A conservation and taxonomic assessment of the least shrew (*Cryptotis parvus)* complex through rangewide phylogeographic analyses and population genomics

by

Thomas Galfano

B.S., University of New Mexico, 2019

A THESIS

submitted in partial fulfillment of the requirements for the degree

MASTER OF SCIENCE

Division of Biology
College of Arts and Sciences

KANSAS STATE UNIVERSITY
Manhattan, Kansas

2021

Approved by:

Major Professor
Andrew Hope

Copyright

Abstract

Cryptotis is a speciose genus of shrews with a broad New World distribution from southern Canada to Northern South America. *Cryptotis* consists of 49 currently recognized species and is traditionally split into six distinct species groups. North American Least shrews (*Cryptotis parvus*) belong to the *parvus* group of shrews which includes six recognized species. *Cryptotis parvus* (sensu stricto) is broadly distributed across central and eastern North America from southern Ontario, Canada, and the Great Lakes, south to Florida and west through tall and short grass prairies to their western extent just east of the Rocky Mountains in the United States. Within *C. parvus* there are two recognized subspecies: *Cryptotis p. floridanus* found throughout the coastal Carolinas south through Florida, and *Cryptotis p. parvus,* found throughout the rest of the range. However, these shrews are locally rare, and little is yet known about their evolutionary history. New populations occurring peripheral to the recorded distribution of *C. parvus* have been repeatedly discovered since the 1980s, progressively extending the known range of this species westward. In the west, potentially isolated populations relying on declining mesic grassland habitats have led to the shrew's listing as threatened within New Mexico, but this listing has elicited additional genetic analyses to qualify existing species, and infraspecific relationships, including rangewide systematics, connectivity/gene flow, and regional genetic diversity and demography. While this taxon is currently recognized as a single species with two subspecies, I hypothesize that cryptic diversity is present that will challenge the existing taxonomy. Preliminary rangewide assessments of *C. parvus* indicate populations in Florida may be distinct from the rest of the range, but these results were based on early genetic techniques, and the evidence for delimiting distinct taxa was not conclusive. Among western peripheral populations in New Mexico, previous research postulated that multiple distinct lineages may

exist, reflecting recent westward range extension in the north but relictual persistence in the south. Given that all peripheral populations are currently considered a single taxon for management, my research aims to inform the most appropriate units of analysis for applied conservation within this species. In this research, I performed a rangewide phylogeographic study, using Cytochrome-B data (n=106) and reduced representation genome sequencing (~10,000 SNP loci; n=64), to 1) assess relationships across the range of *C. parvus* with the inclusion of additional closely related species, and 2) investigate the genetic legacy of New Mexico populations within the broader context of the species complex. Both mitochondrial and nuclear data indicate populations in Florida are highly divergent and support the hypothesis that *C. parvus* is made up of multiple cryptic taxa. Phylogenomic analyses of nuclear data support the high divergence seen in Florida populations and also indicate that southern New Mexico populations constitute a distinct and endemic infraspecific taxon of *C. parvus* reflecting both long-term isolation and adaptive divergence. Although this taxon may warrant recognition as a unique subspecies, it does not align with any currently recognized subspecies of *C. parvus,* and without the inclusion of diagnostic morphology, genomic data alone are not at this point robust for designating taxonomy below the level of species. Instead, I have assessed the relative contributions of adaptive versus neutral evolution for defining different evolutionary units of analysis that reflect locally adaptive diversity or population divergence as a consequence of independent (allopatric) differentiation, respectively. My results highlight a complex history of diversification of least shrews that lends support for shared faunal (and floral) evolution across the Great Plains, and more broadly through North America.

Table of Contents

List of Figures

List of Tables

Acknowledgments

I would like to begin by acknowledging that none of the research or field work for this project would have been possible without access to numerous research sites and samples from land that belongs to Indigenous communities that were forcibly removed from these lands. Firstly, where I spent much of my time, the land that Kansas State University and other field sites throughout Kansas reside on land that belongs to the Arapahoe, Kickapoo, Kaw, Kiowa, Comanche, and Pawnee people of Kansas. Next, land that provided numerous field sites and least shrew samples in New Mexico that belongs to the Jicarilla Apache, Kiowa, and Comanche people of New Mexico. And finally, land in the Rio Grande Valley of southern Texas, that provided multiple field sites that yielded one very important *Cryptotis berlandieri* specimen, that belongs to the Lipan Apache and Coahuiltecan people.

I would also like to acknowledge that none of this work would have possible without funding from the New Mexico Department of Game and Fish Share with Wildlife program. These funds provided ample opportunity for me to perform field surveys and gain experience using next generation DNA sequencing methods. Without these opportunities for hands on experience, I would not be half the scientist I am today. Leaving New Mexico to pursue graduate education was a hard choice, but I am very grateful I was able to work on a project that will benefit the mammals of the state I am so proud to call home. Additionally, many thanks to employees from NMDGF that performed field surveys in the summer of 2020, collecting critical least shrew specimens that were invaluable to my research. In addition to these field surveys, I would like to acknowledge the natural history collections that made my research project possible. Many thanks to Robert Dowler and Dianna Kresja, Angelo State University Natural History Collections; John Demboski, Denver Museum of Nature and Science; Jennifer Frey, New

Mexico State University; Kayce Bell, Los Angeles County Museum; Robb Brumfield, Donna Dittmann, Jake Esselstyn, and Frederick Sheldon, Louisiana State University; Joe Cook and Mariel Campbell, Museum of Southwestern Biology; Jessica Light, Texas A&M University; and Verity Mathis and David Reed, University of Florida. These biological repositories are invaluable infrastructure for biodiversity research, and my project would not have been possible without the generous tissue loans from each of these natural history collections.

I would not have gotten to this stage of my degree without the guidance and direction provided by my advisor, Andrew Hope. Andrew has provided unwavering support since the day I moved to Kansas, and I am so thankful I made the decision to join Andrew and the Hope Lab. He has taken me from the prickly, sharp, and spiny deserts of the US-Mexico border to every corner and small town in the state of Kansas, to the rolling Black Hills of South Dakota, creating memories I will never forget. Andrew has taught me so much about how to be a well-rounded scientist, providing ample opportunities to develop and refine my skills in a field, lab, and classroom setting. He has demonstrated a level of dedication that is honorable and a level of tenacity in his work that is unmatched, and even in the midst of a global pandemic, that did not change. Even when the world shuts down, Andrew Hope doesn't. I am also beyond fortunate to have worked with the lab group he assembled in my time there. Beginning with Fraser Combe, I really cannot say thank you enough for the loads of help and guidance you provided as I began to navigate the world of bioinformatics and data analysis, I would not be where I am now without you answering all of my ridiculous questions, so for that I want to say thank you.

The remainder of the Hope Lab were comprised of some of the most genuine people I have ever met, and I know we will maintain a bond for life, a bond that can only come from

camping together without showers for weeks at a time. Ben, Kailey, Litsa, Mary, and Tommy thanks for being the best group of mammals I could have asked for!

I would also like to thank so many wonderful family and friends that have shown nothing but undying support for me pursuing my graduate education. Beginning with my Mom and Dad, thank you both for your endless love and support, you have supported my dream of pursuing a career working with mammals since day one, and I am so blessed and thankful to have had parents that have supported me to the degree you have. To my sister, Sabrina, and Grandparents: Grandma, Geezer, and Gram thank you for always being there for me and being a source of happiness and relief when I needed it most. Your support means everything to me and I am so thankful you've been there for me through all of this. And of course, thank you to Darwin and Bailey for being cuddliest canines around and providing me with, some may argue, too much love.

Lastly, I would like to acknowledge my friends that have supported me along the way. To my friends that made my time in Manhattan memorable, Molly, Matt, Anne, among others, thank you for the ice cream runs, dinner parties, and binge watching sessions that saved my mental health over these last couple of years. And finally, to my friends near and far, thank you for always being there for me, you all prove that true friends are never apart, maybe in distance, but never in heart. Courtney, Bella, Maddie, Chelsee, Edward, Dom, and countless others, you are the best friends a mammal like me could ask for.

I really cannot thank the amazing people in my life enough for the endless stream of support and love that they have continued to bless me with. This would not have been possible without you!

Chapter 1 - Introduction

Cryptic species present a unique challenge to recognizing the full extent of existing biodiversity, impeding efforts towards effective conservation and wildlife management. Identifying certain taxa may prove difficult using traditional approaches that involve comparisons between morphological characters that cannot resolve differences between cryptic yet genetically distinct species (Gordon & Watson 1985). Cryptic species are defined as two or more taxa that have been classified as one ostensible species because they are superficially morphologically indistinguishable (Bickford et al. 2007). Cryptic species complexes are being recognized with increasing frequency across the tree of life, in part as a consequence of developing analytical technologies such as genomic sequencing (Witt et al. 2006). The proportion of cryptic species is similar across metazoan taxa, and even within mammals, one of the most resolved classes of biodiversity, there remains high levels of unrecognized cryptic diversity (Pfenninger & Schwenk 2007). Members of the family Soricidae, the shrews, are notoriously hard to identify based on morphology because of their diminutive size and conserved body plan, coupled with high rates of diversification (Hope et al. 2012). Recognition of cryptic species complexes, such as are found among soricids, is important because such a group may be considered and managed as a single species that is of no conservation concern, when, in reality, it may be multiple cryptic species with unique evolutionary histories, ecological associations, and facing different levels of threat, thus requiring different conservation strategies (Niemiller et al. 2013).

The genus *Cryptotis* offers a clear example of repeated crypsis and the issues surrounding recognition and description of species. Although at one morphological extreme species of *Cryptotis* can be large and robust, reflecting adaptations for fossorialism, and at the other

extreme, tiny and gracile, reflecting divergent adaptive evolution, often these distinct ecological forms constitute species groups or complexes of closely related and morphologically cryptic taxa that continue to make systematic and taxonomic assessment difficult (Woodman & Morgan 2005; Woodman 2011; Woodman & Gaffney 2014; He & Woodman 2015). *Cryptotis* species are collectively referred to as the small-eared shrews, constituting a diverse genus with a New World distribution. For the purpose of this study, I follow the taxonomic distinctions within the genus *Cryptotis* provided in "American Recent Eulipotyphla" (Woodman 2018). Under this framework, the genus is comprised of 49 recognized species, with most occurring through Central (15 species) and South America (26 species; Woodman 2018). These 49 species are broken down into six species groups based on a combination of biogeographical proximity, morphological factors (Choate 1970; Woodman & Timm 1993; 1999; 2017) and molecular genetic analyses (Guevara & Cervantes 2013; He et al. 2015; Baird et al. 2017). Species ranges are often small, and multiple species within the genus may occur in sympatry. This, coupled with relatively poor sampling, particularly of tissues with adequate quality for robust genetic analysis, has resulted in continued taxonomic uncertainty across the genus *Cryptotis* (Woodman & Timm 1999).

While *Cryptotis* shrews are diverse at lower latitudes, and comprise the only group of shrews to be found in South America (Churchfield 1990), only a single putative species occurs in North America north of Mexico. The least shrew, *Cryptotis parvus,* like other members of its genus, is currently considered part of a species complex (hereafter called the *parvus* group), with members exhibiting variably dark grey-brown pelage with a paler, slightly counter-shaded, underside, small ears and eyes, and a relatively small bi-colored tail (Whitaker 1974; Hall 1981). These indistinct characters make designating and differentiating between species difficult. The *parvus* group occurs throughout the central and eastern United States, and south through much of

Mexico and Central America. Members include *C. berlandieri* occurring from southern Texas through central Mexico, *C. soricinus*, with a limited distribution through the State of Mexico, *C. pueblensis* in the vicinity of Oaxaca and Chiapas, *C. tropicalis* from Chiapas, Mexico through Guatemala, *C. orophilus* through Central America, and *C. parvus* (sensu stricto) occurring broadly through the central and eastern United States (Choate 1970; Hutterer 2005; Woodman 2018). The specific status of these taxa is still contentious given a lack of comprehensive genetic and/or morphometric evidence. But, at one time or another, each member of the *parvus group* has been recognized as a subspecies under *C. parvus,* before ultimately being elevated to full species status. Within *C. parvus* (sensu stricto) there have been four recognized subspecies: *C. p. elasson,* found in Ohio, *C. p. floridanus,* found in Florida and Georgia, *C. p. harlani,* found in eastern Illinois and western Indiana, and *C. p. parvus* which is found throughout the central Great Plains (Whitaker 1974). The status of *C. p. harlani* as a viable subspecies has been questioned. Assessments performed by Mumford and Whitaker (1982) and Hoffmeister (2002) determined that there were no significant differences between populations of least shrews in eastern Illinois and western Indiana and populations from other parts of Illinois and Nebraska, which would be representative of *C. p. parvus.* Based on recent assessments performed by Hoffmeister (2002), and Hutchinson (2010), *C. p. harlani and C.p. elasson* are no longer recognized as subspecies of *C. parvus* because of the lack of differentiation seen between populations in Illinois, Indiana, and Ohio, and populations constituting the rest of the range. This leaves only *C. p. floridanus* with a limited distribution through the southeast United States and *C. p. parvus*, with a much broader distribution extending from the Great Lakes region in the northeast, east coast from New York south to Florida, most of the Gulf Coast, over to western limits in eastern New Mexico, the foothills of the Front Range in Colorado, eastern Wyoming,

and western South Dakota. In the north, *C. p. parvus* is limited to southern South Dakota, northern Iowa, southern Wisconsin, and northern Michigan.

Not only are *Cryptotis* shrews morphologically cryptic, but they are also rarely sampled, and as such, little is still known of their specific evolutionary history, ecology, or ongoing population dynamics. New populations occurring peripheral to the recognized distribution of *C. parvus* have been repeatedly discovered since the 1980s, progressively extending the known range of this species, although whether this reflects recent range extensions or simply more comprehensive sampling is unclear. The current IUCN Red List distribution map does not include populations from New Mexico, Colorado, South Dakota, and Wyoming, which have been identified as recently as 2006 (Marquardt et al. 2006). Records in eastern New Mexico are the earliest of these documented range extensions (Hoditschek 1985). Given the multi-faceted uncertainty surrounding the biology of the *parvus* group, my thesis research strives to address two primary objectives: 1) to clarify systematic relationships among species and sub-species and diagnose any potentially cryptic taxa, and 2) to perform a conservation genomics assessment of of *C. p. parvus* to provide insight into the genetic diversity, connectivity, and demographic trends among little known peripheral populations, with a particular focus on the southwest U.S.

Extant populations of *C. parvus* in New Mexico were only recently discovered. However, this shrew is recorded based on fossil evidence from late-Pleistocene deposits at 2 localities in New Mexico: Dry Cave in Eddy County and Howells Ridge Cave in Grant County (Harris 1973). The earliest documented modern record is represented by a single specimen collected in a snap trap in 1961 from Eddy County in southeastern New Mexico. This sample was only recently discovered at the University of Colorado-Boulder Museum, and it remains the southernmost recent record of *C. parvus* in New Mexico, although the Eddy County population

is currently considered extirpated due to severe habitat degradation (New Mexico Department of Game and Fish [NMDGF] 2020). Subsequent populations of least shrews were identified in New Mexico in 1981 at Tucumcari Lake in Quay County (Hoditschek 1985), in 1982 at Grulla National Wildlife Refuge in Roosevelt County (Owen & Hamilton 1986) and in 1985 at Bitter Lake National Wildlife Refuge in Chaves County (Shuster 1989). These were the only known populations of the taxon within the state, which subsequently led to it being added to the state threatened and endangered species list in 1985, as well as being identified as a species of greatest conservation concern (SGCN) by the New Mexico Department of Game and Fish (NMDGF 1988). The decision to list the least shrew was due to its patchy distribution as well as susceptibility to habitat loss because of its presumed obligate association with mesic grassland. In the late 1980s, immediately following its listing, the state made the only attempt, until present day, to assess the conservation status of *C. parvus* in New Mexico (Frey 2005). Sampling efforts in recent years have greatly expanded the known range of this taxon within New Mexico to include numerous localities within the High Plains and Pecos River valley (Frey 2005; NMDGF 2020). The NMDGF, through the Share with Wildlife Program, provided partial support for this portion of my thesis research. This was specifically to use genomic analyses of existing specimens to better describe the distribution of genetic diversity among *C. parvus* throughout New Mexico, and to resolve the taxonomic status and evolutionary history of these peripheral populations. The ultimate goal here, in alignment with the objectives of the NMDGF, is to help ensure long-term persistence of robust, representative, and secure populations of *C. parvus* in New Mexico, such that into the future they no longer require protection under the New Mexico Wildlife Conservation Act (NMDGF 2020). Recently, the NMDGF published a Least Shrew Recovery Plan (NMDGF 2020) within which they also

outlined a number of criteria for species recovery. This includes identification of four core recovery areas where least shrew populations are present, each consisting of at least four separate habitat patches where least shrew presence is documented, and at least one habitat patch at each locality must be on land that is conserved for wetland or wildlife management (NMDGF 2020).

The evolutionary history and relationships among peripheral populations of *C. parvus* was first assessed by Hafner and Shuster (1996) who performed a preliminary comparative assessment, based on a combination of morphometric measurements and allozyme loci, on the three previously identified populations in New Mexico, as well as one population in northern Texas (Wichita Falls, Wichita Falls County), which was the closest known population at the time, ~225 km further east. They tested two alternate hypotheses for the source of populations of *C. parvus* in New Mexico: 1) that populations represent a recent westward colonization event made possible by increased irrigation networks created in the region beginning in the 1960's and 70's; or 2) that the populations are relictual and have persisted in isolated pockets throughout the region since the late-Pleistocene. This study analyzed four geographic populations including the Tucumcari drainage in northeast New Mexico, Grulla National Wildlife Refuge (NWR; "Salt Lake", an isolated saline lake situated on the Llano Estacado) in eastern Roosevelt County, Bitter Lake NWRin Chaves County, and Wichita Falls in the Red River drainage of northern Texas. Preliminary genetic evidence based on 24 allozyme loci, showed similarity between specimens from Tucumcari, Grulla NWR and Wichita Falls, collectively representing the recognized subspecies *C. p. parvus*. However, the population from Bitter Lake NWR, was genetically distinct, with four private (unshared) alleles (Tucumcari samples had two private alleles, and other sites had zero). Morphological comparisons again showed congruence between Tucumcari, Grulla NWR, and Wichita Falls, but divergence of Bitter Lake NWR shrews, the latter

population distinguished by relatively shorter palatal length (Hafner & Shuster 1996). This shorter palatal length is consistent with measurements of *C. berlandieri* the most geographically close species to *C. parvus*, found further south from the Rio Grande Valley of southern Texas into Mexico. Conclusions from this work were that populations in northern New Mexico likely represent recent westward expansions mediated by increased agricultural irrigation, while populations in southern New Mexico exhibit both genetic distinctness from northern populations, and morphology consistent with *C. berlandieri,* as such representing relict populations of *C. berlandieri* persisting in the Pecos River Valley.

Despite seemingly two distinct *Cryptotis* taxa occurring in eastern New Mexico, these early analyses provided only limited genetic and morphological resolution and, as such, all of these peripheral populations have remained categorized as *C. p. parvus*. Due to the threatened status of this unresolved taxon, a deeper of understanding of the evolutionary history and taxonomic relationships is warranted.

My thesis considers the evolution of *C. parvus* from two complementary genetic perspectives, and based on shrew samples collected broadly from throughout their combined distribution. The first considers sequences of the mitochondrial Cytochrome b (Cytb) gene (full gene length of 1,140 base pairs [bp]). This locus was selected because mtDNA is haploid (evolving relatively quickly compared with diploid nuclear DNA), and maternally inherited (Giles et al. 1980). Benefits of these characteristics include relatively fine temporal and spatial scales of inference for delineating regionally distinct lineages, levels of divergence, and demographic effective population change through time. Additionally, providing evolutionary insight from a uni-parentally inherited locus allows for a preliminary assessment of ongoing gene flow through hybridization as this may be reflected by discordance between mitochondrial and

nuclear inheritance (Toews & Brelsford 2012). Finally, the Cytb gene has, over the last three decades, become the workhorse genetic locus for mammals and an effective barcode gene for assessing relative levels of divergence and identity, for instance, among cryptic species, such as *Cryptotis* (Avise et al. 1987).

The second dataset is comprised of ~11,000 single nucleotide polymorphisms (SNPs) derived from a double-digest restriction site associated DNA sequencing approach (ddRADseq). This relatively recently developed genomic sequencing method (Peterson et al. 2012) was chosen because it is a reduced-representation sequencing approach that allows for consistent sampling across many individuals of incised short reads from across the nuclear genome (using a pair of restriction enzymes), without the computational intensity of a whole-genome sequencing approach. This in turn allows for greater depth of coverage per locus, increasing confidence in genotype calls. After filtering reads, loci can then be divided into neutral and non-neutral datasets and analyzed to consider alternative evolutionary processes of genetic drift or natural selection, respectively. The traditional evolutionary perspective for conservation of distinct diversity was through consideration of all available genetic data combined, to designate evolutionarily significant units (ESUs; Moritz 1994). However, among other benefits, differently evolving loci can inform other distinct kinds of conservation units. Neutral loci inform us of divergent populations due to random genetic drift that may be considered as management units (MUs), which reflect populations that are demographically distinct from each other, most likely as a result of extended isolation. Divergence of populations at non-neutral loci, or outlier loci, can indicate adaptation to local environments, either in the past or ongoing, and as such may be considered adaptive units (AUs) experiencing divergence through natural selection (Funk et al. 2012). MUs inform short-term conservation decisions, while AUs are intended to supplement

consideration of MUs to conserve populations undergoing adaptive processes, as well as their local environments (Funk et al. 2012). The combination of mtDNA and a large nuclear dataset will allow me to explore the evolutionary history of least shrews at both recent and deeper timescales.

My research chapters both use evidence from Cytb sequences as well as ddRADseq loci, but focus on different geographic and taxonomic scales, to address multiple questions associated with the evolutionary history of the *C. parvus*. In my first research chapter, I present the results of a rangewide phylogeographic study that focuses on understanding the distribution and systematic relationships among well-resolved lineages of *C. parvus* across most of their range in the North America, including sequence data from other species of the *parvus* group. My primary questions here include: 1) does the described taxonomy reflect the existing genetic differentiation among lineages? And 2) are there additional cryptic lineages within the *parvus* group that warrant recognition? Briefly, my results indicate that *C. parvus* is a cryptic complex of taxa that warrants taxonomic revision of existing designations, as well as recognition of additional undescribed taxa. In both nuclear and mitochondrial phylogenies, samples cluster into three distinct lineages, a western lineage, including samples from New Mexico and central United States west of the Mississippi River, an eastern lineage, consisting of populations east of the Mississippi, not including Florida, and a Florida lineage. However, these lineages are not a monophyletic group with respect to other recognized species occurring through Mexico and Central America, strongly supporting the presence of multiple species within *C. parvus*. Importantly, New Mexico *Cryptotis* also represent two highly distinct taxa.

In my second research chapter I focused more regionally with an in-depth population genomic analysis of New Mexican populations of least shrews. This chapter addressed the

questions: 1) how evolutionarily distinct (isolated) are *Cryptotis* shrews from Chaves County, New Mexico? And 2) do measures of genomic diversity and demography of New Mexico shrew populations support conservation listing and potentially additional recognition of adaptive and management units? Briefly here, based on genomic data, samples from Chavez County New Mexico form a distinct monophyletic lineage with no evidence of contemporary gene flow. Demographically, High Plains populations have experienced range expansion but may currently be in decline, while populations in Chavez County have clearly experienced recent population contraction. Together, these results coupled with analysis of both neutral and potentially adaptive loci indicate that *Cryptotis* in southern New Mexico are unique, highly isolated, and constitute both management and adaptive units for targeted future conservation efforts. Based on the combined results of my research chapter I finish by providing a synopsis of management recommendations for *C. parvus* in New Mexico that will help to ensure that the NMDGF is able to achieve its goal of maintaining robust populations across both the High Plains and the Pecos River Valley that are at reduced risk of further population deterioration.

Chapter 2 - Rangewide Phylogeographic and Phylogenomic Assessment of the *Cryptotis parvus* Group of small-eared shrews

Introduction

The scientific underpinnings of successful applied conservation biology depend on knowledge of Earth's existing biodiversity. It is estimated that global biodiversity may exceed 8.7 million species, but the total number of discovered and described species is 1.2 million, only a fraction of the estimated total diversity (Mora et al. 2011). Even among mammals, a relatively small but highly charismatic and human-relatable group of species, the recognized species diversity has continued to rise in recent decades (~6,500 species; Burgin et al. 2018). For mammals, as well as other faunal groups, a major obstacle for accurate species identification is crypsis (ambiguous phenotypes) among closely related species that obscure true phylogenetic relationships, recognition of existing diversity, and sometimes highly divergent gene pools. Revealing this cryptic diversity would allow for more accurate management of independently evolving taxa, including recognition of additional species or infraspecific lineages (Bickford 2007). Often cryptic species are closely related and constitute complexes or groups of sibling species (Pfenninger & Schwenk 2007). Cryptic species groups that contain potentially threatened or endangered taxa require special consideration and a unique approach towards conservation because 1) species already considered threatened or endangered may be composed of more than one taxon, each of which may be more or less at risk than if they were considered as a single species, and 2) rare cryptic taxa likely have incorrect nomenclature and, as such, cannot be legally managed or protected as unique entities (Bickford et al. 2007). Given that species have traditionally been diagnosed based on morphology, it is unsurprising that recognition of cryptic

diversity has greatly increased with the advent of modern genetic techniques (Sites & Marshall 2003; 2004).

The fields of phylogenetic systematics and phylogeography have provided insightful lines of additional evidence for relationships among taxa and within morpho-species, by allowing for quantification of genetic/genomic divergence among taxa, and for relating the distribution of independent evolutionary lineages to their geography and the processes of environmental change that have led to their formation through time. Phylogenetics uses morphological traits or genetic loci to infer relationships on the branches of the tree of life. By combining inference of evolutionary relationships with geographic distributions and known changes in geology and climate, Avise (2000) coining the term "phylogeography" and developed both theory and practice for the associated discipline. As the field of phylogeography developed, it became evident that associations between the distribution of genetic diversity and geography across many species were spatially concordant, and repeatable through time (Avise 2008). Taking a phylogeographical approach will therefore provide predictive power and analytical resolution for discovery of cryptic lineages across widespread, diverse taxa, which makes it a strong approach for improving our understanding of global biodiversity.

As the name suggests, the genus *Cryptotis* is a cryptic genus of "small-eared" shrews found through the New World. Constituent species designations within the genus have fluctuated through time, reflecting the inherent difficulties in recognizing distinct taxa. As an example of this, the number of recognized species from Mexico and Central America has historically varied from 25 species (Hall & Kelson 1959) to as low as 8 species (Choate 1970). Current species richness for this region is as high as 30 species (Woodman 2018). The genus *Cryptotis* in its entirety comprises 49 distinct species, with the vast majority of the diversity existing from

Mexico southwards through Central and northern South America. There are 13 species found in Mexico, 17 in Central America, 17 in South America, and only 2 exist north of Mexico, one of which barely reaches the northern extent of its range in southern Texas (*Cryptotis berlandieri*). Distinguishing between members of the genus is a difficult task, with little visible morphological differences existing. Species comprising the genus have traditionally been divided into six species groups based on external morphological characteristics as well as cranial measurements (Woodman 2018). These species groups have been further refined by incorporating post-cranial measurements, in particular considering morphological variation of the humerus and the manus related to functional differentiation of life history traits (He & Woodman 2015). It is perhaps ironic that levels of morphological crypsis within this group are high, and yet morphological features such as forelimb have proven to be a reliable metric for identifying species (Hutterer 1985; Goldman 1912). Choate (1970) showed that morphological characters among *Cryptotis* species often exhibit phylogenetic signal as well, indicating they may be useful in understanding both evolutionary and ecological relationships (Woodman & Morgan 2005). He & Woodman (2015) provided evidence that morphological and genetic relationships align for species groups throughout Central and South America, inferring from this that fossorial lifestyles have evolved multiple times within the genus across relatively short time scales. However, within any of these morpho-groups, there are numerous cryptic taxa that are evolutionarily distinct and occupy discrete geographic distributions. Numerous studies have focused on morphological systematics of Central and South American members of the genus (Woodman & Timm 1993; 1999; 2005; Woodman & Gaffney 2014; He & Woodman 2015), yet very few studies have focused on the sole widely-distributed member of the genus found north of Mexico: *Cryptotis parvus.*

Two primary research efforts have addressed phylogenetic/phylogeographic relationships within *C. parvus* by analyzing genetic data in combination with morphological evidence. The earliest focused explicitly on peripheral populations of *C. parvus* within New Mexico (Hafner & Schuster 1996); a primary focus of my second research chapter. The second study, performed by Hutchinson (2010), identified three distinct lineages of shrews within *C. parvus* distributed across North America, two geographically concordant with the subspecies *C. p. parvus* in the western US, and eastern US, respectively, and the other consistent with the recognized subspecies *C. p. floridanus* distributed in the southeastern United States. However, these analyses were based on low resolution genetic markers and only a few loci. In addition, there was no overlap in sampling or data analyses between the work of Hafner and Schuster (1996) and Hutchinson (2010). Conclusions provided by Hutchinson (2010) included a need for further targeted sampling to collect fresher tissues allowing for more advanced genetic analyses. This prior work provided a baseline for my thesis research, and preliminary evidence from which to develop testable hypotheses.

In this first research chapter of my thesis, I focus on investigating cryptic diversity within *C. parvus*, using a combination of one mitochondrial (maternally inherited and haploid) locus and thousands of nuclear (bi-parentally inherited and diploid) loci. The ultimate goal of this chapter is to use phylogenetic analyses from a phylogeographic perspective to gain a deeper understanding of the evolutionary history of the *C. parvus* group across the breadth of its distribution in the United States and south into Mexico. Based on previous research, I predicted that *C. parvus* will constitute multiple distinct cryptic species as opposed to a single taxon. As such, my null hypothesis is that all *Cryptotis* in North America currently recognized as *C. parvus* are indeed a single species. Alternative hypotheses would include recognition of a cryptic species

currently considered *C. p. floridanus*, with everything else being *C. parvus*, or recognition of multiple species congruent with distinct genetic lineages as outlined previously. As part of these investigations, I will discuss the criteria for recognition of species through genetic analyses, and similarly, the difficulties associated with recognizing taxonomy (especially infraspecific ranks) based on genomic datasets.

Methods

Sampling and DNA extraction

Least shrew samples for this project were acquired from a combination of field surveys and samples housed in research natural history collections. Additional field sampling across New Mexico was performed by the New Mexico Department of Game & Fish (NMDGF). Six species were represented in our tissue samples: *C. parvus, C. berlandieri, C. soricinus, C. goldmani, Blarina carolinensis,* and *Notiosorex crawfordii,* with the latter three as outgroups. *Cryptotis goldmani*, *B. carolinensis* and *N. crawfordii* were selected as appropriate outgroups due to their close but differential systematic placement with respect to *C. parvus*. Within *C. parvus,* both of the putative subspecies (*C. p. floridanus,* and *C. p. parvus*) were well-represented in my sample pool. Samples represent most of the least shrew range in North America, including New Mexico, Colorado, Texas, Kansas, South Dakota, Louisiana, Florida, and Virginia (Figure 2-1; Supplementary Materials Appendix 1). Tissues used in the analysis were preserved in either 95% EtOH or stored at -80°C. DNA was extracted from either muscle or liver tissue for each sample using the Qiagen DNeasy Blood and Tissue kit and accompanying protocol (Qiagen, Valencia, California). DNA extractions were then visualized on a 0.8% agarose gel, 5µL of each extraction was mixed with 3µL of Gel Red to stain and aid in visualizing DNA (Huang et al. 2010). Samples that passed visual evaluation of a strong band of genomic DNA were moved along to

PCR amplification. Samples deemed not to have lots of quality DNA were re-extracted, and re-evaluated for DNA yield. Poor quality samples were excluded from genomic sequencing but where possible were sequenced for the mitochondrial locus.

Mitochondrial Cytochrome-B sequencing

The full Cytochrome-B (cytb) gene (1140 bp) from the mitochondrial genome was amplified for each sample using mammalian cytb-specific primers MSB05 and MSB014 developed previously (Hope et al, 2010). Polymerase chain reaction conditions were: 6 minutes at 94.0℃, 50x cycles of 94.0℃ for 25 seconds, 48.0℃ for 30 seconds, 72.0℃ for 1 minute, 72.0℃ for 15 minutes, and then held at 10.0℃. After performing PCR for each sample, products were assessed on 0.8% agarose gels and quantified visually. Successful PCR products were cleaned using the ExoSAP-it protocol, where 1μL of exo-SAP was added to 5μL of each PCR product, followed by 15 minutes at 37℃ and 15 minutes at 80℃ (Bell 2018). Samples were then loaded into a clean 96-well plate for Sanger sequencing, which was performed by Genewiz LLC on an ABI 3730 (South Plainfield, NJ). Post-sequencing, raw reads were uploaded, cleaned and aligned using Geneious Prime (Kearse et al. 2012). Additional cytb sequences for *C. parvus* samples were downloaded from GenBank, including sequences for two additional species of the *parvus* group, *C. orophilus* and *C. tropicalis.*

ddRADseq library prep and sequencing

Reduced representation genome sequencing using restriction enzymes under various protocols generally requires 100ng of high-quality DNA per sample from extractions. To quantify extractions, 1mL of Quant-iT PicoGreen (Life Technologies, Grand Island, NY) was added to

each sample. This reagent binds to double stranded DNA (dsDNA) and fluoresces, while unbound dye does not fluoresce, allowing dsDNA to be accurately quantified against a standard concentration curve (Ahn et al. 1996). The samples were left to incubate at room temperature, protected from light, for 5 minutes, and then placed into the FilterMax Multimode F5 Microplate Reader (Molecular Devices, San Jose, CA). Samples with sufficient yields of genomic DNA were retained (n=68). Selected samples were submitted to the University of Minnesota Genomics Center (UMGC, Minneapolis, MN) for double digest restriction site associated DNA sequencing (ddRADseq; Peterson et al. 2012). For each sample, the following protocol was followed: 100ng of genomic DNA was subjected to two restriction enzymes (SbfI and TaqI) and incubated at 37℃ for two hours. Samples were then heated at 80℃ for 20 minutes to ensure deactivation of all restriction enzymes. Samples were subsequently ligated using T4 ligase and phased adaptors with CG and CRYG overhangs and then held at 22℃ for 1 hour, before raising the temperature again to deactivate. Post-ligation, each sample was cleaned with SPRI beads, and then amplified for 18 cycles, using 2x NEB Taq Master Mix. These libraries were then pooled and then size selected for reads ranging between 300-744 base pairs. UMGC then sequenced 150 base pair single end reads on a NextSeq 550 High-Output FlowCell (Illumina, USA). This sequencing produced a series of Fastq files that were demultiplexed using a combination of Illumina bcl2fastq software, Trimmomatic (Bolger et al. 2014), and CutAdapt (Martin 2011) to remove extraneous padding at the 5' end of the read and the Illumina adapter at the 3' end of the read, as well as remove any sequences that were <93 base pairs long. These trimmed reads then underwent filtration using *process_radtags*, a module in Stacks v2.54 (Rochette et al. 2019). This filtration removed reads that contained uncalled bases, and then using a sliding window

approach, removed reads with overall low-quality scores (phred score <10). This resulted in 75,223,650 final reads.

Denovo RAD locus and SNP identification

Due to the lack of a *C. parvus* reference genome, the single nucleotide polymorphisms (SNPs) in this analysis were identified de novo, using the denovo_map.pl pipeline, which is part of the Stacks pipeline. Loci formation and polymorphism parameters were modified to match recommendations provided by the developers of Stacks, then a catalog of loci was created using these optimum parameters. For each locus, Single nucleotide polymorphisms were identified against this catalog. To filter these loci, loci were required to be present within 75% of individuals, and had to be present in at least 3 of the designated population groups. Using the R packages *vcfR* (Knaus & Grünwald 2017) and *sambaR* (de Jong et al. 2021), percent missing data and average read depth was calculated for each individual sample. For individual loci, any locus that was missing >25% of the data was removed from the dataset. Finally, within each population designation, individual loci were tested for deviation from Hardy-Weinberg Equilibrium (HWE). This test was implemented in the R package "pegas" (Guol & Thompson 1992; Paradis 2010) using the exact test and 100 Monte Carlo permutations. These filtering steps left 11,323 SNPs for subsequent analyses.

Phylogenetic analyses: Mitochondrial DNA

For the phylogenetic analysis, I created a Bayesian phylogeny of cytb sequences using BEAST v2.6.3 (Bouckaert et al. 2019) The site substitution model, rate of heterogeneity, and proportion of invariant sites were all determined using JModelTest (Posada 2008). I used a log-

normal relaxed clock model and mutation rate of 5.5% (Hope et al. 2010). I also used coalescent constant population priors and Markov Chain Monte Carlo with chain length of 10,000,000, sampling every 5,000 iterations. I used Tracer v1.7.1 to monitor ESS values for parameters (pass requirement of ESS>300; Rambaut 2018). I then used TreeAnnotator v2.6.2 to remove the initial 10% as burn in and annotate the final consensus tree (Rambaut & Drumond 2020). The resulting tree file was viewed using FigTree v1.4.4 (Rambaut 2018). The tree was rooted at the midpoint and bootstrap support values were added to branches to indicate statistical support. Lineages were colored to help visually distinguish independent groups.

Diversity statistics: Mitochondrial DNA

Based on the results from the mitochondrial phylogeny, multiple well supported lineages of least shrews were identified. Using DnaSP v6.12 (Rozas et al. 2017), genetic diversity statistics for each well supported lineage of shrews were calculated. These lineages are generally concordant with geography, including Florida, Southern Texas, and two lineages representing the general rangewide population, one east (representing Virginia and Maryland), and one west of the Mississippi river (representing New Mexico, Texas, Colorado, South Dakota, Kansas, Missouri, Nebraska, and Louisiana). Statistics included pairwise sequence divergence, nucleotide diversity, haplotype diversity, and Tajima's D, the latter used specifically to test for signals of demographic expansion of lineages through changes in effective population size (N_e).

Phylogenetic analyses: Nuclear data

Two different tree-building analyses were performed based on the nuclear dataset. The first method used a maximum likelihood approach and inferred a phylogeny in RAxML v8.2.12

(Stamatakis 2014). In accordance with recommendations in Leaché et al. (2015), an ascertainment bias was applied to the likelihood calculation in the algorithm, which helps to correct for the lack of invariant sites in a SNP-based phylogeny. To accomplish this, the SNP dataset was converted from VCF to phylip format, and potentially invariant sites were filtered out. The maximum likelihood phylogeny was then constructed using 100 bootstrap replicates and the GTR+Gamma substitution model. The Lewis method was applied for ascertainment bias correction (Lewis 2001). The resulting RAxML output was loaded directly into FigTree and visualized. The tree was midpoint rooted, with bootstrap values were placed on the branches to indicate statistical support. Lineages were then colored according to mitochondrial identity of the sample, in order to visualize any discordance of lineage assignments between mitochondrial and nuclear genomes.

The second tree-building approach used SVDquartets which infers relationships among quartets of taxa using a coalescent approach (Chifman & Kubatko 2014). SVDquartets is run through PAUP (Phylogenetics Analysis Using Parsimony; Wilgenbush & Swofford 2003). This approach sampled 1,000,000 quartets and used 200 bootstraps. Similar to the maximum likelihood phylogeny, this tree was viewed using FigTree and rooted at the midpoint.

Diversity statistics: Nuclear DNA

Based on the results of the nuclear phylogenetic analysis, I recovered multiple well supported lineages. These lineages generally aligned with the same relationships recovered from the mitochondrial phylogeny, an exception being that southern New Mexican samples form a statistically well supported lineage. Genetic diversity statistics were obtained from SNP data for populations based on their phylogenetic clustering, including all groups analyzed for mtDNA

data. Additionally, considering the southern New Mexico lineage independently from the west of Mississippi lineage. F_{IS} and nucleotide diversity values were calculated in the Stacks pipeline, *population*s module. Estimates of observed heterozygosity (H_o), expected heterozygosity (H_e), and allelic richness (Ar), were obtained using the R packages "hierfstat" (Goudet 2005), "adegenet" (Jombart 2008), and "dartR" (Gruber et al. 2018), respectively. Pairwise F_{ST}, $G_{ST,}$ and Jost's D values for each represented lineage were also generated in R using the packages "hierfstat" and "mmod" (Winter 2012).

Population Structure

All clustering analyses were performed for the full SNP dataset. To better understand structure within our designated populations, I performed a principal components analysis performed in the R package *dartR* (Gruber et al. 2021). The number of principal components retained was determined by a screeplot developed for the dataset, performed through visual interpretation of the plot, where the optimal number of principal components can be identified by retaining the first value on the plot where the slope of the line drastically flattens, generally resembling an "elbow" in the figure.

Results

Phylogenetic analyses: mtDNA

The Cytb phylogeny (Figure 2-3) shows evidence for multiple well-supported lineages that are coincident with the geographic distribution of these lineages. The largest of these lineages was most broadly distributed through the range of least shrews, and is coincident with the currently recognized distribution of the subspecies *C. p. parvus* but including only those

samples west of the Mississippi River (hereafter West-of-Mississippi lineage). Strong support for additional population structure within this lineage was lacking. This group contains samples from Colorado, South Dakota, Kansas, Nebraska, Texas, Missouri, Louisiana as well as samples from both northern and southern New Mexico. Another well-supported lineage coincident with the eastern distribution of the subspecies *C. p. parvus* contains all samples from east of the Mississippi River (hereafter East-of-Mississippi). Additional well-supported lineages include samples from southernmost Texas that are consistent with the species *C. berlandieri*, grouped with, but divergent from, one sample from the Sierra Madre Occidental of Mexico, the latter of which was consistent with the species *C. soricinus*. Samples of least shrews from Florida form a very distinct well-supported lineage that reflects the distribution of the subspecies *C.p. floridanus* (hereafter Florida lineage). Of interest, samples of the species *C. orophilus* and *C. tropicalis*, both members of the broader *C. parvus* group from Central America, fell out intermediate between Florida and the remainder of the ingroup of putative *C. parvus*, rendering *C. parvus* paraphyletic with respect to *C. orophilus* and *C. tropicalis*.

Diversity statistics: mtDNA

Mitochondrial divergences between lineages support relationships recovered from the Cytb phylogeny (Table 2-1). Florida samples were ~11% divergent from both East-of-Mississippi and West-of-Mississippi samples. Samples of *C. berlandieri* were ~10% divergent from the West-of-Mississippi lineage. Florida and *C. berlandieri* lineages were ~9% divergent from each other. The species *C. orophilus* and *C. tropicalis* were 7-9% divergent from all other lineages in the analysis, supporting their species-level taxonomic distinction. Divergence between West-of-Mississippi and East-of-Mississippi was only 2.9%.

Genetic diversity metrics for Florida reflected high haplotype diversity but low nucleotide diversity, with a negative but non-significant value of Tajima's D. West-of-Mississippi exhibited relatively high haplotype diversity, and nucleotide diversity with a strongly significant negative value for Tajima's D, likely signifying demographic expansion. Sample sizes for other groups were generally too small to infer additional dynamics.

Phylogenetic analyses: Nuclear DNA

The maximum likelihood phylogeny based on SNPs exhibited similar relationships among most of the lineages recovered from the Cytb phylogeny (Figure 2-4). Importantly, Florida formed a well-supported monophyletic lineage that was highly divergent from all other *Cryptotis* samples. Similarly, the sample from Mexico representing the putative species *C. soricinus* was intermediate between Florida and southern Texas, again rendering *C. parvus* paraphyletic. The samples representing *C. berlandieri* again formed a well-supported monophyletic lineage but an additional sample from Kenedy County in southeastern Texas is included in this lineage although it was aligned with the West-of-Mississippi lineage based on the Cytb locus. This latter sample was collected outside the recognized distribution of *C. berlandieri,* and the mito-nuclear discordance exhibited provides strong evidence of sympatry between *C. parvus* and *C. berlandieri* in southern Texas. Similarly, multiple samples collected from the west bank of the Mississippi River in Louisiana were grouped within the West-of-Mississippi lineage according to the Cytb phylogeny but were well supported within the East-of-Mississippi lineage according to SNP data. Again, this indicates mito-nuclear discordance, and episodes of gene flow across the Mississippi River.

The West-of-Mississippi lineage was well-supported as monophyletic and as opposed to the Cytb phylogeny, there is strong internal support for a lineage containing all individuals collected from Chaves County. A single individual from Chaves County was genetically aligned with samples from further north in New Mexico (hereafter High Plains populations), which supports potential connectivity between northern and southern populations.

Diversity Statistics: Nuclear DNA

Genetic diversity metrices indicated that Florida samples showed the highest values across all populations for nucleotide diversity (π) and observed heterozygosity, and one of the lowest values for F_{IS}, the inbreeding statistic, where a value close to zero would indicate panmixia and lack of inbreeding (Table 2-2). These indicate populations in Florida are genetically distinct, but also are not suffering from isolation or inbreeding depression. Chaves County New Mexico samples showed among the lowest values for observed heterozygosity and nucleotide diversity, this combined with evidence from the mitochondrial genetic diversity statistics indicate that Chaves County New Mexico shrews may be undergoing a demographic contraction and experiencing a loss of genetic diversity.

Across the three pairwise genetic differentiation measures and diversity statistics, I saw variable levels of genetic differentiation, but relative levels generally stayed consistent across all three measures (Table 2-3),. Across all analyses, Florida samples were the most genetically distinct of all the populations, and were more divergent from other lineages within *C. parvus* than they were from other recognized species. Likewise, both *C. berlandieri* and *C. soricinus* samples were highly differentiated from *C. parvus*. Both Chaves County New Mexico and East-

of-Mississippi lineages showed low overall differentiation from the broader West-of-Mississippi lineage, but were well differentiated from each other.

Population Structure

In the analysis of internal population structure across the range, the DAPC plot (produced a very clear visual interpretation of the internal population structure across the range Figure 2-6). The Scree Plot indicated that the optimal number of principal components to retain was 4 (Figure 2-7). This corresponds to the number of genetic clusters seen in the DAPC plot, with Florida and southern New Mexico grouping out separately and distinctly from the remainder of the lineages. The other two clusters were comprised of *C. berlandieri* and *C. soricinus,* with the final cluster containing the East and West lineages, including the High Plains of New Mexico.

Discussion

My phylogeographic assessment using a combination of the mitochondrial Cytb gene and approximately 11,300 nuclear SNPs provided high resolution insight into the evolutionary history of the *parvus* group of least shrews through North and Central America. My results confirm the distinction of a number of shrew taxa that are currently considered independent species. In addition, relationships within what has previously been considered as *C. parvus* (sensu stricto) firmly support the presence of additional cryptic diversity that approaches or exceeds species-level differentiation, as well as distinct regional diversity that supports presence of multiple infraspecific lineages, although these are not congruent with the distribution of recognized subspecies.

The *parvus* group within the genus *Cryptotis* contains six distinct species and two subspecies of shrews that range from as far north as Ontario, Canada and as far south as Costa Rica. As the name suggests, this group contains *C. parvus,* as well as five other species that were all originally considered subspecies of *C. parvus* (Whitaker 1974). Subsequent morphological considerations resulted in progressive revision of the taxonomy to raise *C. orophilus* and *C. tropicalis* from Central America to species status (Choate 1970; Hutterer 2005), followed by *C. soricinus*, *C. pueblensis*, and *C. berlandieri* (Woodman 2018). The pattern of previously unrecognized diversity being assigned species-level distinction within this broadly distributed group also raised the possibility that additional cryptic diversity existed within *C. parvus*. Yet none of these relationships until now have actually been considered from an evolutionary perspective until now. Based on all of my genetic analyses, I can qualify strong evolutionary divergence between samples representing all recognized species of the *parvus* group (except for *C. pueblensis* given a lack of available samples). Further, Florida samples that were previously considered as a subspecies of *C. parvus* are the most deeply divergent taxon within the entire *parvus* group. Based on the available evidence, the question then becomes: are the observed levels of divergence indicative of species-level recognition?

From the perspective of the biological species concept (Mayr 2000), I have little evidence of the relative ability of these taxa to hybridize. However, there does appear to have been historic gene flow between West-of-Mississippi and East-of-Mississippi lineages, based on samples from Louisiana. Additionally, a single sample from southern Texas exhibits mito-nuclear discordance between *C. parvus* and *C. berlandieri*. However, my large nuclear SNP dataset firmly allies this specimen with *C. berlandieri*, and it is not possible to infer how recently mitochondrial capture may have occurred. It is well established that ancient hybridization between ancestral

matrilineages may manifest as mito-nuclear discordance in modern species (e.g., Good et al. 2008; Sullivan et al. 2014; Sarver et al. 2021). More pertinent to my question of specific status of among least shrews, inter-specific hybridization is increasingly recognized as a common phenomenon among mammals and may in fact be integral to reinforcing species boundaries at the latest stages of the process of speciation (Abbott et al. 2013). I have instead chosen to assess species based on the unified species concept that relies on the accumulation of multiple lines of evidence, one or more of which may be ambiguous, but together providing a preponderance of evidence for recognizing species (De Queiroz 2007). Here, I have established multiple lines of evidence that representative samples from Florida, concordant with the current subspecies *C. p. floridanus,* represent a distinct taxon, deserving full species designation. The first line of evidence established is divergence between lineages based on the Cytb gene. Florida samples showed very high levels of divergence from all other populations in the analysis in pairwise comparisons. Florida was approximately 11% divergent from the western and eastern lineages, but only about 9% divergent from recognized species *C. berlandieri* and *C. orophilus*. There is strong evidence in mammals that high levels of divergence (>5%) based on mtDNA is a strong proxy for interspecific diversity (Bradley & Baker 2001; 2006). However, mitochondrial DNA is maternally inherited, and therefore incorporating nuclear genomic data provided insight into the bi-parental evolutionary history of this group. Both the nuclear SNP phylogeny and the Cytb phylogeny indicate that Florida samples are reciprocally monophyletic with respect to all other well-supported lineages within the *parvus* group, including the remainder of *C. parvus* (sensu stricto). This observation supports recognition of species according to the phylogenetic species concept (Baum & Donoghue 1995). Finally, all of my clustering analyses provided statistical support for Florida as a unique cluster, with greatest distance from any other cluster within

principle components space. Finally, there was no evidence of gene flow (based on my sampling) between Florida specimens and any other group, although this should be investigated with more extensive sampling through the southeastern United States. Often, additional evidence for taxonomic distinction can be gained from comparative phylogeographic perspectives, where other well-differentiated taxa share the same geographic distributions that are presumed to reflect shared histories of diversification (Bermingham & Moritz 1998). For instance, the deep divergence of Florida least shrews is seen among multiple other species of mammals. These include the meadow vole, *Microtus pennsylvanicus* (Jackson & Cook 2020); short-tailed shrews, genus *Blarina* (Benedict et al. 2006); raccoon, *Procyon lotor* (Trujillo & Hoffman 2016); the white tailed deer, *Odocoileus virginianus* (Combe et al. in revision); and the Florida panther, *Puma concolor* (Ellsworth et al. 1994). Based on all available lines of evidence, I recognize least shrews within Florida as a distinct species, *C. floridanus*. However, future work including analysis of additional sampling through the Southeast US, coupled with a comprehensive morphometric examination and ecological assessment would help to determine the true distribution and ecology of this taxon, and further support species designation.

With the same set of taxonomic considerations, the remainder of least shrew samples distributed across the United States are representative of a single species *Cryptotis parvus* (sensu stricto). This group contains two reciprocially monophyletic lineages geographically concordant with divergence across the Mississippi River. These lineages cluster as distinct based on both nuclear and mitochondrial phylogenies. But as already noted, there is evidence of historic gene flow among samples collected from the eastern bank of the Mississippi River that clustered with the West-of-Mississippi lineage based on Cytb data but with the East-of-Mississippi lineage based on nuclear SNPs. The Mississippi river is recognized as a barrier to dispersal and gene

flow among terrestrial species, both at the present time and through the last, and previous, glacial climate phases, when this corridor represented the outflow of glacial ice sheets covering much of northern North America (Hope et al. 2012). At these times, the Mississippi outflow was likely a much more formidable barrier than even the current river. As such, gene flow across this mid-continental transition zone has likely been low through time, although may be ongoing, and greater sampling, particularly through eastern United States would qualify the extent of contemporary genetic mixing.

Within the West-of-Mississippi lineage, there exists additional distinct genetic diversity. All samples from Chaves County in southeastern New Mexico form a well-supported lineage as well as a statistically distinct cluster based on the SNP dataset although not based on the Cytb dataset. These samples were considered to potentially represent a relictual population of *C. berlandieri* (Hafner & Schuster 1996). My data firmly reject this hypothesis (as discussed in greater detail in Chapter Three). Instead, these shrews represent a distinct lineage within *C. parvus*. A lack of phylogenetic resolution based on the Cytb gene, but well resolved distinction based on nuclear SNPs is generally uncommon within phylogeographic studies, given that the mitochondrial gene is haploid and as such tends to evolve much more rapidly than the nuclear genome, leading to relatively more pronounced differentiation between taxa (Brown et al. 1979). However, the Cytb data represent only a single locus, and at shallow evolutionary timescales, such loci are often recovered as incompletely sorted with respect to lineage relationships (Ballard & Whitlock 2004). Instead, use of thousands of putatively unlinked SNP loci reduces the influence of incomplete lineage sorting and helps to resolve even shallow evolutionary divergence among populations that have been obscured by earlier studies focused on mtDNA (Scornavacca & Gautier 2017). As such, with phylogenomic studies of wildlife becoming

increasingly common, we can expect nuclear differentiation coupled with matrilineal ambiguity to become more frequent (e.g., Wiens 2021). Ultimately, the lineage associated with Chaves County New Mexico is genetically distinct, and may, like both the remainder of West-of-Mississippi lineage and East-of-Mississippi lineage, warrant recognition as distinct subspecies.

Subspecies distinction is commonly used when taxa are geographically isolated, morphologically distinct, and have distinct evolutionary potential (Sackett et al. 2014). These three taxa are mostly geographically isolated, even though signals of gene flow persist. And, based on my results, they also possess distinct evolutionary potential, considering they form distinct clusters based on nuclear population clustering analyses. The issue with assigning these lineages subspecies identities is two-fold. First, all three of these groups are currently recognized as a single subspecies *C. p. parvus.* Second, electing subspecific nomenclature requires a number of criteria to be met. Subspecies level distinctions require some level of consistent morphological distinction between taxa. Genetic and genomic data alone are not sufficient to identify and name novel subspecies without the incorporation of morphometric analyses (reviewed by Hope and Frey 2021). Subspecies should also be geographically discrete (Patten 2010). Although I can qualify the latter requirement, without more detailed morphological analyses of these populations, designation of additional subspecies is not reasonable at this time. But, genomic data at the infraspecific level *is* capable of designating evolutionarily significant units, as well as finer-scale adaptive or management units (Funk et al. 2012; see Chapter Three). I therefore recognize three distinct evolutionarily significant units within *C. parvus* that make up the entirety of its range in the central and eastern United States, which with further morphological analysis, may justify novel subspecies level distinction.

Biogeographic history of least shrews

Evolutionary divergence within species is most often associated with major episodes of environmental change that impacted the distribution, diversity, and demography within discrete (but potentially interacting) populations through time (Hewitt 2004). Results of my phylogeographic assessment of the *parvus* group of shrews indicate that there are currently multiple distinct evolutionary lineages of least shrews that are associated with discrete geographic ranges, even though their combined distribution is seemingly continuous across much of North America. Specifically, in addition to distinct Florida lineage in southeastern United States, there are three distinct lineages: Chaves County, New Mexico, everything else west of the Mississippi River, and everything east of the Mississippi River. These three lineages likely diversified from each other due to allopatric isolation during the last glacial maximum, where each lineage would have persisted within a different glacial refugium. Given their present distributions, coupled with known phylogeographic breaks (barriers to dispersal and gene flow), and based on evidence from other species, I can postulate that *C. floridanus* has likely persisted in the southeastern United States associated with a Florida refugium across multiple Pleistocene glacial cycles. During the last glacial period (~90-18 thousand years ago), the East-of-Mississippi lineage likely also persisted and diverged in a refugium in the Southeast US around Louisiana, Mississippi, and Alabama, while the West-of-Mississippi lineage likely persisted in a refugium located in south-central Texas. The Chaves County New Mexico lineage was likely associated with a refugium located throughout southern New Mexico, much larger than the current distribution of *C. parvus* in the state currently. This is evident based on fossil records of *C. parvus* recovered from localities in southern New Mexico well outside of the current distribution of this taxon (Harris 1973). As such, "refugia" should not necessarily be considered

as confined and extreme hold-outs for biodiversity, but simply as alternative distributions from the present time that emphasized geographic allopatry (isolation) between persisting populations (Stewart & Dalen 2008). These refugia may not have been small in area, but effective for diversification along unique evolutionary trajectories. As an example, genetic diversity and demographic statistics for the Chaves County lineage reflect high diversity coupled with recent signals of population contraction, indicating that refugial isolation was likely coupled with a much larger distribution than at present and more extensive availability of mesic habitats through the southwestern United States, followed by extreme population contraction into isolated Cienega wetlands only recently. Conversely, populations within the other lineages of *C. parvus* have greatly expanded their range since the last glacial maximum to much more widespread current distributions as grasslands and mesic woodlands across the central and eastern United States expanded as a result of climate warming. These dynamics are similarly supported by combined diversity and demographic statistics.

Conclusions and Future Directions

This work has supported and reinforced that the use of genetics for recognizing distinct diversity can help to uncover cryptic species that reflect much more complicated evolutionary histories than previously presumed. In addition, it is then possible to more effectively diagnose units of potential conservation concern, or minimally, the most appropriate focal taxa for applied management. Going forward, I recommend increased sampling for *C. parvus* across its range to expand the available specimen pool for future analyses and associated tissues available for conservation and genomic research. This sampling is especially important east of the Mississippi River, which has been identified to be genetically distinct (Hutchinson

2010), to ensure there are ample samples available for both further genetic analyses, but also full specimens available for morphometric analyses. Furthermore, I recommend that Florida populations of *C. parvus* are recognized with species-level distinction, pending an official taxonomic revision. My results also highlight the value of genomic tools that can resolve cryptic diversity and pinpoint geographically distinct taxa with distinct histories and unique local adaptations.

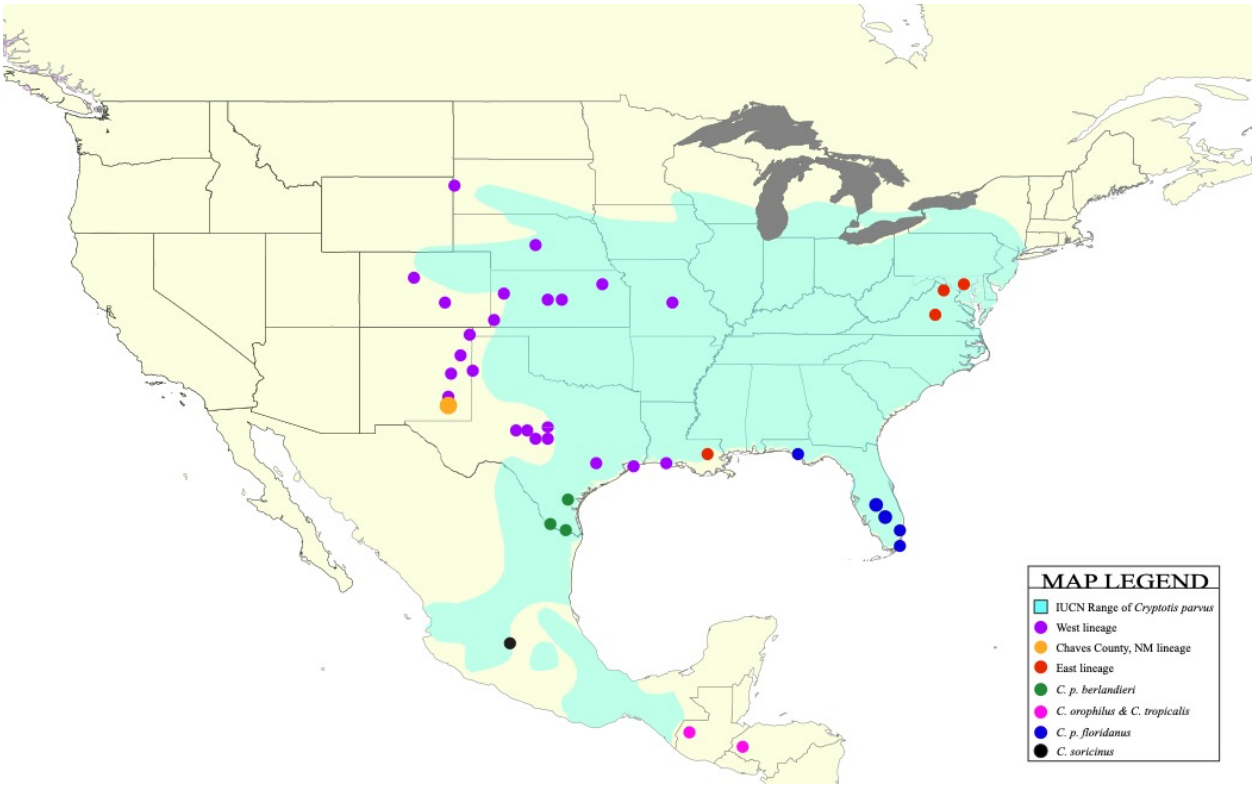

Figure 2-1 Sampling Distribution

Map showing distribution of *C. parvus* samples included in this analysis. Samples are colored by nuclear lineage membership, and if not included in nuclear analysis, are colored according to mitochondrial identity. The blue represents the distribution of *C. parvus* designated by the IUCN.

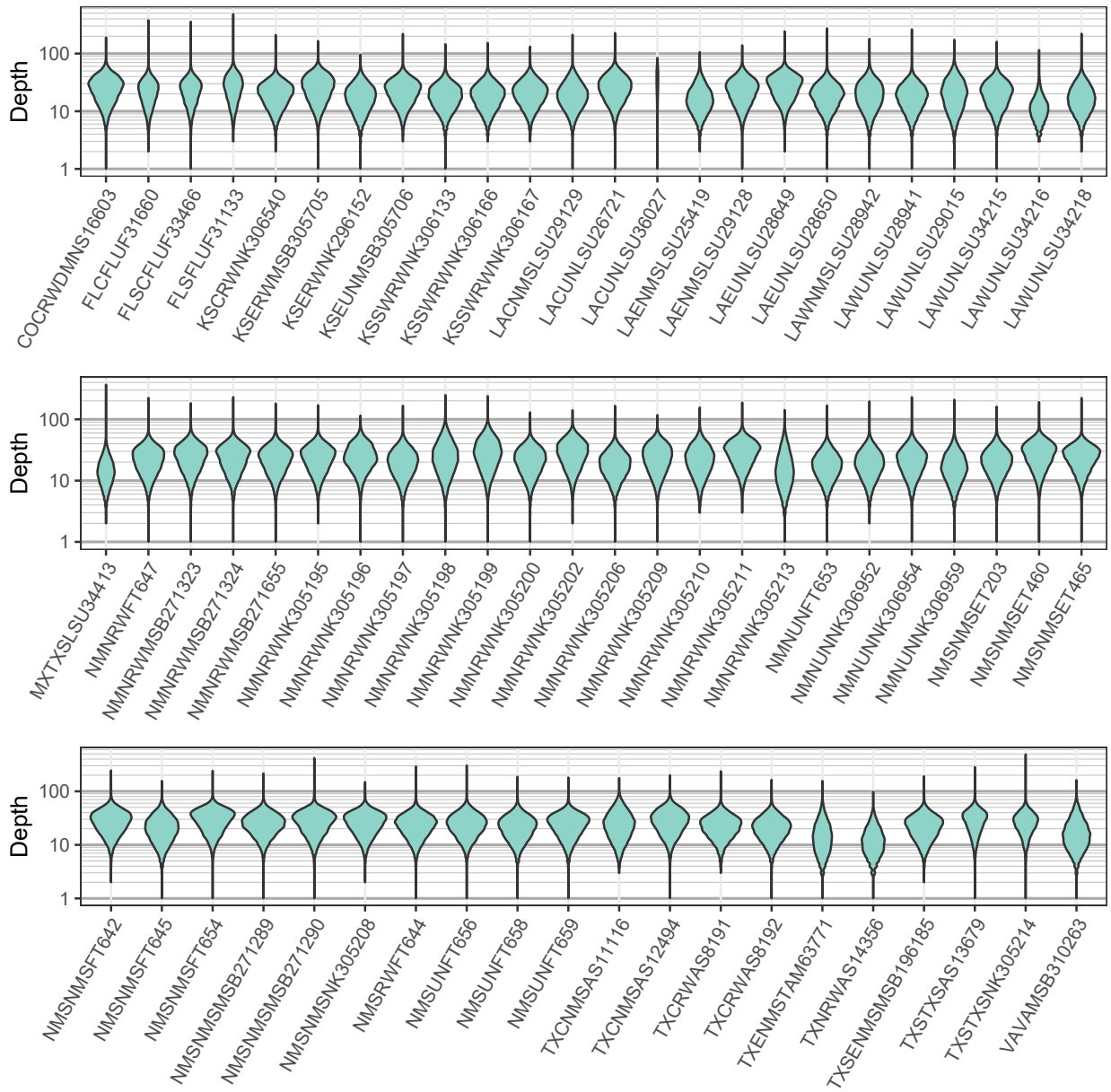

Figure 2-2 Sample Read Depth

Distribution of read depth per individual sample after quality filtering steps for samples included in the nuclear SNP dataset. The first two letters indicate state of origin (CO=Colorado, FL=Florida, NM= New Mexico, etc.) and the remaining letters indicate geographic portion of the state (N=North, S=South, W=West, E=East), mitochondrial lineage identity (RW=Rangewide, FL=Florida, etc.) and the final code represents the unique museum collection code and sample ID number.

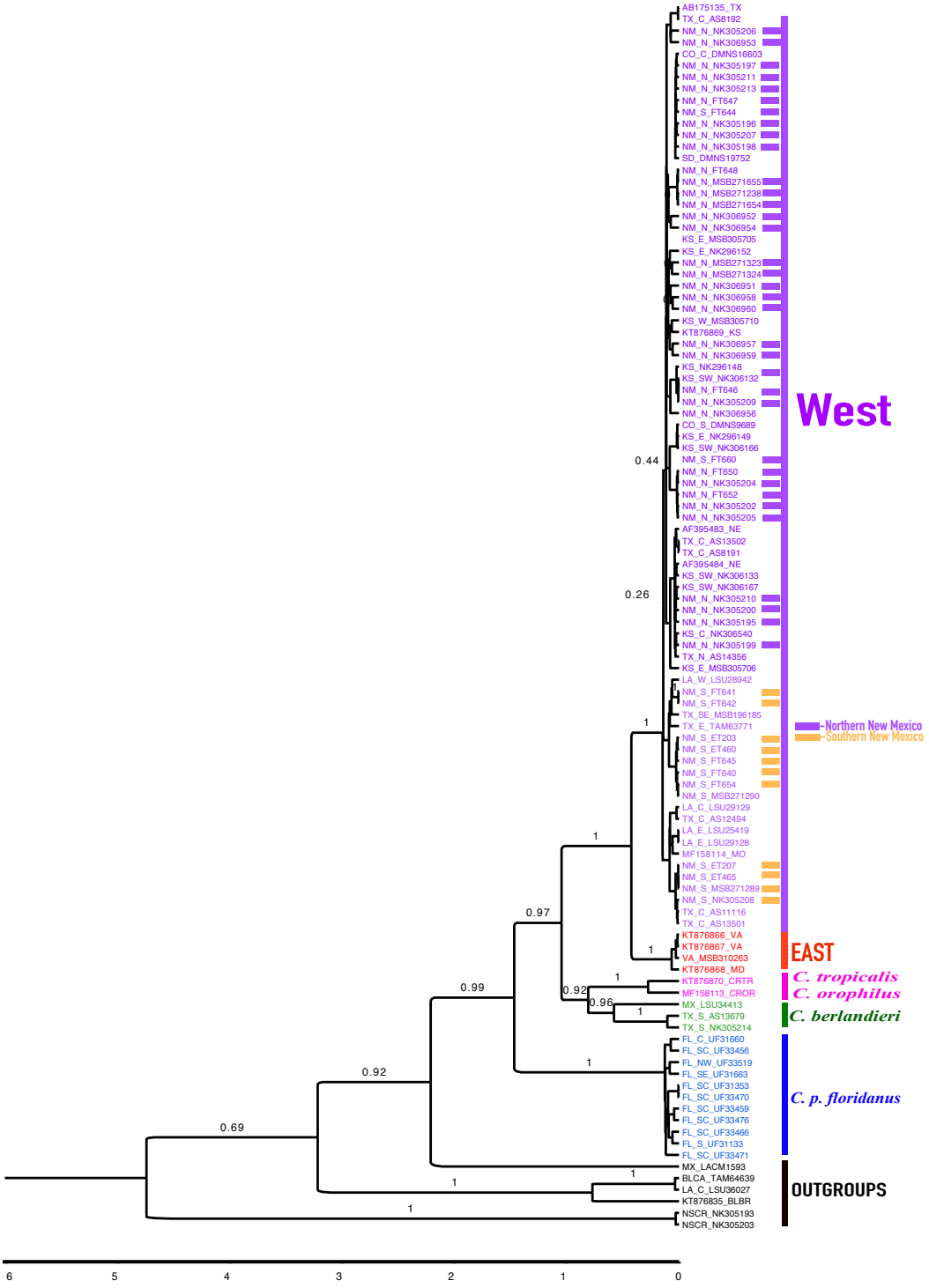

Figure 2-3 Cytochrome-B Phylogeny

Phylogeny using the cytochrome-b gene, created using BEAST2. Branches are labeled with posterior probability values. Posterior probability values < 0.70 were removed for minor lineages.

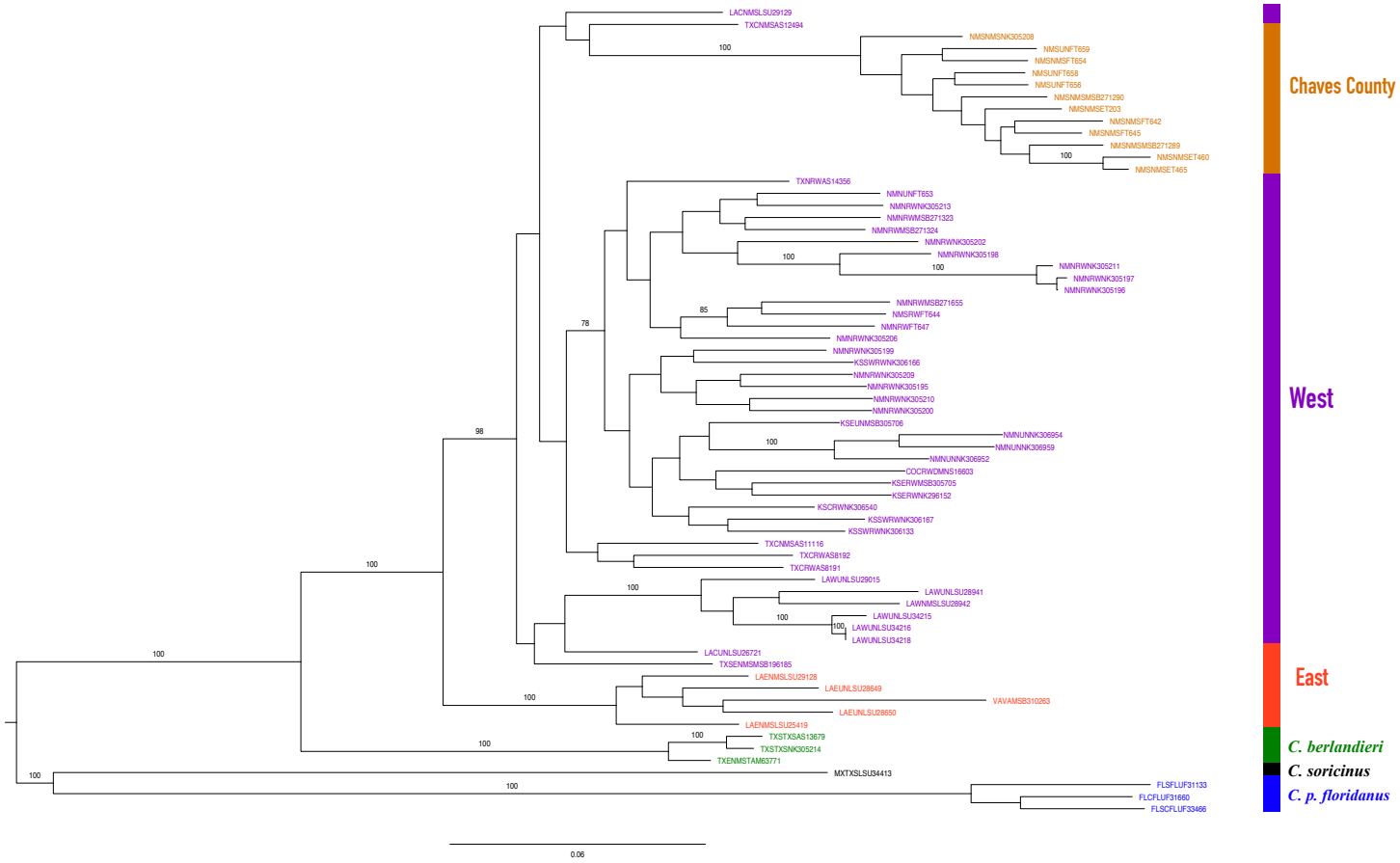

Figure 2-4 Nuclear SNP RAxML Phylogeny

Phylogeny produced for the nuclear SNP dataset using a maximum likelihood approach. Phylogeny was created using RAxML program and bootstrap values were placed on branches to indicate statistical support for the following lineage. Bootstrap values < 70 were removed from the branches.

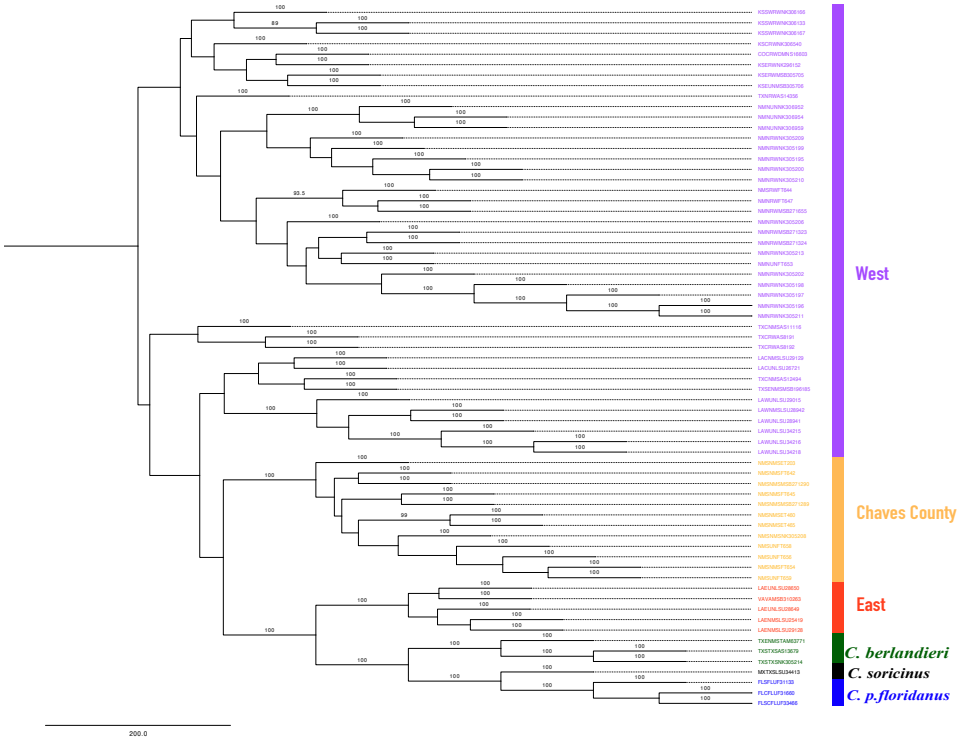

Figure 2-5 Nuclear SNP SVDQuartet Phylogeny

Phylogeny for nuclear SNP data set produced using SVDquartets in PAUP. Branch labels indicate bootstrap support for the following lineages. Bootstrap values < 70 were removed from the branch labels.

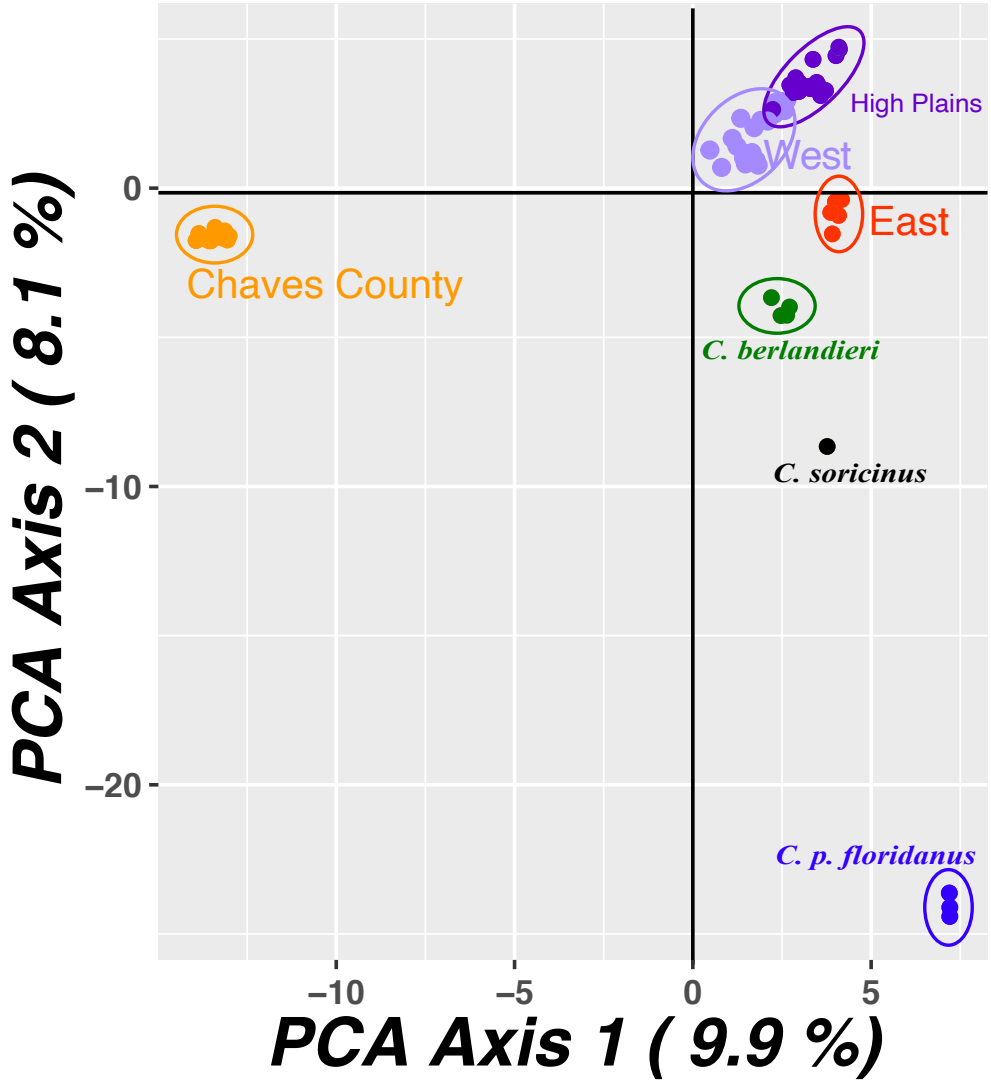

Figure 2-6 DAPC Plot- Nuclear SNP Data

DAPC plot based on Nuclear SNP data. Samples that cluster closer together are more genetically similar whilst samples that are more distant are more genetically distinct.

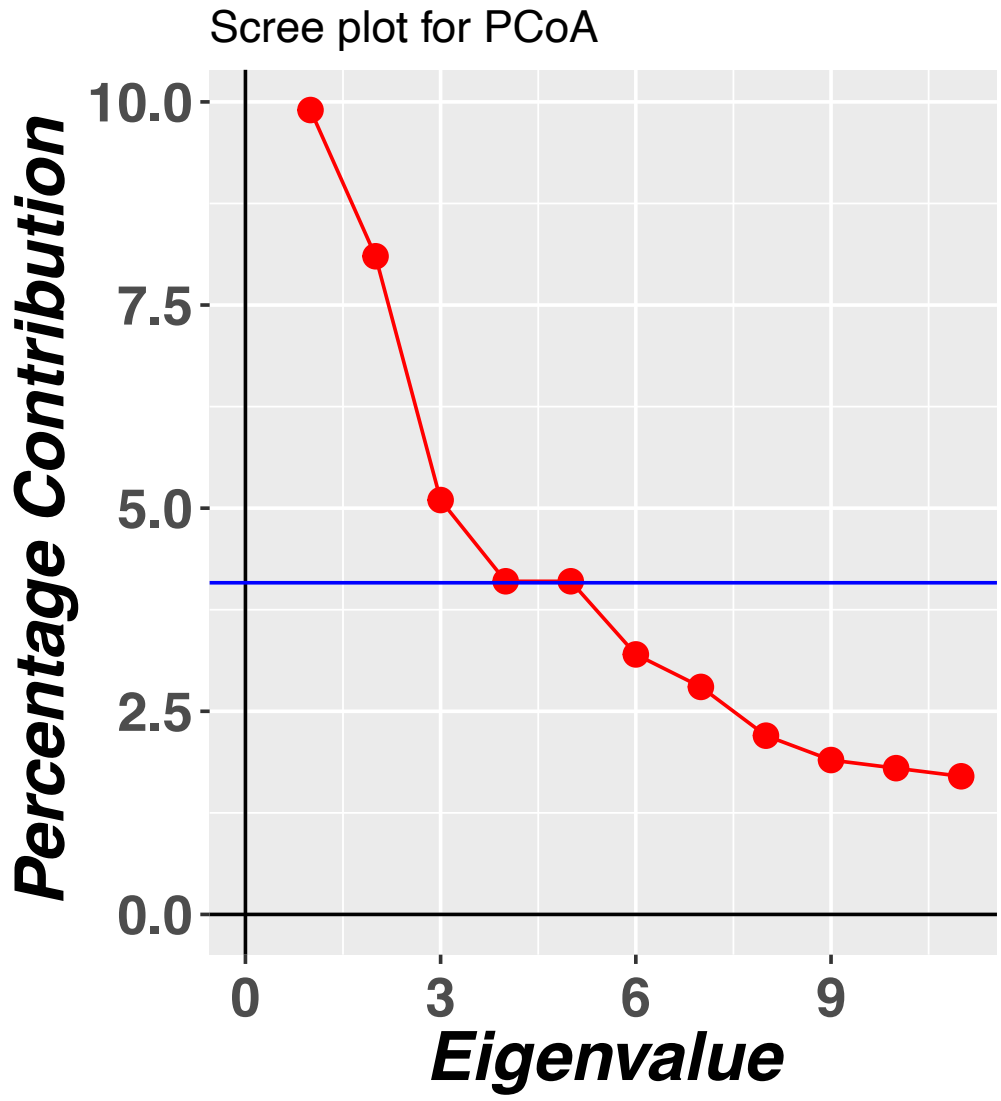

Figure 2-7 Scree Plot- Nuclear SNP Dataset

Screeplot showing percent contribution of principal components retained for explaining the variance on the DAPC plot. Only the first four principal components were retained.

Table 2-1 Cytochrome-B Genetic Differentiation:
Genetic differentiation measures based on the mitochondrial cytochrome-b dataset. **(A)** pairwise population divergences; (B) genetic diversity metrices. Bolded values are statistically significant

(A) Population Divergence

	Florida	*C. berlandieri*	West	East	*C. orophilus*	*C. tropicalis*	*C. goldmani*
Florida	-	0.099	0.107	0.106	0.0982	0.0967	0.127
C. berlandieri	-	-	0.092	0.092	0.0732	0.0745	0.132
West	-	-	-	0.029	0.0912	0.0875	0.130
East	-	-	-	-	0.0856	0.0813	0.114
C. orophilus	-	-	-	-	-	0.0284	0.115
C. tropicalis						-	0.114
C. goldmani	-	-	-	-	-	-	-

(B) **Genetic Diversity Metrics**

	N	*Hd*	π	Tajima's D
Florida	11	0.982	0.00444	-1.41686
C. berlandieri	2	1.000	0.01094	-
East	4	0.500	0.00205	-
West	81	0.833	0.0282	**-2.71555**

Table 2-2 SNP Genetic Diversity Statistics

Genetic diversity statistics based on the whole SNP dataset

	Observed Het.	Expected Het.	Allelic Richness	F_{IS}	π
Florida	0.10151	0.10837	1.10585	0.05244	0.13004
C. berlandieri	0.06145	0.05657	1.057237	0.01178	0.06789
Chaves County	0.06951	0.07948	1.070939	0.03868	0.0832
West	0.09325	0.12192	1.116046	0.13257	0.12354
East	0.09121	0.10348	1.102592	0.05536	0.1163
C. soricinus	0.0343	0.01715	1.026495	0	0.0343

Table 2-3 SNP Pairwise Genetic Differentiation

Genetic differentiation metrices based on the nuclear SNP dataset. **(A)** Pairwise Fst values, all values are statistically significant; **(B)** Pairwise Gst values; **(C)** Pairwise Jost's D values.

(A)	**Florida**	***C. berlandieri***	**West**	**Chaves County**	**East**	***C. soricinus***
Florida	-	0.6226	0.5303	0.6188	0.5425	0.5739
C. berlandieri		-	0.3652	0.5288	0.4308	0.7172
West			-	0.1965	0.1636	0.5195
Chaves County				-	0.3435	0.6876
East					-	0.5562
C. soricinus						-

(B)	**Florida**	***C. berlandieri***	**West**	**Chaves County**	**East**	***C. soricinus***
Florida	-	0.4649	0.3697	0.4560	0.3857	0.4947
C. berlandieri		-	0.2323	0.1120	0.2869	0.5987
West			-	0.3654	0.0959	0.4113
Chaves County				-	0.2139	0.5637
East					-	0.4587
C. soricinus						-

(C)	**Florida**	***C. berlandieri***	**West**	**Chaves County**	**East**	***C. soricinus***
Florida	-	0.1451	0.1326	0.1513	0.1346	0.1477
C. berlandieri		-	0.0571	0.0769	0.0683	0.1419
West			-	0.0261	0.0259	0.1228
Chaves County				-	0.0512	0.1468
East					-	0.1310
C. soricinus						-

Chapter 3 - Conservation Genomics of *Cryptotis parvus* in the High Plains and Pecos drainage basin of New Mexico

Introduction

Use of genetic data, and particularly of multi-locus genomic data, for conservation research and applied management of biodiversity has increased dramatically over recent decades, leading to maturation of the field of conservation genetics (Allendorf & Luikart 2007) and rapid development of the field of population genomics (Luikart et al. 2018). But, as the generation of large-scale genomic data has become increasingly more accessible and less expensive, the traditional field of conservation genetics has faced difficulty keeping pace with the rapid dissemination of big data. Incorporating genetic data into a conservation framework originally relied on fewer loci, ranging from crude allozyme (protein variants) data, to a single digit number of mitochondrial loci, to a few dozen microsatellite loci (Frankham et al. 2001). But, the advent of high throughput DNA sequencing and genotyping technology has vastly expanded the potential resolution of evolutionary relationships among taxa and inference of complex evolutionary processes that reflect real-world population trajectories among wildlife. The emerging field of "conservation genomics" is now explicitly being developed to use modern molecular methods and big data to inform decisions that affect threatened and endangered species, as well as infraspecific levels of conservation concern. A genomic approach allows for more accurate estimation of values for genetic differentiation, genetic diversity, relative inbreeding coefficients, effective population sizes and population trends through time, as well as levels of gene flow and investigation of prevailing evolutionary processes such as selection and genetic drift (DeSalle & Amato 2009).

It has been argued that the biggest advantages of using genomic data for biodiversity conservation is to increase understanding of past demographic changes that have influenced remaining genomic diversity, coupled with knowledge of the nature of local and contemporary evolutionary processes (e.g., adaptative pressures, or relative isolation) leading to the observed genomic trends (Allendorf et al. 2010). Earlier genetic methods relying on only a few loci are not able to disentangle the complexity inherent in population histories across extended time frames; for example, diagnosing population bottlenecks versus selective sweeps (McMahon et al. 2014). It has been shown that genomic data can be useful in providing insight into demographic histories of endangered species, even with a small number of samples (Zhan et al. 2013). The first integral step of applied conservation is diagnosis of conservation units. Conservation units are defined as "population units identified within species (or consideration of whole species) that are used to guide management and conservation efforts" (Funk et al. 2012; Allendorf & Luikart 2007; Fraser & Bernatchez 2001). From an ecological perspective, the three most commonly defined conservation units are whole species, subspecies, or distinct population segments, all of which can be considered for listing in the United States under the Endangered Species Act. From an evolutionary perspective, conservation units may again consider whole species as defined by multiple operational species concepts, most often including the biological species concept, phylogenetic species concept or the unified species concept (Fraser & Bernatchez 2001). However, from an infraspecific perspective, genetic/genomic analyses generally consider evolutionarily significant units (ESU). There have been numerous attempts to define ESU's, but generally, these units can be defined as a population, or metapopulations, that warrants separate management or conservation priority due to genetic distinctiveness (Funk et al. 2012; Crandall et al. 2000; Mortiz 1994; Ryder 1986;). Under this ESU umbrella, typically existing at a smaller

spatial scale or more narrow taxonomic focus are two other conservation-oriented units of analysis, including management units (MU) and adaptive units (AU). Management units are defined as populations that are demographically independent from a genetic perspective (Moritz 1994). Defining management units is good for short-term conservation initiatives, (e.g., monitoring genetic diversity of a population associated with critical habitat) and focuses on the fact that populations are demographically and evolutionarily isolated from populations elsewhere (Palsbøll et al. 2007). Funk et al. (2012) recommend delineating MUs by performing phylogenetic and demographic analyses on only neutrally-evolving loci. This allows population structure to be identified in the absence of adaptive variation. In this situation, focal populations will be recovered as genetically differentiated from one another as a consequence of neutral evolutionary processes such as genetic drift, which is strongly reflective of isolation between focal populations. Multiple such isolated populations may be experiencing similar environments (selective pressures) but are nonetheless on their own demographic and evolutionary trajectories and deserving of independent management. A separate but equally important goal of conservation genomics has been to identify differences in local adaptation between populations. This has only recently become feasible due to big data from genome-wide sequencing. Generally speaking, parts of the genome that code for gene functions are a relatively small proportion of the total genome. Within these coding regions, only a small proportion of genes are experiencing adaptive selection that result in unique population attributes. By filtering these outlier (hereafter non-neutral) loci from the complete dataset we can investigate divergence between populations as a function of selection, and populations exhibiting such evolutionary change are considered AUs (Funk et al. 2012; Barbosa et al. 2018). Identification of AUs also reflect important geographic regions or habitats that are promoting local adaptations (Hohenlohe 2010; Reznick &

Endler 1982). In this way, we can consider the evolutionary uniqueness of a focal population or set of populations from a distinct lineage perspective (ESUs), a distinct genetic demography perspective (MUs) or from the perspective of distinct selective regimes (AUs).

The least shrew (*Cryptotis parvus)* is a prime example of a wide-ranging taxon occurring across multiple ecosystems through its broad range. The distribution of *C. parvus* extends from the northeastern United States and southern Ontario, Canada, across most of the central and eastern United States and the western periphery of this species occurs just east of the Rocky Mountains. This distribution spans habitat types ranging from deciduous and coniferous forests as far north as the Great Lakes, coastal marshes across Florida and the southern United States, tall and short grass prairie habitats across the central United States, and even semi-arid desert grasslands at the western extent of their range. Along the westernmost periphery, populations appear to be discrete and often, but not always, associated with isolated ciénega wetlands, surrounded by more arid habitats. These western peripheral populations are potentially of conservation concern due to their apparent isolated nature, but there is still scant knowledge of the true distribution of *C. parvus* through this region. Multiple populations have been identified along the western periphery of *C. parvus* since the 1980s. As an example, populations occurring in New Mexico, Colorado, Wyoming, and South Dakota were not included in the 2008 IUCN range map for *C. parvus*, and new populations or unique records of locality are still being recorded.

The first record of *C. parvus* in New Mexico is a specimen collected in 1961, from Eddy County, New Mexico. In the 1980s, three more populations were identified within New Mexico, one at Tucumcari Lake in Quay County, one at Grulla National Wildlife Refuge (NWR) in Roosevelt County, and one at Bitter Lake NWR in Chaves County. Recent field surveys by the

New Mexico Department of Game and Fish (NMDGF) have expanded the known distribution of these shrews to include additional localities in the northern High Plains, Tucumcari drainage basin and the Pecos drainage basin and associated tributaries. Limited information about current distribution and population demographics due to data deficiency, has caused the taxon to be listed as threatened in New Mexico, and has been identified as a species of greatest conservation need (SGCN). These threatened New Mexico populations of *C. parvus* make a good candidate system for implementing a framework for diagnosing conservation units, as described in Funk et al. (2012) and implemented by Barbosa et al. (2018). Although *C. parvus* is seemingly widespread across northeastern New Mexico, southernmost populations are most strictly isolated in the vicinity of Bitter Lake NWR. This provides an ideal system to examine the roles of neutral versus adaptive divergence between northern and southern populations and explore their uniqueness from a conservation standpoint.

Preliminary work done by Hafner and Shuster (1996) provided the first conservation assessment of peripheral populations of *C. parvus* in eastern New Mexico. They focused on, at the time, the only three known populations persisting in the state. These populations occur at Tucumcari Lake and Grulla NWR, in the northern part of the state, while the third population was found along the Pecos river basin at Bitter Lake NWR in southern New Mexico. Their work combined morphometric and allozyme analyses to understand the origin of, and relationships among, the three New Mexican populations of *C. parvus.* They measured palatal and post-palatal lengths, as these measurements were determined to distinguishable between the two recognized taxa of least shrews geographically closest to eastern New Mexico (*C. parvus* and *C. berlandieri*). For the allozyme analysis, they assayed 24 loci using horizontal starch-gel electrophoresis. Using distance-based algorithms, they created independent phylogenies based on

morphometric and allozyme relationships among the three New Mexican populations, and including a reference population from northern Texas (Wichita Falls). All data supported a sister relationship between samples from the northern Tucumcari and Salt lake localities and samples from Wichita Falls in Texas. Similarly, samples from Bitter Lake NWR were consistently divergent from all other least shrews sampled, despite geographic proximity to Salt and Tucumcari Lakes. Based on these results, Hafner and Shuster (1996) hypothesized that populations at Tucumcari and Salt lakes represent populations experiencing recent westward expansion, aided by an increase in agricultural drainage networks providing a westward colonization route. Furthermore, morphological assessments of populations from Bitter Lake indicated they most closely resembled the species *C. berlandieri* of southern Texas due to their larger body size and relatively shorter palatal measurements. Based on this morphological resemblance, it was hypothesized that populations at Bitter Lake NWR represent a Wisconsinan glacial period relict population of *C. berlandieri* that persisted throughout the Pecos drainage basin, but as grasslands underwent aridification, populations became fragmented along the Pecos and Rio Grande drainage basins, leaving Bitter Lake populations highly isolated. But, preliminary genetic results were based on allozymes, which are considered only modestly informative of relative relationships among taxa. In this chapter, I tested these two hypotheses by including a robust genomic dataset including thousands of loci, to help better understand the origins, demographic trends, and conservation status of New Mexican populations of *C. parvus*.

My first research chapter (Chapter Two) focused on phylogenomics and phylogeography of the *C. parvus* group. I established that what were previously considered *C. parvus* (sensu stricto) actually constitutes multiple taxa that warrant species-level recognition. In addition, I established that on the western periphery of *C. parvus*, populations within New Mexico

constitute genetically distinct groups. For my final research chapter, I make use of both mitochondrial cytochrome-B (Cytb) and nuclear ddRADseq datasets by performing additional analyses to investigate in finer detail the evolutionary history of western peripheral populations of *C. parvus* through New Mexico. Here I consider the demographic history, relative isolation, and adaptive differentiation of these populations to diagnose units of conservation from both evolutionary and potentially ecological perspectives. The ultimate goal of this work from an applied management perspective is to increase understanding of the predominant evolutionary processes of change that influence these peripheral populations, that will in turn reflect population trajectories and the relative importance of different potential conservation measures to account for the unique localized diversity across the periphery of this species' range.

Methods

Sampling

Shrew samples analyzed in this chapter are the same as the samples presented in Chapter two. Detailed methods of sample acquisition, DNA extraction, amplification, and sequencing are likewise presented in full detail in Chapter two. Here I analyzed the cytb and Nuclear SNP at finer resolution in order to focus explicitly on the population genomics of peripheral populations of least shrews in Eastern New Mexico.

Detection of loci under selection

In order to identify different conservation units in a population as outlined in Funk et al. (2012) and implemented in Barbosa et al. (2018) the dataset containing Single Nucleotide Polymorphisms (SNPs) needed to be separated into two datasets: one containing only neutrally evolving loci, and one that contains only outlier loci, or non-neutrally evolving SNP's. To

identify Evolutionary Significant Units (ESU's), which are the most encompassing of the intraspecific conservation units, it is recommended that both neutral and non-neutral loci are used, because ESU's can be shaped by both neutral processes (e.g., as a result of population isolation and genetic drift in the absence of gene flow) and adaptive processes (e.g., divergent selection), or a combination of both of these processes. Neutrally evolving loci are used to identify management units (MUs). MUs are defined as populations that are demographically independent, with restricted gene flow between them. Conversely, non-neutral loci can help identify populations that are locally adapted to their environment, constituting adaptive units (AUs). As such, depending on the evolutionary processes acting on a given population, that population may be considered from the perspective of ESU, MU, or AU, or some combination of these units. For instance, if two populations for genetically discrete clusters when considering neutral loci but not when considering non-neutral loci, these could be considered as MUs given that they are most likely experiencing genetic drift, a neutral evolutionary process, as opposed to natural selection. In this example, populations have diverged as a result of isolation rather than local selective pressures and would warrant population-specific management considering independent demographics. Similarly, we might infer that these populations are genetically unique and geographically independent but are still experiencing similar environmental conditions, considering there are no signals of adaptive divergence. This process of analyzing neutral versus non-neutral loci may therefore group populations differently, depending on regionally unique evolutionary forces. For the purpose of this study, outlier loci refer to loci that are not evolving under a neutral model, and therefore will be referred to as non-neutral loci. To subset the SNP data, I employed two approaches that identify and separate non-neutral from neutral loci. The first approach used was Bayescan v2.54 (Foll & Gaggiotti 2008), where 20 pilot

runs of 5,000 generations were performed. The starting prior odds were run at values of 10, 100, 1000, and 10,000 and then visualized to determine best starting odds. Bayescan outputs were visualized using the R package "Radiator" (Gosselin 2020).

The second approach used was the R package "PCadapt" (Luu et al. 2017). Rather than a Bayesian approach, PCadapt relies on principal component analyses to identify non-neutral loci. To determine the relevant number of principal components to be kept, scree and PCA plots were examined. These plots indicated the optimal number of components was four, considering all SNPs. Non-neutral loci were then identified and filtered out. Results from both Bayescan and PCadapt were compared, and all loci identified by both processes were filtered from the whole dataset. To minimize risk of incorporating a false positive, loci detected by only one approach were removed. Non-neutral loci were filtered out and used to create a non-neutral dataset using VCFtools (Danachek et al. 2011).

Population Structure

All clustering analyses were performed for the full SNP dataset and separately for neutral and non-neutral datasets, each assigned the same parameters as described. To better understand structure within our designated populations, two different approaches were employed. The first approach was a principal components analysis (PCA) performed in the R package *dartR* (Gruber et al. 2021). The number of principal components retained was determined by a screeplot developed for the dataset, performed through visual interpretation of the plot, where the optimal number of principal components can be identified by retaining the first value on the plot where the slope of the line drastically flattens, generally resembling an "elbow" in the figure.

The second approach to analyze population structure was performed in Structure v2.3.4 (Pritchard et al. 2000), exactly as presented in Chapter 2. I tested for a range of K values, from two to 20, and ran ten iterations for each K value. Running iterations for each of these K values added statistical support to determining the best K value, or the most likely number of clusters within a set of populations. I ran the analysis for 1,000,000 Markov Chain Monte Carlo (MCMC) iterations with a prior burnin of 500,000 MCMC iterations. To decrease computation time, this analysis was parallelized using EasyParallel (Zhao et al. 2020). To determine the most likely value of K for each run, the Evanno method (Evanno et al. 2005) was used in the program Structure Harvester (Earl & vonHoldt 2012). The resulting plots generated by the Structure analysis were viewed using the program Structure Plot (Ramasamy et al. 2014). The resulting population structure was color coded by nuclear lineage identity for continuity.

The role of speciation and drift on population structure

To understand the role that adaptation and genetic drift played in the internal population structure of the New Mexican taxa, I analyzed species structure for both the neutral and non-neutral datasets. These analyses will allow me to determine what is the main driver of the divergence between these populations. If genetic drift is the main driver of the divergence between the northern and southern taxa, then we would expect to see statistically supported population structure when analyzing only neutrally evolving loci. On the contrary, if the driver of divergence is natural selection, or local adaptation, the expectation would be to see stronger signal of population structure in the non-neutral loci dataset. If signals of population structure exist in both neutral and non-neutral loci, that would provide evidence that a combination of both drift and selection contributed to the divergence between these taxa. To quantify these

contributions, genetic diversity metrices were measured for the full SNP dataset and separately for neutral loci. Considering that selection may significantly skew estimates of genetic diversity and differentiation, the non-neutral SNP data were not analyzed for diversity statistics. For the total dataset, F_{IS} and nucleotide diversity were calculated in the Stacks pipeline population module, as presented in Chapter 2. Likewise, estimates of observed heterozygosity (H_o), expected heterozygosity (H_e), and allelic richness (Ar), estimates were obtained using the R packages "hierfstat" (Goudet 2005), "adegenet" (Jombart 2008), and "dartR" (Gruber et al. 2018), respectively. Pairwise F_{ST}, $G_{ST,}$ and Jost's D values were also generated in R using the packages "hierfstat" and "mmod" (Winter 2012). For the neutral loci dataset, F_{IS} values were estimated using the R package "hierfstat".

Demographic histories based on Effective Population Size

To better understand the evolutionary history of least shrew populations within New Mexico, I investigated how effective population size (N_e) across regional groups has changed over time. Understanding these fluctuations may provide evidence of demographic changes that support hypotheses that northern samples represent a recent expansion and southern samples represent relictual populations. Stairway plots that visualize population size changes through evolutionary time were generated using the Stairway Plot v2.0 (Liu & Fu 2020). This approach is beneficial when working with a non-model system because it uses a folded SNP frequency spectrum (SFS), which does not have to take ancestral allelic conditions into consideration, and does not assume selectively neutral markers. To generate the SNP frequency spectrum, BAM outputs (full locus sequence data) from the Stacks *process_radtags* module, were aligned to a reference genome. The most closely related species with a reference genome available is the

European shrew, *Sorex araneus.* In the SAMtools module, using ANGSD v0.933 (Korneliussen et al. 2014), maximum likelihood estimates were generated for each population subset: southern New Mexico samples, northern New Mexico samples, and rangewide samples west of the Mississippi, including northern New Mexico. All reads were subjected to a set of quality filters: presence in 80% or more of the members of the populations, have a mapping quality score of at least 25, and a base quality of at least 13. This produced 555,743 sites for southern New Mexico, 648,753 sites for northern New Mexico, and 750,173 sites for Rangewide samples. The SFSs were then used to generate Stairway plots, which estimate population size change through time. In accordance with Hope et al. (2010), a generation time of one year was used. A mutation rate of $1.2x10^{-8}$ per site, per generation was applied, which is a generally accepted mutation rate for mammals when taxon-specific mutation rates are unavailable (Benazzo et al. 2017; Beynon et al. 2015; Hansen et al. 2018; MacLeod et al. 2013). Plots reported median population sizes as well as 95% confidence intervals over time, which are calculated based on 200 bootstrap replicates.

Results

Diversity statistics: mtDNA

Mitochondrial pairwise divergence (Table 3-1) indicated that High Plains versus Chaves County populations within New Mexico were approximately 0.7% divergent from each other according to the cytochrome-b gene. By comparison, divergence between High Plains samples and the remainder of West-of-Mississippi samples was only ~0.12%, whereas Chaves County samples were ~0.9% divergent from the remainder of West-of-Mississippi. From a comparative perspective, this indicates that Chaves County populations are collectively divergent from other least shrews, despite not forming a well-resolved mitochondrial lineage.

Genetic diversity metrics indicated that High Plains populations have very low values for both nucleotide diversity and haplotype diversity. When testing for divergence from a neutral evolution model, populations in High Plains exhibited a negative value of Tajima's D, although not significant. Together these results indicate that these High Plains populations form the western leading edge of a broader episode of demographic range expansion by the West-of-Mississippi lineage. By contrast Chaves County populations had much larger values for both haplotype diversity and nucleotide diversity than either High Plains or the remainder of West-of-Mississippi. When testing for divergence from a neutral model of evolution, Chaves County populations returned a positive value of Tajima's D. Together these results reflect high ancestral diversity followed by recent and severe demographic population contraction of these relictual populations.

Diversity Statistics: Nuclear DNA

Genetic diversity statistics based on SNP loci showed general trends of observed heterozygosity being lower than expected values (Table 3-2). Coupled with this, relatively high F_{IS} values for all samples within the West-of-Mississippi lineage and in particular the subset of samples representing High Plains of New Mexico, are suggestive of potential non-random mating in these populations that would reflect low population sizes. Chaves County samples had no signal of potential inbreeding. Pairwise genetic diversity comparisons for the neutral data showed a similar trend across all three differentiation statistics (pairwise Fst, Gst, and Jost's D) where the highest level of genetic differentiation was seen between South Texas samples and the West-of-Mississippi lineage, the latter including both northern and southern New Mexico, again indicative of the recognized specific status of these individuals as *C. berlandieri* (Table 3-4).

Chaves County New Mexico specimens were moderately differentiated from High Plains samples. The very low differentiation between High Plains samples and the remainder of the West-of-Mississippi lineage again supports broad genetic connectivity across the Great Plains.

Detection of Loci under Selection

For the dataset containing all individuals, out of the 11,323 total SNPs identified, Bayescan identified 41 outlier loci, while the R package PCAdapt identified 749 outlier loci. Twenty-four outlier loci identified by both methods were filtered out to create the non-neutral dataset. The remaining 10,533 loci not identified by either method were made into the neutral dataset.

This process was repeated for a dataset containing only *C. parvus* from New Mexico. The Bayescan approach identified zero loci and the "PCAdapt" approach identified 251 loci. Due to Bayescan not identifying any non-neutral loci, so that it is still possible to test for local adaptation in New Mexican populations, the non-neutral dataset will be composed only of the loci identified using PCadapt.

Population Structure: All C. parvus

In the analysis of internal population structure across the range, the DAPC plot and the Structure plot were generally congruent. The structure plot was created using the neutral SNP dataset. The best K determined by the Evanno Method for the Structure plot was K=3, meaning the most likely number of distinct evolutionary clusters across the range is three (Figure 2-6). These three clusters consist of Florida and *C. soricinus*, Chaves County New Mexico, and the

remainder of the samples (including the East-of-Mississippi lineage, West-of-Mississippi lineage, and *C. berlandieri*).

Two DAPC plots were created, one using only neutral loci, and one using only non-neutral loci (Figures 3-1; Figure 3-3). This approach allows to identify distinct management units and adaptive units respectively. Both the DAPC plots show similar trends within populations, but there are some noticeable differences. The neutral DAPC plot shows a cluster of the West-of-Mississippi lineage, East-of-Mississippi lineage, and *C. berlandieri*, with Florida and Chaves County New Mexico clustering separately, very distant from the large cluster. In the non-neutral plot however, the large cluster is made of the West-of-Mississippi lineage, the Chaves County New Mexico lineage, and the East-of-Mississippi lineage, with Florida and *C. berlandieri* separately clustering distant from the broad central cluster. Across both plots, Florida was genetically distinct, but based on these results, Chaves County New Mexico represents a distinct MU and *C. berlandieri* represents a distinct AU, while Florida represents both an MU and AU. The principal components retained explain about 10% of the variance between neutral loci and almost 17% of the variance seen in the non-neutral loci for each population. The scree plot for the neutral loci indicated the optimal number of principal components retained was three, while for the outlier loci, the optimal number of components retained was four.

Population Structure: New Mexico populations

In the analysis of internal population structure within New Mexican populations, the DAPC plot and the Structure plot were generally congruent. The structure plot was created using the neutral SNP dataset. The best K determined by the Evanno Method for the Structure plot (Figure 3-9) was K=2, meaning the most likely number of distinct evolutionary clusters within

New Mexico is two, with samples conforming unambiguously to either High Plains or Chaves County lineages recovered from phylogenetic analyses. The single sample that was collected in Chaves County New Mexico (FT644) but is genetically aligned with High Plains samples based on all analyses, indicating sympatry between these two lineages at this locality.

Two DAPC plots were created, one using only neutral loci, and one using only non-neutral loci (Figure 3-5; Figure 3-7). This approach allows to identify distinct management units and adaptive units respectively. Both the DAPC plots show the same internal population structure, with High Plains and Chaves County populations with strong spatial separation from each other on the plot. The principal components retained explain about 18% of the variance between neutral loci and almost 55% of the variance seen in the non-neutral loci for each population. For both the neutral and outlier loci, the scree plots indicated the optimal number of principal components to retain was two.

Demographic Histories:

Demographic histories (Figures 3-10; Figure 3-11) reported as change in effective population size through time were inferred using the neutral SNP dataset in Stairway Plot v2.0 for both High Plains and Chaves County New Mexico lineages. Both lineages exhibited the same general trend of a recent and severe decline in effective population size. High Plains populations of *C. parvus* increased up to 50,000 years ago coincident with onset of last glacial phase environments, followed by a population plateau through much of the glacial period and subsequent decline coincident with warming climate following the Last Glacial Maximum. Chaves County populations exhibited similar trends of glacial phase increase and post-glacial decline.

Discussion

High resolution genomic analyses of isolated populations of *C. parvus* at the western periphery of its known range confirm results from Chapter 2. First, all New Mexican populations sampled were consistent with recognition as *C. parvus* at the species level. Evidence from the matrilineal Cytb gene suggests there is little differentiation between populations in the High Plains and Chaves County, New Mexico.. However, results from nuclear analyses of ~11,300 SNPs lend strong support that samples from Chaves County New Mexico represent a distinct taxon from populations elsewhere in New Mexico. Minimally they represent distinct ESUs. Second, from a conservation perspective, High Plains peripheral populations are closely genetically related to populations distributed across the Great Plains, reflecting ongoing connectivity. Conversely, the Chaves County population is genetically distinct based on both neutrally evolving and potentially adaptive loci and collectively exhibit high genetic diversity; all signals that this lineage constitutes important evolutionary potential for the future of *C. parvus* in general. Finally, as a point of conservation concern, all peripheral populations exhibit signals of population contraction that reflect ongoing loss of genetic diversity.

It has been established that maintaining genetic diversity is vital to the long-term preservation of small and declining populations (Lande 1988; Frankham 1995; Barbosa et al. 2018). Until recently, peripheral populations of least shrews in New Mexico were considered to represent small and isolated populations of high conservation concern (NMDGF 2018). Preliminary analyses performed on initial samples collected from New Mexico further established that populations in the north and south were genetically and morphologically distinct from each other (Hafner & Shuster 1996). As such these populations meet criteria for critical evaluation of their evolutionary potential, to aid in enhancing conservation measures. The

evolutionary distinction between High Plains and Chaves County populations was strongly supported by the results of my nuclear analyses. However, the hypothesis of Hafner and Schuster (1996) that Chaves County shrews represented a relict population of *C. berlandieri* was not supported. *Cryptotis berlandieri* is clearly genetically differentiated from all other least shrews sampled. As such, Chaves County shrews represent a distinct taxon without a taxonomic designation of finer resolution than its inclusion within *C. parvus* (sensu stricto). It also appears that this lineage occurs entirely within Chaves County, New Mexico, associated with Bitter Lake NWR. Conversely, additional recent field surveys have detected least shrews from much more broadly through eastern New Mexico than previously considered, and based on my analyses, all samples outside Chaves County are closely genetically aligned with the West-of-Mississippi lineage. These observations strongly suggest that High Plains *C. parvus* constitute interconnected populations with those further east. One exception of this is a moderately divergent population occurring in the Pecos River floodplain in the vicinity of Fort Sumner (Bosque Redondo Park), that suggests at least partial isolation from other regional populations. And importantly, additional shrews associated with the High Plains lineage were collected in sympatry with Chaves County specimens in the Salt Creek unit of Bitter Lake NWR. These latter samples (FT644 and FT660) show no evidence of hybridization with Chaves County lineage specimens, indicating either that 1) they have expanded into this region only very recently and have not yet experienced gene flow with shrews of the Chaves County lineage, 2) they are reproductively isolated from the Chaves County lineage, or 3) the sampling from this region has been insufficient to detect gene flow among the two lineages. Regardless, additional sampling will be crucial to understanding ongoing inter-lineage dynamics.

Units of Conservation

Given that I cannot at this stage assign traditional Linnaean trinomial classifications to an un-named subspecies (see Chapter 2 for a discussion of criteria), the primary benefit of using thousands of unlinked genomic SNPs is thorough assessment and assignment of evolutionary units, based on different evolutionary legacies (Funk et al. 2012). Neutrally evolving loci are most strongly associated with divergence through genetic drift among isolated populations, whereas non-neutrally evolving loci are most strongly associated with divergence through adaptive evolution. Either one would represent independent evolutionary trajectories of lineages and warrant management as a discrete unit, but the evolutionary process responsible for divergence matters when considering what measures might be taken to enhance these populations. If neutral processes predominate, then this likely reflects small or declining populations where the ideal management would emphasize increasing population sizes. If selective processes predominate, then management might focus on preserving and enhancing the environments and habitats that populations are locally adapted to. In either instance, managing to preserve distinct evolutionary diversity will maximize viability of the species as a whole.

Population clustering analyses for all SNPs combined (Chapter 2) indicate that populations across the High Plains of New Mexico are only minimally distinct from the remainder of the West-of-Mississippi lineage, meaning that all populations of *C. parvus* from western peripheral populations eastward to the Mississippi River (except for the Chaves County lineage) should be considered a single ESU. Based on non-neutral potentially adaptive loci this ESU clustered closely with the East-of-Mississippi lineage meaning that based on the loci sampled, West-of-Mississippi and East-of-Mississippi lineages should be considered a single inclusive AU, suggesting that least shrews broadly distributed across the Great Plains and

Eastern United States are experiencing similar selective pressures. But, based on only neutral loci, the West-of-Mississippi and East-of-Mississippi lineages cluster separately and as such should be considered as separate MUs. When considering the Chaves County lineage, whether based on all SNP loci, only neutral loci or only non-neutral loci, the lineage forms a distinct genetic cluster, and as such should be considered a distinct ESU, MU, and AU.

Genetic demography of peripheral populations

Populations of *C. parvus* through the High Plains of New Mexico show evidence that supports the hypothesis that they represent a westward expansion (Hafner & Schuster 1996). However, the extension of this hypotheses that the expansion was very recent and due to increased irrigation networks throughout the Panhandle of Texas and eastern New Mexico could not be substantiated based on my data. Although there is strong evidence for population expansion, I infer that this occurred during or coincident with the end of the last glacial period. Conversely, a lack of genetic diversity, lack of significant Tajima's D, relatively high inbreeding coefficient, and a signal of dramatic decline in effective population size from the Stairway Plot all support recent population declines across the High Plains (Gibson et al. 2009). The scarcity of samples of this species across the western periphery of its range therefore reflects both the inherent rarity of this species, rangewide, and also the notion that new records of locality represent contemporary range expansions. I note that increase sampling effort targeting these shrews in recent years is strongly correlated with increased incidence (NMDGF 2020).

Chaves County populations showed a similar decrease in population size, but a Tajima's D value that was positive, very low inbreeding coefficient, and relatively high levels of genomic diversity. This combination of evidence strongly supports a recent and severe demographic

contraction in effective population size, and given that despite increased sampling efforts, no additional populations of this lineage have been discovered, this decline is also likely associated with severe range contraction. These dynamics suggest that Bitter Lake NWR does indeed support a relictual lineage that once persisted much more broadly within a glacial refugium throughout southern NM during the last glacial maximum. This is also supported based on fossil evidence discovered well outside the species current range (Harris 1973), although the specific lineage to which these fossils should be assigned is not easily determined.

Management recommendations

My results support the hypothesis that least shrew populations in New Mexico are represented by one species: *C. parvus,* but within this taxon exist two distinct infraspecific lineages occurring in the state. These lineages reflect historical isolation and divergence, variably through neutral genetic drift and/or through adaptive divergence. The observation that both lineages have been collected from the same locality in Chaves County New Mexico is concerning from a conservation standpoint, because there exists potential for hybridization and loss of apparently local adaptive diversity within the Chaves County lineage, or minimally, these separate lineages may experience competitive interactions. Knowledge of the effects from any kind of inter-lineage interactions is completely lacking, and it is imperative to further investigate population connectivity through Eastern New Mexico. For instance the slightly divergent samples from the High Plains lineage collected from Bosque Redondo near Ft. Sumner are associated with a mesic floodplain of the Pecos River. The Bosque Redondo park is located 120km north of Bitter Lake along the Pecos river, and has similar characteristics to the Bitter Lake NWR, although the latter is more strongly associated with saline wetlands and isolated

Cienegas (NMDGF 2020). Both of these sites offer relatively intact and year-round mesic habitats, promoting persistence of least shrew habitats year round. The associated Pecos River may act as a corridor for periodic dispersal and subsequent gene flow between populations, although how effective this corridor might be remains unsubstantiated.

I recommend that the New Mexico Department of Game and Fish recognize populations in Chaves County as a distinct AU, and prioritize conservation of this isolated taxon through maintenance and, where possible, expansion of the associated habitat. With regards to High Plains populations, although they constitute part of a more broadly distributed MU (West-of-Mississippi), they appear to be in decline at the western periphery and therefore should be managed for increasing population size within New Mexico, which likely will require additional ecological research to better understand their habitat requirements, considering human associated land-use change, as well as their association with particular climate trends, given progressive warming and aridification throughout the southwestern United States. In terms of taxonomy of these peripheral populations, it is critical to perform a detailed morphometric analysis of the existing shrew samples within museum collections. The Chaves County lineage may well warrant recognition as a distinct subspecies but without morphological diagnosability, such recognition is not reasonable based on genomic data alone (Patten 2010). Similarly, the West-of-Mississippi lineage likely also warrants recognition as a separate sub-species from the East-of-Mississippi lineage, again pending morphological analysis which has to date been inconclusive (Hutchinson 2010). Within the West-of-Mississippi lineage, the High Plains populations may be justifiably designated as a Distinct Population Segment, but my data do not provide enough evidence to definitively delineate these populations from elsewhere across the Great Plains.

Limitations of this work lie in the scarcity of samples within museum collections. I recommend further sampling across eastern New Mexico as well as more broadly throughout North America. Despite the seeming rarity of this species, even within peripheral populations, shrews in general have rapid turnover and high fecundity. Limited sampling should not significantly affect natural populations, and certainly scientific collections in general have negligible effects compared with other natural or anthropogenic factors that influence population trajectories (Hope et al. 2018). The multitude of benefits that might be realized from specimen collections for affecting positive changes in how we manage these shrews and their environments will hopefully be realized through both traditional and emerging analyses (McLean et al. 2016; Galbreath et al. 2019). My evolutionary assessment of least shrews was possible through the development of reduced genome sequencing methods that provide much greater resolution of both relationships and trajectories among populations of wildlife. However, persistent difficulties of working with non-model organisms like shrews that do not, for instance, have reference genomes available, still limit our ability to assess functional change, or to analyze the genetics of old and degraded specimens, including fossils. As such, future analyses of these enigmatic mammals should include more detailed assessment of local adaptive evolution, considering our rapidly changing world.

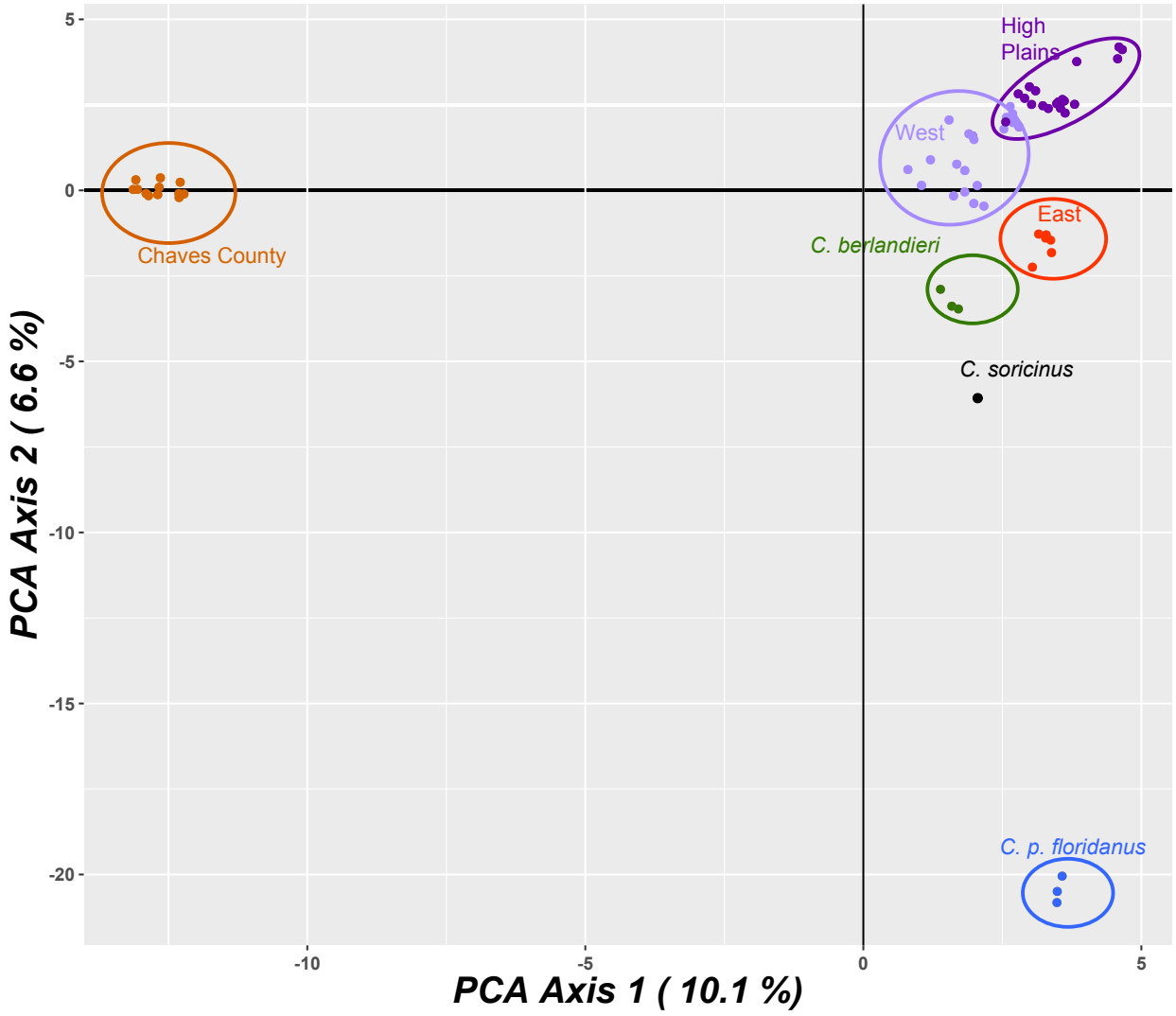

Figure 3-1 Neutral SNP DAPC Plot

DAPC scatter plot based on the neutral SNP dataset for all samples. Samples that are closer together share more genetic similarity than those that are distant on the scatter plot.

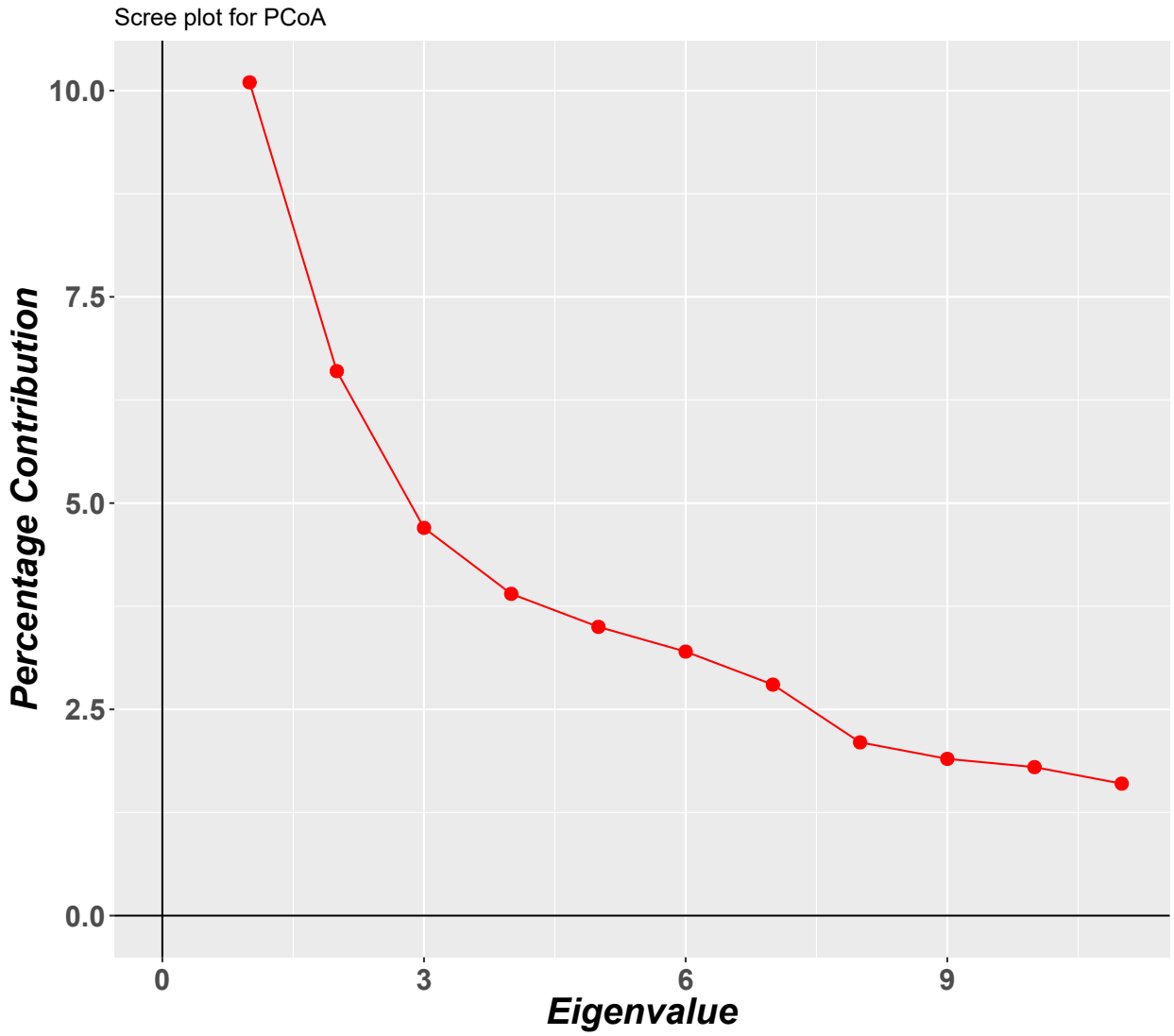

Figure 3-2 Scree Plot- Neutral SNPs

Screeplot showing percent contribution of principal components retained for explaining the variance on the DAPC plot. Only the first three principal components were retained.

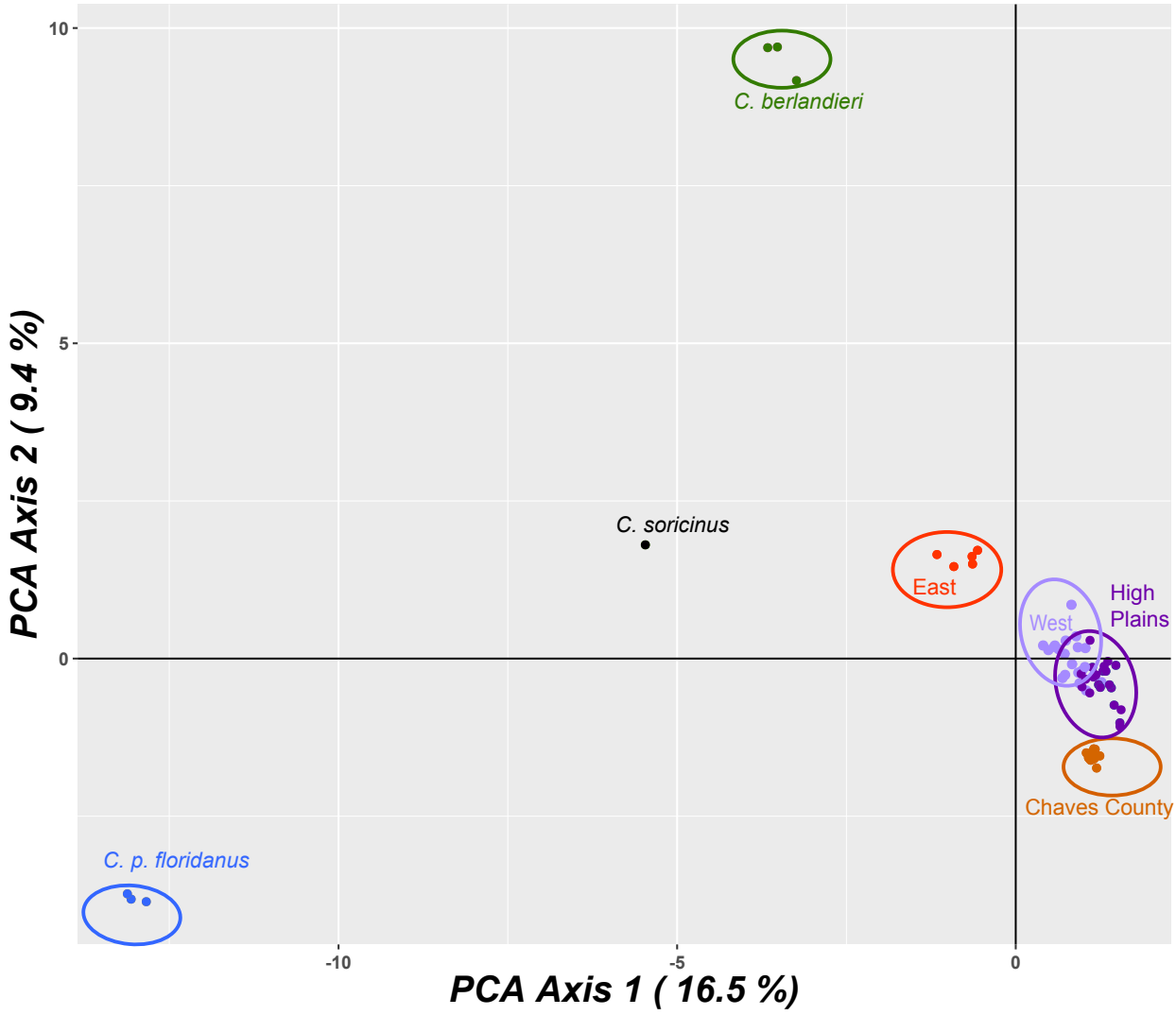

Figure 3-3 Non-Neutral SNP DAPC Plot

DAPC scatter plot based on the non-neutral SNP dataset for all samples. Samples that are closer together share more genetic similarity than those that are distant on the scatter plot.

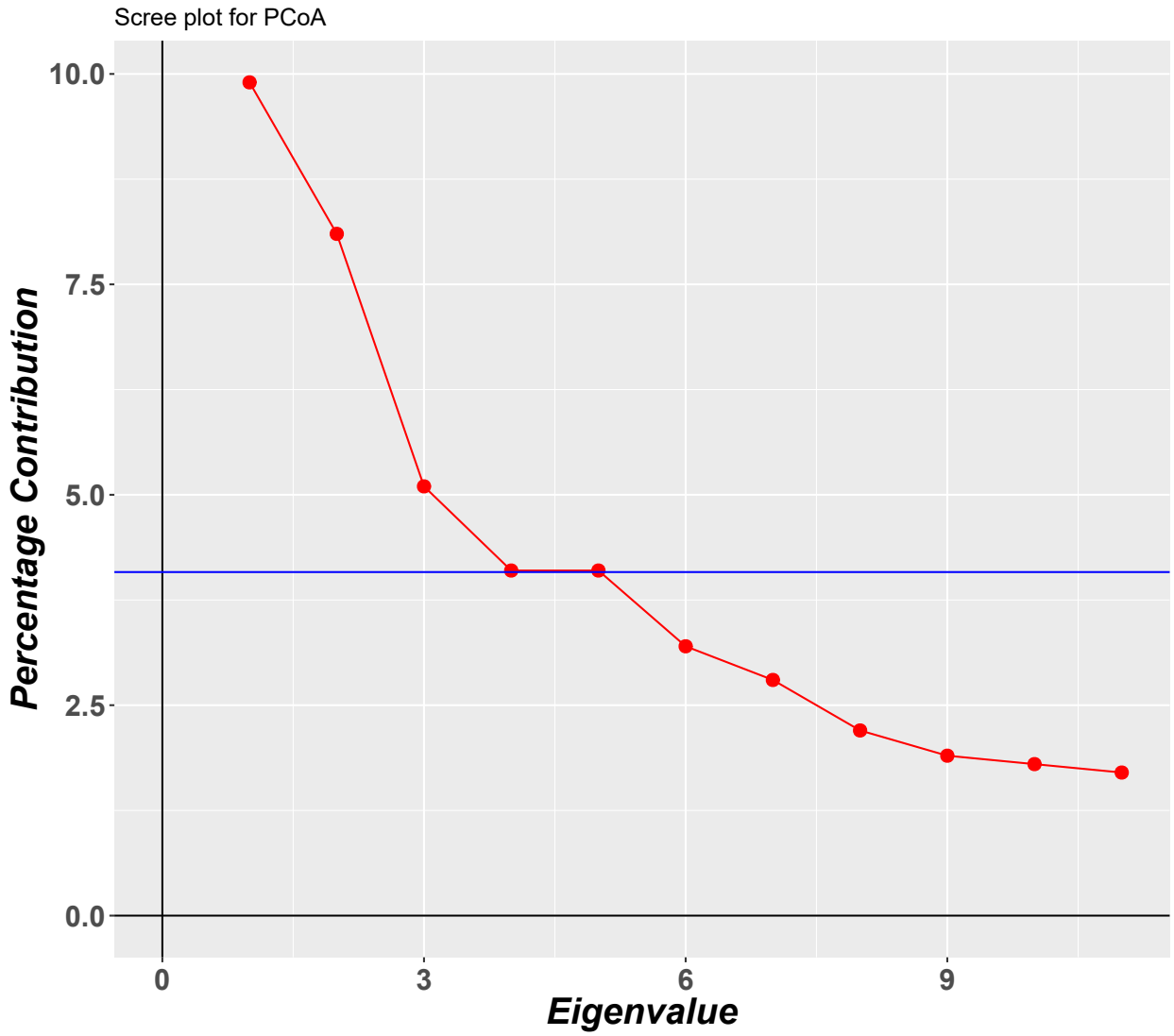

Figure 3-4 Scree Plot- Non-Neutral SNP's

Screeplot showing percent contribution of principal components retained for explaining variance on the DAPC plot. Only the first four principal components were retained.

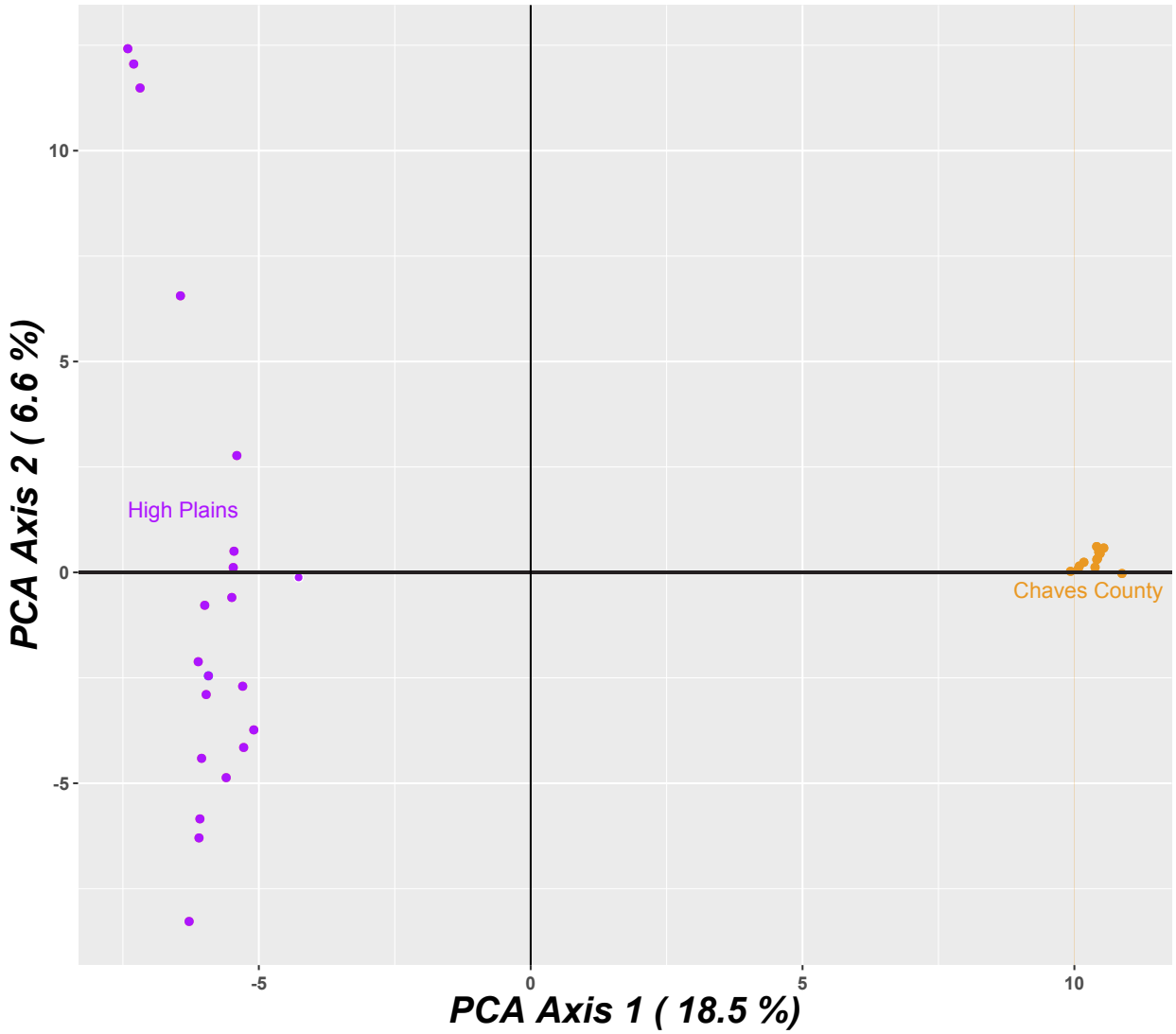

Figure 3-5 Neutral SNP DAPC Plot

DAPC scatter plot based on the neutral SNP dataset for only New Mexican samples. Samples that are closer together share more genetic similarity than those that are distant on the scatter plot.

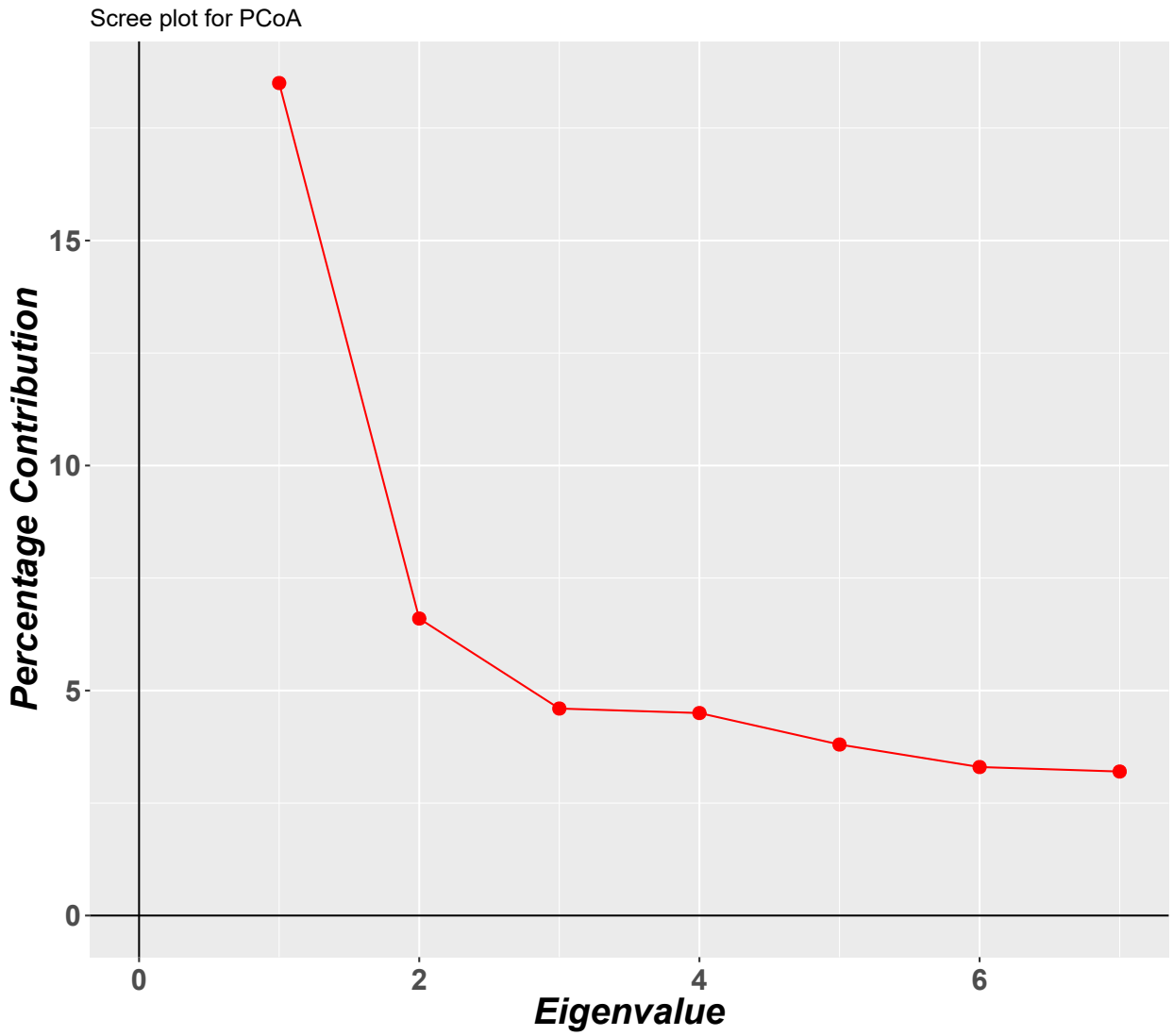

Figure 3-6 Scree Plot- Neutral SNPs

Screeplot showing percent contribution of principal components retained for explaining variance on the DAPC plot. Only the first two principal components were retained.

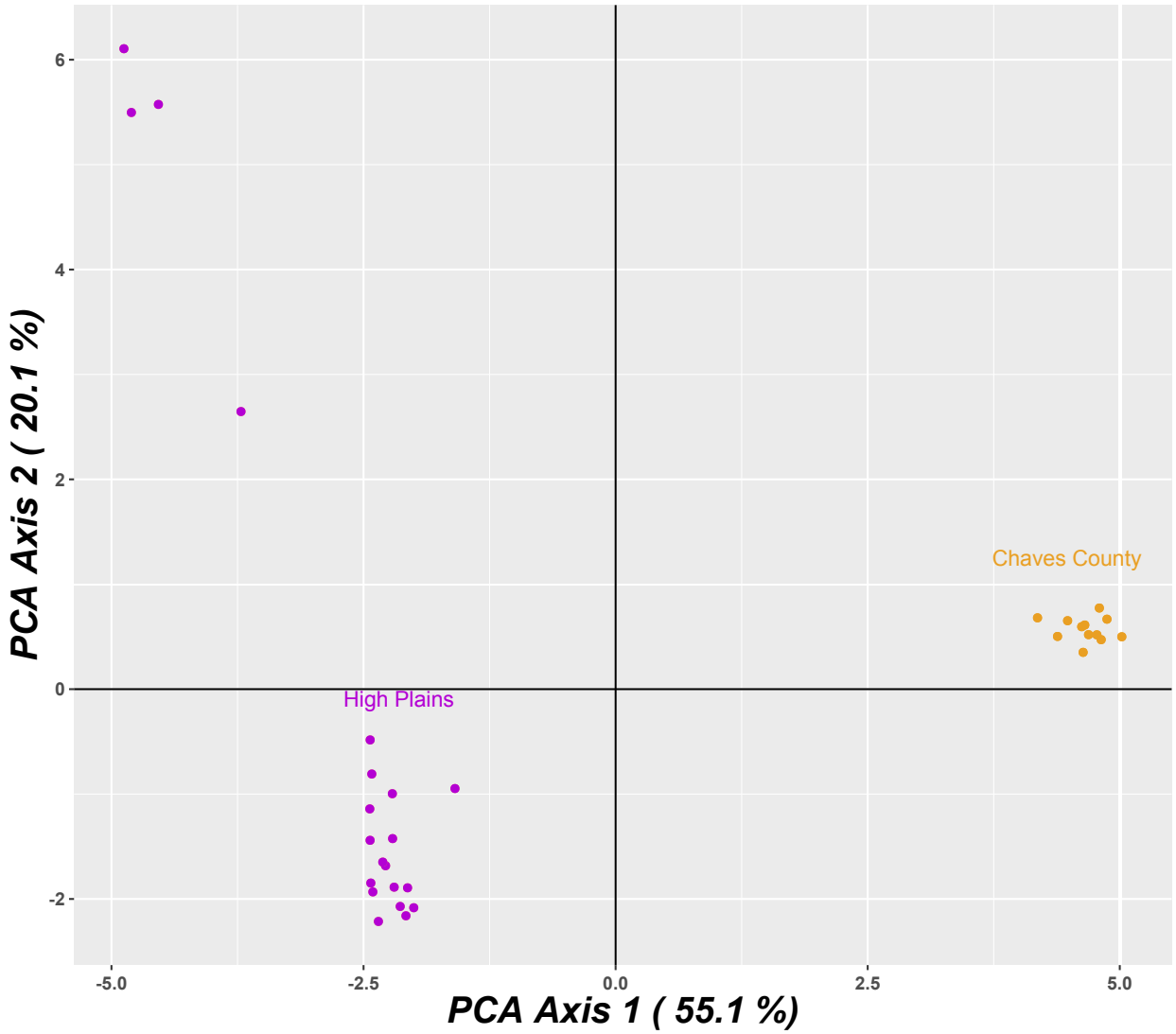

Figure 3-7 Non-Neutral SNP DAPC Plot

DAPC scatter plot based on the non-neutral SNP dataset for only New Mexican samples. Samples that are closer together share more genetic similarity than those that are distant on the scatter plot.

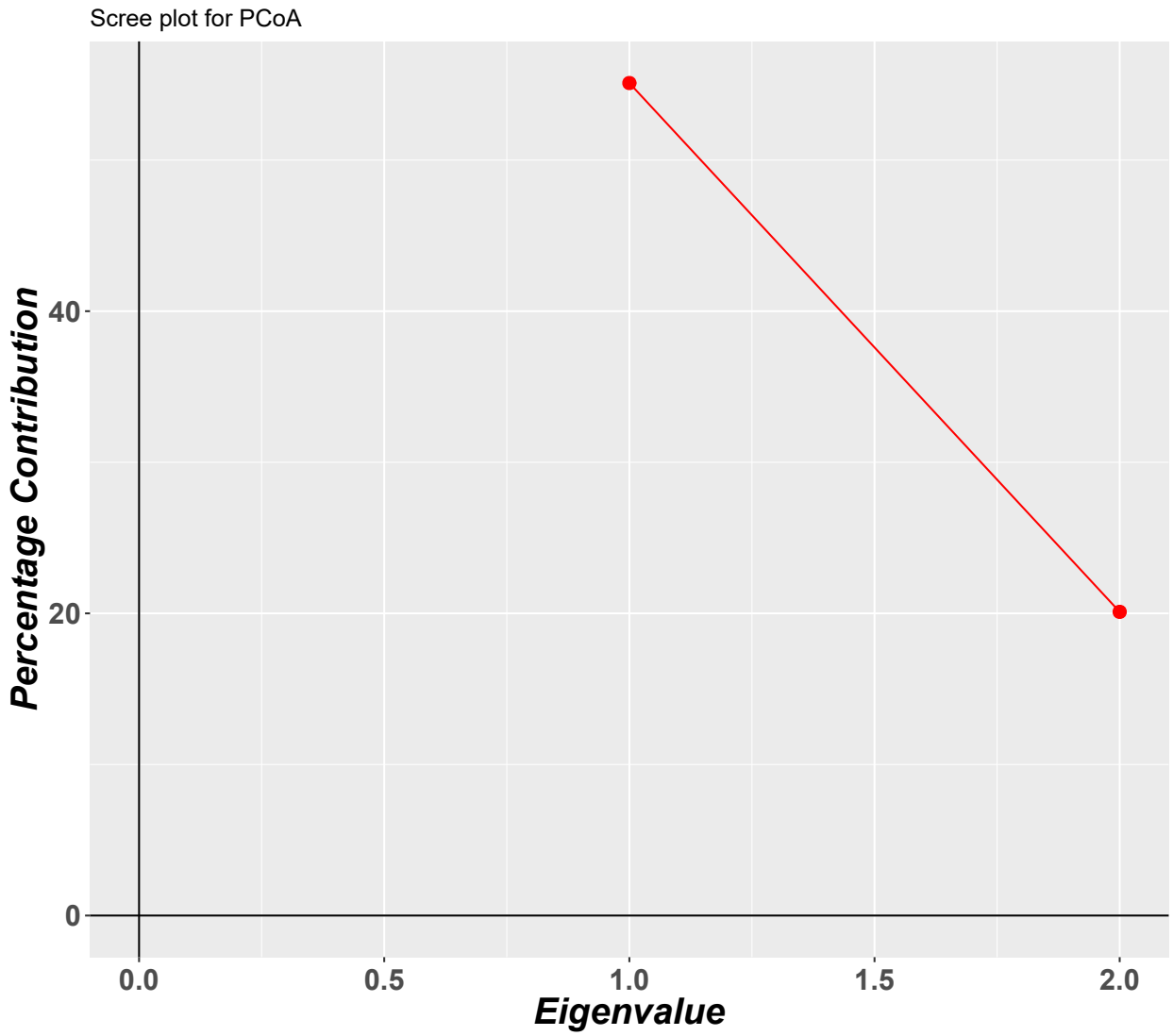

Figure 3-8 Scree Plot- Non-Neutral SNPs

Screeplot showing percent contribution of principal components retained for explaining variance on the DAPC plot. Only the first two principal components were retained.

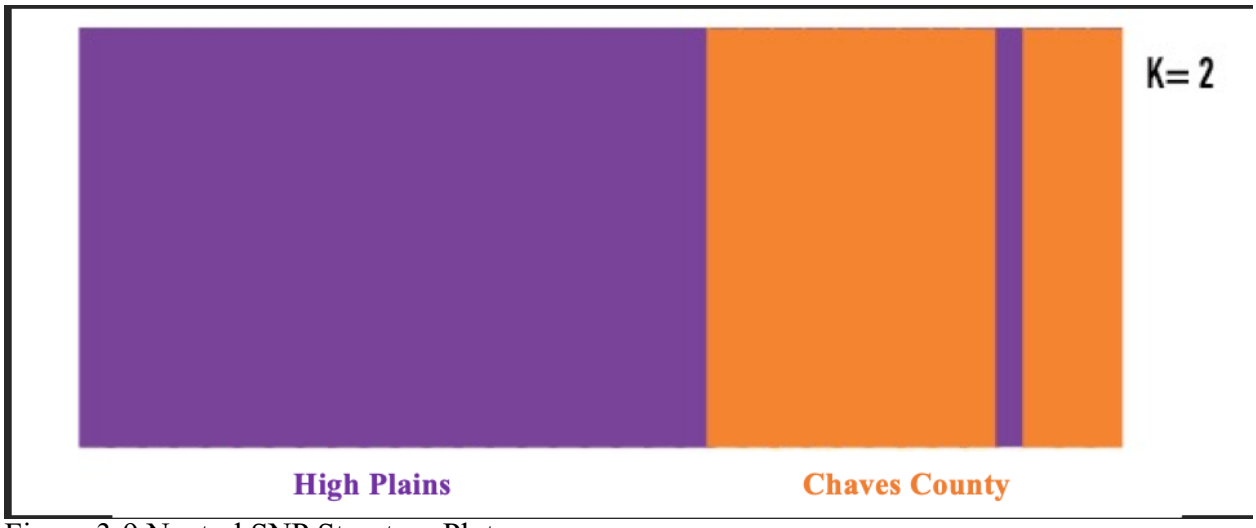

Figure 3-9 Neutral SNP Structure Plot

Structure plot based on neutral SNP data. Colors in plot coordinate to a distinct population cluster identified by the Structure program. The best K for this plot was determined to be 2.

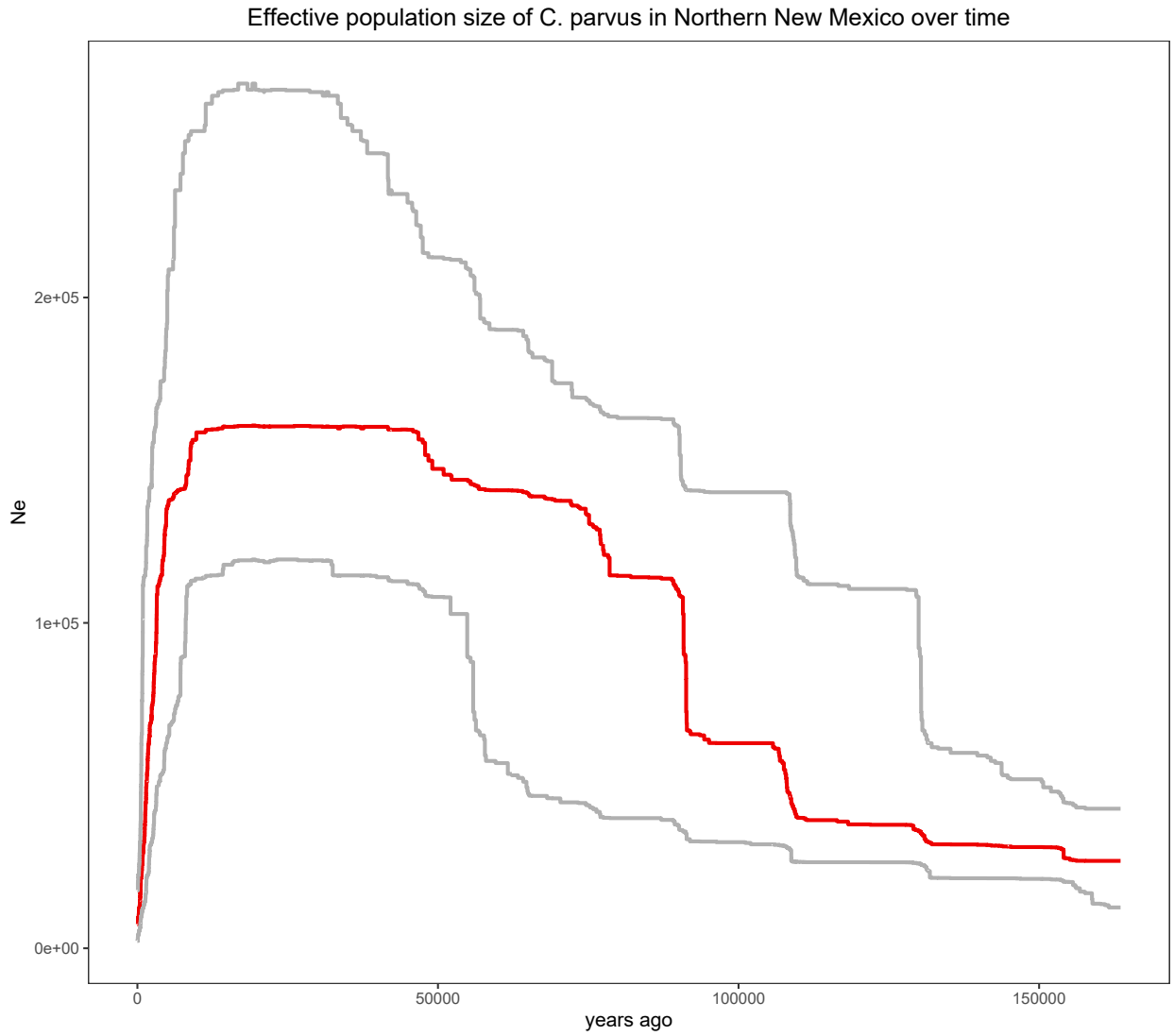

Figure 3-10 Northern New Mexico Stairway Plot

The effective population size of *C. parvus* in northern New Mexico over the last 100,000 years. Grey lines represent the 95% confidence interval.

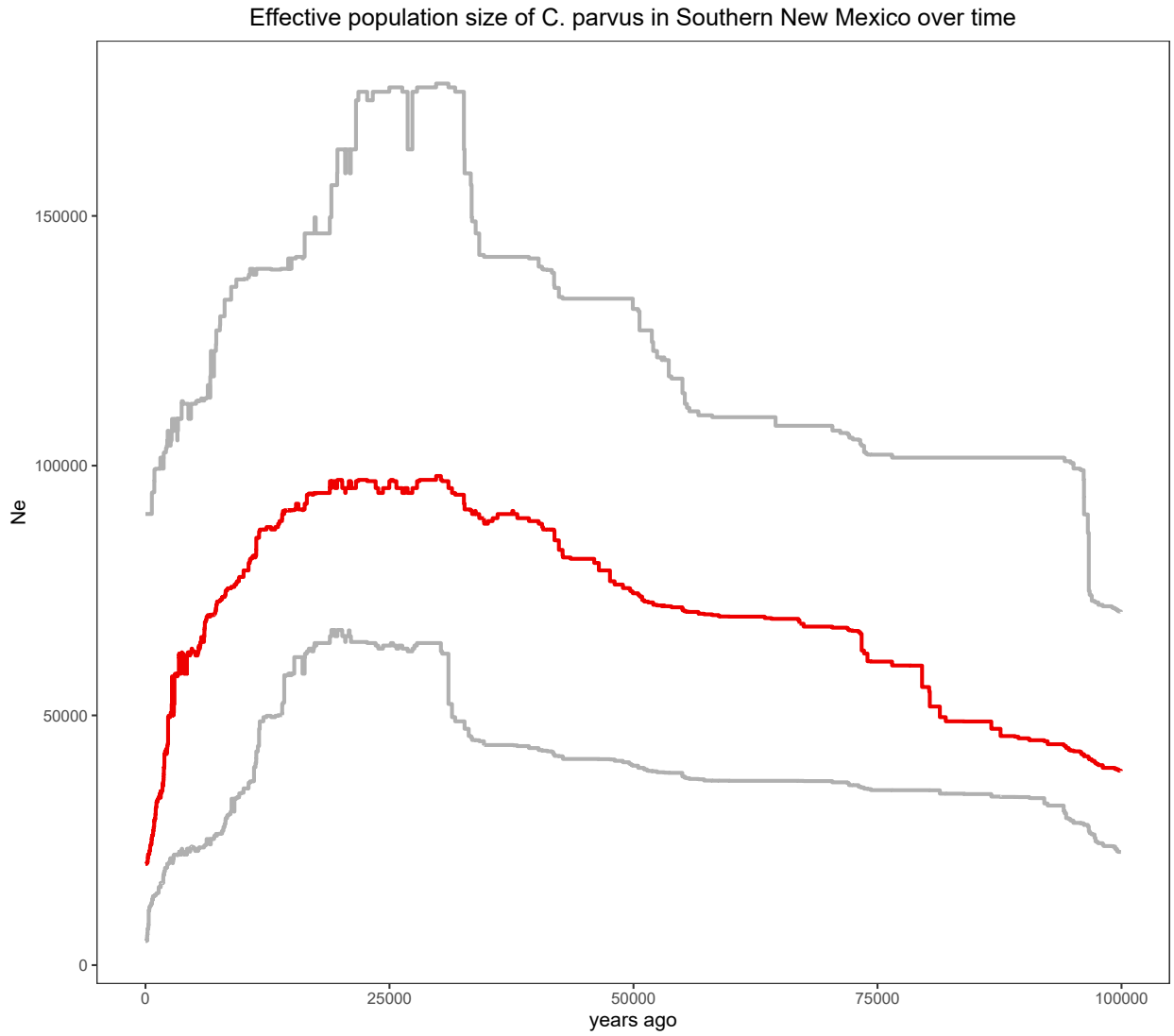

Figure 3-11 Southern New Mexico Stairway Plot

The effective population size of *C. parvus* in southern New Mexico over the last 100,000 years. Grey lines represent the 95% confidence interval.

Table 3-1 Cytochrome-B Genetic Differentiation

Genetic differentiation measures based on the mitochondrial cytochrome-b dataset. **(A)** pairwise population divergences; (B) genetic diversity metrices. Bolded values are statistically significant at $\alpha<0.05$.

(A) **Population Divergence**

	West	High Plains	Chaves County	*C. berlandieri*
West	-	0.0012	0.0070	0.0909
>High Plains	-	-	0.0071	0.0889
>Chaves County	-	-	-	0.0909
C. berlandieri	-	-	-	-

(B) **Genetic Diversity Metrics**

	N	*Hd*	π	Tajima's D
West	20	0.833	0.0282	**-2.71555**
>High Plains	36	0.571	0.00079	-1.26659
>Chaves County	12	0.788	0.00789	1.37446
C. berlandieri	2	1.000	0.0113	-

Table 3-2 SNP Genetic Diversity Statistics

Genetic diversity statistics based on the whole SNP dataset

	Observed Het.	Expected Het.	Allelic Richness	F_{IS}	π
C. berlandieri	0.06145	0.05657	1.057237	0.01178	0.06789
West	0.09325	0.12192	1.116046	0.13257	0.12354
>High Plains	0.08387	0.10506	1.107799	0.18647	0.10247
>Chaves County	0.06951	0.07948	1.070939	0.03868	0.0832
East	0.09121	0.10348	1.102592	0.05536	0.1163

Table 3-3 Neutral SNP Genetic Diversity Statistics

Genetic diversity statistics based on the neutral SNP dataset

	Observed Het.	Expected Het.	Allelic Richness	F_{IS}
C. berlandieri	0.09059	0.11081	1.056922	0.18248
West	0.05254	0.05802	1.110279	0.09449
>High Plains	0.08387	0.10506	1.101973	0.18155
>Chaves County	0.06323	0.07413	1.073626	0.14695
East	0.02782	0.1704	1.099048	-

Table 3-4 Neutral SNP Pairwise Genetic Differentiation

Genetic differentiation metrices based on the nuclear SNP dataset containing only neutral loci. **(A)** Pairwise Fst values, all values are statistically significant; **(B)** Pairwise Gst values; **(C)** Pairwise Jost's D values.

(A)	***C. berlandieri***	**West**	**High Plains**	**Chaves County**
C. berlandieri	-	0.3056	0.3401	0.4631
West		-	0.0379	0.1911
High Plains			-	0.2252
Chaves County				-

(B)	***C. berlandieri***	**West**	**High Plains**	**Chaves County**
C. berlandieri	-	0.1892	0.2136	0.3084
West		-	0.0217	0.1086
High Plains			-	0.1299
Chaves County				-

(C)	***C. berlandieri***	**West**	**High Plains**	**Chaves County**
C. berlandieri	-	0.0419	0.0461	0.0606
West		-	0.0053	0.0247
High Plains			-	0.0287
Chaves County				-

References

Abbott, R., Albach, D., Ansell, S., Arntzen, J. W., Baird, S. J., Bierne, N., ... & Zinner, D. (2013). Hybridization and speciation. *Journal of Evolutionary Biology*, *26*(2), 229-246.

Ahn, S. J., Costa, J., & Rettig Emanuel, J. (1996). PicoGreen quantitation of DNA: effective evaluation of samples pre-or psost-PCR. Nucleic acids research, 24(13), 2623-2625.

Alexa L. Trujillo & Eric A. Hoffman (2017) Uncovering discordance between taxonomy and evolutionary history in Florida raccoons, *Systematics and Biodiversity*, 15:1, 74-85.

Allendorf, F. W., Hohenlohe, P. A., & Luikart, G. (2010). Genomics and the future of conservation genetics. *Nature Reviews Genetics*, *11*(10), 697-709.

Allendorf, F. W., & Luikart, G. (2009). *Conservation and the genetics of populations*. John Wiley & Sons.

Avise, J. C., Arnold, J., Ball, R. M., Bermingham, E., Lamb, T., Neigel, J. E., ... & Saunders, N. C. (1987). Intraspecific phylogeography: the mitochondrial DNA bridge between population genetics and systematics. *Annual Review of Ecology and Systematics*, *18*(1), 489-522.

Backlund, D. C. (2002). The expanding distribution of the least shrew, *Cryptotis parvus*, in South Dakota. *In Proceedings of the South Dakota Academy of Science* (Vol. 81, p. 153).

Baker, R. J., & Bradley, R. D. (2006). Speciation in mammals and the genetic species concept. *Journal of Mammalogy*, *87*(4), 643-662.

Ballard, J. W. O., & Whitlock, M. C. (2004). The incomplete natural history of mitochondria. *Molecular Ecology*, *13*(4), 729-744.

Barbosa, S., Mestre, F., White, T. A., Paupério, J., Alves, P. C., & Searle, J. B. (2018). Integrative approaches to guide conservation decisions: using genomics to define conservation units and functional corridors. *Molecular Ecology*, *27*(17), 3452-3465.

Barbosa, S., Paupério, J., Herman, J. S., Ferreira, C. M., Pita, R., Vale-Gonçalves, H. M., ... & Searle, J. B. (2017). Endemic species may have complex histories: Within-refugium phylogeography of an endangered Iberian vole. *Molecular Ecology*, 26(3), 951-967.

Bell, J. R. (2008). A simple way to treat PCR products prior to sequencing using ExoSAP-IT®. *Biotechniques*, *44*(6), 834.

Benazzo, A., Trucchi, E., Cahill, J. A., Delser, P. M., Mona, S., Fumagalli, M., ... & Bertorelle, G. (2017). Survival and divergence in a small group: The extraordinary genomic history of the endangered Apennine brown bear stragglers. *Proceedings of the National Academy of Sciences*, 114(45), E9589-E9597.

Benedict, R. A., Genoways, H. H., & Choate, J. R. (2006). Taxonomy of short-tailed shrews (genus *Blarina*) in Florida. *Mammalogy Papers: University of Nebraska State Museum*, 103.

Bermingham, E., & Moritz, C. (1998). Comparative phylogeography: concepts and applications. *Molecular Ecology*, *7*(4), 367-369.

Beynon, S. E., Slavov, G. T., Farré, M., Sunduimijid, B., Waddams, K., Davies, B., ... & Larkin, D. M. (2015). Population structure and history of the Welsh sheep breeds determined by whole genome genotyping. *BMC Genetics*, 16(1), 1-15.

Bickford, D., Lohman, D. J., Sodhi, N. S., Ng, P. K., Meier, R., Winker, K., ... & Das, I. (2007). Cryptic species as a window on diversity and conservation. *Trends in Ecology & Evolution*, 22(3), 148-155.

Birney, E. C. (1983). Mumford, RE, and JO Whitaker, Jr. Mammals of Indiana. *Indiana University Press, Bloomington,* xvii+ 537 pp., 1982.

Bolger, A. M., Lohse, M., & Usadel, B. (2014). Trimmomatic: a flexible trimmer for Illumina sequence data. *Bioinformatics*, *30*(15), 2114-2120.

Bouckaert, R., Vaughan, T. G., Barido-Sottani, J., Duchêne, S., Fourment, M., Gavryushkina, A., ... & Drummond, A. J. (2019). BEAST 2.5: An advanced software platform for Bayesian evolutionary analysis. *PLoS Computational Biology*, *15*(4), e1006650.

Bradley R. D. Baker R. J.. 2001. A test of the Genetic Species Concept: cytochrome-b sequences and mammals. *Joumal of Mammalogy* 82:960–973.

Brown, W. M., George, M., & Wilson, A. C. (1979). Rapid evolution of animal mitochondrial DNA. *Proceedings of the National Academy of Sciences*, *76*(4), 1967-1971.

New Mexico Department of Game and Fish. 2021. Least Shrew (*Cryptotis parvus*) Recovery Plan. New Mexico Department of Game and Fish, Wildlife Management Division, Santa Fe, New Mexico. 27 pp.

Chifman, J., & Kubatko, L. (2014). Quartet inference from SNP data under the coalescent model. Bioinformatics, 30(23), 3317-3324.

Choate, J.R. (1970). Systematics and zoogeography of Middle American shrews of the genus *Cryptotis*. *University of Kansas Publications, Museum of Natural History* 19:195–317.

Coates, D. J., Byrne, M., & Moritz, C. (2018). Genetic diversity and conservation units: dealing with the species-population continuum in the age of genomics. *Frontiers in Ecology and Evolution*, *6*, 165.

Churchfield, S. (1990). The natural history of shrews. *Cornell University Press*.

Crandall, K. A., Bininda-Emonds, O. R., Mace, G. M., & Wayne, R. K. (2000). Considering evolutionary processes in conservation biology. *Trends in Ecology & Evolution*, *15*(7), 290-295.

Culver, M., Johnson, W. E., Pecon-Slattery, J., & O'Brien, S. J. (2000). Genomic ancestry of the American puma (*Puma concolor*). *Journal of Heredity*, *91*(3), 186-197.

Danecek, P., Auton, A., Abecasis, G., Albers, C. A., Banks, E., DePristo, M. A., ... & 1000 Genomes Project Analysis Group. (2011). The variant call format and VCFtools. *Bioinformatics*, *27*(15), 2156-2158.

de Jong, M. J., de Jong, J. F., Hoelzel, A. R., & Janke, A. (2021). SambaR: An R package for fast, easy and reproducible population-genetic analyses of biallelic SNP data sets. *Molecular Ecology Resources*, 21(4), 1369-1379.

De Queiroz, K. (2007). Species concepts and species delimitation. *Systematic Biology*, *56*(6), 879-886.

DeSalle, R., & Amato, G. (2009). 1. The Expansion of Conservation Genetics. *Conservation Genetics in the Age of Genomics*, 1-24.

Earl, D. A. (2012). STRUCTURE HARVESTER: a website and program for visualizing STRUCTURE output and implementing the Evanno method. *Conservation Genetics Resources*, *4*(2), 359-361.

Ellsworth, D. L., Honeycutt, R. L., Silvy, N. J., Bickham, J. W., & Klimstra, W. D. (1994). Historical biogeography and contemporary patterns of mitochondrial DNA variation in white-tailed deer from the southeastern United States. *Evolution*, *48*(1), 122-136.

Evanno, G., Regnaut, S., & Goudet, J. (2005). Detecting the number of clusters of individuals using the software STRUCTURE: a simulation study. *Molecular Ecology*, *14*(8), 2611-2620.

Foll, M., & Gaggiotti, O. (2008). A genome-scan method to identify selected loci appropriate for both dominant and codominant markers: a Bayesian perspective. *Genetics*, *180*(2), 977-993.

Frankham, R., Ballou, S. E. J. D., Briscoe, D. A., & Ballou, J. D. (2002). Introduction to conservation genetics. *Cambridge University Press*

Fraser, D. J., & Bernatchez, L. (2001). Adaptive evolutionary conservation: towards a unified concept for defining conservation units. *Molecular Ecology*, *10*(12), 2741-2752.

Frey, J. K. (2005). Status and habitat of the least shrew (*Cryptotis parva*) in New Mexico. Professional services contract, 05-516.

Funk, W. C., McKay, J. K., Hohenlohe, P. A., & Allendorf, F. W. (2012). Harnessing genomics for delineating conservation units. *Trends in Ecology & Evolution*, 27(9), 489-496.

Galbreath, K. E., Hoberg, E. P., Cook, J. A., Armién, B., Bell, K. C., Campbell, M. L., ... & Hope, A. G. (2019). Building an integrated infrastructure for exploring biodiversity: field collections and archives of mammals and parasites. *Journal of Mammalogy*, *100*(2), 382-393.

Gibson, S. Y., Van der Marel, R. C., & Starzomski, B. M. (2009). Climate change and conservation of leading-edge peripheral populations. *Conservation Biology*, *23*(6), 1369-1373.

Giles, R. E., Blanc, H., Cann, H. M., & Wallace, D. C. (1980). Maternal inheritance of human mitochondrial DNA. *Proceedings of the National Academy of Sciences*, *77*(11), 6715-6719.

Goldman, E. A. (1912). New mammals from eastern Panama (Vol. 2142). *Smithsonian Institution.*

Good, J. M., Hird, S., Reid, N., Demboski, J. R., Steppan, S. J., MARTIN-NIMS, T. R., & Sullivan, J. (2008). Ancient hybridization and mitochondrial capture between two species of chipmunks. *Molecular Ecology*, *17*(5), 1313-1327.

Gordon, D. H., & Watson, C. R. B. (1986). Identification of cryptic species of rodents (*Mastomys, Aethomys, Saccostomus*) in the Kruger National Park. *African Zoology*, 21(1), 95-99.

Gosselin, T., Lamothe, M., Devloo-Delva, F., & Grewe, P. (2020). radiator: RADseq Data Exploration, Manipulation and Visualization using R. R package version 1.1. 5.

Gruber, B., Unmack, P. J., Berry, O. F., & Georges, A. (2018). DARTR: An R package to facilitate analysis of SNP data generated from reduced representation genome sequencing. *Molecular Ecology Resource.*

Hafner, D. J., & Shuster, C. J. (1996). Historical biogeography of western peripheral isolates of the least shrew*, Cryptotis parva. Journal of Mammalogy*, 77(2), 536-545.

Hall, E. R. 1981. The mammals of North America. Second ed*. John Wiley & Sons*, New York, 1:1-600 + 90.

Hall, E., & Kelson, K. R. (1959). The mammals of North America. 2 Vols. *Ronald Press.*.

Hansen, C. C. R., Hvilsom, C., Schmidt, N. M., Aastrup, P., Van Coeverden de Groot, P. J., Siegismund, H. R., & Heller, R. (2018). The Muskox Lost a Substantial Part of Its Genetic Diversity on Its Long Road to Greenland. *Current Biology*, 28(24), 4022-4028.e5.

Harris, A. H., Smartt, R. A., & Smartt, W. R. (1973). *Cryptotis parva* from the Pleistocene of New Mexico. *Journal of Mammalogy*, *54*(2), 512-513.

Harrison, R. G., 1989 Animal mitochondrial DNA as a genetic marker in population and evolutionary biology. *Trends Ecol. Evol.* 4: 6–11.

He, K., Woodman, N., Boaglio, S., Roberts, M., Supekar, S., & Maldonado, J. E. (2015). Molecular phylogeny supports repeated adaptation to burrowing within small-eared shrews genus of Cryptotis (Eulipotyphla, Soricidae). *PloS one*, *10*(10), e0140280.

Hewitt, G. M. (2004). The structure of biodiversity–insights from molecular phylogeography. *Frontiers in Zoology*, *1*(1), 1-16.

Hoditschek, B., Cully, J. F., Best, T. L., & Painter, C. (1985). Least shrew (*Cryptotis parva*) in New Mexico. *The Southwestern Naturalist*, 30(4), 600-601.

Hohenlohe, P. A., Bassham, S., Etter, P. D., Stiffler, N., Johnson, E. A., & Cresko, W. A. (2010). Population genomics of parallel adaptation in threespine stickleback using sequenced RAD tags. *PLoS Genetics*, *6*(2), e1000862.

Hope, A. G., & Frey, J. K. (2021). Misinterpretation of genomic data matters for endangered species listing: the subspecific status of the Peñasco least chipmunk (*Neotamias minimus atristriatus). Ecology and Evolution Archives.*.

Hope, A. G., Sandercock, B. K., & Malaney, J. L. (2018). Collection of scientific specimens: benefits for biodiversity sciences and limited impacts on communities of small mammals. *BioScience*, *68*(1), 35-42.

Hope, A. G., Speer, K. A., Demboski, J. R., Talbot, S. L., & Cook, J. A. (2012). A climate for speciation: rapid spatial diversification within the *Sorex* cinereus complex of shrews. *Molecular Phylogenetics and Evolution*, *64*(3), 671-684.

Hope, A. G., Waltari, E., Dokuchaev, N. E., Abramov, S., Dupal, T., Tsvetkova, A., ... & Cook, J. A. (2010). High-latitude diversification within Eurasian least shrews and Alaska tiny shrews (Soricidae). *Journal of Mammalogy*, *91*(5), 1041-1057.

Huang, Q., Baum, L., & Fu, W. L. (2010). Simple and practical staining of DNA with GelRed in agarose gel electrophoresis. *Clinical Laboratory Journal For Clinical Laboratories And Laboratories Related*, *56*(3), 149.

Hutchinson, S. J. (2010). Phylogeography of Cryptotis Parva in the United States Using Morphometrics and Population Genetics (*Doctoral dissertation, University of North Carolina Wilmington*).

Hutterer, R. (2005). Order soricomorpha. Mammal species of the world: a taxonomic and geographic reference, 1, 220-311.

IUCN (International Union for Conservation of Nature) 2008. *Cryptotis parva*. The IUCN Red List of Threatened Species. Version 2021-1

Jackson, D. J., & Cook, J. A. (2020). A precarious future for distinctive peripheral populations of meadow voles (*Microtus pennsylvanicus*). *Journal of Mammalogy*, *101*(1), 36-51.

Kearse, M., Moir, R., Wilson, A., Stones-Havas, S., Cheung, M., Sturrock, S., ... & Drummond, A. (2012). Geneious Basic: an integrated and extendable desktop software platform for the organization and analysis of sequence data. *Bioinformatics*, *28*(12), 1647-1649.

Knaus, B. J., & Grünwald, N. J. (2017). vcfr: a package to manipulate and visualize variant call format data in R. *Molecular Ecology Resources*, *17*(1), 44-53.

Korneliussen, T. S., Albrechtsen, A., & Nielsen, R. (2014). ANGSD: analysis of next generation sequencing data. *BMC Bioinformatics*, 15(1), 1-13.

Lande, R. (1988). Genetics and demography in biological conservation. *Science*, *241*(4872), 1455-1460.

Lewis, P. O. (2001). A likelihood approach to estimating phylogeny from discrete morphological character data. *Systematic Biology*, 50, 913–925

Liu, X., & Fu, Y. X. (2020). Stairway Plot 2: demographic history inference with folded SNP frequency spectra. *Genome Biology*, *21*(1), 1-9.

Luikart, G., Kardos, M., Hand, B. K., Rajora, O. P., Aitken, S. N., & Hohenlohe, P. A. (2018). Population genomics: advancing understanding of nature. In *Population Genomics* (pp. 3-79).

Luu, K., Bazin, E., & Blum, M. G. (2017). pcadapt: an R package to perform genome scans for selection based on principal component analysis. *Molecular Ecology Resources*, *17*(1), 67-77.

MacLeod, I. M., Larkin, D. M., Lewin, H. A., Hayes, B. J., & Goddard, M. E. (2013). Inferring Demography from Runs of Homozygosity in Whole-Genome Sequence,

with Correction for Sequence Errors. *Molecular Biology and Evolution*, 30(9), 2209–2223.

Marantz, C. A., & Patten, M. A. (2010). Chapter 11: Quantifying Subspecies Analysis: A Case Study of Morphometric Variation and Subspecies in the Woodcreeper Genus *Dendrocolaptes*. *Ornithological Monographs*, *67*(1), 123-140.

Marquardt. 2006. First record of the least shrew in Wyoming. *Prairie Naturalist*, 38:195-196.

Martin, M. (2011). Cutadapt removes adapter sequences from high-throughput sequencing reads. *EMBnet.Journal*, *17*(1), 10-12.

Mayr, E. (2000). The biological species concept. Species concepts and phylogenetic theory: a debate. *Columbia University Press, New York*, 17-29.

McLean, B. S., Bell, K. C., Dunnum, J. L., Abrahamson, B., Colella, J. P., Deardorff, E. R., ... & Cook, J. A. (2016). Natural history collections-based research: progress, promise, and best practices. *Journal of Mammalogy*, *97*(1), 287-297.

McMahon, B. J., Johansson, M. P., Piertney, S. B., Buckley, K., & Höglund, J. (2012). Genetic variation among endangered Irish red grouse (*Lagopus lagopus hibernicus*) populations: implications for conservation and management. *Conservation Genetics*, *13*(3), 639-647.

Merlino, J. N., Frohberg, A. R., Harmon, J., & Geluso, K. (2012). Least and Merriam's shrews from Banner County, Nebraska.

Mora, C., Tittensor, D. P., Adl, S., Simpson, A. G., & Worm, B. (2011). How many species are there on Earth and in the ocean?. *PLoS Biology*, *9*(8), e1001127.

Moritz, C. (1994). Defining 'evolutionarily significant units' for conservation. *Trends in Ecology & Evolution*, 9(10), 373-375.

Mumford, R. E., & Whitaker, J. O. (1982). Mammals of Indiana. *Indiana University Press.*

New Mexico Department of Game and Fish (NMDGF). 1988. *Handbook of species endangered in New Mexico*, G-108:1-2.

New Mexico Department of Game and Fish [NMDGF]. 2018. Threatened and Endangered Species of New Mexico: 2018 Biennial Review. *NMDGF, Santa Fe*. 155 pp

Niemiller, M. L., Graening, G. O., Fenolio, D. B., Godwin, J. C., Cooley, J. R., Pearson, W. D., ... & Near, T. J. (2013). Doomed before they are described? The need for conservation assessments of cryptic species complexes using an amblyopsid cavefish (Amblyopsidae: *Typhlichthys*) as a case study. *Biodiversity and Conservation*, *22*(8), 1799-1820.

Nixon, K. C., & Wheeler, Q. D. (1990). An amplification of the phylogenetic species concept. *Cladistics*, *6*(3), 211-223.

Owen, R. D., & Hamilton, M. J. (1986). Second record of *Cryptotis parva* (Soricidae: *Insectivora*) in New Mexico, with review of its status on the Llano Estacado. *The Southwestern Naturalist*, 31(3), 403-405.

Palsbøll, P. J., Berube, M., & Allendorf, F. W. (2007). Identification of management units using population genetic data. *Trends in Ecology & Evolution*, *22*(1), 11-16.

Patten, M.A. (2010). Chapter 3: Null expectations in subspecies diagnosis. *Ornitholigcal Monogr*ams 67, 35-41.

Peterson, B. K., Weber, J. N., Kay, E. H., Fisher, H. S., & Hoekstra, H. E. (2012). Double digest RADseq: an inexpensive method for de novo SNP discovery and genotyping in model and non-model species. *PloS One*, *7*(5), e37135.

Pfenninger, M., & Schwenk, K. (2007). Cryptic animal species are homogeneously distributed among taxa and biogeographical regions. *BMC Evolutionary Biology*, 7(1), 1-6.

Posada, D. (2008). jModelTest: phylogenetic model averaging. *Molecular Biology and Evolution*, *25*(7), 1253-1256.

Pritchard, J. K., Stephens, M., & Donnelly, P. (2000). Inference of population structure using multilocus genotype data. *Genetics*, 155(2), 945-959.

Ramasamy, R. K., Ramasamy, S., Bindroo, B. B., & Naik, V. G. (2014). STRUCTURE PLOT: a program for drawing elegant STRUCTURE bar plots in user friendly interface. *SpringerPlus*, *3*(1), 1-3.

Rambaut, A. (2018). Tracer (1.7.1) [Computer software].

Rambaut, A., & Drummond, A. J. (2012). FigTree (1.4.4) [Computer software].

Rambaut, A., & Drummond, A. J. (2020). TreeAnnotator (2.6.2) [Computer software].

Reznick, D., & Endler, J. A. (1982). The impact of predation on life history evolution in Trinidadian guppies (*Poecilia reticulata*). *Evolution*, 160-177.

Rochette, N. C., & Catchen, J. M. (2017). Deriving genotypes from RAD-seq short-read data using Stacks. *Nature Protocols*, *12*(12), 2640-2659.

Rochette, N. C., Rivera-Colón, A. G., & Catchen, J. M. (2019). Stacks 2: Analytical methods for paired-end sequencing improve RADseq-based population genomics. *Molecular Ecology*, *28*(21), 4737-4754.

Rozas, J., Ferrer-Mata, A., Sánchez-DelBarrio, J. C., Guirao-Rico, S., Librado, P., Ramos-Onsins, S. E., & Sánchez-Gracia, A. (2017). DnaSP 6: DNA sequence polymorphism analysis of large data sets. *Molecular Biology and Evolution*, 34(12), 3299-3302.

Ryder, O. A. (1986). Species conservation and systematics: the dilemma of subspecies. *Trends in Ecology and Evolution.*, *1*, 9-10.

Sackett, L. C., Seglund, A., Guralnick, R. P., Mazzella, M. N., Wagner, D. M., Busch, J. D., & Martin, A. P. (2014). Evidence for two subspecies of Gunnison's prairie dogs (*Cynomys gunnisoni*), and the general importance of the subspecies concept. *Biological Conservation*, *174*, 1-11.

Sarver, B. A., Herrera, N. D., Sneddon, D., Hunter, S. S., Settles, M. L., Kronenberg, Z., ... & Sullivan, J. (2021). Diversification, introgression, and rampant cytonuclear discordance in rocky mountains chipmunks (sciuridae: *Tamias*). *Systematic Biology*.

Scornavacca, C., & Galtier, N. (2017). Incomplete lineage sorting in mammalian phylogenomics. *Systematic Biology*, *66*(1), 112-120.

Shuster, C. J. (1989). Genetics and historical biogeography of western peripheral populations of the least shrew (*Cryptotis parva*) *(Doctoral dissertation, University of New Mexico).*

Siemers, J. L., Chen, Y. R., Canestorp, K. M., Sovell, J. R., & Cornelisse, K. L. (2006). Range expansion of the least shrew (*Cryptotis parva*) in Colorado. *The Southwestern Naturalist*, 51(2), 267-269.

Sites Jr, J. W., & Marshall, J. C. (2003). Delimiting species: a Renaissance issue in systematic biology. *Trends in Ecology & Evolution*, *18*(9), 462-47

Sites Jr, J. W., & Marshall, J. C. (2004). Operational criteria for delimiting species. *Annu. Annual Review of Ecology, Evolution and Systematics.*, *35*, 199-227.

Stamatakis, A. (2014). RAxML version 8: a tool for phylogenetic analysis and post-analysis of large phylogenies. *Bioinformatics*, *30*(9), 1312-1313.

Stewart, J. R., & Dalén, L. (2008). Is the glacial refugium concept relevant for northern species? A comment on Pruett and Winker 2005. *Climatic Change*, *86*(1-2), 19.

Sullivan, J., Demboski, J. R., Bell, K. C., Hird, S., Sarver, B., Reid, N., & Good, J. M. (2014). Divergence with gene flow within the recent chipmunk radiation (Tamias). *Heredity*, *113*(3), 185-194.

Toews, D. P., & Brelsford, A. (2012). The biogeography of mitochondrial and nuclear discordance in animals. *Molecular Ecology*, *21*(16), 3907-3930.

Whitaker, J. O. (1974). Cryptotis parva. Mammalian Species, 43, 1-8.

Wiens, B. (2021). A multi-locus perspective reveals connections between island biogeography and evolutionary history of an endangered shrew *(Sorex pribilofensis)* (Master Thesis).

Wilgenbusch, J. C., & Swofford, D. (2003). Inferring evolutionary trees with PAUP. *Current Protocols in Bioinformatics*, (1), 6-4.

Wilson, D. E., & Reeder, D. M. (Eds.). (2005). Mammal species of the world: a taxonomic and geographic reference (Vol. 1). *JHU Press.*

Witt, J. D., Threloff, D. L., & Hebert, P. D. (2006). DNA barcoding reveals extraordinary cryptic diversity in an amphipod genus: implications for desert spring conservation. Molecular ecology, 15(10), 3073-3082.

Woodman, N. (1993). The correct gender of mammalian generic names ending in-otis. *Journal of Mammalogy*, 74(3), 544-546.

Woodman, N. (2002). A new species of small-eared shrew from Colombia and Venezuela (Mammalia: Soricomorpha: Soricidae: Genus *Cryptotis*). Proceedings of the Biological Society of Washington.

Woodman, N. (2011). Patterns of morphological variation amongst semifossorial shrews in the highlands of Guatemala, with the description of a new species (Mammalia, Soricomorpha, Soricidae). *Zoological Journal of the Linnean Society*, *163*(4), 1267-1288.

Woodman, N. (2018). American recent Eulipotyphla: Nesophontids, solenodons, moles, and shrews in the New World.

Woodman, N., & Gaffney, S. A. (2014). Can they dig it? Functional morphology and semifossoriality among small-eared shrews, genus Cryptotis (Mammalia, Soricidae). *Journal of Morphology*, 275(7), 745-759.

Woodman, N., & Morgan, J. J. (2005). Skeletal morphology of the forefoot in shrews(Mammalia: Soricidae) of the genus *Cryptotis*, as revealed by digital X-rays. *Journal of Morphology*, *266*(1), 60-73.

Woodman, N., & Timm, R. M. (1999). Geographic variation and evolutionary relationships among broad-clawed shrews of the *Cryptotis goldmani*–group (Mammalia: Insectivora: Soricidae).

Woodman, N., Matson, J. O., McCarthy, T. J., Eckerlin, R. P., Bulmer, W., & Ordóñez-Garza, N. (2012). Distributional records of shrews (Mammalia, Soricomorpha, Soricidae) from Northern Central America with the first record of Sorex from Honduras. *Annuals of Carnegie Museum*, *80*(3),

Woodman, N., Matson, J., Cuarón, A. D., & de Grammont, P. C. (2016). Cryptotis parva (errata version published in 2017). The IUCN Red List of Threatened Species, 2016.

Zhao, H., Beck, B., Fuller, A., & Peatman, E. (2020). EasyParallel: A GUI platform for parallelization of STRUCTURE and NEWHYBRIDS analyses. *PloS one*, *15*(4), e02321.

Appendix A - Specimen Locality Data

Museum	Museum ID	Species	State	County	Date	Latitude	Longitude
LACM	1593	Cryptotis goldmani	Mexico	Guerrero	7/25/86	17.5566667	-99.685556
LSU	25419	Cryptotis parvus	Louisiana	East Baton Rouge	5/17/82	30.377	-91.223
LSU	26721	Cryptotis parvus	Louisiana	Vernon	11/12/82	31.06	-93.091
LSU	28942	Cryptotis parvus	Louisiana	Cameron	3/1/86	29.666	-92.761
LSU	29128	Cryptotis parvus	Louisiana	East Baton Rouge	11/18/85	30.377	-91.223
LSU	34216	Cryptotis parvus	Louisiana	Cameron	3/17/86	29.677	-92.732
LSU	34413	Cryptotis parvus	Mexico	Mexico	5/6/93	-	-
LSU	36027	Cryptotis parvus	Louisiana	Vernon	5/14/96	31.081	-93.069
TA&M	63771	Cryptotis parvus	Texas	Kenedy	6/26/15	27.1873	-97.829
TA&M	64482	Cryptotis parvus	Texas	Polk	11/7/15	30.679654	-94.698394
TA&M	64483	Cryptotis parvus	Texas	Polk	11/7/15	30.678763	-94.697512
TA&M	64491	Cryptotis parvus	Texas	Polk	11/9/15	30.679654	-94.698394
TA&M	64514	Cryptotis parvus	Texas	Tyler	11/21/15	30.459068	-94.386361
TA&M	64515	Cryptotis parvus	Texas	Tyler	11/22/15	30.459068	-94.386361
TA&M	64538	Cryptotis parvus	Texas	Tyler	1/6/16	30.719964	-94.220962
TA&M	64573	Cryptotis parvus	Texas	Tyler	1/5/16	30.719964	-94.220962
TA&M	64608	Cryptotis parvus	Texas	Polk	1/30/16	30.658372	-94.681101
TA&M	64609	Cryptotis parvus	Texas	Polk	1/30/16	30.658372	-94.681101
TA&M	64619	Cryptotis parvus	Texas	Polk	1/30/16	30.658372	-94.681101
TA&M	64620	Cryptotis parvus	Texas	Polk	1/30/16	30.658372	-94.681101
TA&M	64639	Cryptotis parvus	Texas	Polk	1/29/16	30.658372	-94.681101
TA&M	64694	Cryptotis parvus	Texas	Polk	1/29/16	30.658372	-94.681101
UF	31133	Cryptotis parvus	Florida	Monroe	4/15/04	25.28613	-80.29212

UF	31353	Cryptotis parvus	Florida	Highlands	11/29/05	27.1338056	-81.327778
UF	31390	Cryptotis parvus	Florida	Highlands	11/3/05	27.2050833	-81.367639
UF	31660	Cryptotis parvus	Florida	Polk	11/17/08	27.6890833	-81.555556
UF	31662	Cryptotis parvus	Florida	Polk	11/21/08	27.6874167	81.5529167
UF	31663	Cryptotis parvus	Florida	Palm Beach	10/29/06	-	-
UF	33456	Cryptotis parvus	Florida	Highlands	11/14/07	27.2230278	-81.387583
UF	33463	Cryptotis parvus	Florida	Highlands	10/26/05	27.3623333	-81.330639
UF	33466	Cryptotis parvus	Florida	Highlands	11/6/07	27.2091667	-81.413611
UF	33469	Cryptotis parvus	Florida	Highlands	3/9/06	27.3741389	-81.345778
UF	33470	Cryptotis parvus	Florida	Highlands	3/11/08	27.2172778	-81.390028
UF	33471	Cryptotis parvus	Florida	Highlands	6/12/07	27.21325	-81.396556
UF	33476	Cryptotis parvus	Florida	Highlands	10/7/08	27.5767222	-81.528194
UF	33519	Cryptotis parvus	Florida	Bay	12/26/12	30.38697	-85.85458
ASNHC	8191	Cryptotis parvus	Texas	Tom Green	8/7/92	31.500481	-100.16609
ASNHC	8192	Cryptotis parvus	Texas	Tom Green	7/22/93	31.441828	-100.38563
ASNHC	11116	Cryptotis parvus	Texas	Concho	10/5/97	31.21611	-99.84425
ASNHC	12494	Cryptotis parvus	Texas	Brown	7/6/02	31.630368	-98.928398
ASNHC	13501	Cryptotis parvus	Texas	Brown	4/26/08	31.642526	-98.937688
ASNHC	13502	Cryptotis parvus	Texas	Brown	4/26/08	31.642526	-98.937688
ASNHC	13679	Cryptotis parvus	Texas	Hidalgo	3/21/09	26.171586	-98.387327
ASNHC	14356	Cryptotis parvus	Texas	Hutchinson	10/22/09	29.582072	-96.444072
DMNS	9689	Cryptotis parvus	Colorado	Pueblo	9/21/99	38.31377	-104.30249
DMNS	16603	Cryptotis parvus	Colorado	Jefferson	8/24/16	39.9112722	-105.22258
DMNS	19752	Cryptotis parvus	South Dakota	Lawrence	10/20/09	44.559626	-104.0171
MSB	196185	Cryptotis parvus	Texas	Galveston	2/13/09	29.3385687	-94.901319
MSB	271289	Cryptotis parvus	New Mexico	Chaves	9/27/86	33.452246	-104.40129
MSB	271290	Cryptotis parvus	New Mexico	Chaves	9/27/86	33.452246	-104.40129
MSB	271323	Cryptotis parvus	New Mexico	Roosevelt	7/12/87	34.089072	-103.07723

MSB	271324	Cryptotis parvus	New Mexico	Roosevelt	7/12/87	34.082847	-103.0534
MSB	271655	Cryptotis parvus	New Mexico	Quay	7/5/87	35.181368	-103.69613
MSB	272899	Cryptotis parvus	New Mexico	Chaves	6/11/87	33.474004	-104.41355
MSB	305705	Cryptotis parvus	Kansas	Pottawatomie	2/8/16	39.19496	-96.40307
MSB	305710	Cryptotis parvus	Kansas	Scott	12/20/14	38.45	-100.91
MSB	310263	Cryptotis parvus	Virginia	Clarke	10/15/15	39.079499	-77.962386
NK	305193	Notiosorex crawfordi	New Mexico	Curry	1/27/18	-	-
NK	305195	Cryptotis parvus	New Mexico	Union	10/3/19	36.47589	-103.12643
NK	305196	Cryptotis parvus	New Mexico	DeBaca	9/4/19	34.43496	-104.21777
NK	305197	Cryptotis parvus	New Mexico	DeBaca	9/5/19	34.43497	-104.21777
NK	305198	Cryptotis parvus	New Mexico	DeBaca	9/5/19	34.43504	-104.21794
NK	305199	Cryptotis parvus	New Mexico	DeBaca	9/5/19	34.43504	-104.21794
NK	305201	Cryptotis parvus	New Mexico	Union	10/4/19	36.697189	-103.20547
NK	305203	Notiosorex crawfordi	New Mexico	Santa Fe	7/8/18	35.393877	-105.94595
NK	305204	Cryptotis parvus	New Mexico	Guadalupe	10/17/18	-	-
NK	305205	Cryptotis parvus	New Mexico	Guadalupe	10/17/18	-	-
NK	305206	Cryptotis parvus	New Mexico	Curry	7/13/19	34.30281	-103.05675
NK	305207	Cryptotis parvus	New Mexico	DeBaca	9/5/19	34.43504	-104.21794
NK	305208	Cryptotis parvus	New Mexico	Chaves	11/5/18	33.19246	-104.34093
NK	305209	Cryptotis parvus	New Mexico	Union	10/4/19	36.69724	-103.20552
NK	305210	Cryptotis parvus	New Mexico	Union	10/2/19	36.69719	-103.20547
NK	305211	Cryptotis parvus	New Mexico	DeBaca	9/4/19	34.43504	-104.21793
NK	305212	Cryptotis parvus	New Mexico	Union	10/3/19	36.697187	-103.20546
NK	305213	Cryptotis parvus	New Mexico	Curry	7/25/19	34.30813	-103.31426
NK	305214	Cryptotis parvus	Texas	Hidalgo	3/11/20	26.11193	-98.01154
NK	306132	Cryptotis parvus	Kansas	Morton	7/13/20	37.12129	-101.89703
NK	306133	Cryptotis parvus	Kansas	Morton	7/13/20	37.12151	-101.89674
NK	306166	Cryptotis parvus	Kansas	Morton	7/14/20	37.12143	-101.89674

NK	306167	Cryptotis parvus	Kansas	Morton	7/14/20	37.12135	-101.89693
NK	306540	Cryptotis parvus	Kansas	Barton	7/20/20	38.47502	-98.6682
JF	ET203	Cryptotis parvus	New Mexico	Chaves	10/30/99	33.4632	-104.40642
JF	ET460	Cryptotis parvus	New Mexico	Chaves	10/30/99	-	-
JF	ET465	Cryptotis parvus	New Mexico	Chaves	10/30/99	-	-
JF	FT640	Cryptotis parvus	New Mexico	Chaves	9/10/05	33.464667	-104.40365
JF	FT641	Cryptotis parvus	New Mexico	Chaves	9/11/05	33.431233	-104.41308
JF	FT642	Cryptotis parvus	New Mexico	Chaves	9/11/05	33.443417	-104.40533
JF	FT643	Cryptotis parvus	New Mexico	Chaves	9/11/05	33.480033	-104.42737
JF	FT644	Cryptotis parvus	New Mexico	Chaves	9/11/05	33.479667	-104.42715
JF	FT645	Cryptotis parvus	New Mexico	Chaves	9/24/05	33.475433	-104.39838
JF	FT646	Cryptotis parvus	New Mexico	Quay	9/24/05	35.093417	-103.6135
JF	FT647	Cryptotis parvus	New Mexico	Quay	9/25/05	35.180033	-103.70497
JF	FT648	Cryptotis parvus	New Mexico	Quay	9/25/05	35.210017	-103.74053
JF	FT650	Cryptotis parvus	New Mexico	Quay	9/25/05	35.145067	-103.6147
JF	FT652	Cryptotis parvus	New Mexico	Quay	9/25/05	35.144567	-103.61368
JF	FT653	Cryptotis parvus	New Mexico	Roosevelt	9/30/05	34.096583	-103.04767
JF	FT654	Cryptotis parvus	New Mexico	Chaves	9/12/05	33.28305	-104.35498
JF	FT655	Cryptotis parvus	New Mexico	Chaves	9/13/05	33.28305	-104.35498
JF	FT656	Cryptotis parvus	New Mexico	Chaves	9/13/05	33.28305	-104.35498
JF	FT658	Cryptotis parvus	New Mexico	Chaves	9/13/05	33.314717	-104.33108
JF	FT659	Cryptotis parvus	New Mexico	Chaves	9/13/05	33.3153	-104.33093
JF	FT660	Cryptotis parvus	New Mexico	Chaves	9/2/05	-	-