. Holmes, et al. (2008). "Genomic Diversity and Evolution of the Lyssaviruses." PLoS ONE **3**(4): e2057.

Drummond, A. and A. Rambaut (2007). "BEAST: Bayesian evolutionary analysis by sampling trees." BMC Evolutionary Biology **7**(1): 214.

Hanada, K., Y. Suzuki, et al. (2004). "A Large Variation in the Rates of Synonymous Substitution for RNA Viruses and Its Relationship to a Diversity of Viral Infection and Transmission Modes." Mol Biol Evol **21**(6): 1074-1080.

Holmes, E. C. (2004). "The phylogeography of human viruses." Mol Ecol **13**(4): 745-56.

Huelsenbeck, J. P., B. Rannala, et al. (2000). "A Bayesian framework for the analysis of cospeciation." Evolution **54**(2): 352-64.

Hughes, G. J. (2008). "A reassessment of the emergence time of European bat lyssavirus type 1." Infection, Genetics and Evolution **8**(6): 820-824.

Hughes, G. J., L. A. Orciari, et al. (2005). "Evolutionary timescale of rabies virus adaptation to North American bats inferred from the substitution rate of the nucleoprotein gene." J Gen Virol **86**(5): 1467-1474.

IUCN (2009). IUCN Red List of Threatened Species. Version 2009.2. <http://www.iucnredlist.org>. Downloaded on **25 February 2010**.

Jackson, A. P. and M. A. Charleston (2004). "A cophylogenetic perspective of RNA-virus evolution." Mol Biol Evol **21**(1): 45-57.

Jones, K. E., P. Andy, et al. (2002). "A phylogenetic supertree of the bats (Mammalia: Chiroptera)." Biological Reviews **77**(2): 223-259.

Kuiken, T., E. C. Holmes, et al. (2006). "Host Species Barriers to Influenza Virus Infections." Science **312**(5772): 394-397.

Kuzmin, I. V., G. J. Hughes, et al. (2005). "Phylogenetic relationships of Irkut and West Caucasian bat viruses within the Lyssavirus genus and suggested quantitative criteria based on the N gene sequence for lyssavirus genotype definition." Virus Res **111**(1): 28-43.

Lichty, B. D., A. T. Power, et al. (2004). "Vesicular stomatitis virus: re-inventing the bullet." Trends in Molecular Medicine **10**(5): 210-216.

Meier-Kolthoff, J. P., A. F. Auch, et al. (2007). "COPYCAT : cophylogenetic analysis tool." Bioinformatics **23**(7): 898-900.

Nel, L. H. and W. Markotter (2007). "Lyssaviruses." Critical Reviews in Microbiology **33**(4): 301-324.

Nylander, J. A. A. (2004). MrModeltest v2, Program distributed by the author. Evolutionary Biology Centre, Uppsala University.

Page, R. D. M. (2003). Introduction. Tangled Trees: Phylogeny, Cospeciation, and Coevolution. R. D. M. Page. Chicago, University of Chicago Press**:** 1-21.

Parrish, C. R., E. C. Holmes, et al. (2008). "Cross-Species Virus Transmission and the Emergence of New Epidemic Diseases." Microbiol. Mol. Biol. Rev. **72**(3): 457-470.

Paterson, A. M. and J. Banks (2001). "Analytical approaches to measuring cospeciation of host and parasites: through a glass, darkly." Int J Parasitol **31**(9): 1012-22.

Paul-Pierre, P. (2009). "Emerging diseases, zoonoses and vaccines to control them." Vaccine **27**(46): 6435-6438.

Posada, D. and K. Crandall (1998). Modeltest: testing the model of DNA substitution, Bioinformatics 14 (9): 817-818.

R Development Core Team (2008). R: A language and environment for statistical computing. R Foundation for Statistical Computing, Vienna, Austria. ISBN 3-900051-07-0, URL http://www.R-project.org.

Ronquist, F. (2003). Parsimony Analysis of Coevolving Species Associations. Tangled Trees: Phylogeny, Cospeciation, and Coevolution. R. D. M. Page. Chicago, University of Chicago Press**:** 22-64.

Simmons, N. B. and T. M. Conway (2003). Evolution and Ecological Diversity in Bats. Bat Ecology. T. H. Kunz and M. B. Fenton. Chicago, University of Chicago Press**:** 493-535.

Stevens, J. (2004). "Computational aspects of host-parasite phylogenies." Brief Bioinform **5**(4): 339-49.

Swofford, D. L. (2003). PAUP*. Phylogenetic Analysis Using Parsimony (*and Other Methods). Version 4. Sinauer Associates, Sunderland, Massachusetts.

Teeling, E. C., M. S. Springer, et al. (2005). "A Molecular Phylogeny for Bats Illuminates Biogeography and the Fossil Record." Science **307**(5709): 580-584.