Comparative Phylogeography Across Multiple Scales: Small Mammals, Their Ecology, Pathogens, and Drivers of Diversification

Schuyler W. Liphardt

Follow this and additional works at: https://digitalrepository.unm.edu/biol_etds

Part of the Biology Commons, Evolution Commons, Population Biology Commons, and the Zoology Commons

Recommended Citation

This Dissertation is brought to you for free and open access by the Electronic Theses and Dissertations at UNM Digital Repository. It has been accepted for inclusion in Biology ETDs by an authorized administrator of UNM Digital Repository. For more information, please contact disc@unm.edu.
Schuyler W. Liphardt  
Candidate

Department of Biology  
Department

This dissertation is approved, and it is acceptable in quality and form for publication:

Approved by the Dissertation Committee:

Dr. Joseph A. Cook, Chairperson

Dr. Helen J. Wearing

Dr. Donald O. Natvig

Dr. Steven B. Bradfute
COMPARATIVE PHYLOGEOGRAPHY ACROSS MULTIPLE SCALES: SMALL MAMMALS, THEIR ECOLOGY, PATHOGENS, AND DRIVERS OF DIVERSIFICATION

by

SCHUYLER W. LIPHARDT

B.S., Biology, University of New Mexico, 2014

DISSertation

Submitted in Partial Fulfillment of the Requirements for the Degree of

Doctor of Philosophy
Biology

The University of New Mexico
Albuquerque, New Mexico

December 2021
COMPARATIVE PHYLOGEOGRAPHY ACROSS MULTIPLE SCALES: SMALL MAMMALS, THEIR ECOLOGY, PATHOGENS, AND DRIVERS OF DIVERSIFICATION

by

Schuyler W. Liphardt
B.S., Biology, University of New Mexico, 2014
Ph.D., Biology, University of New Mexico

ABSTRACT

Comparative phylogeography has historically been defined as the study of how genetic variation of co-distributed species has been shaped by biogeographical history. This is a mature field of study out of which several techniques have been developed directed at identifying the role ecology and geography in diversification patterns across time. I employ the tools developed in classical comparative phylogeography across multiple taxonomic scales and across regions that historically have been understudied. My first two chapters study the complex history of host-switching, codiversification, and reassortment of hantaviruses in their mammal hosts across North America. By taking a host-centric comparative phylogeographic approach to pathogen divergence, I highlighted complex evolutionary processes in host-pathogen systems and the role of host history in shaping the distribution and diversity of pathogens. My final chapter uses a comparative approach to examine the interplay between ecology and climate in shaping divergence and contact within four East Asian pika species during the Quaternary. I show that small mammals in this vast, poorly studied region responded to Pleistocene climatic cycling on finer geographic scales when compared to the North American fauna that is distributed at similar latitudes.
TABLE OF CONTENTS

INTRODUCTION .................................................................................................................1

  Literature Cited ............................................................................................................4

CHAPTER 1: Complex history of codiversification and host switching of a newfound
Soricid-borne orthohantavirus in North America .........................................................5

  Abstract .........................................................................................................................6
  Introduction ....................................................................................................................7
  Materials and Methods ...............................................................................................9
  Results ..........................................................................................................................12
  Discussion ....................................................................................................................17
  Acknowledgements ......................................................................................................23
  References ....................................................................................................................24
  Figure Legends ..........................................................................................................32
  Table ..........................................................................................................................34
  Figures .........................................................................................................................35
  Supplemental .............................................................................................................40

CHAPTER 2: Reassortment between divergent strains of Camp Ripley virus
(Hantaviridae) in the northern short-tailed shrew (Blarina brevicauda) ......................46

  Abstract .........................................................................................................................47
CHAPTER 3: Asynchronous divergence in Inner Asian pikas is consistent with predictions of how niche preference shapes differential response to climate fluctuations
INTRODUCTION

Why organisms are distributed where they are, how they got there, and what happened to them along the way are questions that have interested scientists for hundreds of years. The field of phylogeography combines phylogenetics and biogeography to attempt to answer these questions using molecular data (Avise 1987, 2000). An extension of classical phylogeography expands on the focus of geographic variation within a single organism to comparisons of closely related and co-distributed species to search for signals of congruent responses to the same biogeographic history, an approach termed comparative phylogeography (Avise, 2009; Gutiérrez-Garica & Vázquez-Domínguez, 2011). The approaches and techniques developed in comparative phylogeography are broadly applicable to comparative studies beyond systems of co-distributed species, such as parasite-host systems.

My dissertation is focused on applying classical tools of comparative phylogeography, combined with newer statistical and coalescent approaches, to systems historically understudied using these techniques. I address questions that range across temporal, spatial, and taxonomic scales. In my first two chapters, I focus on the evolutionary dynamics of host-pathogen systems and address how host phylogeographic history may inform us about the maintenance, distribution, and evolution of hantaviruses within Eulipotyphlan (shrews and voles) mammal hosts. Historically, interest in host biology in virological studies has not extended beyond the probability of spillover to human populations. However, understanding the phylogeography of pathogens in the context of the host phylogeography can provide key insights into pathogens persistence, diversify, and dispersal across landscapes. My final chapter shifts from organism-organism interactions to
the impact of Pleistocene climatic cycling on four species of pika in East Asia, an area underrepresented in phylogeographic literature compared to European and North American regions of similar latitude.

The first chapter of my dissertation, *Complex History of Codiversification and Host Switching of Newfound Soricid-borne Orthohantavirus in North America*, examines the phylogeographic history of Jemez Springs hantavirus in the context of the small mammals that serve as hosts, shrews in the *Sorex vagrans* species complex. This system presents an opportunity to examine how host response to Pleistocene glacial cycling and conspecific contact has shaped the distribution and diversification of a pathogen through deep evolutionary time. We found three distinct clades of virus distributed across western North America. Two of these clades appear to have codiverged with a single species of shrew, *Sorex monticolus*, largely mirroring host colonization events of the late Quaternary, while a third clade is widespread across multiple species of shrews along the northern Pacific coastal region, a proposed Pleistocene refugia. This pattern of viral divergence likely results from a complex history of processes spanning temporal, population, and species scales.

My second chapter, *Reassortment Between Divergent Strains of Camp Ripley Virus (Hantaviridae) in the Northern Short-tailed Shrew (Blarina brevicauda)*, tests hypotheses of classical host codiversification across eastern North America. In this system, I investigated the genetic consequences of host divergence and secondary contact for an obligate virus. While I did recover a signal of codiversification with its host in one clade of Camp Ripley virus, I also discovered evidence of viral reassortment across the *Hantaviridae* phylogeny. The phylogeographic history of Camp Ripley virus is discordant with that of its host leading to hypotheses of possibly multiple intermediate hosts prior to colonization of *B. brevicauda*. 

2
My last chapter, *Asynchronous Divergence in Eastern Asian pikas is Consistent with Predictions of How Niche Preference Shapes Differential Response to Climate Fluctuations*, compares the phylogeographic history of four species of pika distributed across East Asia. This chapter combines classical approaches of comparative phylogeography with species distribution modeling and coalescent simulations to unravel how species ecology, when coupled with Pleistocene climatic cycling, shaped patterns of divergence in the relatively understudied northern latitudes of East Asia. We found that species responded to climatic shifts differentially according to ecologies exhibited. The two montane species examined, *Ochotona alpina* and *O. hyperborea*, show a stepping-stone pattern of diversification with populations associated with distinct geographic regions with colonization events likely tied to periods of cooling and plateau glaciation. *Ochotona daurica*, a purely steppe-adapted species, presents a phylogeographic history suggesting a large and stable core population that persisted through time with possible pulses of expansion during interglacial warming. Meanwhile, *O. pallasi*, a species that exhibits an intermediate ecology between montane- and steppe-adapted, appears to have responded to climatic cycling in an intermediate way, with allopatric divergence of a core population during cooling periods, but followed by retreat and secondary contact during warming cycles.
References


Chapter 1

COMPLEX HISTORY OF CODIVERSIFICATION AND HOST SWITCHING OF A NEWFOUND SORICID-BORNE ORTHOHANTAVIRUS IN NORTH AMERICA

Schuyler W. Liphardt, Hae Ji Kang, Laurie J. Dizney, Luis A. Ruedas, Joseph A. Cook and Richard Yanagihara

1 Department of Biology and Museum of Southwestern Biology, University of New Mexico, Albuquerque, NM 87131, USA
2 Department of Pediatrics, John A. Burns School of Medicine, University of Hawaii at Manoa, Honolulu, HI 96813, USA
3 Department of Biology, University of Portland, Portland, OR 97203, USA
4 Department of Biology and Museum of Vertebrate Biology, Portland State University, Portland, OR 97207-0751, USA
ABSTRACT

Orthohantaviruses are tightly linked to the ecology and evolutionary history of their mammalian hosts. We hypothesized that in regions with dramatic climate shifts throughout the Quaternary, orthohantavirus diversity and evolution are shaped by dynamic host responses to environmental change through processes such as host isolation, host switching, and reassortment. Jemez Springs virus (JMSV), an orthohantavirus harbored by the dusky shrew (*Sorex monticola*) and five close relatives distributed widely in western North America, was used to test this hypothesis. Total RNAs, extracted from liver or lung tissue from 164 shrews collected from western North America during 1983–2007, were analyzed for orthohantavirus RNA by reverse transcription polymerase chain reaction (RT-PCR). Phylogenies inferred from the L-, M-, and S-segment sequences of 30 JMSV strains were compared with host mitochondrial cytochrome *b*. Viral clades largely corresponded to host clades, which were primarily structured by geography and were consistent with hypothesized post-glacial expansion. Despite an overall congruence between host and viral gene phylogenies at deeper scales, phylogenetic signals were recovered that also suggested a complex pattern of host switching and at least one reassortment event in the evolutionary history of JMSV. A fundamental understanding of how orthohantaviruses respond to periods of host population expansion, contraction, and secondary host contact is the key to establishing a framework for both more comprehensive understanding of orthohantavirus evolutionary dynamics and broader insights into host–pathogen systems.
INTRODUCTION

Comparative phylogeographic study of parasites and their hosts can provide valuable insights into evolutionary history and processes of diversification for both host and parasite. These analyses also provide a predictive framework for emerging disease under scenarios of changing environmental conditions (Brooks & Ferrao, 2005; Geoghegan & Holmes, 2017; Nieberding et al., 2004). The classic model of strict host-parasite codiversification as the leading process driving parasite evolution has been thrown into question over the past several decades with studies suggesting that host switching is not only more prevalent than originally thought, but that it may be the most common pattern throughout parasite evolutionary history (Araujo et al., 2015; Geoghegan et al., 2017).

Viruses are hypothesized to follow processes of diversification in a manner similar to organisms more traditionally considered to fulfil the ecological role of parasites, such as helminths and bacteria. Members of the recently reclassified genus Orthohantavirus (order Bunyavirales, family Hantaviridae) (Maes et al., 2019) can cause severe disease in humans. These viruses have a complex history of both codiversification with their mammal hosts (Torres-Pérez et al., 2011) and host switching across mammalian species (Kang et al., 2009; Nemirov et al., 2002). Viral reassortment—the exchange of viral material between divergent strains (Arai et al., 2012; Briese et al., 2013; Klempa, 2018)—is another process that can rapidly lead to novel variants (e.g., 2009 human influenza A H1N1 virus (Trifonov et al., 2009)). Reassortment may be frequent in orthohantaviruses; however, work to date has mainly focused on reassortment within orthohantavirus strains hosted by single species (Klempa, 2018).
Jemez Springs virus (JMSV), originally discovered in the dusky shrew (*Sorex monticola*) in New Mexico (Arai et al., 2008), has subsequently been detected in other closely related members of the *Sorex vagrans* complex. The broad distribution of JMSV across western North America provides an ideal system to begin to examine the prevalence of viral codiversification, host switching, and reassortment within a mammalian species complex.

*Sorex monticola* (Order Eulipotyphla, Family Soricidae) is endemic to North American montane forests in the western United States, Canada, and Mexico, and is a member of the *Sorex vagrans* complex (Hennings & Hoffmann, 1977). Demboski and Cook (2001) suggested that rapid diversification early in the history of these mammals was stimulated by repeated population isolation and expansion events due to cyclical glaciation events in the Late Quaternary and the Last Glacial Maximum. This pattern is reflected in the geographic structure of clades elucidated by phylogenetic analyses of *S. monticola* and close relatives in western North America (Demboski & Cook, 2001), including the American water shrews, *Sorex palustris* and *Sorex navigator* (Hope et al., 2014), of which *S. palustris* also hosts JMSV.

Seewis virus in Finland also shows a strong signal of post-glacial expansion and codiversification of the virus with its host, the common shrew, *Sorex araneus* (Ling et al., 2018). Phylogeographic studies of Seewis virus in Siberia suggest a similar pattern of codiversification (Yashina et al., 2010), but with instances of host switching followed by localized variation in closely related species (Gu et al., 2014; Kang et al., 2009; Ling et al., 2014). Similar patterns are found in Puumala virus with preliminary evidence for post-glacial colonization of Finland coinciding with expansion of the mammalian host, the bank vole.
(Myodes glareolus) (Asikainen et al., 2000). However, subsequent studies that were more geographically extensive showed little evidence supporting strict codiversification of Puumala virus and the bank vole (Nemirov et al., 2010).

Phylogeographic study has helped to uncover the complex forces and interactions responsible for the maintenance, transmission, and diversification of viruses, especially those capable of causing disease in humans (Holmes, 2004). Here we examine a species complex broadly distributed throughout North America to help bridge the gap between deep-time phylogenetic analyses of viral diversification across major mammalian clades and studies that examine relatively recent viral evolution within a single host species. We extend our understanding of processes and patterns of virus evolution and ecology by taking a comparative phylogeographic approach to gain insights into the evolutionary history of JMSV in S. monticola and its close relatives.

MATERIALS AND METHODS

Specimens

We used liver or lung tissues from 69 dusky shrews, 10 American water shrews, 17 vagrant shrews, 9 Trowbridge shrews (Sorex trowbridgii) and 1 prairie shrew (Sorex haydeni), archived in the Museum of Southwestern Biology at the University of New Mexico in Albuquerque (Table 1). Also tested were lung tissues of 28 vagrant shrews, 26 Trowbridge shrews, 3 Pacific water shrews (Sorex bendirii) and 1 Baird’s shrew (Sorex bairdi), captured in Multnomah and Washington counties, near Portland, Oregon, between April 2003 and October 2005 and archived at the Museum of Natural History, at the Portland State University (Table 1), as part of an epizootic study of Sin Nombre virus infection in deer mice.
(Peromyscus maniculatus) (Dizney & Ruedas, 2009). All tissue samples were frozen at -70°C until tested by RT-PCR.

**RNA Extraction, cDNA Synthesis and RT-PCR Amplification**

Total RNA was extracted from 20–50 mg of each tissue using the PureLink Micro-to-Midi total RNA purification kit (Invitrogen, San Diego, CA). cDNA was prepared using the SuperScript III First-Strand Synthesis System (Invitrogen) and oligonucleotide primer (5’-TAGTAGTAGACTCC-3’) designed from the conserved 3’-end of the S, M and L segments of orthohantaviruses.

Gene amplification was carried out in 20-μL reaction mixtures containing 250 μM dNTP, 2 mM MgCl₂, 1 U of AmpliTaq polymerase (Roche, Basel, Switzerland), and 0.25 μM of oligonucleotide primers designed from highly conserved regions of previously identified soricid-borne orthohantaviruses. Initial denaturation was followed by touchdown cycling (two-degree step-down annealing from 48°C to 38°C for 40 sec) and elongation at 72°C for 1 min, then 32 cycles of denaturation at 94°C for 40 sec, annealing at 42°C for 40 sec, and elongation at 72°C for 1 min, in a GeneAmp PCR 9700 thermal cycler (Perkin-Elmer, Waltham, MA). Amplified products were separated by electrophoresis on 1.5% agarose gels and purified using the QIAQuick Gel Extraction Kit (Qiagen, Hilden, Germany). DNA was sequenced directly using an ABI Prism 377XL Genetic Analyzer (Applied Biosystems Inc., Foster City, CA).

**Sequence Dataset**

Virus sequences, either generated in this study or downloaded from GenBank (https://www.ncbi.nlm.nih.gov/genbank/), of all currently known strains of JMSV and their respective hosts spanning the range of JMSV, as well as outgroups for rooting of the trees,
were studied (Figure S1A). The final dataset, with outgroups, was composed of partial coding regions for the L (n=31), M (n=10), and S (n=29) segments of orthohantaviruses. In addition, mitochondrial cytochrome \( b \) (cyt \( b \)) gene sequences, generated for each mammalian host across the range of JMSV variation, were included to independently examine host phylogenetic relationships (Figure S1B). GenBank accession numbers for all sequences used in this study are in the figure captions for Figures 1, 2, S2, and S6.

**Phylogenetic Analysis**

Phylogenetic trees were generated from alignments of each individual genomic segment (S, M, and L) and Ash River virus from *Sorex cinereus* (Arai et al., 2008) as an outgroup using Muscle version 8.1.9 (Edgar, 2004), as implemented in Geneious 8.1 (https://www.geneious.com). To detect instances of intergenic rearrangement and assure we were analyzing homologous genomic regions, these independent alignments were tested for recombination using the Phi (T. C. Bruen et al., 2006), NSS (Jakobsen & Easteal, 2007), and Max \( \chi^2 \) (Maynard, 1992) tests as implemented in PhiPack (T. Bruen, 2005). Both a maximum likelihood approach using RAxML version 8.2.12 (Stamatakis, 2014), and Bayesian probability implemented in MrBayes version 3.2.5 (Ronquist et al., 2012), were employed for tree inference. A general time reversible (GTR) model of nucleotide evolution with gamma-distributed rate heterogeneity and invariant sites (RaxML option GTRGAMMAI) was determined to be the best fit model for this dataset by JmodelTest (Darriba et al., 2012). RAxML generated 1,000 bootstrap replicates to determine the best-fit maximum likelihood tree and associated nodal support. MrBayes was run for 10 million generations using the priors from JmodelTest by sampling trees every 1,000 generations.
After the first 25% of trees were discarded as a recommended burn-in, the remaining trees were used to calculate a 50% majority rule consensus tree.

**Tanglegrams, Diversity Analyses, Codiversification Tests, and Reconciliation**

Tajima’s D statistic for both virus and host alignments were computed in R (R Core Team, 2019) with the package Pegas (Paradis, 2010). Negative values for Tajima’s D can be indicative either of population expansion or of a selective sweep (Lessa et al., 2003). Individual pairwise distances and between group mean distances were calculated in Mega7 (Kumar et al., 2016) with a Kimura 2-parameter model and 500 bootstrap replicates. To test codivergence versus host switching, several metrics that measure the extent of similarity between phylogenetic trees were employed. The nPH85 metric (Geoghegan et al., 2017) is a normalized version of the Robinson Foulds tree topology distance that incorporates branch lengths and returns a value between 0 and 1, with 0 indicative of strict codiversification between identical tree topologies or 1 indicative of cross species transmission with completely incongruent tree topologies. Trip (Critchlow et al., 1996) and TripL (Kuhner & Yamato, 2015) measure the number of shared triplets between two trees, with TripL incorporating branch lengths and returning the proportion of triplets not shared between two trees. In this metric, 0 indicates identical trees and 1 indicates no shared triplets. Phylogenetic reconciliation was performed for the L and S segment with host cyt b phylogeny in Jane 4.01 (Conow et al., 2010) with equal costs of 1 assigned to all possible events. Finally, to visualize tree discordance, tanglegrams were generated for each tree comparison in R with the package phytools (Revell, 2012).

**RESULTS**

*Phylogenetic Analysis*
Significant values for recombination analyses differed among segment and test used but identified no regions of intragenic rearrangement in any segment analyzed. The JMSV strains showed a pattern of diversification with defined clades that closely mapped to geographic regions similar to that seen in the phylogeographic structure of the hosts, with the exception of the virus recovered from *S. palustris*, which was previously identified as Fox Creek virus (Bennett et al., 2014; Figure 1). Geographic regions correspond to Northern Continental (NC) and Southern Continental (SC) clades composed of viruses recovered strictly from *S. monticola*. The NC clade contained JMSV strains from British Columbia and Yukon Territory, as well as Alaska, while the SC clade was found in New Mexico and Colorado. A third clade was recovered from the Pacific coast (Coastal) which, in comparison to the NC and SC clades, was hosted by several species of the *S. vagrans* complex, including *S. vagrans*, *S. trowbridgii*, *S. bairdi*, and *S. palustris*. Phylogenies inferred using the L and S segment differed in regard to the placement of JMSV strains recovered from *S. vagrans* in Washington (Figure S2). For host cyt *b*, *S. trowbridgii* and *S. vagrans* were supported as monophyletic; however, *S. monticola* was shown to be paraphyletic, with a coastally distributed clade, including the sole *S. bairdi* sample, that is sister to *S. palustris*, and a larger continental clade that is composed of northern and southern subclades containing viruses from the NC and SC clades, respectively (Figure 2).

Population Demographics

Tajima’s D statistic for all three viral genomic segments, as well as host cyt *b*, were -3.8, -4.5, -3.8, and -3.1 for the S, M, L segments and host cyt *b*, respectively, with all values being highly significant (*P < 0.01*). Overall mean diversity, as computed in MEGA using a Kimura 2-parameter model, was 0.256. This relatively high level of nucleotide diversity,
common in RNA viruses with high mutation rates, is driven mainly by the Coastal clade and between group divergence. Within clade distances were highest in the Coastal clade at 0.196 followed by the NC clade at 0.125. The SC clade had the lowest diversity at 0.053; however, this relatively low value could be an artifact of smaller sample size.

*Cophylogeny Tanglegrams*

Cophylogeny of the L segment and host tree reflected a pattern of codiversification within continental *S. monticola*, but a complex pattern of host switching across multiple species within the Coastal clade. Host phylogeny based on cyt *b* placed the NC and SC clades as sister lineages forming a larger continental *S. monticola* group that matched the pattern of diversification seen in the L segment phylogeny (Figure 3). Included in the JMSV L segment Coastal clade were viruses harbored by *S. vagrans*, *S. trowbridgii*, and *S. bairdi* that corresponded to taxa spread across the host phylogeny. Of note was the placement of *S. bairdi* within the coastal *S. monticola* clade, which represented the only sample of JMSV from that clade. Additionally, JMSV recovered from *S. trowbridgii* was distributed across the JMSV Coastal clade, yet *S. trowbridgii* was only distantly related to the *S. vagrans* complex. *Sorex cinereus*, the host of Ash River virus, which was used as the outgroup to JMSV in this study, was more closely related to the *S. vagrans* complex than *S. trowbridgii*.

Cophylogenies based on the M segment (Figure S3) and S segment (Figure S4) of JMSV showed similar patterns to those recovered with the L segment. However, phylogeny reconstruction based on a single gene for the hosts, in this case cyt *b*, can be misleading and should be further tested with additional, independent loci for this complex of shrew species.

In our study, all segments of JMSV supported monophyletic NC and SC clades that mapped to divergent, reciprocally monophyletic *S. monticola* clades in the host phylogeny
and corresponded to the same geographic regions. While the shallow topology (i.e., branching near the terminal tips) seen within JMSV and the NC and SC clades of *S. monticola* differed and do not map one-to-one, the geographic clades were reciprocally monophyletic and generally supported a pattern of host-parasite codiversification within the continental group.

In contrast, the Coastal clade of JMSV showed a complex pattern of host switching among sympatric, yet often deeply divergent species of shrews along the Pacific coast. While the Coastal clade is monophyletic in all segments of JMSV, this clade is distributed across four separate host species, hence indicative of widespread host switching. Of note among the JMSV Coastal clade was the placement of a virus recovered from *S. palustris* that was well supported within the Coastal clade. The host specimen was, however, collected in the Yukon Territory, well within the range of the NC clade of both JMSV and *S. monticola*. This pattern of host switching was mirrored by the JMSV strains hosted by *S. trowbridgii* and *S. vagrans* that were spread across the Coastal clade and did not form a single species-specific clade.

Virus-virus cophylogenies based on the L and S segments largely mirrored each other with the exception of the placement of two JMSV strains recovered from *S. trowbridgii* from Washington, which likely was an artifact of sequence coverage for those two segments. A single virus recovered from a *S. vagrans* specimen from Vancouver Island, British Columbia, was not supported as being a member of any of the three geographic clades within JMSV for the S segment, although it was supported in the L segment on a long branch within the NC clade (Figure 3 and Figure S4).

*Codivergence and Phylogenetic Reconciliation*
We compared metrics that test for codivergence as they fundamentally differ in their methods of calculating tree similarity. The nPH85 statistic was similar for both S and L segments: their relatively high values, 0.80 and 0.76 respectively, indicated host switching rather than codiversification as the leading pattern of diversification within JMSV. Codiversification previously has been reported as the more prevalent coevolutionary pattern within Bunyavirales (Geoghegan et al., 2017). That pattern, however, is not necessarily reflected in the TripL and Trip metrics. For the L segment, the TripL and Trip metrics were 0.41 and 0.44, respectively, indicating that the L segment and host phylogenies shared roughly 60% of triplets. This is a rather stark comparison to the TripL and Trip metrics for the S segment at 0.23 and 0.19 respectively relating to roughly 80% of shared triplets between the S segment and host phylogenies.

In contrast to the metrics of codivergence between the L and S segment with the host phylogeny are the metrics comparing the L segment to the S segment, which in some cases show greater similarity between virus and host then between segments of JMSV. While the nPH85 metric for the S and L comparison is less than that for either comparison with the host, it is still relatively elevated at 0.5 indicating that the two segments share roughly half of their internal structure to each other. Also striking is the TripL and Trip scores for the comparison between segments at 0.46 and 0.49 respectively, also indicating that in addition to the variation on internal tree structure the S and L segment only share about half of their triplets, less than the S segment shares with the host phylogeny. Differences in evolutionary history between the S and L segments of JMSV are reflected by phylogenetic reconciliation for each segment with the host phylogeny (Figure 5). These segments suggest that a distinctive set of host switching, codivergence, and local extinction events, are necessary to
reconcile the virus segment phylogeny with the host phylogeny. This pattern of independence between segments is consistent with a history of reassortment among the L and S segments of JMSV.

**DISCUSSION**

*Codiversification Processes*

Comparative phylogeography of viruses and their associated mammal reservoir hosts can shed light on the processes driving patterns of coevolutionary diversification (Brooks et al., 2014; Geoghegan & Holmes, 2017) by revealing the role of historical events that shaped contemporary diversity. Contact between divergent hosts may facilitate transmission of viruses to novel hosts, or reassortment of divergent viral components. Comparative phylogeographic studies provide the spatial and temporal foundation necessary for understanding viral evolution, transmission, and disease emergence; these are essential tools for researchers and public health agencies to proactively approach disease emergence and mitigation. In this study, we addressed the evolutionary history that shaped modern diversity within JMSV and associated mammals hosting this virus.

Diversification within JMSV largely reflects the recent biogeographic history of the shrew host species. The phylogeny inferred from cyt *b* for the *S. vagrans* complex supports previously reported species designations and relationships (Demboski & Cook, 2001). Demboski and Cook (2001) recovered substantial geographic structure within *S. monticola*, identifying distinct clades distributed in northern and southern continental North America, and a third distributed along the Pacific coast. Representative clades in JMSV match the NC and SC clades respectively, while the Coastal clade is comprised of viruses recovered from several other species in the *S. vagrans* complex, including *S. bairdi* which falls within the
Pacific coastal *S. monticola* clade. That pattern ostensibly parallels two suggested evolutionary diversification events within the *S. vagrans* species complex (Demboski & Cook, 2001). To date, the absence of JMSV in *S. monticola* specimens representative of the Coastal clade (Oregon and Washington) may reflect true absence or simply low sampling coverage. The validity of *S. bairdi* as a species separate from coastal *S. monticola* is questionable and points to other poorly defined species limits in this shrew complex that complicate our assessment (Demboski & Cook, 2001). Expanded shrew sampling and viral screening that aims to refine the geographic extent of host limits and viral diversity in western North America are necessary.

The hypothesized initial divergence event within the *S. vagrans* complex occurred in the Pacific Northwest coast resulting in current inter-species diversity seen within the complex. Subsequent post-glacial expansion followed the Last Glacial Maximum and produced the currently recognized geographic structure within montane shrews (e.g., NC and SC clades) and their close relatives (e.g., *Sorex palustris* (Hope et al., 2014)) (Figure 6). With JMSV largely mirroring the host pattern, codiversification between JMSV and its shrew hosts appears likely within the NC and SC clades. JMSV emergence tentatively can be dated to the initial diversification of the *S. vagrans* complex ca. 2 MYA (Demboski & Cook, 2001), with subsequent, possibly multiple independent, post-glacial expansion events within *S. monticola*, and likely *S. palustris*, during the mid to late Quaternary. Whether the virus recovered from *S. palustris* within the Coastal clade is the result of a recent host-switching event is unclear, and more sampling is necessary to refine our understanding of the evolutionary history of JMSV within *S. palustris*. The single virus recovered from an *S. vagrans* on Vancouver Island does not align within a defined mainland clade in the *S.
segment and exists on a long branch in the L segment. This finding raises the question of whether there is an endemic insular clade of JMSV similar to that identified for insular *S. monticola* (Sawyer et al., 2019). Highly negative Tajima’s D values for all three viral segments, coupled with observed patterns of sequence divergence centered in the hypothesized source population of the Coastal clade, strengthens the hypothesis of post-glacial expansion for JMSV (Figure 6B). While the timing of orthohantavirus diversification remains elusive (Duffy et al., 2008; Torres-Pérez et al., 2011; Zhang & Holmes, 2014), this shared pattern of diversification, both spatially and temporally, between JMSV and its mammalian hosts is consistent with a cophylogeographic history that is much deeper than several thousand years.

While phylogenies for JMSV and its host species largely mirror each other in deeper phylogenetic structure, codiversification only is partially responsible for contemporary diversity. Elevated codivergence metrics calculated between trees suggest host switching also plays an important role. Geoghegan and colleagues (Geoghegan et al., 2017) applied the nPH85 metric to family level phylogenies for several RNA and DNA viruses. Our results for this metric largely match the general pattern seen for other RNA viruses (Geoghegan et al., 2017). When comparing topologies between the host phylogeny and either the S or L segment, we calculated an nPH85 value of 0.8 and 0.76, indicating scant codivergence, which is consistent with the values seen at the family level. Furthermore, when calculated for the comparison between the L and S segments, we obtained a value of 0.5, indicating a mix of codivergence and host switching between each segment. This is in contrast with the tanglegram for the L and S segment comparison (Figure S5) which indicated the phylogenies largely mirrored each other. However, the level of similarity between tree topology and
groupings of terminal taxa do not necessarily tell the same story as indicated by drastically
different TripL and Trip values compared to nPH85. Internal branching structure is driving
the difference in codivergence metrics calculated between phylogenies inferred from the L
and S segments. However, our sampling of the L segment is more complete in terms of
number of specimens and coverage of the segment. This fact, coupled with our phylogenetic
reconciliation analysis, suggests that in addition to a complex history of post-glacial
codivergence and local host switching between JMSV and the S. vagrans complex as a
whole, there is additional complexity due to independent evolutionary histories associated
with distinct viral segments. Increased sampling and full-length genomic sequencing would
help test whether incongruity between the S and L segments, both tree topology and
similarity metrics, reflects deeper phylogenetic patterns or is merely an artifact of sampling
and sequencing bias.

Viral Reassortment

Viral reassortment is possible for viruses with segmented genomes and can be a
catalyst for driving diversification and pathogenesis. The 2009 influenza outbreak is a prime
eexample of reassortment among multiple divergent strains resulting in a pandemic (Trifonov
et al., 2009). Reassortment produces unique combinations of viral segments that have the
potential to influence pathogenicity due to presentation of novel virions to an
immunologically naïve population. Co-circulation of divergent viruses within a single cell is
hypothetically necessary for reassortment and calls for a more detailed understanding of the
range of ecological circumstances that can lead to viral switching between hosts if we hope
to predict disease emergence (Geoghegan & Holmes, 2017).
Historically, reassortment events within orthohantaviruses were thought to be relatively rare compared to other members of Bunyavirales due to method of transmission (i.e., direct host-host contact versus arthropod transmission, respectively) (Briese et al., 2013). However, an increasing number of studies of orthohantaviruses have shown instances of reassortment (Black et al., 2009; Razzauti et al., 2008, 2009; Zou et al., 2008), both ancient and modern. The extent of reassortment and its contribution to modern-day orthohantavirus diversity is a new avenue of research and there is still much to learn.

Implicit in orthohantavirus reassortment events is the necessity for contact between hosts for transmission of divergent viruses, or what are called spillover events. For instance, there is evidence in Finland for two distinct strains of Puumala virus co-circulating with active and ongoing reassortment resulting from contact, both modern and historic, between divergent populations of the host species, *M. glareolus* (Razzauti et al., 2009). A more contemporary example of co-circulation was reported in Belgium with two deeply divergent strains of *Hantaviridae* co-circulating in the European mole, *Talpa europaea* (Laenen et al., 2018). This system is unique in that each strain of virus has been recovered from a single individual host sample, potentially representing a prime situation for reassortment. Such scenarios, where reassortment and host-switching events can occur, highlight how crucial studies like these are to gaining a better understanding of the role of host ecology and evolutionary history on viral spread, emergence, and development of pathogenesis.

Incongruence in tree topologies between the L and S segments likely reflects host-switching and reassortment in the evolutionary history of JMSV. The large geographic distance between the NC and SC clades (Canada/Alaska and New Mexico/Colorado, respectively) containing the possible reassortant strain suggests that this event likely predates
the post-glacial expansion of *S. monticola*. However, the lack of sampling spanning the
distance between the NC and SC clades, minimal sampling for the M segment to date, and
only a single instance of reassortment, precludes full elucidation of its role in the
evolutionary history of JMSV. To attain a more complete history of JMSV, and shrew-borne
orthohantaviruses in general, a far more complete sequence dataset of all three genomic
segments is needed, spanning the breadth of host diversity. Klempa showed that reassortment
among orthohantaviruses is more prevalent than originally believed (Klempa, 2018),
however, the study of orthohantavirus reassortment is relatively young and the extent of
reassortment in orthohantavirus diversification remains largely unknown.

Shi and colleagues (2018) provided a better understanding of the deeper phylogenetic
and evolutionary history responsible for the current diversity in vertebrate viruses at the order
and family level, revealing that an overall trend of codivergence is coupled with a complex
history of host switching between distantly related taxa. Bennett and co-workers (Bennett et
al., 2014) showed that these trends of host switching and codivergence hold true within a
single virus genus, *Orthohantavirus*, in their exploration of the relationships and
phylogeographic history of several strains of orthohantavirus. In addition, Torres-Pérez and
colleagues (Torres-Pérez et al., 2011) examined the phylogeographic history and patterns of
divergence within a single strain and host, Andes virus hosted by the South American rodent,
*Oligoryzomys longicaudatus*. Their results revealed that while there was a general similarity
in spatial structure between virus and host phylogenies, the timing of diversification was
incongruent. That work highlights the difficulty of accurately dating the evolution of
hantaviruses, with estimates ranging broadly from several thousand to several million years
before present (Castel et al., 2017; Guo et al., 2013; Souza et al., 2014; Zhang & Holmes,
2014). Our study of the patterns and processes of one virus among multiple host species begins to explore the role of host history that is temporally intermediate between population-level diversification and much deeper evolutionary events.

JMSV does not show a history of strict codiversification, but rather multiple host-switching events at both broad and fine geographic scales, and at least one instance of reassortment of divergent strains. More sampling is necessary to elucidate the evolutionary history of JMSV within *S. palustris, S. bairdi*, and other close relatives of this shrew complex. Nonetheless, it is evident that JMSV has a complex and relatively deep evolutionary history in North America. That history remains complicated by uncertainty of host taxonomy such as polyphyletic assemblages in both *S. monticola* and *S. palustris* (Demboski & Cook, 2001; Hope et al., 2014). Such uncertainty illuminates the necessity for a solid understanding of host relationships and history when discerning the evolutionary history of obligate viruses and parasites in general.

ACKNOWLEDGMENTS

We thank Fernando Torres-Pérez for helpful critique and suggestions. For the collection of specimens, we acknowledge the logistical and permit support from the USDA Forest Service (Tongass National Forest and Pacific Northwest Laboratory), Alaska Department of Fish and Game, Oregon Department of Fisheries and Wildlife, US Fish and Wildlife Service, US Geological Survey, Bureau of Land Management, Yukon Department of Environment, and multiple state natural resource agencies. We also thank the University of New Mexico Center for Advanced Research Computing, supported in part by the National Science Foundation, for providing the high-performance computing resources used in this study.
REFERENCES


FIGURE LEGENDS

Figure 1. Maximum likelihood phylogeny of Jemez Springs virus L segment inferred using RAxML and rooted at the midpoint. Nodes with bootstrap support values greater than 70% are indicated with an asterisk. Color-coded branches correspond to geographic clades: Green = Southern Continental; blue = Northern Continental; red = Pacific Coastal. State abbreviations: AK, Alaska; BC, British Columbia; CA, California; MN, Minnesota; NM, New Mexico; OR, Oregon; WA, Washington; YT, Yukon Territory. GenBank accession numbers for the L-segment sequences are provided in Table S2.

Figure 2. Maximum likelihood phylogeny of the Sorex vagrans complex cytochrome b inferred using RAxML and rooted at the midpoint. Bootstrap support values greater than 70% are indicated with an asterisk. Color-coded branches correspond to geographic clades of the virus: Green = Southern Continental; blue = Northern Continental; red = Pacific Coastal. State abbreviations: AK, Alaska; AZ, Arizona; BC, British Columbia; CA, California; CO, Colorado; ID, Idaho; MN, Minnesota; MT, Montana; NM, New Mexico; NV, Nevada; OR, Oregon; PA, Pennsylvania; WA, Washington; YT, Yukon Territory. GenBank accession numbers for the shrew cytochrome b sequences are available in Table S2.

Figure 3. Tanglegram for Sorex vagrans complex on the left, and Jemez Springs virus L segment on the right. Host individuals and their associated virus strains are illustrated with a black line. Color-coded branches correspond to geographic clades of the virus: Green = Southern Continental; blue = Northern Continental; red = Pacific Coastal. State abbreviations are as shown in Figure 2. Bootstrap support values greater than 70% are indicated with an asterisk.
Figure 4. Phylogenetic reconciliation for the Jemez Springs virus L and S in blue and the *Sorex vagrans* complex phylogeny in black. Open circles represent codivergence events, solid circles followed by an arrow indicating host-switching events and direction, and dashed lines represent virus extinction events. State abbreviations are as shown in Figure 1.

Figure 5. (A) Extent of ice sheets across western North America during the Last Glacial Maximum and the modern distribution of *Sorex monticola*. (B) Current glacial extent of western North America according to the Global Land Ice Measurements from Space (GLIMS) [50], modern distribution of *Sorex monticola*, and the hypothesized direction of post-glacial expansion of *S. monticola* to the North and South.
### Table 1. RT-PCR detection of orthohantavirus RNA in tissues of soricine shrews.

<table>
<thead>
<tr>
<th>Species</th>
<th>Country</th>
<th>State</th>
<th>Year</th>
<th>RT-PCR Positive/Tested</th>
</tr>
</thead>
<tbody>
<tr>
<td><em>Sorex monticola</em></td>
<td>Canada</td>
<td>British Columbia</td>
<td>2006</td>
<td>3/10</td>
</tr>
<tr>
<td></td>
<td></td>
<td>Yukon Territory</td>
<td>2004</td>
<td>0/1</td>
</tr>
<tr>
<td></td>
<td></td>
<td>Yukon Territory</td>
<td>2005</td>
<td>3/17</td>
</tr>
<tr>
<td></td>
<td></td>
<td>Yukon Territory</td>
<td>2006</td>
<td>0/1</td>
</tr>
<tr>
<td></td>
<td>USA</td>
<td>Alaska</td>
<td>1998</td>
<td>0/1</td>
</tr>
<tr>
<td></td>
<td>USA</td>
<td>Alaska</td>
<td>2001</td>
<td>3/9</td>
</tr>
<tr>
<td></td>
<td>USA</td>
<td>Alaska</td>
<td>2002</td>
<td>0/7</td>
</tr>
<tr>
<td></td>
<td>USA</td>
<td>Alaska</td>
<td>2005</td>
<td>0/6</td>
</tr>
<tr>
<td></td>
<td>USA</td>
<td>Colorado</td>
<td>1994</td>
<td>1/1</td>
</tr>
<tr>
<td></td>
<td>USA</td>
<td>New Mexico</td>
<td>1995</td>
<td>0/1</td>
</tr>
<tr>
<td></td>
<td>USA</td>
<td>New Mexico</td>
<td>1996</td>
<td>2/5</td>
</tr>
<tr>
<td></td>
<td>USA</td>
<td>New Mexico</td>
<td>1998</td>
<td>1/4</td>
</tr>
<tr>
<td></td>
<td>USA</td>
<td>New Mexico</td>
<td>2000</td>
<td>1/4</td>
</tr>
<tr>
<td></td>
<td>USA</td>
<td>Utah</td>
<td>1994</td>
<td>0/1</td>
</tr>
<tr>
<td></td>
<td>USA</td>
<td>Utah</td>
<td>1997</td>
<td>0/1</td>
</tr>
<tr>
<td><em>Sorex palustris</em></td>
<td>Canada</td>
<td>Yukon Territory</td>
<td>2005</td>
<td>1/4</td>
</tr>
<tr>
<td></td>
<td>USA</td>
<td>New Mexico</td>
<td>1998</td>
<td>0/3</td>
</tr>
<tr>
<td></td>
<td>USA</td>
<td>New Mexico</td>
<td>2007</td>
<td>0/2</td>
</tr>
<tr>
<td></td>
<td>USA</td>
<td>Wyoming</td>
<td>2007</td>
<td>0/1</td>
</tr>
<tr>
<td><em>Sorex trowbridgii</em></td>
<td>USA</td>
<td>Washington</td>
<td>1996</td>
<td>2/3</td>
</tr>
<tr>
<td></td>
<td>USA</td>
<td>Oregon</td>
<td>2003</td>
<td>2/24</td>
</tr>
<tr>
<td></td>
<td>USA</td>
<td>Oregon</td>
<td>2004</td>
<td>1/8</td>
</tr>
<tr>
<td><em>Sorex vagrans</em></td>
<td>Canada</td>
<td>British Columbia</td>
<td>1996</td>
<td>1/9</td>
</tr>
<tr>
<td></td>
<td>USA</td>
<td>California</td>
<td>1983</td>
<td>1/3</td>
</tr>
<tr>
<td></td>
<td>USA</td>
<td>New Mexico</td>
<td>1994</td>
<td>0/1</td>
</tr>
<tr>
<td></td>
<td>USA</td>
<td>New Mexico</td>
<td>1996</td>
<td>0/1</td>
</tr>
<tr>
<td></td>
<td>USA</td>
<td>Oregon</td>
<td>1994</td>
<td>0/3</td>
</tr>
<tr>
<td></td>
<td>USA</td>
<td>Oregon</td>
<td>2003</td>
<td>3/17</td>
</tr>
<tr>
<td></td>
<td>USA</td>
<td>Oregon</td>
<td>2004</td>
<td>1/5</td>
</tr>
<tr>
<td></td>
<td>USA</td>
<td>Oregon</td>
<td>2005</td>
<td>2/6</td>
</tr>
<tr>
<td><em>Sorex bairdi</em></td>
<td>USA</td>
<td>Oregon</td>
<td>2005</td>
<td>1/1</td>
</tr>
<tr>
<td><em>Sorex bendirii</em></td>
<td>USA</td>
<td>Oregon</td>
<td>2003</td>
<td>0/1</td>
</tr>
<tr>
<td></td>
<td>USA</td>
<td>Oregon</td>
<td>2005</td>
<td>0/2</td>
</tr>
<tr>
<td><em>Sorex haydeni</em></td>
<td>USA</td>
<td>New Mexico</td>
<td>1996</td>
<td>1/1</td>
</tr>
<tr>
<td><strong>Total</strong></td>
<td></td>
<td></td>
<td></td>
<td>30/164</td>
</tr>
</tbody>
</table>
Figure 1.
Figure 2.
Figure 3.
Figure 4.
SUPPLEMENTAL

Figure S1. (A) Geographic distribution of Jemez Springs virus represented by sampling of the L segment. Shapes correspond to host taxa with colors representing the geographically defined clade recovered from the virus phylogeny. (B) Geographic sampling of the *Sorex vagrans* complex used for phylogenetic reconstruction. Host taxa is represented by different shapes with the color of the shape corresponding to geographically defined Jemez Springs virus clades.
Figure S2. Maximum likelihood phylogeny of the Jemez Springs virus S segment inferred using RAxML and rooted at the midpoint. Bootstrap support values for major nodes are labeled. Color-coded branches correspond to geographic clades as listed for the L segment. GenBank accession numbers for the S segment used in this study are: MSB87746 (MK992757); MSB900752 (FJ686859); MSB89332 (EF619962); MSB76658 (MK992756); MSB90111 (FJ686860); MSB144475 (FJ599499); MSB145726 (MK992759); MSB145713 (FJ686862); MSB147675 (FJ686863); MSB147745 (FJ686864); MSB145441 (MK992758); MSB144181 (KF96384); MSB83395 (MK992761); MSB281100 (KF963288); MSB281850 (MK992760); UAM58255 (FJ868865); MSB58308 (MK992769); UAM58449 (MK992770); PSU742 (KF963286); PSU800 (MK992762); PSU1622 (MK992765); PSU2150 (MK992766); PSU4694 (MK992767); PSU4638 (KF963282); PSU902 (MK992763); PSU1307 (MK992764); MSB53136 (K962807).
Figure S3. Tanglegram for Sorex vagrans complex on the left, and Jemez Springs M segment on the right. Host individuals and their associated virus isolate are illustrated with a black line. Branches are color-coded according geographically defined viral clades. Bootstrap support values for major clades are labeled.
Figure S4. Tanglegram for *Sorex vagrans* complex on the left, and Jemez Springs S segment on the right. Host individuals and their associated virus isolate are illustrated with a black line. Branches are color-coded according geographically defined viral clades. Bootstrap support values for major clades are labeled.
Figure S5. Tanglegram for Jemez Springs L segment on the left and the S segment on the right. Identical viruses from each segment are illustrated with a black line. Branches are color-coded according geographically defined viral clades. Bootstrap support values for major clades are labeled.
Figure S6. Maximum likelihood phylogeny of the Jemez Springs virus M segment inferred using RAxML and rooted at the midpoint. Bootstrap support values for major nodes are labeled. Color-coded branches correspond to geographic clades as listed for the L segment. GenBank accession numbers for the M segment used in this study are: MSB87746 (MK992747); MSB89332 (MK992748); MSB144475 (FJ593500); MSB147533 (MK992750); MSB147745 (MK992751); PSU742 (MK992752); PSU1622 (MK992755); PSU1307 (MK992754); MSB99213 (MK992749); PSU800 (MK992753).
CHAPTER 2

ANCIENT REASSORTMENT BETWEEN DIVERGENT STRAINS OF
CAMP RIPLEY VIRUS (HANTAVIRIDAE) IN THE NORTHERN SHORT-TAILED
SHREW (BLARINA BREVICAUDA)

Schuyler W. Liphardt¹, Hae Ji Kang², Satoru Arai³, Se Hun Gu², Joseph A. Cook¹,
Richard Yanagihara²

¹Museum of Southwestern Biology and Department of Biology, University of New Mexico,
Albuquerque, New Mexico, USA

²Department of Pediatrics, John A. Burns School of Medicine, University of Hawaii at
Manoa, Honolulu, HI, USA

³Infectious Disease Surveillance Center, National Institute of Infectious Diseases, Tokyo,
Japan
ABSTRACT

Genomic reassortment of segmented RNA virus strains is an important evolutionary mechanism that can generate novel viruses with profound effects on human and animal health, such as the H1N1 influenza pandemic in 2009 arising from reassortment of two swine influenza viruses. Reassortment is not restricted to influenza virus and has been shown to occur in members of the order Bunyavirales. The majority of reassortment events occurs between closely related lineages purportedly due to molecular constraints during viral packaging. In the original report of Camp Ripley virus (RPLV), a newfound hantavirus in the northern short-tailed shrew (Blarina brevicauda), phylogenetic incongruence between different genomic segments suggested reassortment. We have expanded sampling to include RPLV sequences amplified from archival tissues of 36 northern short-tailed shrews collected in 12 states (Arkansas, Iowa, Kansas, Maryland, Massachusetts, Michigan, Minnesota, New Hampshire, Ohio, Pennsylvania, Virginia, Wisconsin), and one southern short-tailed shrew (Blarina carolinensis) from Florida, within the United States. Using Bayesian phylogenetic analysis and Graph-incompatibility-based Reassortment Finder, we identified multiple instances of reassortment that spanned the Hantaviridae phylogenetic tree, including three highly divergent, co-circulating lineages of the M segment that have reassorted with a conserved L segment in multiple populations of B. brevicauda. In addition to identifying the first known mobatvirus-like M-segment sequences from a soricid host and only the second from a eulipotyphlan mammal, our results suggest that reassortment may be common between divergent virus strains and provide strong justification for expanded spatial, temporal, and taxonomic analyses of segmented viruses.
INTRODUCTION

Genomic reassortment is an important evolutionary mechanism by which phylogenetically distinct, segmented viral strains that co-infect a host cell may shuffle gene segments to generate new viral genotypes (Lowen, 2018; Vijaykrishna et al., 2015). Reassortment is possible for any segmented virus and occurs during the packaging stage of viral replication where individual segments from two or more strains are combined into a single virion. While a majority of reassortment events are thought to result in deleterious combinations and subsequent reduction in fitness, reassortment on occasion can lead to novel combinations that can result in increased fitness (McDonald et al., 2016). However, the extent of reassortment and associated clinical implications, such as increased pathogenicity, are not well defined. In influenza A virus, reassortant strains have been repeatedly documented to increase virulence compared to the respective parental strains (White and Lowen, 2018). For example, the H1N1 strain responsible for the 2009 influenza A epidemic was a reassortant composed of several divergent strains of influenza virus, including at least two from swine, and one each from birds and humans (Trifonov et al., 2009). Reassortment is not restricted to influenza virus. It has been suggested that a large number, possibly most, currently recognized members of the order Bunyavirales are reassortants (Briese et al., 2013). This includes evidence of reassortment in disease-causing viruses, such as Crimean-Congo hemorrhagic fever virus (Hewson et al., 2004), Rift Valley fever virus (Liu et al., 2017), Ngari virus (Lowen, 2018) and many hantaviruses (Klemra, 2018).

Possible implications to human health posed by viral reassortment may be profound, given that global movement of people and changing species distributions due to environmental disruption may increase novel interactions of divergent viruses (Carlson et al.,...
Understanding the extent of viral reassortment among wild reservoir host populations, as opposed to in vitro experiments, and the effectiveness of segment molecular incompatibilities in constraining successful reassortment, is dependent on spatial and temporal screening of diverse wild hosts at the community and population levels, as well as concerted efforts to produce genomic level sequences for comparative analyses.

Viruses of the family Hantaviridae contain three genomic segments named for their relative size, small (S), medium (M), and large (L), which encode the nucleocapsid protein, glycoprotein precursor, and RNA-dependent RNA polymerase, respectively (Abudurexiti et al., 2019; Plyusnin et al., 1996). Klempa (2018) provides a summary of the role reassortment has played in the evolution of hantaviruses across multiple temporal scales. For example, Bruges virus in the European mole (Talpa europaea) is likely the result of one or more ancient reassortment events (Laenen et al., 2018). In contrast, other studies suggest recent reassortment events in Sin Nombre virus harbored by the deer mouse (Peromyscus maniculatus) (Black et al., 2009; Henderson et al., 1995; D. Li et al., 1995) and Puumala virus in the bank vole (Myodes glareolus) (Razzauti et al., 2008, 2009). Briese and colleagues (2013) suggested that reassortment may be quite common and that all members of the order Bunyavirales, of which Hantaviridae is a member, could be reassortants or descendants of historical reassortment events.

Camp Ripley virus (RPLV) is a hantavirus originally discovered in the northern short-tailed shrew (Blarina brevicauda) (Order Eulipotyphla, Family Soricidae) (Arai et al., 2007). Bennett and colleagues (2014) suggested that the S segment of RPLV could be the result of recombination due to its ambiguous placement in the hantavirus phylogenetic tree; however, that assessment was based on sequence coverage from a single sample that limited insight
into the evolutionary history of RPLV. With expanded sampling of RPLV from multiple populations and new statistical approaches, we have been able to develop a clearer picture of host switching and reassortment within RPLV. However, much more extensive sampling across populations and geography of *B. brevicauda* is still necessary to achieve a truly complete history.

**MATERIALS AND METHODS**

*Ethics statement*

All field procedures for trapping of shrews and well-established protocols for processing and preserving their tissues were reviewed and approved by the Institutional Animal Care and Use Committee of the University of New Mexico, under protocol number 19-200908-MC.

*Sampling and sequencing*

Total RNA was extracted from liver and/or lung tissue, using the PureLink Micro-to-Midi total RNA purification kit, from 101 *B. brevicauda* and 10 southern short-tailed shrews (*B. carolinensis*) archived at the Museum of Southwestern Biology at the University of New Mexico (Table 1). Shrews were collected between 1980 and 2001 and represent sampling across a large portion of the distributions of both species in the United States (Figure 1). Complementary DNA (cDNA) was synthesized using the SuperScript III First-Strand Synthesis System (Invitrogen, San Diego, CA) with a universal oligonucleotide primer (5’–TAGTAGTAGACTCC–3’) designed from the conserved 3’–end of the S, M, and L segments of hantaviruses (Song et al., 2007).

Gene amplification was carried out in 20-µL reaction mixtures containing 250 µM dNTP, 2 mM MgCl₂, 1 U of AmpliTaq polymerase (Roche, Basel, Switzerland), and 0.25
µM of oligonucleotide primers, designed from highly conserved regions of previously identified soricid-borne hantaviruses. A listing of the oligonucleotide primers used to amplify the S, M, and L segments is provided in Supplementary Table 1. Initial denaturation was followed by touchdown PCR cycling (two-degree step-down annealing from 48°C to 38°C for 40 sec) and elongation at 72°C for 1 min, then 32 cycles of denaturation at 94°C for 40 sec, annealing at 42°C for 40 sec, and elongation at 72°C for 1 min, in a GeneAmp PCR 9700 thermal cycler (Perkin-Elmer, Waltham, MA, USA). Amplified products were separated by electrophoresis on 1.5% agarose gels and purified using the QIAQuick Gel Extraction Kit (Qiagen, Hilden, Germany). DNA was sequenced directly using an ABI Prism 377XL Genetic Analyzer (Applied Biosystems Inc., Foster City, CA, USA).

**Phylogenetic analysis**

Additional sequences representative of all currently recognized species of *Hantaviridae* were downloaded from GenBank (Supplementary Table 2). Nucleotide alignments were generated for each segment using MAFFT v7.402 (Multiple Alignment using Fast Fourier Transform) (Katoh and Standley, 2013) with the Smith-Waterman algorithm (--localpair and --maxiterate 1000). Alignments were visually inspected in Geneious v. 8 (https://www.geneious.com). The 3’– and 5’–ends that were poorly aligned and had limited coverage across samples were trimmed to ensure analysis of homologous regions. This process was repeated for the amino acid translation of each segment to address possible artifacts introduced by high levels of nucleotide variation and potential homoplasy between deeply divergent sequences. Bayesian phylogenetic inference was performed for nucleotide and amino acid alignments in MrBayes 3.2.6 (Ronquist et al., 2012) with a mixed model of evolution and a gamma distribution with the command 'iset nst=mixed
rates=gamma’ and run for 10,000,000 generations with trees sampled every 1,000
generations. Convergence was verified by inspecting that standard deviation of split
frequencies between runs was under 0.01, as recommended, with trees sampled every 1,000
generations. Convergence was tested using Tracer v.1.7.1 (Rambaut et al., 2018) to verify the
effective sample size for the posterior was above 200. A burn-in of 25%, as recommended
(Ronquist et al., 2012), was used prior to generating a 50% majority rule consensus tree.
Means of raw p-distances between groups and within groups were calculated on the amino
acid alignment in MEGA7 (Kumar et al., 2016).

Tests of reassortment

Reassortment was tested for the M and L segments using the Graph-incompatibility-
based Reassortment Finder program (GiRaF version 1.01) (Nagarajan and Kingsford, 2011)
with the tree files produced by MrBayes as input. The tree files for each segment were used
for graph-mining to determine gene segments with incompatible phylogenies. A reduced
dataset was used for analysis in GiRaF to accommodate the requirement of identical taxa
between trees. A 25% burn-in, 50% cull, all candidate set, non-star bicliques, and single
bicliques options were used for the GiRaF analysis. While GiRaF was originally written for
reassortment analysis of influenza viruses, it is broadly applicable to any segmented virus
capable of reassortment. A tanglegram of proposed reassorted sequences was visualized in R
with the package ‘phytools’ (Revell, 2012). As a secondary test of reassortment the program
RDP4 was used (Martin et al., 2015). Samples with sequences for both the L and M segment
were concatenated with the S segment, when available. Reassortment breakpoints were tested
by using the RDP, GENECONV, Bootscan, Maxchi, Chimaera, SiSscan, and 3Seq methods.
The highest possible $p$-value for accepting a reassortment event was set to 0.05 and all other parameters set to default.

**RESULTS**

*Phylogenetic analysis and tests of reassortment*

Virus sequences were recovered for 36 *B. brevicauda* and one *B. carolinensis* (Figure 1 and Supplementary Table 3). Phylogenies generated independently for all three segments showed differing levels of incongruence. Trees inferred from the M segment recovered three distinct, deeply paraphyletic clades across the *Hantaviridae* (Figures 1 and 2). The largest clade recovered was predictably part of the *Orthohantavirus* clade that includes other eulipotyphlan-borne hantaviruses (harbored by shrews and moles) from both the Old World and New World and represents the majority of our samples (labeled Ripley1). The Ripley1 clade was further subdivided into two well-supported subclades that roughly encompassed the geographic distribution of *B. brevicauda*, with an eastern subclade of shrews from Ohio, Massachusetts, and Pennsylvania, and a western subclade from Minnesota, Iowa, Wisconsin, and Kansas. This East-West division was also reflected in previously generated host phylogenies inferred from cytochrome $b$ (Brant and Orti, 2003). The second RPLV clade (Ripley2) was composed of M-segment sequences from *B. brevicauda* collected in Minnesota and Massachusetts, but this segment unexpectedly formed a monophyletic clade with Bruges virus from *Talpa europaea* from Germany. The third RPLV clade (Ripley3) was quite distinct from other RPLV sequences and even more unexpectedly was placed in the genus *Mobatvirus*, sister to Quezon virus from the Geoffroy’s rousette (*Rousettus amplexicaudatus*), a pteropodid bat from the Philippines. The Ripley3 clade consisted of M-segment sequences from *B. brevicauda* collected in Arkansas, Wisconsin, Minnesota,
Kansas, Michigan, and Virginia, and included the only sequences recovered from *B. carolinensis* collected in Florida. These sequences represented the first known mobatvirus-like M segments recovered from a North American soricid host and only the second from a eulipotyphlan, with the other being Nova virus that co-circulates with Bruges virus in the European mole (Kang et al., 2009; Laenen et al., 2018). Mean within-group amino acid p-distances for each clade in the M segment ranged from 2% to 4% divergence. Mean between-group distances were much larger, however, with 70% amino acid sequence identity between Ripley1 and Ripley2. Ripley3 shared 47% amino acid sequence similarity with Ripley1 and 49% with Ripley2.

The L-segment phylogenies did not match the pattern of diversification seen in the M segment, with all L-segment samples forming a single, well-supported monophyletic clade within *Orthohantavirus* (Figure 3). That clade was composed of virus segments representative of each clade seen in the M segment, however, the L-segment phylogeny placed MSB49712, the single M segment Ripley3 representative, as basal to the rest of RPLV. RPLV formed a sister relationship with Oxbow virus from the American shrew mole (*Neurotrichus gibbsii*) and Tigray virus from the Ethiopian white-footed mouse (*Stenocephalemys albipes*), and in turn were part of the well-supported clade of the genus *Orthohantavirus*. Within the larger clade of RPLV variation was rather low with a mean p-distance about 1.5% divergence. However, when compared to MSB49712, the mean distance was 11%.

Although our sampling for the S segment was less complete than for the M and L segments, we recovered two distinct RPLV clades (Figure 3), with the major S-segment clade containing representatives of both Ripley1 and Ripley2 from the M-segment
phylogeny. The one representative from Ripley3 in the S-segment phylogeny, MSB49712 from Arkansas, was related to the Hainan oriental leaf-toed gecko virus (HOLGV), a non-mammalian borne hantavirid-like virus from China, sharing only 43% amino acid sequence similarity with the rest of RPLV. Interestingly, the placement of HOLGV in both the M- and L-segment phylogenies disagreed with the findings in the original report of HOLGV (Shi et al., 2018).

GiRaF analyses identified 11 candidate sets of reassortment with high confidence (>0.95). Of these, nine contained only sequences from RPLV (Supplementary Table 4). Additionally, all RPLV included in the reduced dataset for GiRaF were consistently identified as reassortants. However, our GiRaF settings were set to the highest sensitivity to detect reassortment across the phylogeny, which also resulted in detection of reassortment among subclades within Ripley1 due to incongruence of internal branching structure towards the tips (Supplementary Figure 1). RDP4 identified six samples as likely reassortants with significant p-values depending on the method used (Supplementary Table 5). The six identified samples represented all three RPLV clades in the M-segment phylogeny. The breakpoints identified as the sites of reassortment corresponded to the concatenation points between the L, M, and S segments (Supplementary Table 6).

**DISCUSSION**

Reassortment among divergent viral strains can catalyze rapid evolution and result in serious consequences for human health. Benign parental strains, when reassorted, can introduce novel interactions with host systems and increase the ability to evade the host immune system or the likelihood of a virion to enter a host cell, leading to increased virulence (McDonald et al., 2016; Vijaykrishna et al., 2015). However, reassortment has been
generally thought to be restricted to closely related viral strains due to molecular incompatibilities that inhibit reassortment among highly divergent viruses in vivo (Klempa, 2018; McDonald et al., 2016; White and Lowen, 2018). In this study, we show a possible example of reassortment between deeply diverged viruses differing by 30–50% in the amino acid sequence of the M segment. While our understanding of how common reassortment between distantly related hantaviruses is limited, this study suggests that such events might not be so restrained. However, our ability to fully explore this for RPLV is limited by the lack of sequence representation for all segments, as well as by only partial coverage across the RPLV genome.

The hantavirus species demarcation criteria, published in the Ninth Report of the International Committee on Taxonomy of Viruses (King et al., 2012), which required at least a 7% amino acid sequence difference in the complete nucleocapsid and glycoprotein complex, have been abandoned. And DivErSity pArtitioning by hieRarchical Clustering (DEmARC) analysis of the coding regions of the complete S- and M-segment sequences has been used to objectively establish taxonomic classification of the family Hantaviridae since 2017 (Laenen et al., 2019). The limitations of our data set (that is, lacking the full-length S- and M-segment sequences of hantavirids from B. brevicauda and B. carolinensis) do not allow definitive conclusions about whether or not the three RPLV clades represent three different hantavirus species.

Co-circulation of distinct hantaviruses has been reported previously in other eulipotyphlan species, including Seewis and Altai-like viruses in the Eurasian common shrew (Sorex araneus) in Finland (Ling et al., 2014), Bruges and Nova viruses in the European mole in western Europe (Laenen et al., 2016), and most recently, Seewis, Altai, and Altai-
like viruses co-circulating in sympatric populations of Eurasian common shrew, Laxmann’s shrew (S. caecutiens), and flat-skulled shrew (S. roboratus) in Hungary and Russia (Kang et al., 2019). Here, we report three distinct M-segment lineages of RPLV co-circulating in *B. brevicauda* in the United States. However, co-circulation of distinct viruses within a single species does not necessarily indicate either current or historical reassortment. For example, despite Bruges and Nova virus co-infections of the same host individual on multiple occasions (Laenen et al., 2018), evidence of reassortment between the two viruses has yet to be documented. That is not the case for Seewis and Altai-like viruses co-circulating in shrews where phylogenetic disagreement between segments suggests an ancient history of host-switching and reassortment although this has yet to be tested statistically (Kang et al., 2019; Ling et al., 2014).

We recovered a similar pattern of disagreement between phylogenies for RPLV. Incongruencies in tree topology were found across all three segments of the RPLV genome. Further, the L-segment monophyly within RPLV was consistent with a model of classical co-diversification with its shrew host, whereas the M segment was composed of at least three unique histories suggesting multiple instances of successful host switches followed by subsequent reassortment of the L and S segments. Our sampling across genes limits our ability to more robustly test the history of reassortment in this system. Two samples that do have sequences for the L, M, and S segments, MSB89866 and MSB89863, were identified by RDP4 as containing the M segment from the Ripley2 clade and the L and S segments from representatives of the Ripley1 clade. This would follow the pattern of reassortment within *Hantaviridae* of the L and S segments packaged together with a divergent M segment, although this is not always the case (Klempea, 2018). Expanded sampling across the entire
RPLV genome and more individual hosts will be necessary to adequately address this question.

Our analysis of the L segment was limited to a relatively short section (347 bp), potentially affecting the reliability of phylogenetic inference. Still, while deeper nodes lack adequate posterior support, RPLV was well supported as distinct and monophyletic in the L-segment phylogeny (Figure 3). This pattern is difficult to explain biogeographically considering the depth of divergence and geographic distance between the mammalian hosts of the viruses that are sister to each RPLV clade; for example, Bruges virus and Ripley2 (Figure 3). It is possible that the Ripley2 and Ripley3 clades represent ancient lineages that predate the migration of the ancestor of *Blarina* to North America and have since been maintained. We find this scenario unlikely given the deep history of host colonization of North America (Brant and Orti, 2003), coupled with the high mutation rate of hantaviruses that should have a monophyletic lineage of hantaviruses in *Blarina*, which is not what we see. Another possibility is that the placement of the Ripley2 and Ripley3 clades in the M-segment phylogeny is strictly stochastic or an artifact of long-branch attraction among deeply divergent viruses. However, this scenario does not address how such deeply divergent strains are maintained within a single species. Lastly, the Ripley2 and Ripley3 clades could represent a relatively recent host switch to *B. brevicauda*. Implicit in this explanation is that because geography precludes direct contact and transmission between host species there is likely a significant amount of hantavirus diversity that has yet to be discovered, which would help us better interpret these biogeographical gaps. It is likely that *B. brevicauda* is the primary host of RPLV (Ripley1 clade) and that the Ripley2 and Ripley3 clades are the result of recent host-switching events and subsequent reassortment from unknown hantaviruses yet
to be discovered. It is important to note, however, that the topologies recovered here, specifically for the L and S segments, could be prone to bias and artifacts during phylogenetic inference due to the relatively short L-segment sequences and the limited S-segment representation.

Overall, our data suggest that reassortment among highly divergent viral strains may be rather common. Evidence of reassortment between deeply divergent viruses highlights the urgency of more extensive sampling of the global virome (French and Holmes, 2020) and hantaviruses, in particular, across diverse mammalian and non-mammalian taxa, including recently discovered hantavirid-like viruses recovered from reptiles, fish, and mosquitoes (Li et al., 2015; Shi et al., 2018). The diversity and evolutionary dynamics of hantaviruses remain limited, as does knowledge of the complete suite of organisms that host them. Continued research focused on increasing breadth of taxa examined, as well as population level studies within individual host species, is required if we are to develop a more complete history of the origin and diversification of the Hantaviridae family.

ACKNOWLEDGMENTS

We would like to thank Jocelyn Colella for helpful critiques and comments during the development of this manuscript. We thank the mammalogists instrumental in collecting specimens in the field and curators and collection managers at the Museum of Southwestern Biology who have managed these specimens and associated data. We also thank the University of New Mexico Center for Advanced Research Computing, supported in part by the National Science Foundation, for providing the high-performance computing resources used in this study.
REFERENCES


FIGURE LEGENDS

Figure 1. Geographic distribution of RPLV M-segment clades according to collection sites of *Blarina brevicauda* and *Blarina carolinensis* in the United States. RPLV M-segment clades are represented by colored circles (green for Ripley1, red for Ripley2 and purple for Ripley3). White circles represent RPLV samples for which L- and/or S-segment sequences are available, but M segment sequences could not be amplified.

Figure 2. Bayesian phylogeny inferred for the M segment. Branch colors represent geographic origin of the reservoir host and associated virus. Host taxa are indicated (right), with proposed hantavirus genera labeled on respective branches. RPLV clades are boxed, labeled, and indicated with an asterisk. Posterior support values over 0.95 are labeled.

Figure 3. Discordance of Camp Ripley virus across genomic segments. Phylogenies were inferred using Bayesian probability in MrBayes. Nodes with posterior probabilities less than 0.8 are marked with an asterisk. Camp Ripley representative samples are boxed in red.
### Table 1.

<table>
<thead>
<tr>
<th>State</th>
<th>County</th>
<th>Year</th>
<th>MSB Number</th>
<th>S</th>
<th>M</th>
<th>L</th>
<th>Tissue</th>
</tr>
</thead>
<tbody>
<tr>
<td>Arkansas</td>
<td>Washington</td>
<td>14-Mar-1982</td>
<td>MSB49712</td>
<td>MT377511 (885)</td>
<td>MT377558 (1480)</td>
<td>MT377528 (347)</td>
<td>liver</td>
</tr>
<tr>
<td>Florida</td>
<td>Leon</td>
<td>18-Jun-1983</td>
<td>MSB53277</td>
<td></td>
<td>MT377562 (205)</td>
<td></td>
<td>liver</td>
</tr>
<tr>
<td>Iowa</td>
<td>Allamakee</td>
<td>24-Aug-1994</td>
<td>MSB73580</td>
<td></td>
<td>MT377535 (1154)</td>
<td></td>
<td>lung</td>
</tr>
<tr>
<td>Kansas</td>
<td>Douglas</td>
<td>11-Oct-2001</td>
<td>MSB151832</td>
<td>MT377555 (1187)</td>
<td></td>
<td></td>
<td>lung</td>
</tr>
<tr>
<td></td>
<td></td>
<td></td>
<td>11-Oct-2001</td>
<td>MSB151836</td>
<td>MT377542 (875)</td>
<td>MT377523 (347)</td>
<td>lung</td>
</tr>
<tr>
<td></td>
<td></td>
<td></td>
<td>11-Oct-2001</td>
<td>MSB151838</td>
<td>MT377556 (1187)</td>
<td>MT377522 (347)</td>
<td>lung</td>
</tr>
<tr>
<td>Maryland</td>
<td>Charles</td>
<td>22-Sep-1997</td>
<td>MSB92373</td>
<td></td>
<td></td>
<td></td>
<td>liver</td>
</tr>
<tr>
<td>Massachusetts</td>
<td>Franklin</td>
<td>10-Jul-1982</td>
<td>MSB47876</td>
<td>MT377549 (1104)</td>
<td></td>
<td></td>
<td>liver</td>
</tr>
<tr>
<td></td>
<td></td>
<td></td>
<td>12-Jul-1982</td>
<td>MSB47879</td>
<td>MT377551 (1082)</td>
<td></td>
<td>MT377533 (347)</td>
</tr>
<tr>
<td></td>
<td></td>
<td></td>
<td>13-Jul-1982</td>
<td>MSB47882</td>
<td>MT307509 (394)</td>
<td></td>
<td></td>
</tr>
<tr>
<td></td>
<td></td>
<td></td>
<td>11-Jul-1982</td>
<td>MSB47883</td>
<td>MT377550 (1099)</td>
<td></td>
<td>MT377532 (347)</td>
</tr>
<tr>
<td></td>
<td></td>
<td></td>
<td>12-Jul-1982</td>
<td>MSB47887</td>
<td>MT307508 (400)</td>
<td>MT377552 (1101)</td>
<td></td>
</tr>
<tr>
<td></td>
<td></td>
<td></td>
<td>12-Jul-1982</td>
<td>MSB47890</td>
<td></td>
<td>MT377561 (1105)</td>
<td></td>
</tr>
<tr>
<td></td>
<td></td>
<td></td>
<td>13-Jul-1982</td>
<td>MSB47891</td>
<td>KF958463 (1028)</td>
<td>MT377553 (1103)</td>
<td></td>
</tr>
<tr>
<td>Michigan</td>
<td>Crawford</td>
<td>9-Jul-1999</td>
<td>MSB2437</td>
<td></td>
<td></td>
<td></td>
<td>MT958467 (2345)</td>
</tr>
<tr>
<td>Minnesota</td>
<td>Clay</td>
<td>31-Jul-1983</td>
<td>MSB53249</td>
<td>MT377554 (886)</td>
<td></td>
<td></td>
<td>MT376519 (347)</td>
</tr>
<tr>
<td></td>
<td></td>
<td></td>
<td>31-Jul-1983</td>
<td>MSB53251</td>
<td>MT377560 (1130)</td>
<td>MT377530 (347)</td>
<td></td>
</tr>
<tr>
<td></td>
<td></td>
<td></td>
<td>28-Jul-1983</td>
<td>MSB53254</td>
<td>MT377507 (783)</td>
<td>MT377548 (1370)</td>
<td>MT37753 (347)</td>
</tr>
<tr>
<td></td>
<td></td>
<td></td>
<td>20-Aug-1998</td>
<td>MSB89858</td>
<td>MT377536 (1155)</td>
<td>MT377513 (347)</td>
<td></td>
</tr>
<tr>
<td></td>
<td></td>
<td></td>
<td>21-Aug-1998</td>
<td>MSB89859</td>
<td>MT377537 (1390)</td>
<td>MT377514 (347)</td>
<td></td>
</tr>
<tr>
<td></td>
<td></td>
<td></td>
<td>20-Aug-1998</td>
<td>MSB89861</td>
<td>MT377538 (1390)</td>
<td>MT377515 (347)</td>
<td></td>
</tr>
<tr>
<td></td>
<td></td>
<td></td>
<td>20-Aug-1998</td>
<td>MSB89862</td>
<td>MT377539 (1153)</td>
<td>MT377534 (347)</td>
<td></td>
</tr>
<tr>
<td></td>
<td></td>
<td></td>
<td>20-Aug-1998</td>
<td>MSB89863</td>
<td>FJ790772 (330)</td>
<td>EF540774 (2755)</td>
<td>EF540771 (6551)</td>
</tr>
<tr>
<td></td>
<td></td>
<td></td>
<td>20-Aug-1998</td>
<td>MSB89864</td>
<td>MT377540 (1155)</td>
<td>MT377516 (347)</td>
<td></td>
</tr>
<tr>
<td></td>
<td></td>
<td></td>
<td>21-Aug-1998</td>
<td>MSB89866</td>
<td>MT377505 (276)</td>
<td>EF540775 (1717)</td>
<td>EF540772 (300)</td>
</tr>
<tr>
<td></td>
<td></td>
<td></td>
<td>19-Aug-1998</td>
<td>MSB89867</td>
<td></td>
<td>MT377541 (1041)</td>
<td>MT377517 (347)</td>
</tr>
<tr>
<td></td>
<td></td>
<td></td>
<td>21-Aug-1998</td>
<td>MSB90845</td>
<td>KF958464 (900)</td>
<td>EF540773 (2077)</td>
<td>KF958465 (6551)</td>
</tr>
<tr>
<td>New Hampshire</td>
<td>Grafton</td>
<td>1-Sep-1998</td>
<td>MSB151869</td>
<td></td>
<td></td>
<td></td>
<td>MT377525 (347)</td>
</tr>
<tr>
<td>Ohio</td>
<td>Summit</td>
<td>21-Aug-2001</td>
<td>MSB151834</td>
<td>MT377543 (1390)</td>
<td></td>
<td>MT377524 (347)</td>
<td></td>
</tr>
<tr>
<td></td>
<td>Ashtabula</td>
<td>30-Jun-1980</td>
<td>MSB43407</td>
<td></td>
<td></td>
<td>MT377546 (1102)</td>
<td></td>
</tr>
<tr>
<td>Pennsylvania</td>
<td>Indiana</td>
<td>2-Oct-2001</td>
<td>MSB151849</td>
<td>MT377506 (425)</td>
<td>MT377527 (347)</td>
<td></td>
<td></td>
</tr>
<tr>
<td></td>
<td>Westmoreland</td>
<td>5-Jul-1983</td>
<td>MSB53264</td>
<td>MT307510 (1100)</td>
<td>MT377547 (884)</td>
<td>MT377529 (347)</td>
<td></td>
</tr>
<tr>
<td>Virginia</td>
<td>Fauquier</td>
<td>2-Apr-1994</td>
<td>MSB76255</td>
<td>MT377559 (1479)</td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>Wisconsin</td>
<td>Dane</td>
<td>21-Sep-2000</td>
<td>MSB151873</td>
<td>MT377557 (1187)</td>
<td>lung</td>
<td></td>
<td></td>
</tr>
<tr>
<td>-----------</td>
<td>------</td>
<td>-------------</td>
<td>-----------</td>
<td>-----------------</td>
<td>------</td>
<td></td>
<td></td>
</tr>
<tr>
<td></td>
<td></td>
<td></td>
<td></td>
<td>MT377545 (1390)</td>
<td>MT377520 (347)</td>
<td>lung</td>
<td></td>
</tr>
<tr>
<td></td>
<td></td>
<td></td>
<td></td>
<td>MT377544 (1390)</td>
<td>MT377526 (347)</td>
<td>lung</td>
<td></td>
</tr>
</tbody>
</table>

Table 1 (cont.)
FIGURES

Figure 1.
Figure 2.
Figure 3.
**Supplementary Table 1. Oligonucleotide primers used for amplification and sequencing of RPLV.**

<table>
<thead>
<tr>
<th>Segment</th>
<th>Primer Name: Oligonucleotide Sequence</th>
</tr>
</thead>
<tbody>
<tr>
<td><strong>S</strong></td>
<td>OSM55F: 5'-TAGTAGTAGACTCC-3'</td>
</tr>
<tr>
<td></td>
<td>HTN-S1F: 5'-TAGTAGTAGACTYCTNMDDA-3'</td>
</tr>
<tr>
<td></td>
<td>RPL-S504F: 5'-GNATTMGAYTNAARGATGA-3'</td>
</tr>
<tr>
<td></td>
<td>Han-S604F: 5'-GCHGADGARHTNACCCNGG-3'</td>
</tr>
<tr>
<td></td>
<td>Han-S694F: 5'-CCNGTNATGRGNGTNATHGGNTT-3'</td>
</tr>
<tr>
<td></td>
<td>Cro2F: 5'-AGYCCNGTNATGRGWGTNYTYGG-3'</td>
</tr>
<tr>
<td></td>
<td>RPL-S700F: 5'-ATGGCWKCAAARTGGGARGA-3'</td>
</tr>
<tr>
<td></td>
<td>Han-S952F: 5'-TGGDTNTTTGCNDRDGSNCC-3'</td>
</tr>
<tr>
<td></td>
<td>RPL-S987F: 5'-AAGTNGCWGAAGATGCWGG-3'</td>
</tr>
<tr>
<td></td>
<td>Han-S1042F: 5'-CAGGAYATGAGNAAYAANAT-3'</td>
</tr>
<tr>
<td></td>
<td>RPL-S1199F: 5'-TGGGGNAAAGAGGCAGTGGA-3'</td>
</tr>
<tr>
<td></td>
<td>RPL-S473R: 5'-ACCCCTAAGGTATACTCATGTA-3'</td>
</tr>
<tr>
<td></td>
<td>Han-S987F: 5'-ACCCCTAAGGTATACTCATGTA-3'</td>
</tr>
<tr>
<td></td>
<td>HTN-S6R: 5'-AGCTCNGGATCCATNTCATC-3'</td>
</tr>
<tr>
<td></td>
<td>Cro2R: 5'-ANGAYTGRTARAANGANGAYTTYTT-3'</td>
</tr>
<tr>
<td></td>
<td>RPL-S522R: 5'-TCATCYTTYNARTCKAATNC-3'</td>
</tr>
<tr>
<td></td>
<td>Hanta-S587R: 5'-CCWGGTGTNANYTCDTC-3'</td>
</tr>
<tr>
<td></td>
<td>SO-S620R: 5'-CGGATCTCAATCCTCAATNCC-3'</td>
</tr>
<tr>
<td></td>
<td>RPL-S754R: 5'-CGGATCTCAATCCTCAATNCC-3'</td>
</tr>
<tr>
<td></td>
<td>SO-S970R: 5'-GGYSACHNGCAAANACCCA-3'</td>
</tr>
<tr>
<td></td>
<td>Han-S974R: 5'-GGYSACHNGCAAANACCCA-3'</td>
</tr>
<tr>
<td></td>
<td>Han-S1260R: 5'-CDGGRTCCATRTCATCHCCNA-3'</td>
</tr>
<tr>
<td></td>
<td>RPL-S1062R: 5'-ATNTTRTNTCARTCTCCTG-3'</td>
</tr>
<tr>
<td></td>
<td>Han-S3R: 5'-TAGTAGTANNCTCCYTRWACA-3'</td>
</tr>
<tr>
<td><strong>M</strong></td>
<td>OSM55F: 5'-TAGTAGTAGACTCC-3'</td>
</tr>
<tr>
<td></td>
<td>M-900F: 5'-CCMRGGGAGAAGAYCAYGA-3'</td>
</tr>
<tr>
<td></td>
<td>HTN-M1190F: 5'-CCMNCCNGGDCWNVHTGTGA-3'</td>
</tr>
<tr>
<td></td>
<td>RPLV-M1434F: 5'-GTTTTYCATGGKTGGGACAA-3'</td>
</tr>
<tr>
<td></td>
<td>HTN-M2570F: 5'-GGYSACHNGCAAANACCCA-3'</td>
</tr>
<tr>
<td></td>
<td>HTN-M2620F: 5'-TCNNGNGCHNGCAAANAHCCA-3'</td>
</tr>
<tr>
<td></td>
<td>Han-S1260R: 5'-CDGGRTCCATRTCATCHCCNA-3'</td>
</tr>
<tr>
<td></td>
<td>RPL-S1062R: 5'-ATNTTRTNTCARTCTCCTG-3'</td>
</tr>
<tr>
<td></td>
<td>Han-S3R: 5'-TAGTAGTANNCTCCYTRWACA-3'</td>
</tr>
<tr>
<td><strong>L</strong></td>
<td>OSM55F: 5'-TAGTAGTAGACTCC-3'</td>
</tr>
<tr>
<td></td>
<td>PHL-173F: 5'-GATWAAGCATGAYTGGTCTGA-3'</td>
</tr>
<tr>
<td></td>
<td>RPLV-L-366F: 5'-RTGCACAGCTGACGAGYWGATGT-3'</td>
</tr>
<tr>
<td></td>
<td>SO-L496F: 5'-TATNGAGTTTGATGTYATGCTGT-3'</td>
</tr>
</tbody>
</table>
Abbreviations: A, Adenine; B, C or G or T; C, Cytosine; D, A or G or T; G, Guanine; H, A or C or T; I, Inosine; K, G or T; M, A or C; N, any nucleotide; R, A or G; S, G or C; T, Thymine; V, A or C or G; W, A or T; Y, C or T.
**Supplementary Table 2.** GenBank accession numbers for *Hantaviridae* sequences used in this study.

<table>
<thead>
<tr>
<th>Virus</th>
<th>S</th>
<th>M</th>
<th>L</th>
</tr>
</thead>
<tbody>
<tr>
<td>AAIV</td>
<td>DQ345764</td>
<td>AY515599</td>
<td>EU424341</td>
</tr>
<tr>
<td>ALTV</td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>AMRV</td>
<td>AB620028</td>
<td>AB620029</td>
<td>AB620030</td>
</tr>
<tr>
<td>ANAV</td>
<td>DQ451829</td>
<td></td>
<td></td>
</tr>
<tr>
<td>ANDV</td>
<td>NC003466</td>
<td>NC003467</td>
<td>NC003468</td>
</tr>
<tr>
<td>ANJV</td>
<td></td>
<td>KC490920</td>
<td></td>
</tr>
<tr>
<td>ARAUV</td>
<td>HQ337907</td>
<td>FJ409557</td>
<td></td>
</tr>
<tr>
<td>ARTV</td>
<td>KF974360</td>
<td>KF974359, MG913807</td>
<td>EU424339</td>
</tr>
<tr>
<td>ARRV</td>
<td>EF650086</td>
<td></td>
<td>EF619961</td>
</tr>
<tr>
<td>ASAIV</td>
<td>EU929070</td>
<td>NC038274</td>
<td>EU929076</td>
</tr>
<tr>
<td>ASIV</td>
<td>KC880343</td>
<td>KC880346</td>
<td>KC880349</td>
</tr>
<tr>
<td>AZGV</td>
<td>JF276226</td>
<td>JF276227</td>
<td>JF276228</td>
</tr>
<tr>
<td>BCCV</td>
<td>L39949</td>
<td>L39950</td>
<td>L39951</td>
</tr>
<tr>
<td>BLLV</td>
<td>U19303, U19301</td>
<td>U19305</td>
<td></td>
</tr>
<tr>
<td>BMJV</td>
<td>AF482713</td>
<td>AF028025</td>
<td></td>
</tr>
<tr>
<td>BOGV</td>
<td>JX990966</td>
<td></td>
<td></td>
</tr>
<tr>
<td>BOWV</td>
<td>KC631782</td>
<td>KC631783</td>
<td>KC631784</td>
</tr>
<tr>
<td>BROV</td>
<td>MF683845</td>
<td>MF683846</td>
<td></td>
</tr>
<tr>
<td>CADV</td>
<td>DQ285566</td>
<td>DQ284451</td>
<td></td>
</tr>
<tr>
<td>CATV</td>
<td>DQ256126</td>
<td>DQ177347</td>
<td></td>
</tr>
<tr>
<td>CBNV</td>
<td>EF543524</td>
<td>NC034474</td>
<td>EF543525</td>
</tr>
<tr>
<td>CHOV</td>
<td>DQ285046</td>
<td>DQ285047</td>
<td>EF397003</td>
</tr>
<tr>
<td>DBSV</td>
<td>AB027523</td>
<td>AB027115</td>
<td>DQ989237</td>
</tr>
<tr>
<td>DOBV</td>
<td>GQ205407</td>
<td>GQ205409</td>
<td>GU904039</td>
</tr>
<tr>
<td>ELMCV</td>
<td>AB620106</td>
<td>AB620104</td>
<td>AB620108</td>
</tr>
<tr>
<td>HOKV</td>
<td>AB010730</td>
<td></td>
<td></td>
</tr>
<tr>
<td>HOLGV</td>
<td>MG599942</td>
<td>MG599941</td>
<td>MG599940</td>
</tr>
<tr>
<td>HTNV</td>
<td>NC005218</td>
<td>NC005219</td>
<td>NC005222</td>
</tr>
<tr>
<td>HUPV</td>
<td>JX473273</td>
<td></td>
<td>JX465369</td>
</tr>
<tr>
<td>ITPV</td>
<td>DQ345765</td>
<td>AY515600</td>
<td></td>
</tr>
<tr>
<td>ISLAV</td>
<td>U19302</td>
<td>U19304</td>
<td></td>
</tr>
<tr>
<td>JABV</td>
<td>JN232078</td>
<td></td>
<td></td>
</tr>
<tr>
<td>JJUV</td>
<td>HQ663933</td>
<td>HQ663934</td>
<td>HQ663935</td>
</tr>
<tr>
<td>JMSV</td>
<td>FJ686859</td>
<td>FJ593500</td>
<td>FJ593501</td>
</tr>
<tr>
<td>KHAV</td>
<td>U35255</td>
<td>AJ011648</td>
<td>AJ011650</td>
</tr>
<tr>
<td>KKMV</td>
<td>GQ306148</td>
<td>KJ857337</td>
<td>GQ306150</td>
</tr>
<tr>
<td>KMJV</td>
<td>JX193698</td>
<td>JX193699</td>
<td></td>
</tr>
<tr>
<td>Virus</td>
<td>Accession 1</td>
<td>Accession 2</td>
<td>Accession 3</td>
</tr>
<tr>
<td>---------</td>
<td>-------------</td>
<td>-------------</td>
<td>-------------</td>
</tr>
<tr>
<td>LAIV</td>
<td>KM102248</td>
<td></td>
<td></td>
</tr>
<tr>
<td>LANV</td>
<td>AF005727</td>
<td>AF005728</td>
<td>AF005729</td>
</tr>
<tr>
<td>LECV</td>
<td>AF482714</td>
<td>AF028022</td>
<td></td>
</tr>
<tr>
<td>LHEV</td>
<td>JX465404</td>
<td>JX465390</td>
<td>JX465370</td>
</tr>
<tr>
<td>LQUV</td>
<td>JX465413</td>
<td>JX465396</td>
<td>JX465379</td>
</tr>
<tr>
<td>LSCV</td>
<td>AF307322</td>
<td>AF307323</td>
<td></td>
</tr>
<tr>
<td>LXV</td>
<td>HM756286</td>
<td>HM756287</td>
<td>HQ404253</td>
</tr>
<tr>
<td>MJNV</td>
<td>EF641805</td>
<td>EF641799</td>
<td>EF641807</td>
</tr>
<tr>
<td>MOYV</td>
<td></td>
<td></td>
<td>JQ287716</td>
</tr>
<tr>
<td>MPRLV</td>
<td>AY267347</td>
<td>AY363179</td>
<td>EU788002</td>
</tr>
<tr>
<td>MTNV</td>
<td>AB620100</td>
<td>AB620101</td>
<td>AB620102</td>
</tr>
<tr>
<td>MUJV</td>
<td>DQ138142</td>
<td>DQ138141</td>
<td></td>
</tr>
<tr>
<td>MULV</td>
<td>U54575</td>
<td></td>
<td></td>
</tr>
<tr>
<td>NVAV</td>
<td>FJ539168</td>
<td>HQ840957</td>
<td>FJ593498</td>
</tr>
<tr>
<td>NYV</td>
<td>U32591</td>
<td>U36802</td>
<td></td>
</tr>
<tr>
<td>ORNV</td>
<td>AF482715</td>
<td>AF028204</td>
<td></td>
</tr>
<tr>
<td>OXBV</td>
<td>FJ539166</td>
<td>FJ539167</td>
<td>FJ593497</td>
</tr>
<tr>
<td>PHV</td>
<td>M34011</td>
<td>X55129</td>
<td>EF646763</td>
</tr>
<tr>
<td>PRGV</td>
<td>AF482717, EF534079</td>
<td>AF028028</td>
<td></td>
</tr>
<tr>
<td>PUUV</td>
<td>EF442087</td>
<td>NC005223</td>
<td>NC005225</td>
</tr>
<tr>
<td>QDLV</td>
<td>GU566023</td>
<td>GU566022</td>
<td>GU566021</td>
</tr>
<tr>
<td>QZNV</td>
<td></td>
<td></td>
<td>KU950714</td>
</tr>
<tr>
<td>RIOSV</td>
<td>U18100</td>
<td></td>
<td></td>
</tr>
<tr>
<td>RIOMV</td>
<td>FJ532244</td>
<td>FJ608550</td>
<td>FJ809772</td>
</tr>
<tr>
<td>RKPV</td>
<td>HM015218</td>
<td>NC038694</td>
<td>HM015221</td>
</tr>
<tr>
<td>SAAV</td>
<td>AJ616854</td>
<td>AJ616855</td>
<td>AJ410618</td>
</tr>
<tr>
<td>SANGV</td>
<td>JQ082300</td>
<td>JQ082301</td>
<td>JQ082302</td>
</tr>
<tr>
<td>SEOV</td>
<td>AF288298</td>
<td>AB618130</td>
<td>AF288297</td>
</tr>
<tr>
<td>SNV</td>
<td>L33683</td>
<td>L37903</td>
<td>L37902</td>
</tr>
<tr>
<td>SOOV</td>
<td>AY675349</td>
<td>AY675353</td>
<td>DQ056292</td>
</tr>
<tr>
<td>SWSV</td>
<td>EF636024</td>
<td>KY651069</td>
<td>EF636026</td>
</tr>
<tr>
<td>TATV</td>
<td>JX316009</td>
<td></td>
<td></td>
</tr>
<tr>
<td>TGNV</td>
<td></td>
<td></td>
<td>EF050454</td>
</tr>
<tr>
<td>THAIIV</td>
<td>AM397666</td>
<td>AM397671</td>
<td>JN116258</td>
</tr>
<tr>
<td>TIGV</td>
<td></td>
<td></td>
<td>JQ956487</td>
</tr>
<tr>
<td>TOPV</td>
<td>AJ011646</td>
<td>AJ011647</td>
<td>AJ011649</td>
</tr>
<tr>
<td>TPMV</td>
<td>NC010704</td>
<td>NC010708</td>
<td>NC010707</td>
</tr>
<tr>
<td>TULV</td>
<td>DQ662090</td>
<td>DQ665814</td>
<td>NC005226</td>
</tr>
<tr>
<td>UULV</td>
<td>JX193695</td>
<td>JX193696</td>
<td></td>
</tr>
<tr>
<td>VLA</td>
<td>EU072480</td>
<td>EU072488</td>
<td>FJ170807</td>
</tr>
<tr>
<td>WEHV</td>
<td>MG599954</td>
<td>MG599953</td>
<td>MG599952</td>
</tr>
<tr>
<td></td>
<td>WEMBV</td>
<td>MG599945</td>
<td>MG599944</td>
</tr>
<tr>
<td>---</td>
<td>-----------</td>
<td>----------</td>
<td>----------</td>
</tr>
<tr>
<td>XSV</td>
<td>KF704709</td>
<td>KJ000538</td>
<td></td>
</tr>
<tr>
<td>XYIV</td>
<td>KF705677</td>
<td>KF705678</td>
<td></td>
</tr>
<tr>
<td>YKSV</td>
<td>JX465423</td>
<td>NC038705</td>
<td></td>
</tr>
<tr>
<td>YUJV</td>
<td>FJ170792</td>
<td>FJ170800</td>
<td></td>
</tr>
</tbody>
</table>
**Supplementary Table 3.** Sampling localities (state, county) and collection year for *Blarina brevicauda* vouchers housed at the Museum of Southwestern Biology. Tested and positive columns represent the number of specimens analyzed by RT-PCR and showing hantavirus RNA, respectively.

<table>
<thead>
<tr>
<th>State</th>
<th>County</th>
<th>Year</th>
<th>Tested</th>
<th>Positive</th>
<th>RPLV Strain*</th>
</tr>
</thead>
<tbody>
<tr>
<td>Arkansas</td>
<td>Washington</td>
<td>1982</td>
<td>1</td>
<td>1</td>
<td>MSB49712</td>
</tr>
<tr>
<td>Indiana</td>
<td>Porter</td>
<td>1994</td>
<td>2</td>
<td>0</td>
<td></td>
</tr>
<tr>
<td></td>
<td>Westchester</td>
<td>1994</td>
<td>1</td>
<td>0</td>
<td></td>
</tr>
<tr>
<td>Iowa</td>
<td>Allamakee</td>
<td>1994</td>
<td>5</td>
<td>1</td>
<td>MSB73580</td>
</tr>
<tr>
<td>Kansas</td>
<td>Douglas</td>
<td>2001</td>
<td>9</td>
<td>4</td>
<td>MSB151832, MSB151836, MSB151838, MSB151840</td>
</tr>
<tr>
<td></td>
<td>Riley</td>
<td>1998</td>
<td>1</td>
<td>0</td>
<td></td>
</tr>
<tr>
<td>Maryland</td>
<td>Charles</td>
<td>1997</td>
<td>3</td>
<td>1</td>
<td>MSB92373</td>
</tr>
<tr>
<td>Massachusetts</td>
<td>Franklin</td>
<td>1982</td>
<td>12</td>
<td>7</td>
<td>MSB47876, MSB47879, MSB47882, MSB47883, MSB47887, MSB47890, MSB47891</td>
</tr>
<tr>
<td>Michigan</td>
<td>Leelanau</td>
<td>1991</td>
<td>1</td>
<td>0</td>
<td></td>
</tr>
<tr>
<td></td>
<td>Benzie</td>
<td>1994</td>
<td>1</td>
<td>0</td>
<td></td>
</tr>
<tr>
<td></td>
<td>Crawford</td>
<td>1999</td>
<td>1</td>
<td>1</td>
<td>MSB92437</td>
</tr>
<tr>
<td>Minnesota</td>
<td>Cass</td>
<td>1983</td>
<td>3</td>
<td>1</td>
<td>MSB53254</td>
</tr>
<tr>
<td></td>
<td>Clay</td>
<td>1983</td>
<td>2</td>
<td>2</td>
<td>MSB53249, MSB53251</td>
</tr>
<tr>
<td></td>
<td>Morrison</td>
<td>1998</td>
<td>12</td>
<td>9</td>
<td>MSB89858, MSB89859, MSB89861, MSB89862, MSB89863, MSB89864, MSB89866, MSB898867, MSB90845</td>
</tr>
<tr>
<td>New Hampshire</td>
<td>Grafton</td>
<td>1998</td>
<td>4</td>
<td>1</td>
<td>MSB151869</td>
</tr>
<tr>
<td>New York</td>
<td>Essex</td>
<td>2001</td>
<td>5</td>
<td>0</td>
<td></td>
</tr>
<tr>
<td>Ohio</td>
<td>Ashtabula</td>
<td>1980</td>
<td>1</td>
<td>1</td>
<td>MSB43407</td>
</tr>
<tr>
<td></td>
<td>Summit</td>
<td>2001</td>
<td>6</td>
<td>1</td>
<td>MSB151834</td>
</tr>
<tr>
<td>Pennsylvania</td>
<td>Westmoreland</td>
<td>1983</td>
<td>5</td>
<td>1</td>
<td>MSB53264</td>
</tr>
<tr>
<td></td>
<td>Indiana</td>
<td>2001</td>
<td>9</td>
<td>1</td>
<td>MSB151849</td>
</tr>
<tr>
<td>Tennessee</td>
<td>Monroe</td>
<td>2001</td>
<td>2</td>
<td>0</td>
<td></td>
</tr>
<tr>
<td>Virginia</td>
<td>Fauquier</td>
<td>1994</td>
<td>1</td>
<td>1</td>
<td>MSB76255</td>
</tr>
<tr>
<td></td>
<td>Page</td>
<td>1994</td>
<td>1</td>
<td>0</td>
<td></td>
</tr>
<tr>
<td></td>
<td>Appomattox</td>
<td>1995</td>
<td>1</td>
<td>0</td>
<td></td>
</tr>
<tr>
<td>West Virginia</td>
<td>Wayne</td>
<td>2001</td>
<td>3</td>
<td>0</td>
<td></td>
</tr>
<tr>
<td>Wisconsin</td>
<td>Dane</td>
<td>2000</td>
<td>7</td>
<td>3</td>
<td>MSB151873, MSB151874, MSB151878</td>
</tr>
<tr>
<td>Florida</td>
<td>Leon</td>
<td>1983</td>
<td>10</td>
<td>1</td>
<td>MSB53277 (<em>Blarina carolinensis</em>)</td>
</tr>
</tbody>
</table>
**Supplementary Table 4.** GiRaF results with taxa (ID numbers corresponding to voucher specimens at the Museum of Southwestern Biology [MSB]) identified as the result of reassortment between the M and L segment of *Hantaviridae* and the associated percent confidence level of each identified set. Other hantaviruses identified as being possible reassortants are referenced by their abbreviation.

<table>
<thead>
<tr>
<th>Candidate ID</th>
<th>Taxa</th>
<th>Confidence</th>
</tr>
</thead>
<tbody>
<tr>
<td>17</td>
<td>MSB89866</td>
<td>&gt;0.99</td>
</tr>
<tr>
<td>21</td>
<td>MSB53254, MSB89858, MSB89895, MSB89861, MSB89867</td>
<td>&gt;0.99</td>
</tr>
<tr>
<td>25</td>
<td>MSB89861</td>
<td>&gt;0.99</td>
</tr>
<tr>
<td>30</td>
<td>MSB89861, MSB89866</td>
<td>&gt;0.99</td>
</tr>
<tr>
<td>41</td>
<td>MSB53254, MSB89858, MSB89895, MSB89867</td>
<td>&gt;0.99</td>
</tr>
<tr>
<td>89</td>
<td>AMRV, ANDV, ASA, ASIV, AZGV, BOWV, BRGV, CBV, CHOV, DBSV, DOBV, ELMC, HTNV, JIU, JMSV, KKMV, LANV, LHEV, LQV, LXV, MJNV, MSB151834, MSB47876, MSB47879, MSB47883, MSB49712, MSB53274, MNTV, NAVA, OXB, PHV, PUUV, QHSV, RIOMV, RKPV, SANGV, SEOV, SNV, SOOV, SWS, TPMV, TULV, VLAV, WEHV, WEMBV, XSV, XYIV, YKSV, YUV</td>
<td>&gt;0.99</td>
</tr>
<tr>
<td>105</td>
<td>MSB53251</td>
<td>&gt;0.99</td>
</tr>
<tr>
<td>121</td>
<td>MSB151834, MSB53251</td>
<td>&gt;0.99</td>
</tr>
<tr>
<td>127</td>
<td>MSB151834, MSB151874, MSB151875, MSB47876, MSB47879, MSB47883, MSB53274, MSB53274, MSB73580, MSB89858, MSB89859, MSB89861, MSB89863, MSB89864, MSB89866, MSB89867</td>
<td>&gt;0.99</td>
</tr>
<tr>
<td>258</td>
<td>MSB151834, MSB151874, MSB151875, MSB47876, MSB47879, MSB47883, MSB53274, MSB73580, MSB89858, MSB89859, MSB89861, MSB89863, MSB89864, MSB89866, MSB89867</td>
<td>&gt;0.99</td>
</tr>
<tr>
<td>260</td>
<td>MSB151834, MSB151874, MSB151875, MSB47876, MSB47879, MSB47883, MSB53274, MSB73580, MSB89858, MSB89859, MSB89861, MSB89864, MSB89867, OXB</td>
<td>&gt;0.96</td>
</tr>
</tbody>
</table>

Abbreviations: AMRV, Amur virus; ANDV, Andes virus; ASA, Asama virus; ASIV, Asikkala virus; AZGV, Azagny virus; BOWV, Bow virus; BRGV, Bruges virus; CBV, Cao Bang virus; CHOV, Choclo virus; DBSV, Dabieshan virus; DOBV, Dobrava-Belgrade virus; ELMC, El Moro Canyon virus; HTNV, Hantaan virus; JIU, Jeju virus; JMSV, Jemez Springs virus; KKMV, Kenkeme virus; LANV, Laguna Negra virus; LHEV, Lianhe virus; LQV, Longquian virus; LXV, Luxi virus; MJNV, Imjin virus; MNTV, Montano virus; NAVA, Nova virus; OXB, Oxbow virus; PHV, Prospect Hill virus; PUUV, Puumala virus; QHSV, Qian Hu Shan virus; RIOMV, Rio Mamore virus; RKPV, Rockport virus; SANGV, Sangassou virus; SEOV, Seoul virus; SNV, Sin Nombre virus; SOOV, Soochong virus; SWS, Seewis virus; TPMV, Thottapalyam virus; TULV, Tula virus; VLAV, Vladivostok Fusong virus; WEHV, Wenling hagfish virus; WEMBV, Wenling minipizza batfish virus; XSV, Xuan Son virus; XYIV, Xinyi virus; YKSV, Yakeshi virus; YUJV, Yuanjiang virus.
**Supplementary Table 5.** RDP4 results. Sequences identified as the result of reassortment and the associated \( p \)-value of each test. NS indicates that the method did find the breakpoint significant.

<table>
<thead>
<tr>
<th>Recombinant Sequence</th>
<th>RDP</th>
<th>GENECONV</th>
<th>Bootscan</th>
<th>Maxchi</th>
<th>Chimaera</th>
<th>SiSscan</th>
<th>3Seq</th>
</tr>
</thead>
<tbody>
<tr>
<td>MSB8986_3_Minnesota</td>
<td>1.07E-33</td>
<td>2.62E-25</td>
<td>4.52E-21</td>
<td>2.55E-20</td>
<td>0.0086206</td>
<td>4.11E-42</td>
<td>7.23E-39</td>
</tr>
<tr>
<td>MSB8986_4_Minnesota</td>
<td>4.34E-26</td>
<td>5.29E-20</td>
<td>2.45E-23</td>
<td>5.47E-12</td>
<td>1.33E-12</td>
<td>4.11E-42</td>
<td>7.23E-39</td>
</tr>
<tr>
<td>MSB47887_Massachusetts</td>
<td>3.96E-18</td>
<td>3.65E-21</td>
<td>8.99E-10</td>
<td>1.79E-10</td>
<td>1.47E-07</td>
<td>6.90E-07</td>
<td>6.90E-07</td>
</tr>
<tr>
<td>MSB49712_Arkansas</td>
<td>NS</td>
<td>NS</td>
<td>1.41E-12</td>
<td>8.22E-07</td>
<td>1.47E-07</td>
<td>6.90E-07</td>
<td>1.47E-07</td>
</tr>
<tr>
<td></td>
<td>NS</td>
<td>NS</td>
<td>1.20E-05</td>
<td>8.82E-07</td>
<td>1.47E-07</td>
<td>6.90E-07</td>
<td>6.90E-07</td>
</tr>
</tbody>
</table>
**Supplementary Table 6.** RDP4 identified breakpoints for each reassortant sequence.

<table>
<thead>
<tr>
<th>In Alignment</th>
<th>In Recombinant Sequence</th>
<th>Recombinant Sequence</th>
</tr>
</thead>
<tbody>
<tr>
<td>Begin</td>
<td>End</td>
<td></td>
</tr>
<tr>
<td>346</td>
<td>1035</td>
<td>346</td>
</tr>
<tr>
<td>1244</td>
<td>338</td>
<td>1024</td>
</tr>
<tr>
<td>1422</td>
<td>343</td>
<td>1300</td>
</tr>
<tr>
<td>1035</td>
<td>1396</td>
<td>679</td>
</tr>
<tr>
<td>106</td>
<td>348</td>
<td>106</td>
</tr>
</tbody>
</table>
Supplementary Figure 1. Tanglegram for the M and L segments across the diversity of Hantaviridae showing reassortment between the two segments. Sequences representing RPLV are highlighted.
CHAPTER 3

ASYNCHRONOUS DIVERGENCE IN INNER ASIAN PIKAS IS CONSISTENT
WITH PREDICTIONS OF HOW NICHE PREFERENCE SHAPES DIFFERENTIAL
RESPONSE TO CLIMATE FLUCTUATIONS

Schuyler W. Liphardt¹, Kayce C. Bell², Jason L. Malaney¹,³, Kurt E. Galbreath⁴, Batsaikhan Nyamsuren⁵, Candace Espinosa¹, Joseph A. Cook¹

¹Department of Biology, The University of New Mexico, Albuquerque, New Mexico
²Natural History Museum of Los Angeles County, Los Angeles, California
³New Mexico Museum of Natural History and Science, Albuquerque, New Mexico
⁴Department of Biology, Northern Michigan University, Marquette, Michigan
⁵National University of Mongolia, Ulaanbaatar, Mongolia
ABSTRACT

The evolutionary relationships, species limits, and geographic distributions of many organisms in Inner Asian biomes remain among the most poorly known worldwide due in large part to inadequate specimen availability and limited research focus on geographic variation. Species delimitation and phylogeographic history of Inner Asian pikas (Mammalia: genus Ochotona) presents an intriguing opportunity to examine historical biogeography. In this region, two species, *O. hyperborea* and *O. alpina*, are rock dwellers and often found in talus slopes, whereas *O. pallasi* and *O. dauurica* are burrowing species that typically occur in arid and semi-arid steppe. We focus on these four species of pika primarily in Mongolia because their distinctive behavior and habitat preferences provide an opportunity to explore how ecological traits may influence response to historical climate fluctuation, thus providing insight into how other regional species may have also responded. By coupling a phylogeographic comparative approach with species distribution models, we address questions regarding structure and demography of extant and historic populations of pikas across the dynamic climates and topographically diverse landscape of Inner Asia. We ask: 1) How was the demographic history of each species impacted by episodic climate change? 2) What role did ecology play in the differential response of species to changing conditions? and 3) How does the history of Mongolian pikas expand our understanding of the historical biogeography of Inner Asia during the Pleistocene? Our study reveals that the burrowing species, *O. pallasi* and *O. dauurica*, largely maintained panmictic populations across their ranges that likely reflects long-term persistence and generally high connectivity. Those histories, however, contrast with the rock-dwelling species, *O. hyperborea* and *O. alpina*, which are geographically structured, a finding consistent with patterns of contraction and
isolation followed by subsequent range expansion that may coincide with climate fluctuations of the Late Quaternary. These distinctive biogeographic histories highlight differential response of close relatives to changing environmental conditions and start to illuminate the dynamic faunal history of Inner Asia.

INTRODUCTION

A key goal of phylogeography is to determine the drivers and timing of diversification of species (Lexer et al., 2013). An extension of this, comparative phylogeography, examines sets of species to determine concordant or discordant patterns of diversification (Avise, 2009). These patterns of diversification, and the underlying mechanisms fueling these patterns across co-distributed species, often are inconsistent either temporally or spatially adding complexity to the already difficult goal of identifying generalizable processes (Edwards et al., 2021). For instance, in North America, intraspecific patterns of variation in extant mammal assemblages were largely molded both directly and indirectly by glacial oscillations throughout the Pleistocene (Galbreath et al., 2009; Hope et al., 2012). Interglacial warming resulted in reduction of ice sheets and expansion of suitable habitat for montane- and cold-adapted species northward, while southern latitudes generally experienced recession of boreal habitat and expansion of grasslands and deserts along with their associated mammal fauna concurrent with elevation shifts to sky islands or extirpation of southernly distributed montane-associated species (Emerson et al., 2010; Galbreath et al., 2010; Hope et al., 2012; Mantooth et al., 2013; Riddle & Hafner, 2006). During periods of glaciation, these trends were reversed with expansion of ice sheets and temperate forest and subsequent contraction of grasslands and deserts. Glacial oscillations and periods of repeated
habitat expansion and contraction are reflected in the genetic structure of the mammals that persisted in, or subsequently colonized, these regions (Colella et al., 2018; Hewitt, 2004).

When considered across multiple co-distributed species, however, the patterns of differentiation can be complex. Due to the cyclical nature of glaciations, the genetic signatures of isolation in glacial refugia and subsequent expansion into ice-free zones following glacial recession are not always congruent. For example, recolonization from multiple refugial sources in some cases led to secondary contact between previously isolated conspecific populations, a pattern that was potentially repeated during each interglacial cycle. Montane-associated species that persisted south of the North American ice sheets experienced repeated elevational or latitudinal range shifts (Galbreath et al., 2010). Conversely, grassland and desert-adapted species show signals of expansion during interglacial aridification of southern latitudes, and contraction during the glacial advances, sometimes leading to isolation and divergence within distinct desert refugia (Ayoub & Riechert, 2004; Riddle et al., 2000; Wilson & Pitts, 2012). Differential response to environmental change may be shaped by distinctive ecology.

While phylogeographic studies of North American and European species have grown steadily over the past few decades, there have been fewer investigations of the drivers of species diversification across Inner Asian biomes (Lebedev et al., 2021; Lv et al., 2016). The distinctive topography and climatic history of Inner Asia complicates the extension of emergent patterns from the well-studied systems of North America and Europe. A key distinction of Inner Asia during the Quaternary that contrasts with North America and Europe is the lack of extensive ice sheets covering most of the northern latitudes (Batchelor et al., 2019). With glaciation instead restricted to montane areas, temperate species of Inner
Asia may have experienced less displacement during the cycles of warming and cooling of the Quaternary. The literature on how climate may have influenced diversification of Inner Asian mammals, however, remains sparse with a majority of studies to date focused primarily on species distributed in the Himalayas and Tibetan Plateau and relatively few focused on more northerly distributed species (Ci et al., 2009; Dahal et al., 2017; McLean et al., 2018; Yu et al., 2012). The mountains and steppe of Mongolia and southeastern Siberia provide an ideal setup to test hypotheses surrounding how climatic cycling and ecological niche preference may have shaped phylogeographic patterns of intraspecific diversification and demographic history across Inner Asia during the dynamic climate of the Pleistocene.

There is some disagreement about the center of origin for central and eastern Asian flora and fauna (Lebedev et al., 2021; Lv et al., 2016). One leading scenario suggests a Tibetan origin (“Out of Tibet”) with dispersal of species from the Qinghai-Tibetan plateau during the late Miocene (Deng et al., 2011; Jia et al., 2012; Wang et al., 2020). This hypothesis postulates that the Tibetan Plateau acted as a training ground for cold-adapted temperate species prior to dispersal to environments farther north after the cooling of the Miocene, but it does not address species responses to the warming and cooling cycles of the Pleistocene. Like North America, the Pleistocene climate of Inner Asia was not static but instead showed repeated cycles of both warm and cold periods. These dynamic climate fluctuations combined with the complex landscape of the region may have produced repeated habitat shifts that hypothetically led to periods of expansion and contraction of organisms. In small mammals, the timing and direction of these dynamic events apparently are not consistent across taxa (Lebedev et al., 2021; Lv et al., 2016; McLean et al., 2018). For example, expansion of steppe habitat during interglacial periods allowed the desert hamster,
Phodopus roborovskii, to extend its range and increase effective population size, but during periods of glaciation this species experienced contraction of habitat and distribution (Lv et al., 2016). Similar patterns appear in the relatively few other phylogeographic studies in Inner Asia (e.g., Lebedev et al., 2021).

Pikas (genus Ochotona) are small mammals (< 1 kg) that provide an elegant system for a comparative approach to testing hypotheses surrounding the interplay of climate cycling and ecological niche in shaping demographic history of Inner Asian fauna. The vast majority of extant pika diversity, 32 of 34 currently recognized species (mammaldiversity.org), occurs in Asia. Ochotona consists of four subgenera, Conothoa, Pika, Ochotona, and Alienauroa, that represent distinct phylogenetic lineages with corresponding shared ecologies within each subgenus with exception of Pika (Yu et al., 2000). While Alienauroa and Ochotona consist of colonial, burrowing pikas that occur in steppe and forest habitats, respectively, Conothoa contains less social pikas that are found in montane talus environments. Pika also contains talus-adapted species, as well as more intermediate and steppe-associated members (Smith et al., 2018). There is an extensive body of work on species description and boundaries within Ochotona, which incorporates karyotype, morphological, and molecular data (Koju et al., 2017; Lissovsky, 2014; Lissovsky et al., 2007, 2016; Melo-Ferreira et al., 2015). By contrast, research on diversification patterns and phylogeographic history remain sparse for Asian pikas (Khalilipour et al., 2014; Mohammadi et al., 2018), especially compared to the two North American species, Ochotona princeps and O. collaris (Galbreath et al., 2009, 2010; Knowles et al., 2016). Using phylogenetic approaches, Wang et al. (2020) aimed to identify the origin and timing of diversification of extant pikas. They placed initial diversification of four subgenera of Ochotona in the late Miocene with subsequent diversification and northern
expansion from the Tibetan Plateau during the Pliocene, a finding that is generally consistent with the “Out of Tibet” hypothesis. However, the recent history, especially the drivers and timing of diversification, of most species are unknown.

Mongolia in particular represents a unique opportunity to address broader questions surrounding Pleistocene diversification of high latitude Inner Asian fauna. Mongolia is topographically heterogeneous, with distinctive biomes that range from the vast deserts of the Gobi, widespread steppe, multiple mountain ranges, and near contiguous taiga forest in the northern part of the country. As such, the landscape of Mongolia is reflective of diverse environments stretching across northern Inner Asia and contains the core ranges of three of the four pika species we studied, allowing us to sample their spatial breadth.

Here we use a multifaceted comparative approach to test hypotheses of asynchronous diversification within four species of pikas distributed across Mongolia, two of which can be tentatively grouped as the montane ecotype, *O. alpina* and *O. hyperborea*, and two as the steppe ecotype, *O. pallasii* and *O. dauurica*. We examine how these ecotypes differentially responded to the dynamic climate of the Pleistocene to shape their phylogeographic structure, demographic history, and niche preference and we compare these patterns to those uncovered in the North American pika. Galbreath et al. (2009, 2010) studied the North America montane-associated species, *O. princeps*, and suggested that Pleistocene glacial cycles drove elevational shifts to sky island refugia during warmer interglacials, whereas during glacial periods there was increased connectivity between pikas associated with discrete mountain ranges, a system potentially analogous to the mountain ranges of western Mongolia. Specifically, we predict that during the warm interglacial periods the montane-associated Asian ecotypes retreated to sky island refugia where populations diverged in isolation.
During periods of plateau glaciation and cooling, these montane-associated pikas responded by shifting downslope where they subsequently dispersed and colonized new, adjacent montane systems. Due to climate cycling during the late Pleistocene, this process likely repeated, producing a pattern of diversification similar to that detected in *O. princeps* with divergent clades associated with geographically separated mountain systems (Galbreath et al., 2009, 2010) and consistent with expectations of the Pleistocene species-pump hypothesis (Bringloe & Saunders, 2019; Papadopoulou & Knowles, 2015; Schoville et al., 2012). We predict that this pattern of diversification will be reflected in the geographic structure of intraspecific clades corresponding to mountain systems, with diversification times roughly corresponding to periods of cooling, and deeper coalescent times reflecting this disjunct distribution of populations.

In contrast to elevational shifts, some species likely shifted geographically with climate cycling and remained largely intact with little population differentiation. We predict that steppe-associated species maintained a core population in the Mongolian Plateau with a relatively stable population that shifted its range latitudinally and longitudinally following shifts of steppe habitat during climatic cycling consistent with the habitat tracking hypothesis (Maresova et al., 2021; Raia et al., 2012; Wake et al., 2009). This pattern should be reflected in relatively stable population size through time, limited geographic structure, and relatively deep coalescence times.

Finally, one might also predict an intermediate response to climatic cycling for species that are less tightly tied to their environmental niche with retreat to steppe habitat leading to secondary contact and panmixia during cooler glacial periods followed by contraction of populations during warm interglacial periods resulting in population
fragmentation and divergence. This response predicts relatively shallow branching structure due to repeated cycles of relatively short periods of vicariance followed by gene flow after secondary contact.

In this study we aim to address key questions surrounding patterns of small mammal diversification across Inner Asia during the Pleistocene. 1) How was the demographic history of each species impacted by episodic climate change? 2) What role did ecology play in the differential response of species to changing conditions? and 3) How does the history of Mongolian pikas expand our understanding of the historical biogeography of Inner Asia during the Pleistocene? We combine divergence dating, demographic reconstruction, coalescent simulations, and species distribution modeling to test hypotheses surrounding climate-driven diversification of the four species of pika to lay a foundation for broader investigations of Inner Asian biogeographic history during the Pleistocene.

**HYPOTHESES**

Driving this study are specific hypotheses on how species with distinct ecologies reacted to dynamic climatic cycling of the Pleistocene in Inner Asia. We expect idiosyncratic responses to climate change with each ecotype responding differently such that these responses were asynchronous but consistent with hypotheses related to their ecology. For steppe-associated species (*O. dauurica* and *O. pallasi*) we expect one large panmictic and stable population through time largely unaffected by climatic cycling consistent with the Pleistocene Habitat Tracking Hypothesis (PHTH). Based on this hypothesis, we predict the following demographic and phylogenetic signals: 1) a signal of population expansion concordant with the last interglacial cycle followed by a population decline during glacial cooling, however overall population size should remain relatively large, 2) phylogenetic
signal indicative of a large and stable population maintained through time resulting in a lack of geographic structure to the phylogeny with relatively deep coalescence of terminal nodes. For montane-associated species (*O. hyperborea* and *O. alpina*), we expect that Pleistocene climatic cycling resulted in multiple vicariant events during interglacial periods followed by pulses of colonization during glacial cycles consistent with the Pleistocene Species Pump Hypothesis (PSPH). Concordant predictions include: 1) older populations within a species will have a signal of a stable population through time with derived populations exhibiting a signal of population expansion following glacial cycles, 2) species phylogenies will show a step-like pattern of diversification with clades structured geographically. Finally, consistent with the Pleistocene Ecological Plasticity Hypothesis (PEPH), we expect to detect a potential intermediate signal between either the predominantly steppe or montane species with migration out of a central population during interglacial warming up in elevation to more montane habitat followed by retraction back to a core population and secondary contact between diverged peripheral lineages. Our specific hypotheses and predictions are outlined in Figure 1.

**MATERIALS AND METHODS**

*Sampling and sequencing*

Tissue samples for 146 pikas (86 *O. pallasi*, 23 *O. dauurica*, 9 *O. alpina*, 24 *O. hyperborea*, two *O. collaris*, two *O. princeps*, one *O. roylei*, and one *O. hoffmanni*) were acquired from the Division of Genomic Resources at the Museum of Southwestern Biology. Pikas were collected during collaborative field expeditions between the Mongolian National University and the University of New Mexico between 1999 and 2016 as part of the Beringian Coevolution Project, Collaborative Integrative Investigations of Biomes of the
Arctic, and Mongolian Vertebrate Parasite Project (IACUC protocol 19-200908-MC) and following guidelines of the American Society of Mammalogists (Sikes et al., 2016). Total DNA was extracted from liver tissue with an E.Z.N.A DNA extraction kit (Omega) or a standard salt extraction protocol (Fleming & Cook, 2002). A 720 base pair fragment of the mitochondrial cytochrome b gene (cyt b) was PCR amplified using primers PIKA1 and PIKA2 with the following conditions: initial denature at 94°C for five minutes, 40 cycles consisting of 30 seconds of denature at 94°C, 30 seconds for annealing at 45°C, one-minute extension at 72°C followed by a final 10-minute extension at 72°C. Individuals representative of well-supported mitochondrial clades were also sequenced for five independent nuclear genes: beta-fibrinogen (BFIB), glutamate ionotropic receptor AMPA type subunit 3 (GRIA3), interphotoreceptor retinoid-binding protein (IRBP), spectrin beta non-erythrocytic 1 (SPTBN1), and Thy-1 cell surface antigen (THY). All primer sequences used in this study can be found on Table S1. Conditions for nuclear PCR followed that for cyt b, but with variable annealing parameters (Table S1). Sanger sequencing of amplified regions was performed on an ABI 3130 Genetic Analyzer at the University of New Mexico’s Molecular Biology Facility following manufacturers protocol for BigDye 3.1 (Applied Biosystems). Forward and reverse sequences were assembled and manually edited in Geneious Prime 2021.1.1.

Phylogenetic analyses, population delimitation, and divergence dating

Sequences for each gene were aligned using MAFFT version 7.475 (Katoh & Standley, 2013). Phylogenies were inferred per gene with all four species included, as well as per gene for individual species under a Bayesian framework with the program MrBayes version 3.2.7 (Ronquist et al., 2012). A mixed model of nucleotide substitution was used to
test all model space during tree search with a gamma distribution of rates using the option ‘lset nst=mixed rates=gamma’. MCMC analysis was run for 20,000,000 generations, two runs and four chains sampling trees every 2,000 generations with all priors left at default values. Convergence of independent runs was verified with the average standard deviation of split frequencies being below 0.01 as recommended by Ronquist et al. (2012). Parameters where summarized using the ‘sump’ command and a 50% majority rule consensus tree was generated using the ‘sumt’ command with a 25% burn-in.

The program bPTP (Zhang et al., 2013) was used to delimit populations based on cyt b clades. A tree file was prepared for input to bPTP by combining the last 500 trees from each independent run ‘.t’ file generated by MrBayes. bPTP was then run on this combined tree file with the ‘-r’ flag to root trees on the longest branch prior analysis and all other values set to default. The best supported populations as determined by posterior values was used to define our intraspecific populations for subsequent analyses.

To estimate inter and intraspecific diversification times for the four species, we used BEAST version 2.6.2 (Bouckaert et al., 2019; Alexei J. Drummond et al., 2012). We applied a multi-locus multi-species coalescent approach with STARBEAST. Alignments containing diversity across four sub-genera of pika species for cyt b and five nuclear loci were used for divergence estimates. Individuals were grouped by population based on those delimited by bPTP and ploidy value was adjusted to 0.5 for cyt b. The site model per loci was determined by ModelFinder (Kalyaanamoorthy et al., 2017) implemented in IQ-TREE (Nguyen et al., 2015) to determine the best-fit model of nucleotide substitution. A relaxed log-normal clock was estimated for all loci with a starting rate of 0.05 substitutions per site per million years for cyt b and 0.005 for nuclear loci. We set a monophyletic MRCA prior of 32 million years
(mya) with a gamma distribution for the root of our tree representing the timing of divergence of *Ochotona* from our outgroup, *Lepus oiostolus*. These dates were taken from the mean dates estimated for these splits from Wang et al. (2020). The STARBEAST analysis was run for 100,000,000 generations storing trees every 5,000. The analysis was run three times and log files were inspected with Tracer version 1.7.1 (Rambaut et al., 2018) to verify ESS values over 200 for each parameter. Trees from each run were combined with LogCombiner version 2.6.6 (Bouckaert et al., 2019) removing the first 10% of trees from each run as a burn-in. A maximum clade credibility tree was then calculated and annotated using TreeAnnotator. We repeated this analysis for the three species representing the sub-genus *Pika*, *O. pallasi*, *O. alpina*, and *O. hyperborea*, to provide greater resolution of intraspecific diversification times. We used two calibration points for this subset, one of 5.3 mya representing the root for these three species, and a second of 4.7 mya as the MRCA between *O. hyperborea* and *O. alpina*. Both calibration points were given a gamma distribution prior to allow flexibility in these estimates.

*Population demographics and Bayesian skyline plots*

To test for neutral evolution absent of selection we calculated Tajima’s D and Fu’s Fs statistics for each species for cyt *b* in Arlequin (Excoffier & Lischer, 2010). To investigate effective population size (*N_e*) through time, Bayesian skyline plots (BSPs) were inferred in BEAST 2.6.2 (Bouckaert et al., 2019; Drummond et al., 2005; Drummond et al., 2012). BSPs were inferred from the cyt *b* dataset for each species as well as within species populations following the same designation as those used for tests of neutrality.

*Species distribution modeling*
Occurrence points for each species were downloaded from GBIF and spatially thinned to 10 km using a randomization algorithm in the R package spThin. Current conditions, mid-Holocene, Last Glacial Maximum (LGM), and Last Interglacial bioclimatic variables were downloaded from Worldclim and Envirem for a total of 35 variables (Fick & Hijmans, 2017; Title & Bemmels, 2018). Bioclimatic and Envirem variables were either downloaded natively at a spatial resolution of 2.5 arcminutes or downsampled to this resolution from 30 arcsecond rasters. Envirem variables for the LIG were calculated from monthly modeled climate data downloaded from Worldclim using the envirem R package. Past climate conditions were simulated under the Community Climate System Model (CCSM4). The R package biomod2 was used to generate distribution models using seven different methods; Maxent, Generalized Boosting Model, Classification Tree Analysis, Artificial Neural Network, Surface Range Envelope, Flexible Discriminant Analysis, and Random Forest. The median value calculated across all models was used as the general model for that time period. We then normalized this continuous prediction and converted it to a binary suitable/not suitable raster with a threshold of 0.5 (Sillero et al., 2021). We then summed binary predictions across time periods to determine areas of predicted stable habitat through time, where values of 2 equates to stable habitat, 1 is habitat suitable either currently or during historic periods, and 0 is not suitable at any time.

**Hypothesis testing**

To test explicit hypotheses of demographic histories, we employed coalescent simulations with the program msprime (Kelleher et al., 2016). Because of the relative lack of population structure within *O. dauurica* we limited our simulations to populations within *O. pallasi*, *O. hyperborea*, and *O. alpina*. Our time calibrated species phylogeny for these three
species was used as a template to model demographic history with a generation time of two years (Galbreath et al., 2010) to transform branch lengths to coalescent time. The genealogical history was simulated 100 times using this demographic model. Genetic sequences were simulated from each phylogeny using the python package pyvolve with an evolutionary model estimated from empirical data during phylogenetic reconstruction as outlined previously. Phylogenies were then inferred from each simulated alignment with IQ-TREE (Nguyen et al., 2015). To test concordance with our species tree inferred from empirical data, both gene and site concordance factors (gCF and sCF) were calculated using IQ-TREE as well. Concordance factors measure either the concordance of branching structure and length of one phylogeny to the other (gCF), or the support of a branch in a reference tree given an alignment (sCF).

RESULTS

Phylogenetic analyses

Our Bayesian and maximum likelihood trees inferred from cyt b are largely consistent in topology and consistent with published phylogenies for pikas (Lissovsky, 2014; Melo-Ferreira et al., 2015; Wang et al., 2020). All four species of pika were recovered as monophyletic (Fig 2).

The two burrowing species (O. pallasi, O. dauurica) showed minimal geographic structure compared to the two montane-associated species, O. hyperborea and O. alpina. Although O. pallasi is subdivided into three well-supported clades, (Fig 2), divergence estimates for these clades are relatively shallow and geographically overlap, suggesting possible recent expansion or maintenance of ancestral variation. Ochotona dauurica also lacks deep structure and appears to be a single, large panmictic population based on lack of
geographic structure. Minimal structure is highlighted by the bPTP analysis recovering a majority of *O. dauurica* as a single clade, with the exception of one sample from China (Fig 2). In contrast, the rock-dwelling species show significantly greater intraspecific variation, although our sampling for *O. hyperborea* covers a much larger geographic area compared to the other three species examined. For *O. alpina*, a species with similar sampling and distributional extent to *O. dauurica* and *O. pallasi*, we recover greater geographic structure and deeper divergence compared to those steppe-associated species. *Ochotona alpina* is composed of geographically defined lineages corresponding to distinct mountain systems (Fig 2). *Ochotona hyperborea* also displays geographic structure, with eight lineages recovered in cyt *b* associated with distinct mountain systems across Siberia, with one lineage, hyperborea5, associated with the Mongolian Khangai and Altai mountains, although there is some spatial overlap in these lineages (Fig 2).

**Timing of divergence**

Divergence time estimates generated in STARBEAST show clade divergences in the rock-dwelling *O. alpina* and *O. hyperborea* that largely mirror each other in timing (Fig 4). Although *O. hyperborea* has a deeper origin than *O. alpina*, subsequent intraspecific diversification is largely concurrent over the last ~500 thousand years. In contrast, *O. pallasi* shows less diversity in general with only three clades recovered by bPTP and divergence of these clades is also much shallower, occurring during the late Pleistocene (Fig 4). While bPTP did split *O. dauurica* into two clades, dauurica1 is represented by one individual, AI87185, from China. Dating divergence within *O. dauurica* using a multi-species, multi-locus approach in the absence of internal structuring is inappropriate.

**Population demographics**
Tajima’s D and Fu’s Fs statistics were not significant for any of the four species suggesting these species have evolved neutrally (Table 1). Our Bayesian skyline plots suggest that when each species is analyzed as one large panmictic population they exhibit a stable population size through time with limited negative or positive change, with the exception of *O. dauurica* which shows a distinct increase in effective population size starting at the beginning of the Holocene (Fig 5). When analyzed by locality, *O. pallasi* from the Mongolian Aimag Bayan Olgii show a brief reduction in population size followed by an increase that dates to around 5 kya and corresponds to a similar peak seen in the genetic clade pallasi3 (Fig 6).

**Coalescent simulation**

Our coalescent simulations resulted in alignments and phylogenies that are highly concordant to the calibrated phylogeny for the subgenus *Pika* (Fig 7, 8). Interspecific divergence was supported by all of the simulated phylogenies with slightly less support for intraspecific divergences, although still relatively high, as indicated by per branch gCF values (Fig 7). We see a similar pattern in sCF values with high interspecific support and lower intraspecific support, although in general support was lower than for gCF values (Fig 8).

**Species distribution modeling**

Across their present range, both *O. pallasi* and *O. dauurica* show areas of suitable habitat from the LIG through the LGM and up to current conditions with minor longitudinal shifts during the LGM (Fig 9, 10). SDMs for the two steppe-associated species suggest that suitable habitat remained largely connected throughout at least the past 120,000 years. In contrast, *O. alpina* is predicted to have multiple small, disjunct patches of suitable habitat.
that were stable through time (Fig 11). While *O. hyperborea* is predicted to have larger patches of stable habitat, those patches are also predicted to be disjunct and patchy across their range (Fig 12). This patchiness in predicted suitable habitat largely reflects the geographic distribution of genetic clades that were tied to distinct and separated mountain systems. SDMs for both *O. hyperborea* and *O. alpina* show a large longitudinal band of connected and highly suitable habitat during the LGM, a time that corresponds to mountain plateau glaciation.

**DISCUSSION**

Conflicting conclusions have emerged regarding the distinctive roles of montane and steppe habitat as refugia during Pleistocene climatic cycling across Inner Asia. For the long-tailed hamster, the steppe of the Mongolian plateau was hypothesized to only be occupied during relatively warm interglacial periods (Lebedev et al., 2021). In contrast, despite ecological similarity between desert and semi-arid steppe habitats, the desert hamster retreated to the Mongolian plateau during periods of glaciation and maintained a stable core population (Lv et al., 2016). Meanwhile montane and boreal habitats were in flux, essentially shifting up and down mountain ranges in elevation, with only limited refugial habitat for associated boreal species (Hais et al., 2015). This dynamic fluctuation in montane and boreal habitat likely impacted diversification in ground squirrels as they followed the expansion and contraction of available habitat (McLean et al., 2018). The Pleistocene of Inner Asia showed a general trend of increased aridification and expansion of steppe and semi-desert habitat with pulses of expansion, surprisingly, during the colder periods when northern latitudes of North America were largely covered by ice sheets (Seidl et al., 2021). The lack of a northern ice field barrier to dispersal for cold-adapted species of Inner Asia likely resulted in more
fine-scale forces at play either restricting or expanding species distributions. Here, we examined four related species that span the ecological breadth of a single genus in an effort to tease apart the components of climate and habitat that shaped demographic and phylogeographic history in the absence of large-scale glaciation at northern latitudes.

**Geographic structure and climate**

Limited intraspecific diversification within *O. pallasi* and *O. daurica* compared to *O. alpina* and *O. hyperborea* (Fig 2, 3), as also reflected in the delimitation analysis which recovered few OTUs in *O. pallasi* (3) and *O. daurica* (2) compared to *O. hyperborea* (9) and *O. alpina* (7), is likely due to extensive geneflow across large, panmictic populations that homogenized variation across the two steppe-associated species. *Ochotona daurica* with a single population across Mongolia (Fig 13) is a pattern also recovered in other arid-adapted species distributed across the Mongolian Plateau (Lebedev et al., 2021; Lv et al., 2016). This finding is consistent with the expansion and contraction of steppe habitat during the Pleistocene in this region (Seidl et al., 2021). Our study suggests that *O. daurica*, a steppe-associated species, maintained a large core population that remained relatively stable throughout the warming and cooling cycles of the Pleistocene, but with instances of expansion on the range periphery during interglacial cycles. This core population scenario is further supported by the expansion of population size starting in the Holocene that is reflected in the Bayesian skyline plots of these pika species (Fig 5), a period with similar environmental conditions to interglacial periods. In contrast to *O. daurica*, both *O. alpina* and *O. hyperborea* are structured geographically by mountain ranges, but with relatively stable populations (Fig 5, 6, 13). Because *O. hyperborea* has a much larger range than the other three species, we may not have sufficient sampling to rigorously explore the extent of
panmixia across its range; however, these analyses suggest that this species is structured geographically, a finding consistent with acoustic races and phylogeny based on cytochrome oxidase c subunit 1 (Lissovsky et al., 2021).

*Ochotona alpina*, with a similar distribution to *O. pallasi* (Fig 13), is geographically structured by mountain ranges within Mongolia and the Russian Altai. While we did not recover a signal of LGM population expansion in either of the montane-associated species, the phylogeographic structure of both species suggests multiple instances of colonization followed by subsequent allopatric divergence. Given the history of glaciation in this region and the recognition that these two species are generally found in mountain talus habitat, but are known to colonize taiga (Smith et al., 2018), and (Batchelor et al., 2019; Böhner & Lehmkuhl, 2005), this pattern of diversification is consistent with our predictions of colonization during glacial maxima followed by elevation shifts up slope during interglacial warming, a pattern shared by their North American cousin with a very similar ecology, *Ochotona princeps*. In contrast to the other three species examined, *O. pallasi* exhibits a unique phylogeographic history, with the current distribution of phylogenetic clades largely sympatric across its range, similar to *O. dauurica*, however with distinct mitochondrial clades suggesting a deeper history of vicariance followed by secondary contact of divergent lineages. The history of diversification in *O. pallasi* resembles an intermediate stage between that shared by the montane-associated species, and that of *O. dauurica*. While it is difficult to determine the factors and timing of divergence and secondary contact with this data set, expanded genomic sampling may help resolve this.

*Species distribution modeling*
Our species distribution modeling is largely consistent with the hypothesis of habitat fragmentation for talus-adapted species during interglacial periods as arid and semi-arid steppe expanded to previously forested mountain valleys, effectively cutting off gene flow between mountain slopes. Although glacial cycling caused pulses of alternating connectivity and allopatry within talus-adapted *Ochotona* species due to habitat fragmentation and expansion, these cycles appear to have a different impact on steppe-associated species. Arid and semi-arid steppe environments expanded during interglacial periods and contracted during glacial cooling; however, this apparently resulted in contraction of habitat, but not extensive fragmentation. Steppe-associated *Ochotona* species likely underwent loss of peripheral populations on the northern edge of their range, as has been seen in long-tailed hamsters (Lebedev et al., 2021), however, the core population likely remained intact with connectivity across their range. While it is likely that cold adaptation was the catalyst for the initial radiation of *Ochotona* from the Tibetan Plateau (Wang et al., 2020), expansion of preferred habitat appears to be a primary factor in driving diversification within pikas during the interglacial periods in the Pleistocene. However, during cooler glacial periods suitable habitat appears to be largely overlapping for all four species of Mongolian distributed pikas (Fig 9-12). This suggests the presence of possible biotic factors (e.g., interspecific competition) at play during these periods restricting the distribution of these species.

*Timing of divergence and climatic cycling*

Although the timing of origin for these four species is relatively congruent, subsequent patterns of intraspecific diversification are not. Much deeper and more numerous instances of divergence occurred in the montane-associated pikas compared to their steppe-associated relatives. *Ochotona hyperborea* for example shows multiple instances of
colonization and divergence throughout the Pleistocene with the youngest clade divergence dating to roughly 80 kya (Fig 4). Divergence in *O. pallasii* in contrast is much younger and with only three extant clades as recovered by bPTP, with divergence dating to roughly 60 kya and 25 kya respectively (Fig 4). The relatively uncommon and shallow instances of divergence in *O. pallasii* may be due to high higher rates of gene flow that have homogenized variation throughout their range, a finding consistent with the hypothesis of relatively stable habitat that persisted throughout the Pleistocene. Due to the large confidence intervals surrounding divergence estimates in *O. alpina* and *O. hyperborea*, it is difficult to tie these events to specific glacial or interglacial periods.

*Hypotheses*

We predicted distinct patterns of phylogenetic structure, demographic history, and geographic distribution of populations based on three ecological niche-based hypotheses for four species of Inner Asian pikas (Fig 1). *Ochotona dauurica* largely responded as predicted for a steppe-associated species with one relatively large population that appears to maintain panmixia across its range and with deeper coalescence events indicating stability through time (Fig 2, 13). We also recovered a signal of recent population expansion for *O. dauurica* that started during the Holocene, the most recent interglacial period of global warming. Both *O. alpina* and *O. hyperborea* display phylogenies that are geographically structured with lineages, populations, associated with discrete mountain systems, although at different scales with *O. hyperborea* having a much larger range (Fig 2, 13). We did not recover signals of increases in population sizes associated with glacial periods for either *O. alpina* or *O. hyperborea* as both species appear to have had relatively stable population sizes through time (Fig 5), which does not follow our predictions, or the pattern of diversification seen in the
ecologically similar *O. princeps* of North America (Galbreath et al., 2010). *Ochotona pallasi*, the arid to semi-arid steppe-associated species, largely tracks predictions for the hypothesis of an intermediate ecology between purely steppe and purely montane ecologies. While we did recover phylogenetic structure that is relatively deep for cyt *b* (Fig 2), this structure is not reflected by geographically defined populations as is seen in *O. alpina* and *O. hyperborea*. Furthermore, the depth of coalescence for these lineages, relatively deep in cyt *b*, but becoming very shallow with addition of the nuclear loci (Fig 4) suggesting that either not enough time passed in isolation for nuclear loci to sort, or that variation accrued in isolation has since been homogenized. Similar to the two montane-associated species, we do not see a signal of an increase in population size associated with cooler glacial periods as predicted, instead, demographic analyses suggest a recent decrease in population size of *O. pallasi* around 5 kya during the Holocene.

Lastly, it is important to note that the four species examined there are no demographic signals indicative of a cyclical pattern of changes in population sizes corresponding to either interglacial warming or glacial cooling. While the structure of phylogenetic divergences for *O. alpina* and *O. hyperborea* is suggestive of a congruent and repeated pattern of vicariance and colonization, it is difficult to associate this with underlying climatic or geologic events. This finding stands in contrast to the strong signal of demographic expansion following glacial expansion seen in North American pika which guided initial demographic hypotheses for the montane-associated Inner Asian pikas.

**Conclusions**

Differences in ecological niche preference, when combined with climatic cycles during the Pleistocene, have differentially shaped diversification of species of *Ochotona* in
Inner Asia. During periods of cooling (glacial maxima), montane-associated species colonized new mountain ranges, likely forced to lower elevations by plateau glaciation. Subsequently, during interglacial cycles, arid and semi-arid steppe habitat expanded and effectively restricted connectivity between newly colonized mountains, with elevational shifts upslope resulting in allopatric divergence of montane-associated species. Those pulses of colonization and divergence are not shared by the steppe-associated species, however. In contrast, there is a pattern of population expansion and contraction during interglacial and glacial periods respectively, but without subsequent divergence of populations. This pattern is similar to systems in North America (Hope et al., 2012), although without interglacial allopatry in steppe-associated species. These processes are not binary, however, with diversification patterns following a spectrum across the four species examined that is in line with species-specific, fine-scale ecological preferences. While the strongly steppe-restricted *O. daururica* appears to exhibit patterns reflective of a large panmictic population that was likely stable throughout the Pleistocene, *O. pallasi*, a species with habitat preferences that span both steppe and semi-steppe foothills, shows phylogenetic and demographic signal of historical isolation despite current geographic overlap.

We addressed three questions related to the biogeographic history of Inner Asia during the Pleistocene: 1) How was the demographic history of each species impacted by episodic climate change? 2) What role did ecology play in the differential response of species to changing conditions? and 3) How does the history of Mongolian pikas expand our understanding of the historical biogeography of Inner Asia during the Pleistocene? While we did not recover a signal of large-scale population changes through time in montane-associated species, we did uncover a signature of recent expansion in *O. daururica*, a steppe-
associated representative. It is also evident from these analyses that montane- and steppe-associated pikas responded differentially following niche preference during the Pleistocene as habitats expanded, contracted, and shifted during periods of warming and cooling, and these responses appear to be asynchronous respective to ecology. However, a strict binary separation between ecotypes might not be appropriate as we recovered an intermediate history in *O. pallasi*. We were not able to strongly support or reject the cyclical nature of these patterns, however. With this data set, signals corresponding to the last cycle of Pleistocene cooling were recovered with deeper phylogenetic structure merely suggestive of a cyclical pattern to divergence. Further sampling across individuals, species, and loci should provide increased resolution to whether or not these patterns were repeated temporally. The lack of extensive ice sheets restricting northern migration of flora and fauna during periods of glaciation in Inner Asia results in a distinctive environment for the evolution of northern latitude distributed species when contrasted with counterparts in Europe and North America. An environment where barriers to gene flow, or lack thereof, appear to be heavily tied to species autecology and biotic variables rather than physical obstacles appears to be the chief factor. Clearly, abiotic and biotic forces at play across Inner Asia are complex and will require extensive spatial sampling across diverse species to decipher whether generalizable processes are producing patterns that are applicable across biota. Despite the latitudinal equivalence between Inner Asia and both Europe and North America, the patterns of diversification and underlying drivers do not appear to be shared across regions. The specific barriers to gene flow, or avenues encouraging genetic exchange across the landscapes of Inner Asia, remain clouded, but the emergent patterns of faunal responses to Pleistocene climatic cycling found in heavily glaciated North America and Europe do not readily apply.
This study begins to tease apart what those barriers may be for the Inner Asian biota broadly and adds to the sparse, yet growing, literature on the biogeographic history of the region.

**ACKNOWLEDGEMENTS**

We thank the field mammalogists that collected the specimens that made this research possible. We would also like to thank the amazing group of individuals in Mongolia and Siberia that have shown endless hospitality to us throughout the years. We especially thank the students of the National University of Mongolia that accompanied the various field expeditions as well as the drivers that kept everyone safe and provided invaluable local support. Lastly, we thank the University of New Mexico Center for Advanced Research Computing, supported in part by the National Science Foundation, for providing the high-performance computing resources used in this study. This research was made possible by funding over the past two decades from the National Science Foundation through the Mongolian Vertebrate Parasite Project, Beringian Coevolution Project, and the Collaborative Integrated Inventory of Biomes of the Arctic under the following NSF grants: 2026377, 2033482, 1561342, 1258010, 0717214.
References


Papadopoulou, A., & Knowles, L. L. (2015). Genomic tests of the species-pump hypothesis: Recent island connectivity cycles drive population divergence but not speciation in


FIGURE LEGENDS

Figure 1. Hypotheses and predictions of phylogeographic and demographic response by distinctive ecotypes of *Ochotona* to climatic cycling of the Pleistocene in Inner Asia.

Figure 2. Bayesian phylogeny of genus *Ochotona* based on the mitochondrial cytochrome b gene. Non-focal species are collapsed to aid visualization.

Figure 3. Multi-species, multi-locus phylogeny for *Ochotona* as a whole, and populations for the four focal species of this study. Colored bars correspond to ecotypes associated with subgenera.

Figure 4. Intraspecific diversification estimates for *O. pallasi*, *O. hyperborea*, and *O. alpina*.

Figure 5. Extended Bayesian skyline plots for geographic and genetic populations of *Ochotona pallasi*. Time is presented in millions of years before present.

Figure 6. Extended Bayesian skyline plots for geographic and genetic populations of *Ochotona pallasi*. Time is presented in millions of years before present.

Figure 7. Gene concordance factor (gCF) per branch representing the number of times the reference branch (empirically generated) was represented in the simulated data.

Figure 8. Site concordance factor (sCF) per branch representing the percentage the reference branch (empirically generated) was supported by the simulated alignment.

Figure 9. Species distribution models for *Ochotona pallasi* for present, LGM, and LIG conditions. Dashed lines represent current species distribution.

Figure 10. Species distribution models for *Ochotona dauurica* for present, LGM, and LIG environmental conditions. Dashed lines represent current species distribution.

Figure 11. Species distribution models for *Ochotona alpina* for present, LGM, and LIG conditions. Dashed lines represent current species distribution.
Figure 12. Species distribution models for *Ochotona hyperborea* for present, LGM, and LIG conditions. Dashed lines represent current species distribution.

Figure 13. Geographic distribution of genetic clades for *Ochotona alpina* (a), *O. dauurica* (b), *O. pallasi* (c), and *O. hyperborea*. 
**TABLE**

<table>
<thead>
<tr>
<th>Statistics</th>
<th>O. pallası</th>
<th>O. alpina</th>
<th>O. dauurica</th>
<th>O. hyperborea</th>
</tr>
</thead>
<tbody>
<tr>
<td>Sample size</td>
<td>85</td>
<td>9</td>
<td>13</td>
<td>24</td>
</tr>
<tr>
<td>S</td>
<td>28</td>
<td>19</td>
<td>14</td>
<td>56</td>
</tr>
<tr>
<td>S</td>
<td>28</td>
<td>19</td>
<td>14</td>
<td>56</td>
</tr>
<tr>
<td>Pi</td>
<td>7.49468</td>
<td>7.63889</td>
<td>3.84615</td>
<td>15.73188</td>
</tr>
<tr>
<td>Tajima's D</td>
<td>1.05332</td>
<td>0.45451</td>
<td>-0.6119</td>
<td>0.19124</td>
</tr>
<tr>
<td>Tajima's D p-value</td>
<td>0.889</td>
<td>0.733</td>
<td>0.311</td>
<td>0.63</td>
</tr>
<tr>
<td>Fu's FS test</td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>Real no. of alleles</td>
<td>34</td>
<td>6</td>
<td>9</td>
<td>19</td>
</tr>
<tr>
<td>Orig. no. of alleles</td>
<td>85</td>
<td>9</td>
<td>13</td>
<td>24</td>
</tr>
<tr>
<td>Theta_pi</td>
<td>7.49468</td>
<td>7.63889</td>
<td>3.84615</td>
<td>15.73188</td>
</tr>
<tr>
<td>Exp. no. of alleles</td>
<td>19.30432</td>
<td>6.22585</td>
<td>6.0872</td>
<td>14.88152</td>
</tr>
<tr>
<td>FS</td>
<td>3.9017</td>
<td>0.9412</td>
<td>-2.62121</td>
<td>-0.42897</td>
</tr>
<tr>
<td>FS p-value</td>
<td>0.896</td>
<td>0.647</td>
<td>0.062</td>
<td>0.439</td>
</tr>
</tbody>
</table>

Table 1. Neutrality statistics for *Ochotona pallası, O. alpina, O. dauurica, and O. hyperborea* as calculated in Arlequin.
Figure 1.
Figure 2.
Figure 3.
Figure 4.
Figure 5.
Figure 6.
Figure 7.
Figure 8.
Figure 9.
Figure 10.
Figure 11.
Figure 12.
Figure 13.
CONCLUSION

Comparative phylogeography is an inherently multidisciplinary field broadly applicable to studies of biodiversity (Avise, 2009; Edwards et al., 2021; Gutiérrez-Garica & Vázquez-Domínguez, 2011). For my dissertation, I combined classic techniques of comparative phylogeography with population genetics, statistical inference, and coalescent simulation to systems that span multiple temporal, spatial, and organismal scales. In addition to elucidating the processes responsible for patterns of diversification in small mammals and associated viral pathogens, I also examine how different ecological niches may influence response to environmental change. While the scope of my research is broad, these questions are fundamental to the underlying goal of phylogeography; what processes govern how organisms are distributed spatially.

By examining viruses in the context of their hosts across North America, I highlighted the role of host phylogeographic history in viral distribution across the landscape. In western North America, I compared virus and host history and showed that, although there are instances of presumed codiversification, this pattern is complicated by rampant host switching. Whether or not this is a result of regional variation in host diversity providing more opportunities for spillover and switching, or a fundamental biological barrier to novel host colonization is a question that needs further sampling and analyses. I also investigated a host-virus system in eastern and central North America and discovered a similar signal of possible codiversification on a fine-scale, however in this case it was coupled with viral reassortment between deeply divergent lineages. While the existence of divergent lineages within a single host species necessitates spillover, the source of that diversity remains to be discovered. Finally, I used a comparative approach that focused on four species within the
genus *Ochotona* that are distributed across East Asia to determine how ecology and Pleistocene climatic cycling shaped diversification processes. I showed that phylogeographic histories of diversification in these species were asynchronous, differential, and tied to ecology with three patterns that followed the spectrum of host ecologies.
References

