Conservation genomic assessment of western peripheral populations of the least shrew (*Cryptotis parvus**)

Final Report

New Mexico Share with Wildlife Program
Grant # W-196-R-1

Submitted to:

Share with Wildlife Program
Ecological and Environmental Planning Division
New Mexico Department of Game and Fish
One Wildlife Way
Santa Fe, NM 87507

Submitted by**:

Andrew G Hope, Ph.D.
Fraser J Combe, Ph.D.
Tommy M Galfano (MS in Progress)

Division of Biology
Kansas State University
116 Ackert Hall
Manhattan, KS 66506

T: 785-477-1876
E: ahope@ksu.edu

30 June 2021

*The scientific name of the least shrew was recently corrected to *Cryptotis parvus* in accord with proper nomenclatural gender assignment (Woodman 2018).

**Author Contributions: This work constitutes the basis of the Master's Thesis research project of Mr. Tommy Galfano (projected to graduate December 2021). AGH designed the study, provided oversight for data collection and analyses, and wrote the final report. FJC contributed to data collection, analyses, and report preparation. TMG contributed to data collection, analyses, and report preparation.

Executive Summary

The least shrew (*Cryptotis parvus*) is of management concern within the State of New Mexico, with a listing of Threatened under the New Mexico Wildlife Conservation Act and recognition as a Species of Greatest Conservation Need. This project report summarizes the results of a genomic assessment of New Mexico least shrew specimens in relation to least shrews from throughout their known distributional range. The primary objectives of this project were to 1) determine the genetic distinction of New Mexico shrew populations from each other and how each is related to populations in the broader North American range; 2) assess the potential for locally-adapted gene pools that would warrant recognition of adaptive units, that reflect potential selection and divergence across non-neutral gene regions that may be associated with local fitness; 3) investigate genetic diversity, connectivity, and demographic trends of New Mexico least shrew populations; and 4) detect potential gene flow through hybridization among genetically distinct least shrew populations. Through a combination of field work conducted by the New Mexico Department of Game and Fish in summer 2020 and specimen tissue loans from multiple public research archives, we obtained specimens and sequenced both the mitochondrial cytochrome b gene (n=95) and 19,400 independent nuclear loci (n=68) and performed a series of standard phylogenetic and population genetic analyses to assess genetic lineage distributions, diversity, and relationships among lineages.

Our results indicate that least shrews within New Mexico form two genetically distinct lineages based on nuclear genomes, despite only minimal divergence at the mitochondrial cytochrome b gene, which is commonly used to infer genetic divergence among mammals. Not only are New Mexico populations differentiated from one another, but least shrews from Chaves County, New Mexico, represent a distinct evolutionarily significant unit (ESU), and are also divergent from all other least shrews sampled from the perspective of both neutral and adaptive (non-neutral) genetic loci, suggesting both long term isolation-and-divergence coupled with local adaptations. Genetic demographic trends for this group indicate both historic and contemporary population contractions, and ecologically, these populations are seemingly limited to only a few remnant Cienega wetlands. Least shrews from elsewhere in New Mexico (High Plains) are genetically continuous with shrews from all localities further north and east to the Mississippi River, the latter which represents a significant biogeographic barrier for this species. Genetic demography suggests that High Plains shrews from New Mexico may well have colonized westward from the central Great Plains during the Holocene but are now experiencing a decline. Other least shrews studied likely constitute multiple cryptic species based on very high genetic divergence among some lineages that are currently recognized subspecies. The subspecies *C. p. floridanus* is highly unique and potentially warrants elevation to species status. Likewise, specimens from the distributional range of *C. p. berlandieri* are also highly divergent and may constitute one or more cryptic species.

Based on combined molecular results, it is clear that least shrews within New Mexico should be considered as belonging to two separate ESUs. Further, the Chaves County specimens are likely locally-adapted to limited available habitat. Added concerns for future inventory and monitoring include the presence of individuals associated with the High Plains lineage within Chaves County, having implications for genetic mixing, loss of local adaptations, and potential competition. High Plains least shrews are of much less conservation concern from a genetic perspective, although signals of recent and continuing population decline warrant continued field efforts to better resolve the distribution of this taxon and monitor inter-annual trends. Continued modest specimen collection, including from Chaves County, should be considered in order to

provide a critical resource for increasing resolution in genomic analyses and materials for other emerging techniques, including isotopic/metagenomic analysis of diet, and parasite/pathogen dynamics that might reflect more complex local community dynamics.

Introduction

The least shrew (*Cryptotis parvus*) is a small insectivorous mammal (Order Eulipotyphla) endemic to North America and widespread through the eastern United States (Hall 1981; Woodman et al. 2016). The range of this species (Fig. 1) extends from the Great Lakes region in the northeast, east coast from New York south to include Florida, most of the Gulf Coast, much of Mexico, including the Sierra Madre Occidental, 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, least shrews are limited to southern South Dakota, northern Iowa, southern Wisconsin, and northern Michigan. While Hall (1981) recognized nine morphologically distinct subspecies of least shrew; recent morphometric and preliminary genetic analyses have challenged the validity of these putative taxa. Hutterer (2005) recognized *C. tropicalis* and *C. orphilus* from Central America as distinct species. The subspecies *C. p. pueblensis* and *C. p. soricina* from Southern Mexico remain data deficient. The subspecies *C. p. harlani,* found in Eastern Illinois and Western Indiana, and *C. p. elasson,* found primarily in Ohio, are of questionable validity, pending better sampling and higher resolution genomic analyses (Hutchinson 2010). As such, three subspecies of *C. parvus* are currently considered valid within the United States including: *C. p. floridanus* distributed through Florida, *C. p. berlandieri* occurring from southern Texas through Mexico, and *C. p. parvus* widespread through the Great Plains and eastern United States (Whitaker 1974). Further, the recognized distributional limits of *C. parvus* (IUCN 2008) are not valid; numerous extralimital specimens of least shrews have been sampled and archived in public research collections in recent decades that extend the range of this species westward (Fig. 1). Western populations of least shrew have variably been considered as peripheral, relictual isolates or as leading-edge populations of recent westward expansion, and the westernmost records of locality along with a hypothesized biogeographic history have been discussed in detail by Hafner and Schuster (1996).

Least shrew populations in New Mexico were first identified from Eddy County (Co.) by a single specimen in 1961 (unpublished and only recently discovered; reviewed in NMDGF 2020), which remains the southernmost record of this species from the state, and this population is currently considered extirpated due to severe habitat degradation. Subsequently, specimens were collected in 1981 from the vicinity of Tucumcari (Hoditschek 1985), in 1982 from Grulla National Wildlife Refuge (Owen and Hamilton 1986), and in 1985 from Bitter Lake National Wildlife Refuge (Hafner and Shuster 1996). Since these initial occurrences, further sampling by Schuster (1989), Hafner and Schuster (1996), Frey (2005), and New Mexico Department of Game and Fish (NMDGF, unpublished data) have greatly expanded the known distribution of these shrews in New Mexico through the High Plains and Pecos Valley ecoregions (NMDGF 2020; Appendix A). Given an initial relative lack of knowledge of the distributional extent, population status, or habitat requirements of least shrews in New Mexico, *C. parvus* was listed as Threatened in 1985 and is currently a Species of Greatest Conservation Need (SGCN; Jones and Schmidt 1997; NMDGF 2016, 2018). However, given the much broader sampling distribution since the species was listed and, at least locally, common occurrence of least shrews through eastern New Mexico, the conservation status of this shrew is being reassessed through recovery plan efforts (NMDGF 2020) and modern molecular methods (present study). The ultimate goal of this reassessment is to ensure long-term persistence of robust, representative, and secure populations of least shrew in New Mexico, such that they no longer require protection under the New Mexico Wildlife Conservation Act (NMDGF 2020).

Initial genetic and morphometric evidence based on allozyme and cranial measurement data, respectively, has indicated that least shrews within New Mexico constitute two lineages conforming to geographic regions, with shrews from the vicinity of Bitter Lake National Wildlife Refuge (all within Chaves Co.) being distinct from shrews further north in the vicinity of Tucumcari and Grulla National Wildlife Refuge (Hafner and Schuster 1996). Northern shrews were also more closely related to samples from western Texas. As such, Hafner and Schuster (1996) hypothesized that High Plains and Pecos Valley (herein Chaves Co.) populations reflected distinct subspecies (*C. p. parvus* and *C. p. berlandieri*, respectively). The High Plains populations were thought to be the result of recent westward dispersal, constituting part of the continuous distribution of *C. p. parvus* and therefore of limited conservation concern (Hafner and Schuster 1996). However, Chaves Co. populations were considered to be relicts of a more widely distributed range, dating to the Pleistocene (supported also by fossil evidence from sites in southern New Mexico and Mexico where shrews no longer occur), that became isolated as the region experienced aridification since ~4000 years ago. As such, these populations are of higher conservation concern (Hafner and Schuster 1996). In addition to this initial systematic assessment, the natural history and habitat requirements of least shrews have been investigated to the extent possible based on existing specimens and early ecological studies (reviewed in detail by Whitaker 1974; Frey 2005; NMDGF 2020). From the perspective of New Mexico populations, the general conclusion is that Chaves Co. populations are limited to remnant Cienega wetlands, whereas High Plains populations are found in mesic areas, riparian corridors, and upland grassland habitats and do not appear to be obligately limited to wetlands (Frey 2005; NMDGF 2020). These contrasting habitat associations further support taxonomic distinction of Chaves Co. populations from High Plains populations.

There remains poor resolution of the distributional limits of least shrew populations within New Mexico, their taxonomic status, and their population trends (NMDGF 2020). A molecular approach to monitoring and model-based population assessments may significantly enhance the information gained from more traditional, field-based population estimates and experimentation, and facilitate more effective and holistic wildlife management. As distinct evolutionary units often confer local fitness advantages, adaptive management should address ongoing local population changes from both ecological and evolutionary perspectives. Barbosa et al. (2018) recently developed a conservation genetics framework based on genomic single nucleotide polymorphism (SNP) data that considers both neutral and non-neutral (potentially adaptive) loci as they pertain to different management considerations. Whole genome divergence (considering all genetic loci examined regardless of neutrality) of lineages is considered when identifying evolutionarily significant units (ESUs) for traditional conservation of discrete genetic diversity (Moritz, 1994). However, more recently, consideration of adaptive potential has resulted in recognition of additional, biologically meaningful units of analysis for management purposes. The first are adaptive units (AUs), which are populations for which non-neutral outlier loci have diverged, suggesting that such units are geographically unique due to local adaptations derived through natural selection in response to divergent environmental pressures (Barbosa et al. 2018). Management units (MUs), on the other hand, are geographically discrete populations that have experienced neutral evolutionary divergence from one another, most often due to extended isolation accompanied by genetic drift. These latter groups may be experiencing similar environmental conditions to other MUs (and as such might belong within a shared AU) but, given their relative geographic isolation and neutral genetic divergence, should be considered independently for ongoing management.

Here we performed a genomic assessment of least shrews to: 1) determine the genetic distinction of New Mexico shrew populations from each other and how each is related to populations in the broader North American range of least shrews that might lend to diagnoses of potential ESUs based on thousands of independently evolving genomic loci; 2) assess the potential for locally-adapted gene pools that would warrant recognition of AUs based on functional genes or, conversely, MUs based on neutrally evolving divergence among populations; 3) investigate genetic diversity, connectivity, and demographic trends of New Mexico least shrew populations to address the extant hypotheses of relict population persistence (i.e., Chaves Co. shrews) and recent westward expansion (i.e., High Plains shrews) and to suggest continuing population demographic trajectories; and 4) detect potential gene flow through hybridization among genetically distinct populations that might have implications for the continued genomic integrity of local adaptations among populations.

Methods

Sampling

Field sampling planned for summer 2020 throughout eastern New Mexico in order to collect additional shrew specimens and associated tissues for genetic analyses was curtailed due to COVID-19 pandemic travel restrictions. However, limited field sampling was conducted in 2020 by NMDGF staff at both previous localities of record for least shrews and at multiple sites where shrews had not previously been detected (NMDGF 2020). Additional samples of least shrews from across their North American distribution were obtained on loan from multiple public research archives (Appendix A). This included genetic data from *C. orophilus* and *C. tropicalis* (mitochondrial cytochrome b [Cytb] sequences retrieved from GenBank) and tissues from *Cryptotis goldmani* (n=1), *Blarina brevicauda* (n=1), *Blarina carolinensis* (n=2), and *Notiosorex crawfordi* (n=2) as outgroup taxa.

Mitochondrial cytochrome b sequencing and analyses

Samples and sequencing – Genomic DNA was extracted from tissue for each sample (Appendix A) following the New England Biolabs (NEB; Ipswich, Massachusetts) Monarch Blood DNA extraction kit using manufacturer's instructions. The full Cytb gene was amplified for most specimens (n=95) using primers MSB05/MSB14 (Hope et al. 2010). PCR reagents and conditions were: 1 μL DNA template (variable concentration); 1.5 μL each of dNTPs (10 mM), MgCl (25 mM), 10x PCR buffer, and Bovine Serum Albumin (1%); 0.2 μL of each primer (10 mM); 0.08 μL AmpiTaq DNA polymerase (Applied Biosystems, Foster City, California); and 7.52 μL ddH_2O to total 15 μL reactions. PCR included initial denaturation at 94°C for 6 min, followed by 40 cycles of denaturation at 94°C for 25 s, annealing at 50°C for 30 s, extension at 72°C for 1 min, and final extension at 72°C for 5 min, with cooling at 15°C for 10 min. PCR products were cleaned using ExoSAP-IT (Affymetrix, Santa Clara, California) diluted 1:10. All PCR products were confirmed on a 2% electrophoresis gel and sequenced in both directions (on an Applied Biosystems (Foster City, California) ABI 3730 sequencer at Genewiz LLC. Raw reads were cleaned and aligned using Geneious (Kearse et al. 2012). The final dataset consisted of an alignment of 1,103 base pairs (bp) of the Cytb gene. All sequences will be deposited in

GenBank. Additional archived Cytb sequence data were downloaded from GenBank for phylogeny reconstruction (n=11).

Phylogeny estimation – We estimated an independent phylogeny for the Cytb locus including incomplete sequences and without assigning haplotypes. We produced chronograms (dated trees) through Markov Chain Monte Carlo (MCMC) searches in the Bayesian phylogeny analytical program BEAST2 (Bouckaert et al. 2019), setting all parameters in BEAUti, a graphical user interface module of the BEAST2 software package, and estimating the substitution model through use of the bModelTest package, also in BEAST2 (Boukaert and Drummond 2017). To account for potential variable evolutionary rates through the history of evolution of least shrews, we applied a relaxed clock: uncorrelated log-normal molecular clock model and set the mutation rate to 0.055, a value for relative speed of evolutionary divergence previously estimated for other shrew species (5.5% per million years; Hope et al. 2010). We used empirical base frequencies and assumed a constant population size tree prior, with other parameters run with default settings. We ran MCMC for 50 million generations, sampling every 1,000 generations, with the first 1,000 trees discarded as burn-in. Stationarity of MCMC runs was assessed in Tracer v1.7 (Rambaut 2018). We annotated tree files in TreeAnnotator (BEAST2 package). Chronograms were visualized with posterior probabilities in FigTree v1.3.1 (Rambaut 2012) and reported as a mid-point rooted tree.

Genetic diversity and population demographics – Cytb genetic diversity and demographic analyses were performed considering distinct genetic lineages retrieved from the Cytb phylogeny (see previous section), as well as considering shrews from Chaves Co. and the High Plains of New Mexico separately. Other groups analyzed included Florida, South Texas, Mexico, all shrews east of the Mississippi River (excluding Florida), and all shrews west of the Mississippi River (excluding South Texas and Mexico). For each group, we calculated summary statistics and assessed haplotype diversity (*Hd*), nucleotide diversity (π), and pairwise sequence divergence in DnaSP v5 (Librado and Rozas, 2009). For tests of demographic expansion, we used DnaSP to calculate Tajima's *D*, a statistic commonly used in a phylogeographic context to assess changes in effective population size (Tajima 1989), assessing significance with 10,000 coalescent simulations.

Genomic DNA sequencing and analyses

Sequencing – DNA was quantified using Quant-iT Picogreen dsDNA Assay (Invitrogen, Waltham, Massachusetts) and gel electrophoresis (2% agarose). Samples with sufficient yields of high molecular weight DNA (>100 ng; n=68) were submitted to the University of Minnesota Genomics Center (UMGC), Minneapolis for double digest restriction-site associated DNA sequencing (ddRADseq) amplification and sequencing. Following an in-silico digest of the reference genome, it was determined in silico which restriction enzymes were optimal using a sub-sample of 8 individuals. UMGC prepared ddRADseq libraries and sequenced samples using the following protocols. For each sample, 100ng of DNA was digested with 10 units each of *Sbf*I and *Taq*I restriction enzymes from NEB and incubated at 37° C for 2 hours before heat inactivating at 80° C for 20 minutes. Samples were then ligated with 200 units of T4 ligase (NEB) and with phased adaptors with CRYG and CG overhangs (reflecting DNA base symbols and standard genetic ambiguity codes) at 22° C for 1 hour before heat killing. The ligated

samples were purified with solid phase reversible immobilization (SPRI) beads and then amplified for 18 cycles with 2x NEB Taq Master Mix to add unique barcodes to each sample. Libraries were purified, quantified, pooled, and size selected for the 300-744bp library region and diluted to 2nM prior to sequencing. UMGC sequenced 150bp single-end reads across 0.25 lanes of a NextSeq 550 High-Output FlowCell (Illumina, San Diego, California). The resulting fastq files were demultiplexed using Illumina bcl2fastq software and Trimmomatic (Bolger et al. 2014) was used to remove adapter sequences (the first 12 bases) from the 3' ends of reads.

Data filtering – Raw Illumina reads were inspected with FASTQC software (available at http://www.bioinformatics.babraham.ac.uk/projects/fastqc). To process ddRADseq data and produce SNP datasets, we used the *process_radtags* module of STACKS 2.5 (Rochette et al. 2019) to filter out low-quality reads. Loci were discovered *de novo* using the *denovo_map.pl* pipeline in STACKS. First, the parameters controlling loci formation and polymorphism were optimized for the dataset following the recommendations provided by the software developers (Rochette and Catchen 2017), and a locus catalog was built using the optimal parameters. Single nucleotide polymorphisms (SNPs) were then called for each locus. A total of 19,676 polymorphic SNPs were identified, with an average read depth of 27.9. We then ran *ref_map* and *populations* pipelines within STACKS, retaining loci found in at least 80% of samples (r = 0.80), with a minor allele frequency of at least 5% (min_maf =0.05), and heterozygosity upper bound of 0.8 (max_het = 0.8) that produced a variant call format (VCF) file. We also only retained one SNP per locus to avoid spurious genetic inference from using multiple variable, but linked, base sites within a given locus (write_single_snp) and therefore meet the assumptions of linkage equilibrium in subsequent analyses (e.g., Discriminant Analysis of Principal Components [DAPC] and STRUCTURE; see sections below). The VCF file was filtered using VCFtools (Danecek et al. 2011) to only include reads with a minimum read depth of 20, remove SNPs out of Hardy-Weinberg equilibrium, and keep individuals with less than 50% missing data. We then used this filtered VCF file for all subsequent analyses. These filters retained a total of 19,400 polymorphic SNPs for 68 individuals across all *Cryptotis parvus* and a single specimen of *Blarina carolinensis* as outgroup.

Genetic diversity and relatedness – Numbers of private alleles (genetic variants unique to a single population), expected (H_E) and observed (H_O) heterozygosity, nucleotide diversity (π), and inbreeding coefficients (F_{IS}) were calculated for each sampled population (reflecting population designations consistent with the Cytb phylogeny) in STACKS software using the *populations* module. Estimates of pairwise F_{ST} were calculated in STACKS on a site-by-site basis between populations and then averaged across loci. F_{ST} statistics were calculated to attempt to understand the degree of variation attributable to putative population structure.

Identification of neutral vs. non-neutral loci – To consider the three tiers of conservation genetic units (ESUs, AUs, and MUs), we performed both clustering analyses and phylogenetic reconstructions (see sections below) using all genomic loci, using only non-neutral (potentially adaptive) loci, and finally using only neutral loci, requiring separation of the full dataset into neutral and non-neutral datasets. This separation was performed using two statistical tools for the identification of loci putatively experiencing selection (outlier loci): BayeScan v2.1 (Foll and Gaggiotti 2008) and *PCadapt* v4.1.0 (Luu et al. 2017). For these analyses, we included all samples of *Cryptotis parvus.* We used BayeScan to estimate the posterior probability that a given

locus is affected by selection. Briefly, populations were designated as Chaves Co., West of Mississippi, East of Mississippi, South Texas, Mexico, and Florida. Prior odds of 10 (prior belief that a selection model is 1/10 as likely as a neutral model for a given SNP), 100, and 1000 were used for identifying the top candidates for loci experiencing selection. A total of 50,000 reversible-jump Markov Chain Monte Carlo chains were run with a thinning interval of 10, following 20 pilot runs of 5,000 iterations each and a burn-in length of 50,000. An R function, provided along with the BayeScan software package, was used to plot and identify outliers using different criteria from the BayeScan output file. Recent development of multivariate methods, such as *PCAdapt* (Luu et al. 2017), allows for the identification of outlier loci in admixed or continuous populations. For this analysis, individuals are not sorted into predefined populations. Instead, *PCadapt* ascertains population structure using principal component analysis (PCA) and then identifies markers under putative selection as those that are excessively correlated with population structure. A scree plot of the first 20 principal components (PCs; termed *K* in *PCadapt*) indicated that the optimal *K* from our data was 5 for computing correlations between loci and *K* principal components. We used Benjamini and Hochberg's (1995) method for correction of the false discovery rate in both BayeScan and *PCadapt* at $\alpha = 0.05$.

Based on the results of both analyses, we separated our ddRADseq data into neutral and non-neutral loci using a custom script (https://github.com/fraser-combe). This uses VCFtools to create separate VCF files for neutral and non-neutral loci. For non-neutral loci, we conservatively included loci found to be under selection from both outlier methods for further analyses. For subsequent analyses including genetic clustering, hybridization, and phylogeny reconstruction (see sections below), we analyzed neutral (17,850 loci) and non-neutral (1,550 loci) SNP datasets separately.

Phylogeny estimation, genetic clustering, and hybridization – In order to estimate systematic relationships based on the SNP dataset, we used a maximum likelihood (ML) method to estimate a phylogeny using RAxML v8.2.12 (Stamatakis 2014). To estimate the ML tree, SNPs were first converted to phylip format using ambiguity codes for heterozygous sites and Ns for uncalled sites following standard ambiguity code notation. We then estimated the ML phylogeny with 100 bootstrap replicates, implementing the GTR+G nucleotide substitution model.

We investigated the presence of population structure and hybridization by separately analyzing the neutral and non-neutral SNP datasets in the software program STRUCTURE version 2.3.4 (Pritchard et al. 2000). Assignment tests using the admixture model in STRUCTURE assume that SNPs are not linked. Therefore, only the first SNP per locus was included in STRUCTURE input data matrices using the write_single_snp option for populations in STACKS to ensure only putatively unlinked SNPs were used. In STRUCTURE, a Bayesian algorithm was used to assign individuals to a value of K clusters. The likelihood that a given individual belongs to a particular cluster is given by a Q-value. Higher Q-values indicate a greater posterior probability that an individual belongs to that cluster. We executed runs with a burn-in of 10 iterations followed by 10 iterations and performed 10 replicate runs for K = 1 through K = 10. For the STRUCTURE analyses, we set the parameters to allow for admixture between clusters and selected the default correlated allele frequency model. The likely number of genetic clusters (K) was selected by evaluating mean likelihood scores and ΔK across all replicate runs implemented in Structure Harvester (Evanno et al. 2005; Earl 2012) and choosing the value of K with the highest likelihood.

DAPC is a multivariate approach that performs a PCA in a first step and then subjects the PC scores to a discriminant function analysis (DFA). Unlike PCA, DFA fits orthogonal discriminant functions that maximize between-group relative to within-group variation, making it well-suited to differentiating between genetic groups (Jombart 2008). A K-means clustering approach can be applied to assess the number and composition of K genetic clusters in the data. The best supported model is identified using the Bayesian Information Criterion (BIC), where the lowest BIC score is preferred. We performed PCA and DAPC in the 'adegenet' package (Jombart 2008) in R (R 3.6.3, R Core Team, 2020). The optimal number of PCs to be retained was determined using the a.optim.score function with 10,000 simulations for each number of PCs retained. In both PCA and DAPC analyses, we retained the first 10 PCs and we retained all the discriminant functions in the DAPC analyses.

We also generated a phylogenetic network analysis using SplitsTree4 version 4.14.6 (Huson and Bryant 2006), which takes account of more realistic models, such as loss and duplication events, hybridization or recombination, and provides an output network tree illustrating inferred relationships among individuals, where potential hybrids would be spatially intermediate to clusters of respective "parent" populations. We used Phylip files based on ddRADseq data. Within the program, we used uncorrected p-distances, NeighborNet, and then EqualAngle to compute an unrooted network for all populations/individuals. In order to test each split, our matrix was bootstrapped with 100 replicates using default parameters.

Inferred demographic histories – The demographic histories of our populations, again grouped to reflect the genetically distinct clades recovered from the Cytb phylogeny, were inferred using STAIRWAY PLOT V2 software (Liu and Fu 2020). This method has a demonstrated utility for ddRADseq data and considered effective population size (N_e) changes through the evolutionary time frames of our study species. In the absence of direct estimates of *Cryptotis* mutation rates for the nuclear genome, we used the default setting of $1.2x10^{-8}$ per site per generation. This mutation rate is similar to those estimated for other mammals (Kumar and Subramanian 2002; Roach et al. 2010; The 1000 Genomes Project Consortium 2010; Campbell et al. 2012; Liu et al. 2014) and those used in other demographic history analyses of mammals with unknown mutation rates (MacLeod et al. 2013; Beynon et al. 2015; Benazzo et al. 2017; Hansen et al. 2018). Median effective population sizes and 95% confidence intervals were estimated based on 200 bootstrap replicate analyses.

Results

Mitochondrial cytochrome b phylogeny, diversity, and demographics

The Cytb phylogeny reflects novel perspectives on the history of diversification among populations of *Cryptotis parvus*, identifying multiple well-supported lineages that are generally coincident with geography (Figs. 1, 2). These include shrews that are currently considered members of the subspecies *C. p. parvus* being represented by two well supported lineages, one consisting of specimens east of the Mississippi River, with the exception of four samples collected from the east bank of the Mississippi River in Louisiana (Appendix A), and a second lineage consisting of all remaining specimens collected from west of the Mississippi River (Fig. 1, 2). Within this latter group, there is no strong support for distinct Cytb lineages of shrews

within New Mexico, although all shrew specimens from Chaves Co. are found in one of two loosely supported lineages including individuals from central/southern Texas and Louisiana, and all shrew specimens from the High Plains of New Mexico are very closely related to shrews from Colorado, South Dakota, Wyoming, Kansas, and northern Texas (Fig. 2). Among the other shrews sampled, two shrews from South Texas and a single specimen from the Sierra Madre Occidental of Mexico form a well-supported lineage coincident with the subspecies distribution of *C. p. berlandieri*. However, the specimen from Mexico is highly divergent from the South Texas specimens. Shrew samples from Florida all form a highly distinct Cytb lineage coincident with the subspecies *C. p. floridanus*. Of interest, the recognized species *C. tropicalis* and *C. orophilus* from Central America fall within the multiple supported lineages of *C. parvus*, rendering the least shrew paraphyletic and suggesting that the current taxonomy does not adequately reflect the existing diversity.

Mitochondrial genetic divergence among well supported groups of *Cryptotis* is generally very high (Table 1) with South Texas and Florida lineages (representing *C. p. berlandieri* and *C p. floridanus*) having between ~9% and 11% average sequence divergence from *C. p. parvus*. Within *C. p. parvus*, shrews from the High Plains of New Mexico are <1% divergent from Chaves Co. shrews (Table 1). In terms of genetic diversity, High Plains shrews exhibit very low genetic diversity, low haplotype diversity, and no significant signal of demographic expansion (from Tajima's *D*). Shrews from Chaves Co. exhibit much higher genetic diversity but a strongly positive value of Tajima's *D* (Table 2). Together these results for Chaves Co. are indicative of a relict population (retaining ancestral genetic diversity) that has experienced sudden and recent population contraction.

Genomic DNA phylogeny, clustering, diversity, and demography

The maximum likelihood phylogeny based on the complete nuclear SNP dataset (19,400 loci) indicates similar relationships to those identified based on analysis of the maternally-inherited Cytb locus, with some notable exceptions (Fig. 3). It should be noted that tissue samples were not available with which to sequence nuclear DNA from *C. tropicalis* or *C. orophilus*, and only one sample from the East of Mississippi lineage, as identified by the Cytb analysis, was available on loan for nuclear sequencing. However, all four specimens from the east bank of the Mississippi River in Louisiana are more closely related to the single sample from Virginia and form a well-supported lineage based on the nuclear genome. These four specimens contained Cytb haplotypes more similar to shrews from the West of Mississippi Cytb lineage, indicating at least limited gene flow and interbreeding between west and east lineages across this river typically considered as a biogeographic dispersal barrier. Likewise, a single specimen from Kennedy Co., Texas that grouped with the West of Mississippi lineage based on Cytb data was more closely related to putative *C. p. berlandieri* (South Texas lineage) from the Rio Grande Valley of southern Texas, indicating another region of gene flow between distinct lineages. Of interest, all specimens from Chaves Co., except one individual (FT644; Appendix A), formed a well-supported lineage, which is divergent from all other shrews within the West of Mississippi lineage (Fig. 3). The single exception (FT644) from Chaves Co. appears to have both Cytb (Fig. 2) and nuclear (Fig. 3) signatures more closely related to shrews from the High Plains of New Mexico, indicating sympatry between the genetically distinct High Plains and Chaves Co. groups in the vicinity of Bitter Lake National Wildlife Refuge (Bitter Creek area) but as yet with no evidence of gene flow (Fig. 4). All shrews from the High Plains of New Mexico were

grouped as West of Mississippi. Additionally, least shrews from Florida and least shrews from Mexico again each form highly divergent lineages from all other specimens of *C. parvus* (Fig. 3).

From clustering analysis using STRUCTURE, the complete nuclear SNP dataset recovered K=4 as the most strongly supported number of groups (Fig. 5). These groups included: Florida; South Texas (inclusive of the single individual from Kennedy Co., Texas); Chaves Co., NM; and West of Mississippi. Shrews from East of Mississippi were clustered with West of Mississippi (mostly purple; Fig. 5). The single specimen from Mexico was ambiguously placed, intermediate between Florida and South Texas. The ordination plot based on DAPC analysis of the complete nuclear SNP dataset was generally congruent with the STRUCTURE results; West of Mississippi (including all shrews from High Plains of NM and the FT644 specimen from Chaves Co.) and East of Mississippi were similar but slightly separated within ordination space (Fig. 6). Florida; South Texas; and Chaves Co., NM were all genetically distinct and the single specimen from Mexico was located mid-way between Florida and South Texas.

Considering only the 17,850 neutrally evolving SNPs (Fig. 7), most spatial genetic relationships are the same as for the complete dataset. Again, shrews from Chaves Co. NM and Florida are highly divergent. West of Mississippi and East of Mississippi are closely related but still distinct. The Mexico specimen is still intermediate between Florida and South Texas. The major change based on neutral loci is that South Texas (putative *C. p. berlandieri*) specimens are very closely related to West of Mississippi shrews representative of *C. p. parvus*. There is also notable spatial structure among groups of shrews within West of Mississippi.

Considering only the 1,550 non-neutral SNPs (Fig. 8), Florida and South Texas are most divergent from other specimens, with the Mexico specimen again intermediate. In this plot, Chaves Co. shrews are again distinct from both West of Mississippi and East of Mississippi, but only minimally, and the West and East of Mississippi groups are genetically very similar. Based on SplitsTree analyses, the DAPC spatial genetic groups are recovered again in the form of phylogenetic networks for both neutral and non-neutral datasets (Figs. 9, 10). The primary observation based on both of these networks is that several distinct genetic clusters exist, when considering either neutral or non-neutral loci, including Florida, Mexico, South Texas, West of Mississippi, East of Mississippi, and Chaves Co., NM.

In relation to genetic diversity and demography, we again analyzed the major phylogenetic groups (Florida, Mexico, South Texas, West of Mississippi, East of Mississippi, and Chaves Co., NM) using several statistics. Pairwise F_{ST} values indicate that Chaves Co. shrews were least differentiated from West of Mississippi specimens, followed by East of Mississippi (Table 3). Differentiation of South Texas, Mexico, and Florida was much higher from all other shrews. The observed heterozygosity of Chaves Co., NM and West of Mississippi (including High Plains, NM) were comparable, but nucleotide diversity of Chaves Co. shrews was lower (Table 4). Given that all F_{IS} values were close to zero, no groups of shrews exhibited signs of significant inbreeding, suggesting random mating within populations (Table 4). Considering only New Mexico shrew specimens, we calculated change in effective population size through time (Figs. 11, 12). High Plains shrews indicated a sharp population decline within the last ~4,000 years and a small contemporary effective population size (Fig. 11). Specimens from Chaves Co. provided lower historic estimates of effective population size than High Plains specimens and a signal of significant population decline coincident with warming following the Last Glacial Maximum (~20,000 years ago; Fig. 12). However, high confidence intervals surrounding contemporary population size trends leaves modern effective size inconclusive for the Chaves Co., NM population.

Discussion

Conservation and management of biodiversity can significantly benefit from understanding the evolutionary relationships of the focal population in the broader context of an entire species, or a group of closely-related species, and considering the distributional limits of distinct genetic units of analysis (Barbosa et al. 2018). Regional populations may reflect local adaptations and/or extended isolation and divergence from other populations. By using advanced molecular methods, such phylogeographic assessments can diagnose distinct regional diversity, in some instances including recognition of morphologically cryptic species (Allendorf et al. 2010).

Our study of least shrews, with a focus on New Mexico populations peripheral to the broader species' distribution, highlights these concepts from multiple perspectives. Our major findings include 1) least shrews from throughout their range likely include multiple species as opposed to a single species; 2) New Mexico populations do indeed represent two distinct intra-specific genetic lineages; High Plains shrews are genetically indistinguishable from shrews further north and east and Chaves Co. shrews are genetically highly unique and appear to be endemic to the Pecos Valley of New Mexico; and 3) Chaves Co. shrews retain relatively high genomic diversity but exhibit genetic signatures of recent population contraction, and High Plains shrews have higher genomic diversity as part of a more widely distributed lineage, but also indicate recent population decline.

Rangewide phylogeography

A preliminary genetic and morphometric study of least shrews with a focus on eastern North America and based only on short, mitochondrial sequences and a single, low resolution nuclear gene recovered three primary genetic lineages associated with shrews from Florida, from east of the Mississippi River, and from west of the Mississippi river, although morphological evidence was less well resolved (Hutchinson 2010). No samples from western peripheral populations of New Mexico were included in this study. Similarly, the study by Hafner and Schuster (1996) had only a regional focus and did not place New Mexico least shrews within a broader context. Given more extensive numbers of samples in the present study and more comprehensive genomic and geographic representation, we have provided further perspectives on these preliminary studies. Based on all of our analyses, including Cytb data and both adaptive and neutral nuclear loci, least shrews from Florida are highly unique. Hutchinson (2010) also found samples from Georgia to belong to this lineage, although we had no samples from Georgia to include to further evaluate this conclusion. From a purely phylogenetic perspective, Florida shrews are strongly indicative of a morphologically cryptic but independent species that warrants taxonomic revision. Our dataset included other highly distinct genetic lineages, specifically shrews from Mexico and shrews from South Texas. Based on Cytb data, these lineages are more distantly related to other *Cryptotis parvus* than to the currently recognized species *C. tropicalis* and *C. orophilus*. Unfortunately, given that we had no nuclear DNA from these latter species from Central America, we cannot confirm these relationships from a nuclear perspective. However, our data do suggest that the single shrew from Mexico and three individual shrews from southern Texas are genetically divergent from all shrews further north. Species limits

among these southern taxa warrant further consideration, including species delimitation analysis, and may represent additional cryptic species.

Our data indicate that the species *C. parvus* (sensu stricto) minimally includes all least shrews from both east and west of the Mississippi River within the United States, excluding our Florida samples and South Texas samples. Within this group and based on a large nuclear dataset, we recognize three distinct lineages including: Chaves Co. specimens, all specimens west of the Mississippi River, and all specimens east of the Mississippi River. Of these three lineages, the Cytb locus indicated that Chaves Co. and West of Mississippi lineages were minimally distinct (i.e., not separate clades) and together were distinct from East of Mississippi specimens. However, the nuclear data indicate that Chaves Co. shrews are markedly more divergent from West of Mississippi and East of Mississippi lineages than the latter two lineages are from each other. Such discordance between mitochondrial and nuclear data is increasingly common as genomic studies become more advanced (Coates et al. 2018). Primary explanations for such discordance include both incomplete lineage sorting of the single mitochondrial gene (where shared ancestral haplotypes are retained among multiple populations even after extended isolation) and hybridization (e.g., Colella et al. 2019; Linck et al. 2019). Both of these dynamics may be occurring in least shrews. For instance, fixation of a few mitochondrial haplotypes within the Chaves Co. population that are still also common among populations west of the Mississippi may have allowed these haplotypes to persist within this population in isolation, while the nuclear genome progressively diverged, resulting in unsorted (non-unique) mitochondrial DNA but divergent nuclear DNA. In terms of hybridization, mitochondrial capture among mammals has been documented in chipmunks (e.g., Good et al. 2008) and other mammals, where infrequent hybridization leaves a signal of the maternally inherited mitochondrial DNA from one lineage, while retaining the nuclear signature from the other lineage through a process of back-cross breeding with the original lineage. In the case of Chaves Co. it is possible that previous limited contact between Chaves Co. shrews and least shrew populations further east resulted in capture and propagation of mitochondrial haplotypes from the West of Mississippi lineage, followed by back-crossing of hybrid shrews with other Chaves Co. shrews. Particularly in small populations, such as Chaves Co., this could quickly lead to the fixation of captured mitochondrial haplotypes and loss of any haplotypes unique to Chaves Co. Mitochondrial capture most often occurs across geographic transition zones between genetic lineages. This is clearly evident from our data, supporting two episodes of mitochondrial capture between least shrew lineages across the Mississippi River in Southern Louisiana and also in southern Texas. The latter transition zone in Texas is a recognized phylogeographic break for multiple other mammal species (Anderson and Light 2012; Hope et al. unpublished data).

The combined results allow for a reconstruction of phylogeographic history for least shrews that have evolved in response to major episodes of environmental change during late-Pleistocene glacial cycles. During glacial episodes, least shrews were likely isolated in multiple refugial areas, including the Florida panhandle, southern Mexico, possibly northern Mexico, and east and west of the Mississippi River. In addition, given the high endemicity of Chaves Co. specimens, we suggest an additional refugium in the lower Pecos Valley. With additional historic records of least shrews from Eddy Co. (NMDGF 2020) and based on fossil evidence from Eddy and Hidalgo Counties (Hafner and Schuster 1996), this Pecos Valley refugial population was likely significantly more widespread through southern New Mexico, followed by recent contraction as postulated elsewhere (Hafner and Schuster 1996). Shrews of the High Plains of New Mexico likely expanded westward into the High Plains during the Holocene from the

refugium west of the Mississippi River, but our results suggest that current trends are of population decline along the westernmost range limits of the least shrew. This decline may reflect degradation of available habitat and environments through local land practices and/or ongoing climate change.

Conservation implications for New Mexico least shrews

As genomic methods become increasingly advanced in terms of resolving genetic relatedness, as well as functional implications for continued evolution, additional analytical frameworks are being developed to consider what these data mean for biodiversity conservation (Barbosa et al. 2018; Coates et al. 2018). Consideration of ESUs was a major advance in recognition of the importance of considering unique evolutionary diversity when developing management actions (Moritz 1994). Given that genomes contain both functional genes on which natural selection can act, and also neutrally evolving regions that diverge, for instance, through genetic drift of isolated populations, we can now consider populations from additional perspectives, depending on conservation priorities. ESUs consider all available genomic data and represent populations or lineages that are simply genetically different from each other, and therefore worthy of focused conservation efforts. However, if populations are more specifically divergent from each other at non-neutral loci, this instead suggests that one or more of these populations is adapted to local conditions, and may be considered an AU. This has strong implications for possibly detrimental impacts from future translocation efforts that bring animals from one AU to another and in particular suggests that local habitats and environments are critically important for future conservation. However, if divergence of populations is the result of neutrally evolving loci, this suggests that populations are experiencing limited or no interaction and are simply diverging through a lack of gene flow. Such populations may be considered as separate MUs given that they are demographically independent. These dynamics are reviewed in detail by Barbosa et al. (2018).

Focusing on least shrews from Chaves Co., and considering the large nuclear dataset, specimens are highly unique based on all loci, only neutral loci, and only non-neutral (adaptive) loci. As such, they should be considered an ESU, MU, and AU, respectively, although from the perspective of non-neutral loci, the Chaves Co. cluster is less divergent, but still distinct, from other lineages within *C. p. parvus*. Considering the other unambiguous *C. parvus* lineages (East of Mississippi, West of Mississippi), each is an ESU based on the Cytb dataset, although there is less strong support for these two lineages being ESUs based on clustering analyses (STRUCTURE and DAPC plots) of the full nuclear dataset. Likewise, based on non-neutral loci, both are genetically very similar, indicating that they could be considered a single AU, and likely suggesting that habitat and environmental conditions for least shrews across much of their range (including the High Plains of New Mexico) are comparable, with little local adaptation. However, these lineages are distinct based on the neutral dataset, indicating different MUs and, again, reflecting isolation and subsequent genetic divergence across the Mississippi River. Of interest, there also exists substructure within the West of Mississippi lineage based on neutral loci (Fig. 7), and future detailed analyses should determine if High Plains shrews themselves warrant designation as a MU.

Conclusions and Future Considerations

All least shrews within New Mexico sampled so far should be considered as *C. parvus*. However, Chaves Co. specimens are unique from High Plains specimens and are endemic to wetlands within Chaves Co. These shrews also should not be considered as the sub-species *C. p. berlandieri* as hypothesized by Hafner and Schuster (1996) based on morphological similarity of two cranial measurements. As such, Chaves Co. specimens are not assignable to an existing sub-species designation, as they are also significantly genetically divergent from other shrews assigned to *C. p. parvus*. He et al. (2015) first indicated that least shrews west and east of the Mississippi River were genetically distinct based on only a few samples, and Woodman (2018) suggested that these likely constituted separate species. Our results do not support clear species-level designation between west and east of the Mississippi, but considering that Chaves Co. shrews are more highly differentiated from either West of Mississippi or East of Mississippi lineages, they would clearly constitute an additional, yet undesignated, sub-species within *C. parvus*. Unfortunately, given that all samples from eastern New Mexico post-date original taxonomic descriptions, no potential sub-specific Holotype from Chaves Co. has yet been formally described. Regardless, the Chaves Co. lineage should be considered locally-adapted and a clear ESU for independent management consideration. High Plains shrews are not unique from least shrews further north or east, but signals of population decline may warrant additional habitat remediation and continued population monitoring. All New Mexico shrew populations (as well as historic localities and potentially new localities) warrant additional field surveys. This should include some modest element of continued specimen acquisition, given that time-series of specimens not only vouch for the existence of these shrews at a given locality and time (without which this sort of phylogeographic assessment would not be possible), but can also be used to examine in more depth the life histories of these shrews through numerous potential avenues (Hope et al. 2018), including diet through metagenomics and isotopic analysis (e.g., Hope et al. 2021), parasite and microbiome/pathogen diversity (e.g., Greiman et al. 2019), and more in-depth morphometric and functional genomic analyses. There is a positive outlook on least shrew conservation within New Mexico, given that genetic diversity within both lineages is relatively high and much of the available habitat at existing localities of species occurrence is currently protected or specifically managed for wildlife (NMDGF 2020). However, a critical need for further survey efforts will be to establish the extent of occurrence of High Plains shrews within Chaves Co.; two High Plains shrews appear to be sympatric with Chaves Co. shrews at Bitter Lake National Wildlife Refuge in the vicinity of Bitter Creek (Figure 4). This presents the potential for outbreeding depression through loss of local adaptations as a result of hybridization or may simply lead to competition among these lineages.

Additional future work should include sequencing individuals from other localities along the Pecos River Valley, since the existing habitat resembles that in Chaves Co. Such spring-fed wetlands, for instance in the vicinity of the Rock Lake fish hatchery in Guadalupe Co., NM, are geographically proximate to Chaves Co. and potentially inter-connected.

References

1000 Genomes Project Consortium. (2010). A map of human genome variation from population-scale sequencing. *Nature*, *467*(7319), 1061.

Allendorf, F. W., Hohenlohe, P. A, Luikart, G., Ouborg, N. J., Pertoldi, C., et al. (2010). Genomics and the future of conservation genetics. *Nature Reviews. Genetics*, *11*(10), 697–709.

Andersen, J. J., & Light, J. E. (2012). Phylogeography and subspecies revision of the hispid pocket mouse, Chaetodipus hispidus (Rodentia: Heteromyidae). *Journal of Mammalogy*, *93*(4), 1195-1215.

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

Benazzo, A., Trucchi, E., Cahill, J. A., Delser, P. M., Mona, S., et al. (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.

Benjamini, Y., & Hochberg, Y. (1995). Controlling the false discovery rate: a practical and powerful approach to multiple testing. *Journal of the Royal statistical society: series B (Methodological)*, *57*(1), 289-300.

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

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

Bouckaert, R. R., & Drummond, A. J. (2017). bModelTest: Bayesian phylogenetic site model averaging and model comparison. *BMC evolutionary biology*, *17*(1), 1-11.

Bouckaert, R., Vaughan, T. G., Barido-Sottani, J., Duchêne, S., Fourment, M., et al. (2019). BEAST 2.5: An advanced software platform for Bayesian evolutionary analysis. *PLoS computational biology*, *15*(4), e1006650.

Campbell, C. D., Chong, J. X., Malig, M., Ko, A., Dumont, B. L., et al. (2012). Estimating the human mutation rate using autozygosity in a founder population. *Nature genetics*, *44*(11), 1277-1281.

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.

Colella, J. P., Lan, T., Schuster, S. C., Talbot, S. L., Cook, J. A., et al. (2018). Whole-genome analysis of Mustela erminea finds that pulsed hybridization impacts evolution at high latitudes. *Communications biology*, *1*(1), 1-10.

Danecek, P., Auton, A., Abecasis, G., Albers, C. A., Banks, E., et al. (2011). The variant call format and VCFtools. *Bioinformatics*, *27*(15), 2156-2158.

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.

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.
Frey, J. K. (2005). Status and habitat of the least shrew (*Cryptotis parva*) in New Mexico. *Professional services contract*, 05-516.
Good, J. M., Hird, S., Reid, N., Demboski, J. R., Steppan, et al. (2008). Ancient hybridization and mitochondrial capture between two species of chipmunks. *Molecular ecology*, *17*(5), 1313-1327.
Greiman, S. E., Cook, J. A., Tkach, V. V., Hoberg, E. P., Menning, D. M., et al. (2018). Museum metabarcoding: a novel method revealing gut helminth communities of small mammals across space and time. *International journal for parasitology*, *48*(13), 1061-1070.
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*. John Wiley & Sons, New York, pp. 1-600.
Hansen, C. C. R., Hvilsom, C., Schmidt, N. M., Aastrup, P., de Groot, P. J. V. C., et al. (2018). The muskox lost a substantial part of its genetic diversity on its long road to Greenland. *Current Biology*, *28*(24), 4022-4028.
He, K., Woodman, N., Boaglio, S., Roberts, M., Supekar, S. et al. (2015). Molecular phylogeny supports repeated adaptation to burrowing within small-eared shrews genus of *Cryptotis* (Eulipotyphla, Soricidae). *PloS One*, *10*(10), p.e0140280.
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.
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., Waltari, E., Dokuchaev, N. E., Abramov, S., Dupal, T., et al. (2010). High-latitude diversification within Eurasian least shrews and Alaska tiny shrews (Soricidae). *Journal of Mammalogy*, *91*(5), 1041-1057.
Hope, A. G., Gragg, S. F., Nippert, J. B., & Combe, F. J. (2021). Consumer roles of small mammals within fragmented native tallgrass prairie. *Ecosphere*, *12*(3), e03441.
Huson, D. H., & Bryant, D. (2006). Application of phylogenetic networks in evolutionary studies. *Molecular Biology and Evolution*, *23*(2), 254–267.
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.
International Union for Conservation of Nature [IUCN]. (2008). Cryptotis parva. The IUCN Red List of Threatened Species. Version 2021-1
Jombart, T. (2008). adegenet: a R package for the multivariate analysis of genetic markers. *Bioinformatics*, *24*(11), 1403-1405.
Jones, C. and C.G. Schmitt. (1997). Mammal species of concern in New Mexico. Pp. 179- 205 in. T.L. Yates, W.L. Gannon and D.E. Wilson (eds.), Life Among the Muses: Papers in Honor of James S. Findley. Special Publications, Museum of Southwestern Biology, University of New Mexico *3*, 1-290.
Kumar, S., & Subramanian, S. (2002). Mutation rates in mammalian genomes. *Proceedings of the National Academy of Sciences*, *99*(2), 803-808.

Librado, P., & Rozas, J. (2009). DnaSP v5: a software for comprehensive analysis of DNA polymorphism data. *Bioinformatics*, *25*(11), 1451-1452.
Linck, E., Epperly, K., Van Els, P., Spellman, G. M., Bryson Jr, R. W., et al. (2019). Dense geographic and genomic sampling reveals paraphyly and a cryptic lineage in a classic sibling species complex. *Systematic biology*, *68*(6), 956-966.
Liu, Q., Zhou, B., Ma, W., Bawa, B., Ma, J., et al. (2014). Analysis of recombinant H7N9 wild-type and mutant viruses in pigs shows that the Q226L mutation in HA is important for transmission. *Journal of virology*, *88*(14), 8153-8165.
Liu, X., & Fu, Y. X. (2020). Stairway Plot 2: demographic history inference with folded SNP frequency spectra. *Genome biology*, *21*(1), 1-9.
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.
Moritz, C. (1994). Defining 'evolutionarily significant units' for conservation. *Trends in ecology & evolution*, *9*(10), 373-375.
New Mexico Department of Game and Fish [NMDGF]. (2020). Least Shrew (*Cryptotis parva*) Recovery Plan. New Mexico Department of Game and Fish, Wildlife Management Division, Santa Fe, New Mexico. 28 pp.
New Mexico Department of Game and Fish [NMDGF]. (2016). Threatened and Endangered Species of New Mexico: 2016 Biennial Review.
New Mexico Department of Game and Fish [NMDGF]. (2018). Threatened and Endangered Species of New Mexico: 2018 Biennial Review. NMDGF, Santa Fe. 155 pp
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.
Pritchard, J. K., Stephens, M., & Donnelly, P. (2000). Inference of population structure using multilocus genotype data. *Genetics*, *155*(2), 945-959.
R Development Core Team. (2018). R: a language and environment for statistical computing. R Foundation for Statistical Computing, Vienna, Austria. version 3.5.1. http://www.R-project.org.
Rambaut, A. (2018). Tracer (1.7.1) [Computer software]. http://beast.community
Rambaut, A., & Drummond, A. J. (2012). FigTree (1.4.4) [Computer software].
Roach, J. C., Glusman, G., Smit, A. F., Huff, C. D., Hubley, R., et al. (2010). Analysis of genetic inheritance in a family quartet by whole-genome sequencing. *Science*, *328*(5978), 636-639.
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.
Shuster, C. J. (1989). Genetics and historical biogeography of western peripheral populations of the least shrew (*Cryptotis parva*). Doctoral dissertation, University of New Mexico.
Stamatakis, A. (2014). RAxML version 8: A tool for phylogenetic analysis and post-analysis of large phylogenies. *Bioinformatics*, *30*(9), 1312–1313.

Tajima, F. (1989). Statistical method for testing the neutral mutation hypothesis by DNA polymorphism. *Genetics*, *123*(3), 585-595.
Whitaker, J. O., Jr. 1974. *Cryptotis parva*. Mammalian Species *43*, 1-8.
Woodman, N. (2018). American Recent Eulipotyphla: Nesophontids, Solenodons, Moles, and Shrews in the New World. Smithsonian Contributions to Zoology. https://doi.org/10.5479/si.1943-6696.650
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*, 2016-3.

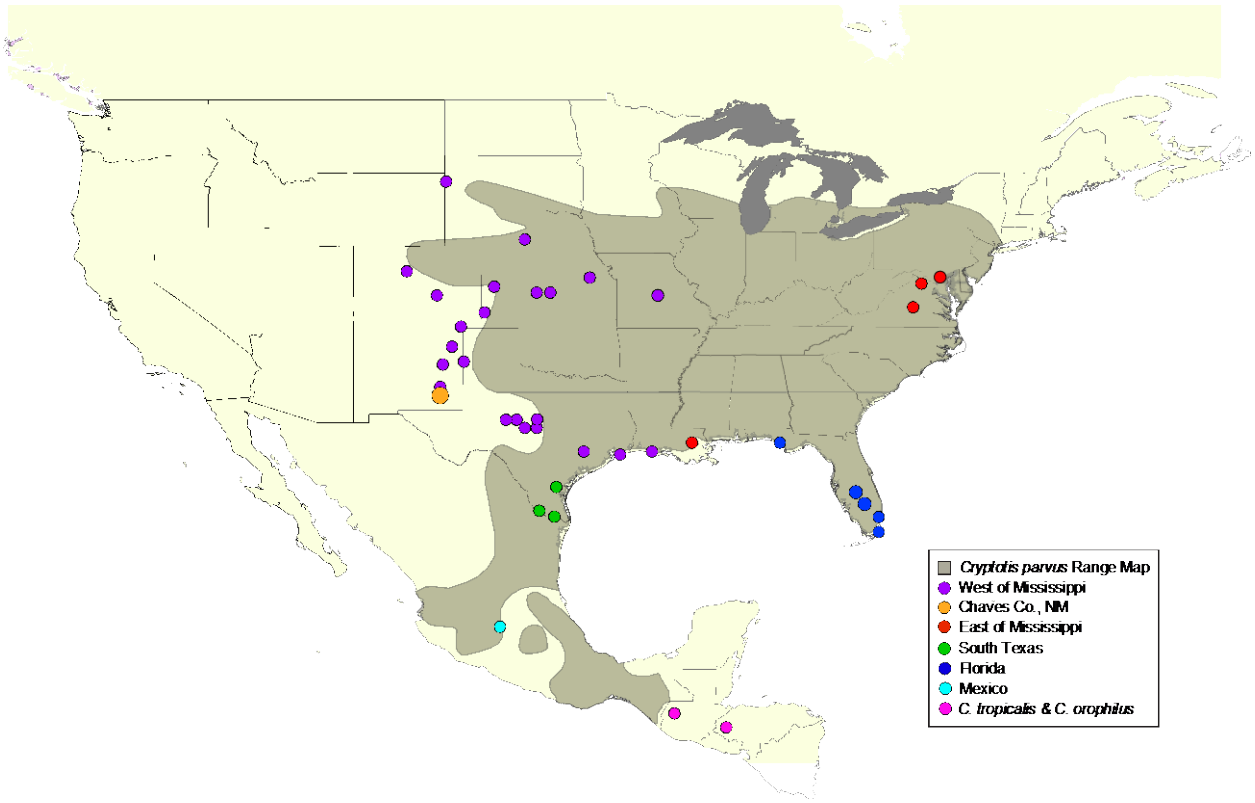

Figure 1. Sampling localities for *Cryptotis parvus* and related species represented by genetic data. Localities are colored by major clade assignment based either on well-supported nodes from the nuclear SNP maximum likelihood phylogeny or on the mitochondrial cytochrome b phylogeny for specimens without nuclear data. Where there was discordant assignment of individuals based on both phylogenies, colors represent nuclear clade membership. Shaded area represents the recognized distributional range downloaded from the IUCN website.

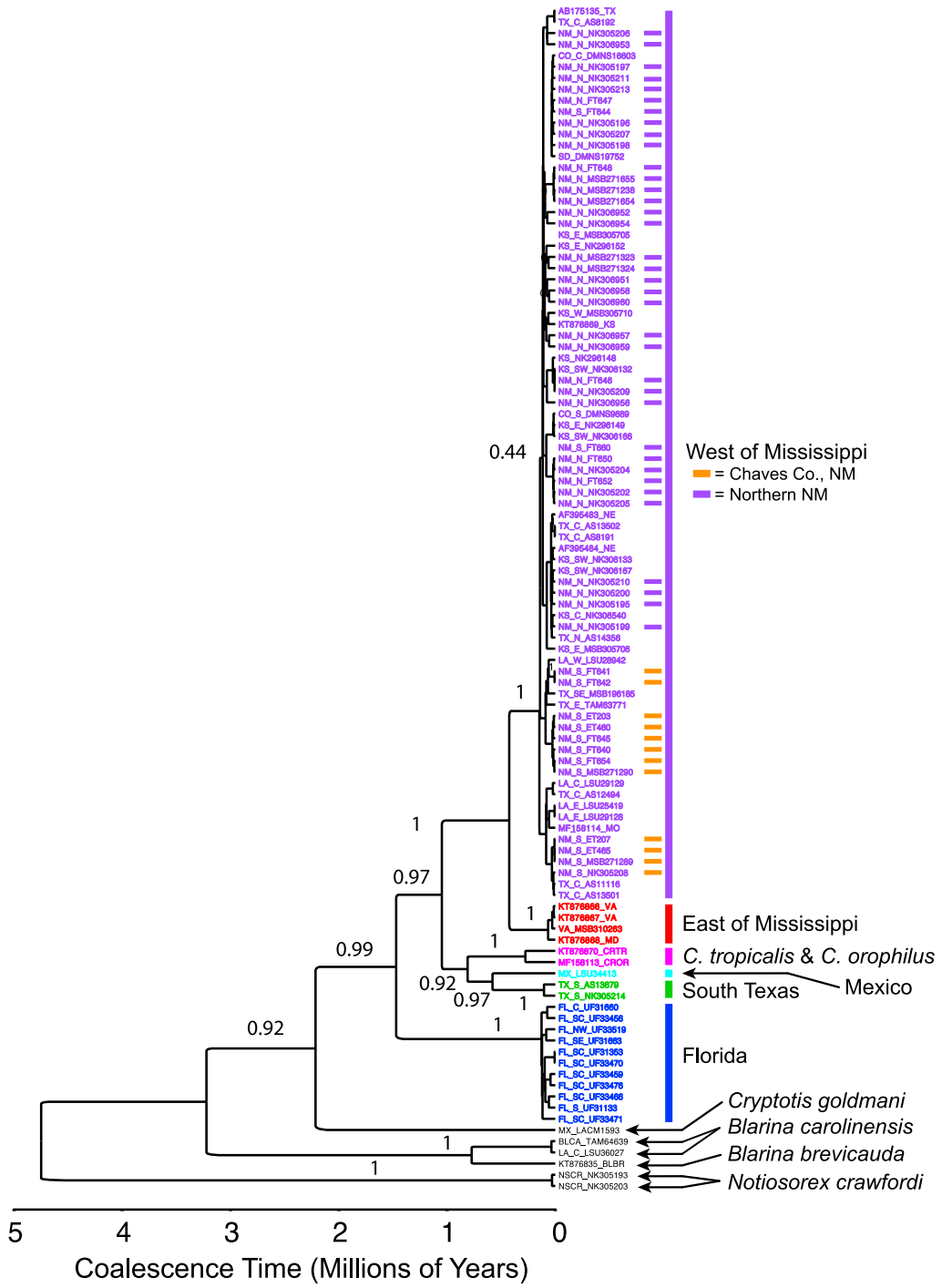

Figure 2. Bayesian mitochondrial cytochrome b gene phylogeny for *Cryptotis parvus* and associated outgroup taxa. Clades are colored according to geographic populations. Orange bars indicate specimens from Chaves Co. New Mexico, which do not constitute a well-supported mitochondrial lineage but form a well-supported lineage based on the nuclear phylogeny. Purple bars indicate specimens from the High Plains of New Mexico nested within the West of Mississippi clade. Numbers at nodes are Bayesian posterior probability values.

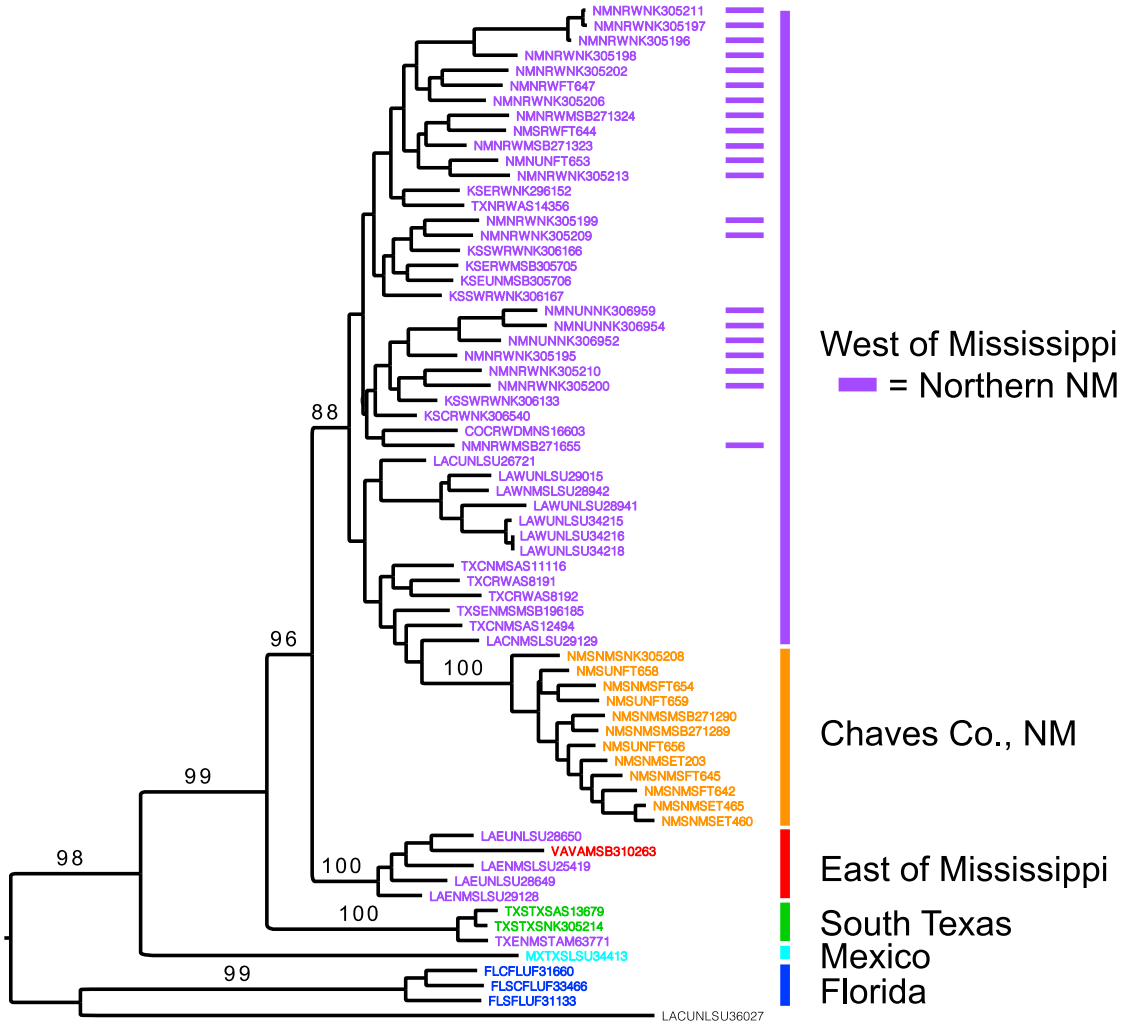

Figure 3. Nuclear SNP phylogeny for *Cryptotis parvus*. Clades represent geographic populations. Where individual shrews are genetically discordant between nuclear and mitochondrial datasets, individuals are colored to reflect their mitochondrial cytochrome b clade membership. Chaves Co. shrews that are genetically distinct across genomic analyses are colored orange. Numbers at nodes are bootstrap support values from maximum likelihood analysis.

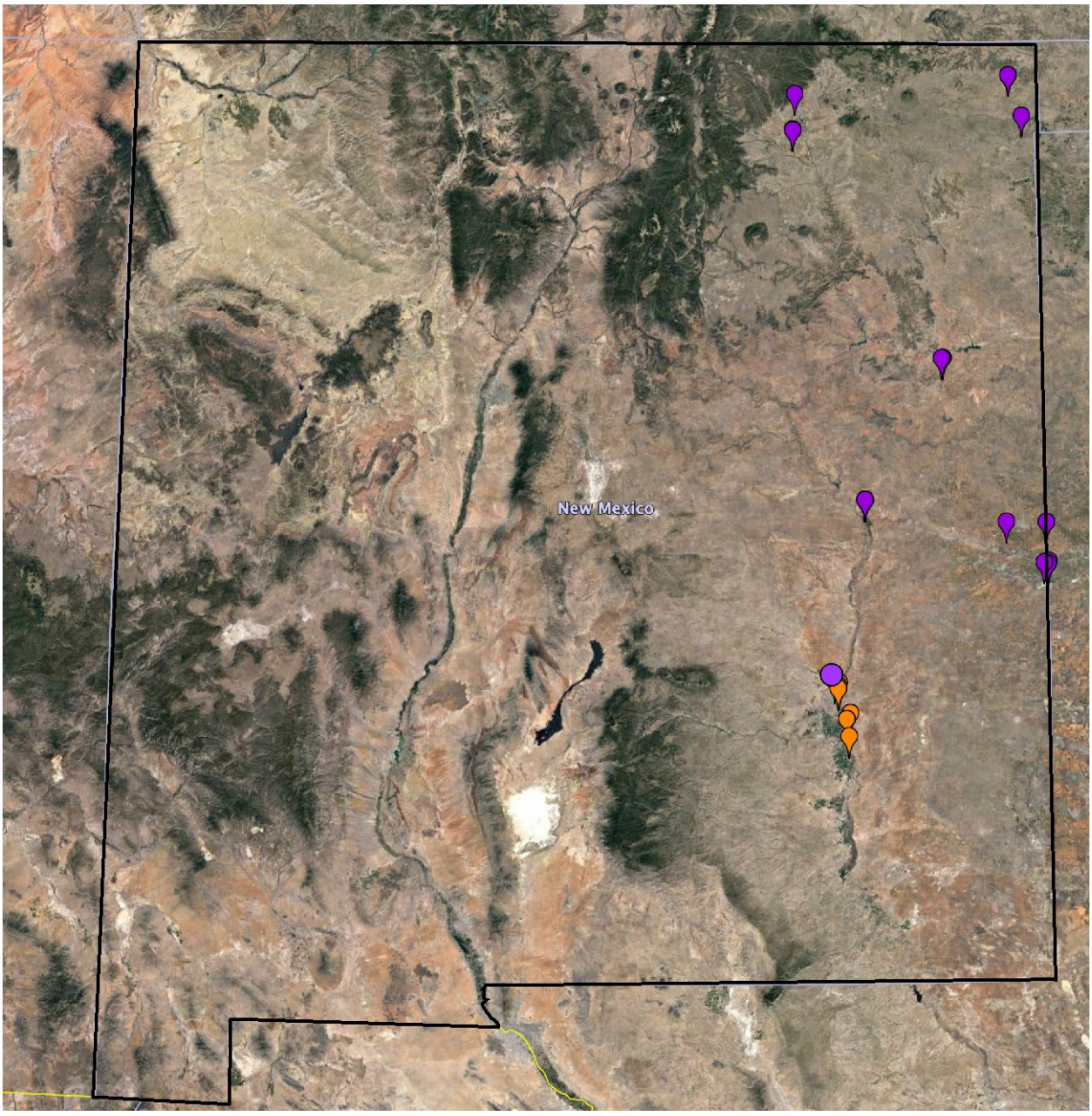

Figure 4. Map of New Mexico indicating localities of record for *Cryptotis parvus* specimens that were included in genetic analyses. Colors reflect distinct genetic groups based on nuclear SNP data. Note two specimens genetically representative of "northern New Mexico" (purple) were collected in Chaves Co., indicating sympatry between the two genetic lineages occurring in New Mexico in the vicinity of the northern unit of Bitter Lake National Wildlife Refuge.

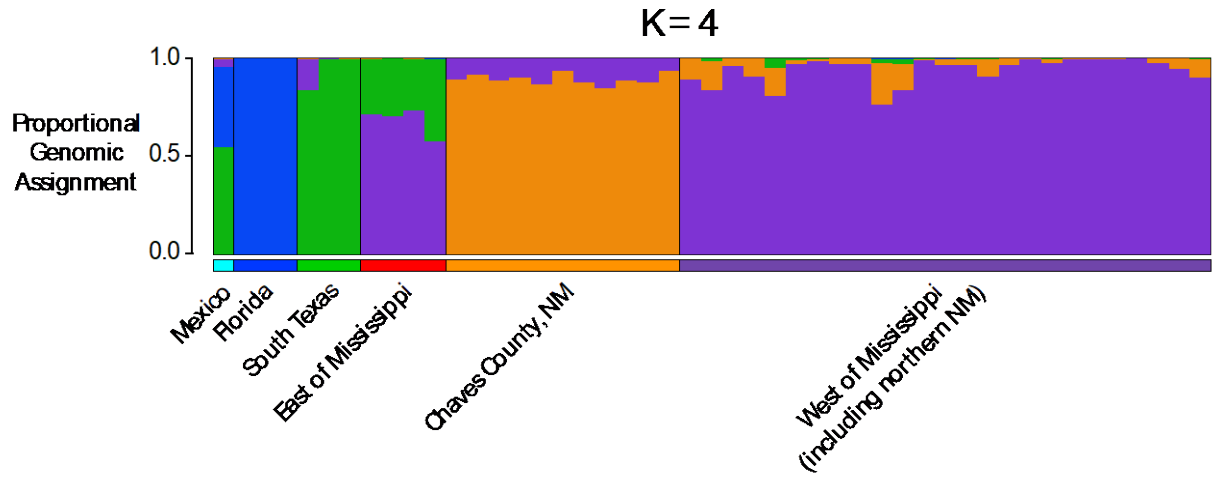

Figure 5. Structure assignment plot indicating major genetic groups of *Cryptotis parvus* based on nuclear SNP data. The most strongly supported number of groups (K) is four, reflected by colored proportional genomic assignment, although individual specimens were ordered in the plot a priori by geographic populations. The colored bar below the plot is congruent with population assignments based on the nuclear maximum likelihood phylogeny.

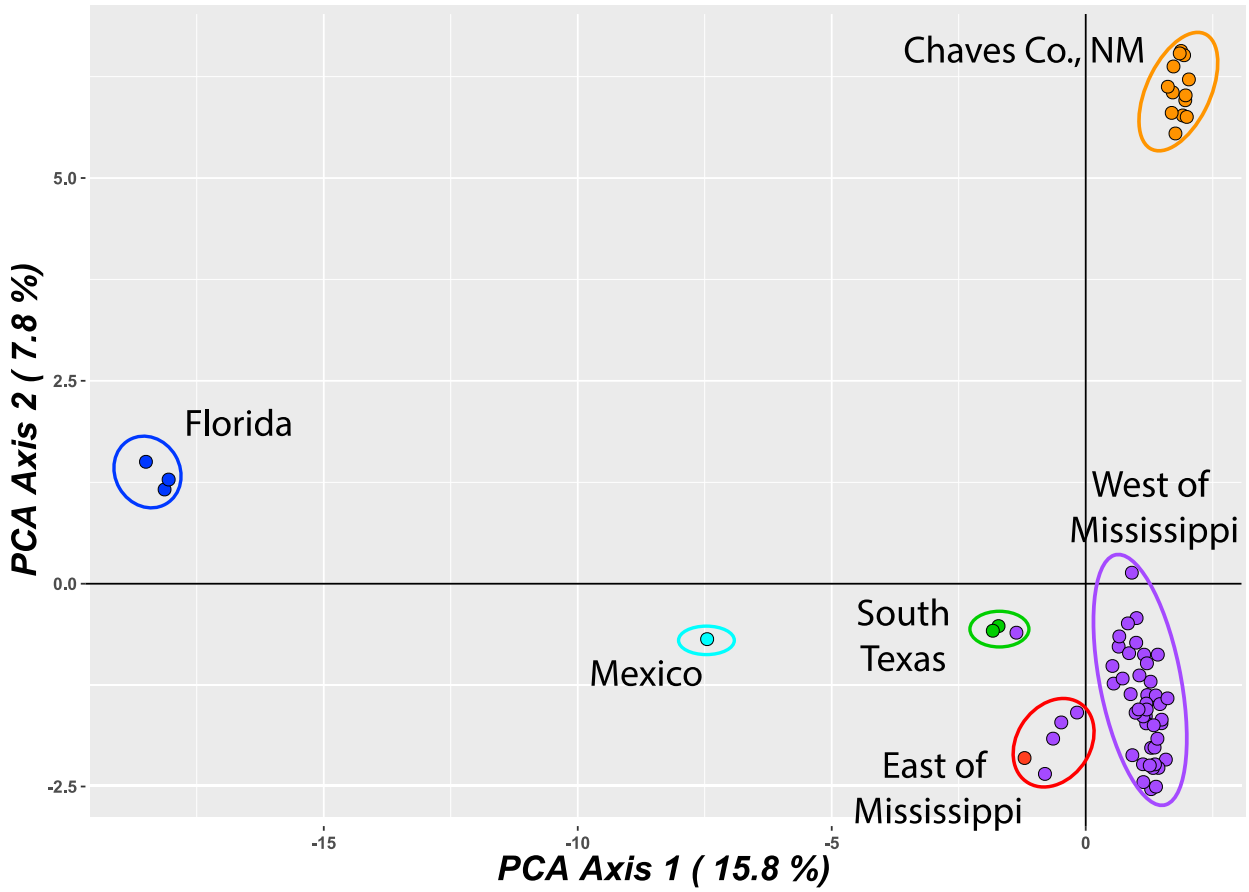

Figure 6. Discriminant Analysis of Principle Components plot based on the complete nuclear SNP dataset (19,400 loci) for *Cryptotis parvus*. Spatial separation between groups reflects genetic divergence. Circled groups represent geographic populations. Where individual shrews are genetically discordant between nuclear and mitochondrial datasets, individuals are colored purple to reflect their mitochondrial cytochrome b clade membership. All Chaves Co shrews are colored orange to reflect divergence of this lineage based on nuclear data.

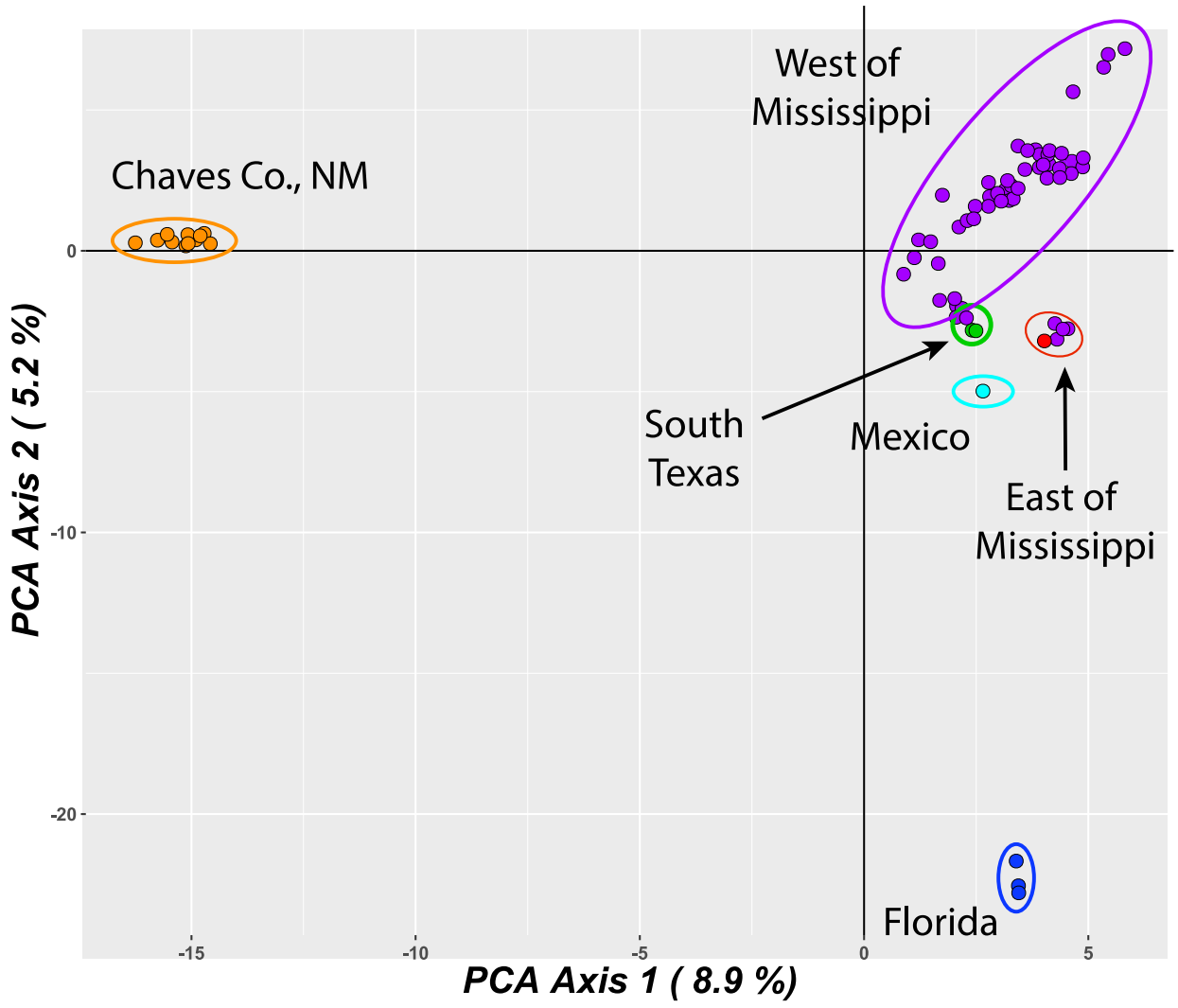

Figure 7. Discriminant Analysis of Principle Components plot based only on the neutral SNP data (17,850 loci) for *Cryptotis parvus*. Spatial separation between groups reflects genetic divergence. Circled groups represent geographic populations. Where individual shrews are genetically discordant between nuclear and mitochondrial datasets, individuals are colored purple to reflect their mitochondrial cytochrome b clade membership. All Chaves Co shrews are colored orange to reflect divergence of this lineage based on nuclear data.

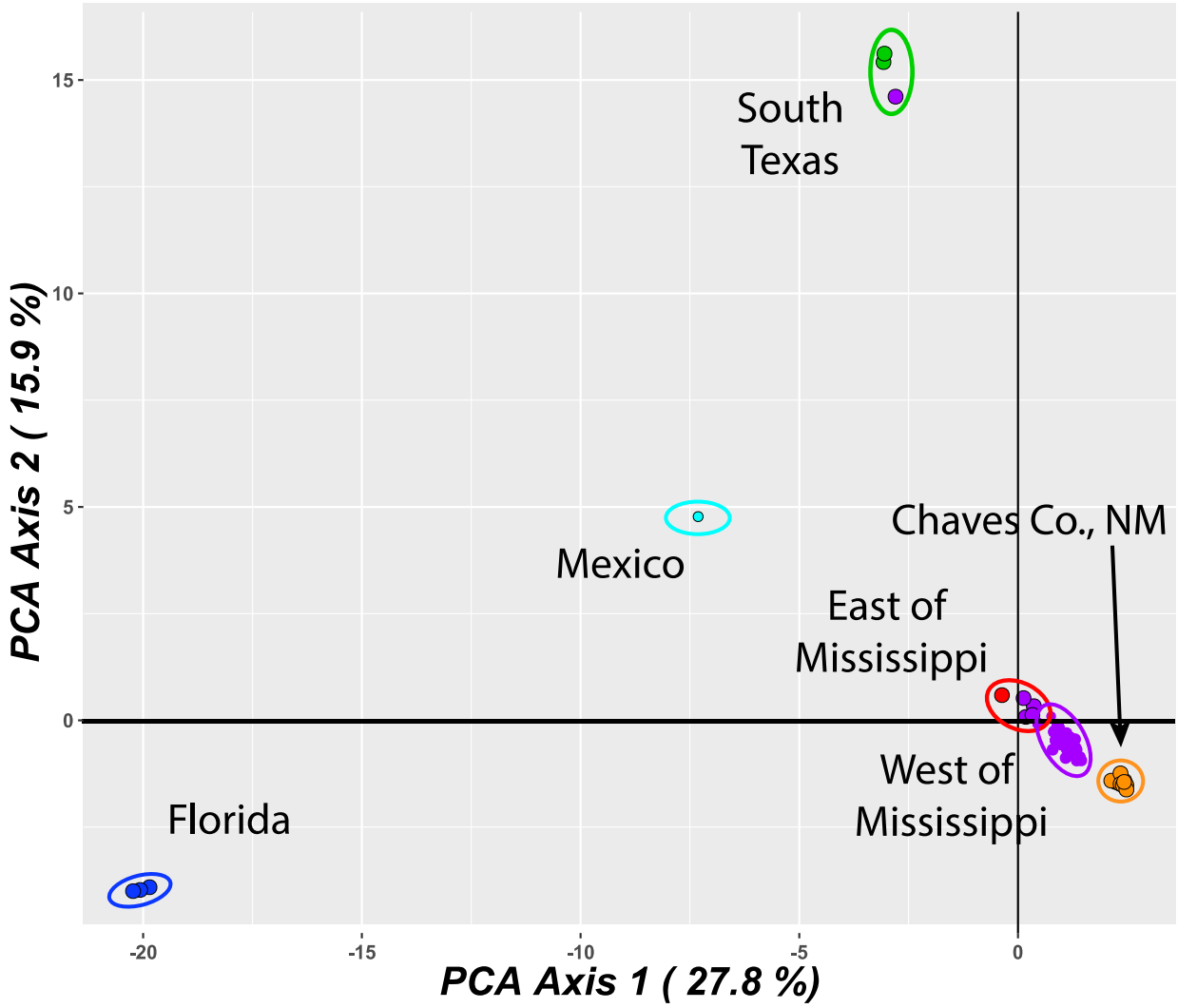

Figure 8. Discriminant Analysis of Principle Components plot based only on the non-neutral potentially adaptive SNP data (1,550 loci) for *Cryptotis parvus*. Spatial separation between groups reflects genetic divergence. Circled groups represent geographic populations. Where individual shrews are genetically discordant between nuclear and mitochondrial datasets, individuals are colored purple to reflect their mitochondrial cytochrome b clade membership. All Chaves Co shrews are colored orange to reflect divergence of this lineage based on nuclear data.

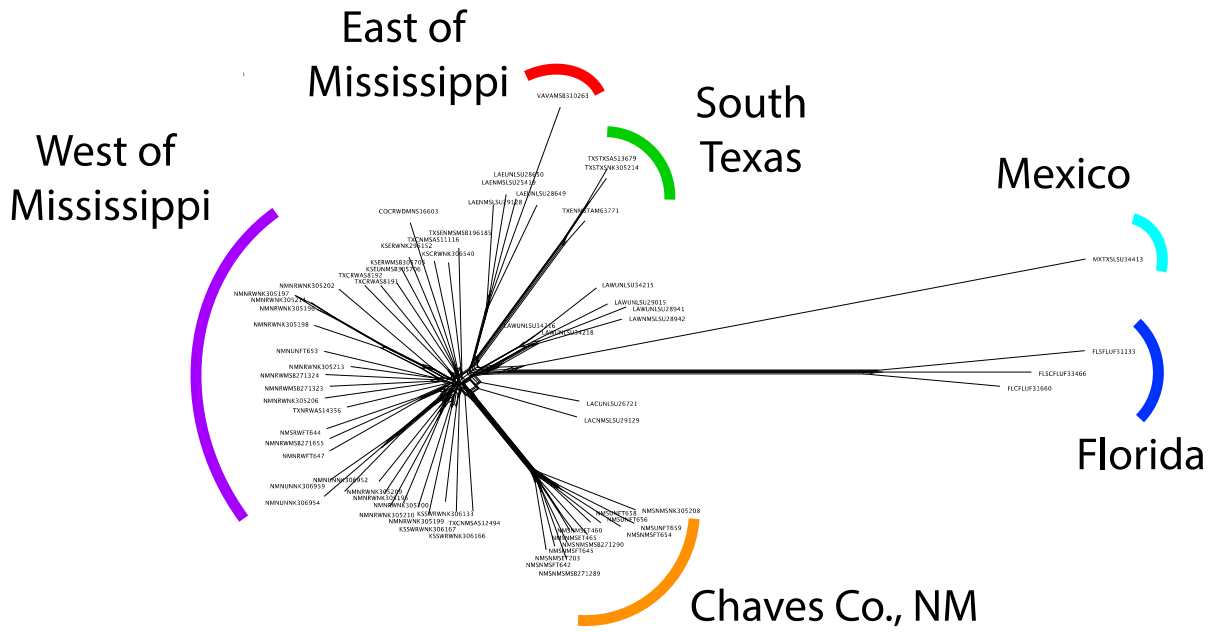

Figure 9. SplitsTree phylogenetic network based on 17,850 neutral SNP loci indicating relationships among major lineages of *Cryptotis parvus*.

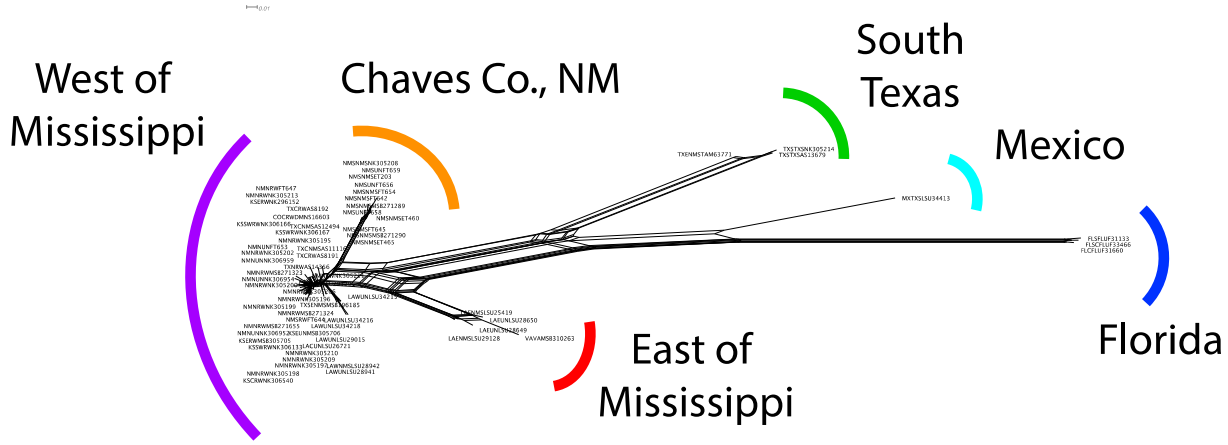

Figure 10. SplitsTree phylogenetic network based on 1,550 non-neutral SNP loci indicating relationships among major lineages of *Cryptotis parvus*.

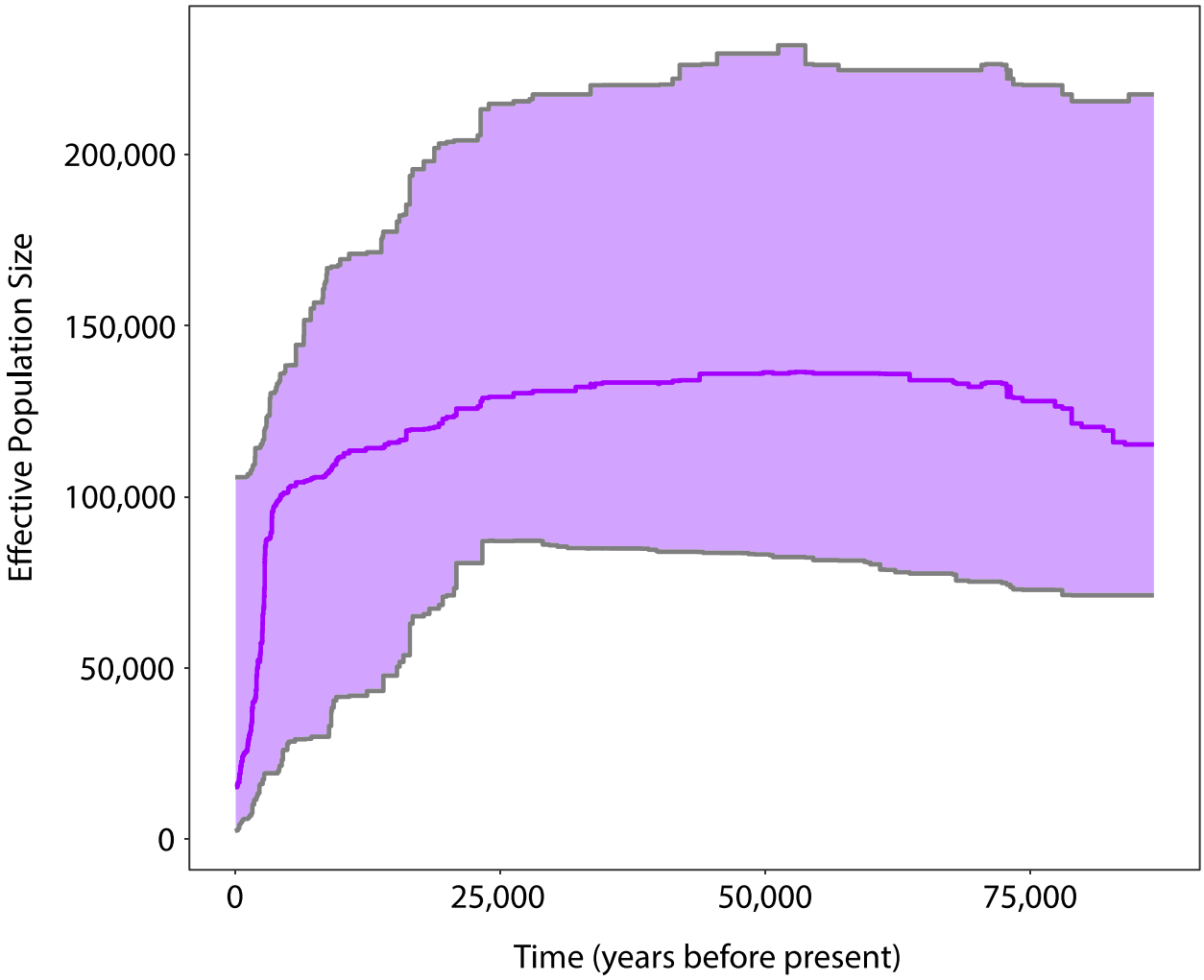

Figure 11. Stairway plot indicating change in genetic effective population size through time for *Cryptotis parvus* collected from localities in High Plains of New Mexico (all localities outside of Chaves Co.), suggesting a recent severe population decline. Filled region reflects 95% confidence intervals.

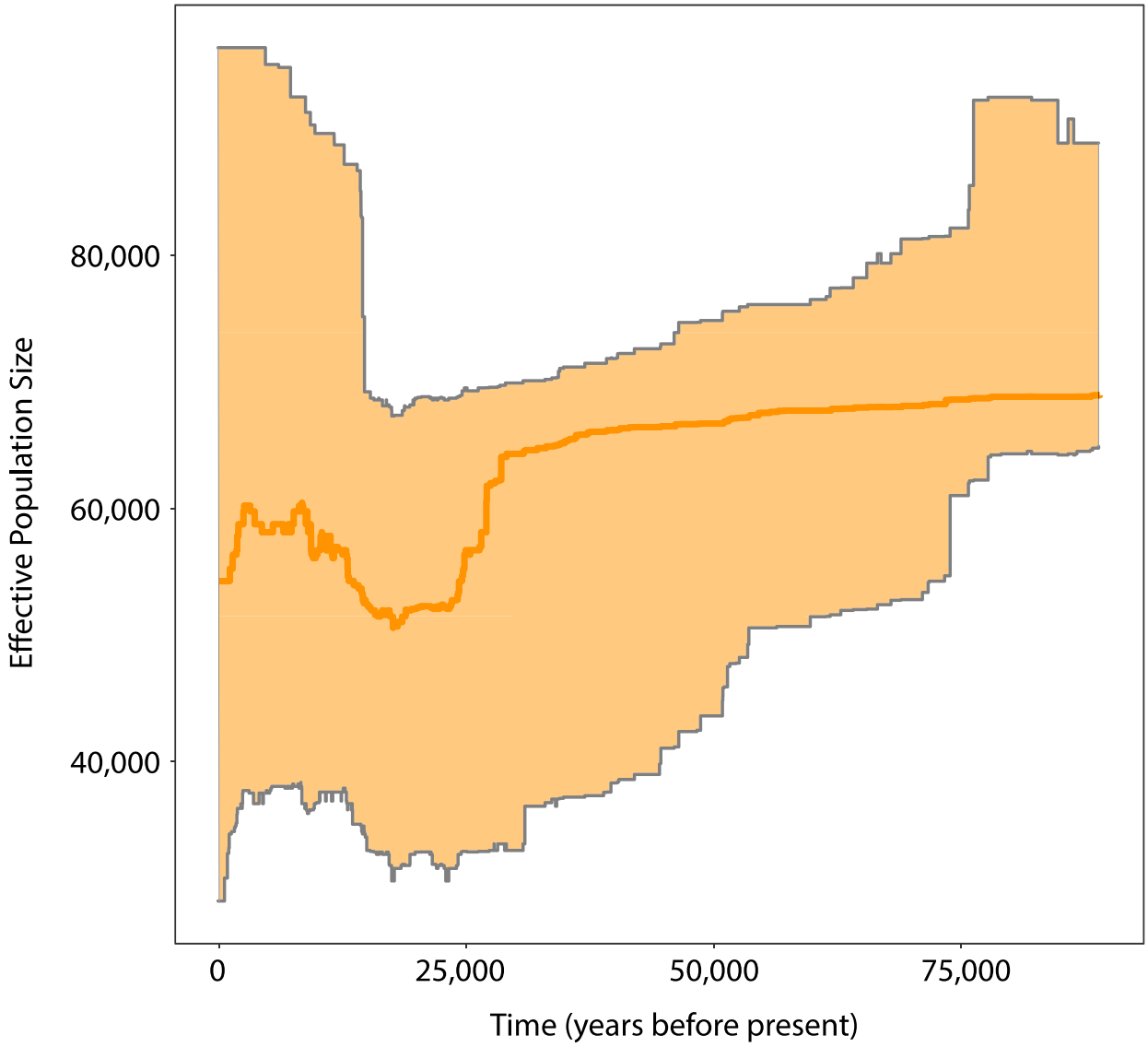

Figure 12. Stairway plot indicating change in genetic effective population size through time for *Cryptotis parvus* collected from localities in Chaves Co., New Mexico, suggesting a population decline coincident with warming since the Last Glacial Maximum (~20 ka). Filled region reflects 95% confidence intervals. Given the large confidence intervals near the present time, the signal of population fluctuations since ~15 ka is not reliable.

Table 1. Average pairwise sequence divergence (percentages) for the mitochondrial cytochrome b gene for selected geographic populations of *Cryptotis parvus*.

	Chaves Co., NM	West of Mississippi	South Texas	Florida
High Plains, NM	0.7	--	8.9	10.9
Chaves Co., NM		0.9	9.1	10.8
West of Mississippi			9.1	11.2
South Texas				9.9

Table 2. Genetic diversity metrices for the mitochondrial cytochrome b gene for selected geographic populations of *Cryptotis parvus*, including sample size (N), haplotype diversity (*Hd*), nucleotide diversity (π), and results of the Tajima's *D* test to assess potential demographic expansion/contraction (bold values are significant at P<0.05). Significantly negative Tajima's *D* values suggest demographic expansion and positive values suggest demographic contraction.

Population	N	*Hd*	π	Tajima's *D*
High Plains, NM	36	0.571	0.0008	-1.27
Chaves Co., NM	12	0.788	0.0079	1.37
West of Mississippi	68	0.770	0.0060	**-2.84**
Florida	11	0.982	0.0053	-1.42

Table 3. Pairwise F_{ST} values for major clades of *Cryptotis parvus* based on nuclear SNP data where clade membership was assigned based on well-supported nodes from the maximum likelihood SNP-based phylogeny. A value of 1 would indicate different alleles at all SNP loci and a value of 0 would indicate identical alleles at every SNP locus.

	Chaves Co., NM	East of Mississippi	South Texas	Florida	Mexico
West of Mississippi	0.047	0.046	0.072	0.169	0.134
Chaves Co., NM		0.183	0.261	0.350	0.403
East of Mississippi			0.241	0.312	0.362
South Texas				0.394	0.559
Florida					0.438

Table 4. Genetic diversity metrices for the nuclear SNP loci for major clades of *Cryptotis parvus* based on nuclear SNP data where clade membership was assigned based on well-supported nodes from the maximum likelihood SNP-based phylogeny. Metrics include number of private alleles (i.e., unique alleles in each clade), sample size (N), observed (H_O) and expected (H_E) heterozygosity, nucleotide diversity (π), and inbreeding coefficient (F_{IS}), where values close to 0 indicate random mating and values close to 1 indicate severe inbreeding.

Population	Private Alleles	N	H_O	H_E	π	F_{IS}
West of Mississippi	5731	43	0.093	0.122	0.124	0.133
Chaves Co., NM	1043	12	0.070	0.079	0.083	0.039
East of Mississippi	1493	5	0.091	0.103	0.116	0.055
South Texas	851	3	0.061	0.057	0.068	0.012
Florida	2577	3	0.102	0.108	0.130	0.052
Mexico	882	1	0.034	0.017	0.034	0

Appendix A. Specimens included in genetic analyses. Unk. = unknown.

Species	Museum	Museum ID	State	County	Date of Collection	Sex
Cryptotis goldmani	LACM	1593	Mexico	Guerrero	7/25/1986	Female
Cryptotis goldmani	LACM	1599	Mexico	Guerrero	7/25/1986	Male
Cryptotis goldmani	LACM	1634	Mexico	Guerrero	7/25/1986	Male
Cryptotis parvus	LACM	1803	Missouri	Dade	10/25/1986	Female
Cryptotis parvus	LACM	1804	Missouri	Dade	10/25/1986	Male
Cryptotis parvus	LACM	1806	Missouri	Dade	10/25/1986	Male
Cryptotis parvus	LSU	25419	Louisiana	East Baton Rouge	5/17/1982	Male
Cryptotis parvus	LSU	26721	Louisiana	Vernon	11/12/1982	Male
Cryptotis parvus	LSU	28942	Louisiana	Cameron	3/1/1986	Male
Cryptotis parvus	LSU	29128	Louisiana	East Baton Rouge	11/18/1985	Female
Cryptotis parvus	LSU	34216	Louisiana	Cameron	3/17/1986	Male
Cryptotis parvus	LSU	34413	Mexico	Mexico	5/6/1993	Male
Cryptotis parvus	LSU	36027	Louisiana	Vernon	5/14/1996	Male
Cryptotis parvus	TA&M	63771	Texas	Kenedy	6/26/2015	Unk.
Cryptotis parvus	TA&M	64482	Texas	Polk	11/7/2015	Unk.
Cryptotis parvus	TA&M	64483	Texas	Polk	11/7/2015	Unk.
Cryptotis parvus	TA&M	64491	Texas	Polk	11/9/2015	Unk.
Cryptotis parvus	TA&M	64514	Texas	Tyler	11/21/2015	Unk.
Cryptotis parvus	TA&M	64515	Texas	Tyler	11/22/2015	Unk.
Cryptotis parvus	TA&M	64538	Texas	Tyler	1/6/2016	Unk.
Cryptotis parvus	TA&M	64573	Texas	Tyler	1/5/2016	Unk.
Cryptotis parvus	TA&M	64608	Texas	Polk	1/30/2016	Unk.
Cryptotis parvus	TA&M	64609	Texas	Polk	1/30/2016	Unk.
Cryptotis parvus	TA&M	64619	Texas	Polk	1/30/2016	Unk.
Cryptotis parvus	TA&M	64620	Texas	Polk	1/30/2016	Unk.
Cryptotis parvus	TA&M	64639	Texas	Polk	1/29/2016	Unk.
Cryptotis parvus	TA&M	64694	Texas	Polk	1/29/2016	Unk.
Cryptotis parvus	UF	31133	Florida	Monroe	4/15/2004	Female
Cryptotis parvus	UF	31353	Florida	Highlands	11/29/2005	Unk.
Cryptotis parvus	UF	31390	Florida	Highlands	11/3/2005	Unk.
Cryptotis parvus	UF	31660	Florida	Polk	11/17/2008	Unk.
Cryptotis parvus	UF	31662	Florida	Polk	11/21/2008	Unk.
Cryptotis parvus	UF	31663	Florida	Palm Beach	10/29/2006	Unk.
Cryptotis parvus	UF	33456	Florida	Highlands	11/14/2007	Female
Cryptotis parvus	UF	33463	Florida	Highlands	10/26/2005	Unk.

Cryptotis parvus	UF	33466	Florida	Highlands	11/6/2007	Unk.
Cryptotis parvus	UF	33469	Florida	Highlands	3/9/2006	Unk.
Cryptotis parvus	UF	33470	Florida	Highlands	3/11/2008	Unk.
Cryptotis parvus	UF	33471	Florida	Highlands	6/12/2007	Unk.
Cryptotis parvus	UF	33476	Florida	Highlands	10/7/2008	Unk.
Cryptotis parvus	UF	33519	Florida	Bay	12/26/2012	Unk.
Cryptotis parvus	ASNHC	8191	Texas	Tom Green	8/7/1992	Male
Cryptotis parvus	ASNHC	8192	Texas	Tom Green	7/22/1993	Female
Cryptotis parvus	ASNHC	11116	Texas	Concho	10/5/1997	Unk.
Cryptotis parvus	ASNHC	12494	Texas	Brown	7/6/2002	Male
Cryptotis parvus	ASNHC	13501	Texas	Brown	4/26/2008	Male
Cryptotis parvus	ASNHC	13502	Texas	Brown	4/26/2008	Female
Cryptotis parvus	ASNHC	13679	Texas	Hidalgo	3/21/2009	Male
Cryptotis parvus	ASNHC	14356	Texas	Hutchinson	10/22/2009	Male
Cryptotis parvus	DMNS	9689	Colorado	Pueblo	9/21/1999	Male
Cryptotis parvus	DMNS	16603	Colorado	Jefferson	8/24/2016	Female
Cryptotis parvus	DMNS	19752	South Dakota	Lawrence	10/20/2009	Male
Cryptotis parvus	MSB	196185	Texas	Galveston	2/13/2009	Female
Cryptotis parvus	MSB	271289	New Mexico	Chaves	9/27/1986	Male
Cryptotis parvus	MSB	271290	New Mexico	Chaves	9/27/1986	Male
Cryptotis parvus	MSB	271323	New Mexico	Roosevelt	7/12/1987	Female
Cryptotis parvus	MSB	271324	New Mexico	Roosevelt	7/12/1987	Female
Cryptotis parvus	MSB	271655	New Mexico	Quay	7/5/1987	Male
Cryptotis parvus	MSB	272899	New Mexico	Chaves	6/11/1987	Female
Cryptotis parvus	MSB	305705	Kansas	Pottawatomie	2/8/2016	Male
Cryptotis parvus	MSB	305710	Kansas	Scott	12/20/2014	Female
Cryptotis parvus	MSB	310263	Virginia	Clarke	10/15/2015	Female
Cryptotis parvus	NK	305195	New Mexico	Union	10/3/2019	Unk.
Cryptotis parvus	NK	305196	New Mexico	DeBaca	9/4/2019	Male
Cryptotis parvus	NK	305197	New Mexico	DeBaca	9/5/2019	Unk.
Cryptotis parvus	NK	305198	New Mexico	DeBaca	9/5/2019	Female
Cryptotis parvus	NK	305199	New Mexico	DeBaca	9/5/2019	Female
Cryptotis parvus	NK	305201	New Mexico	Union	10/4/2019	Unk.
Cryptotis parvus	NK	305214	Texas	Hidalgo	3/11/2020	Male
Cryptotis parvus	JF	ET203	New Mexico	Chaves	10/30/1999	Female
Cryptotis parvus	JF	ET460	New Mexico	Chaves	10/30/1999	Female
Cryptotis parvus	JF	ET465	New Mexico	Chaves	10/30/1999	Female
Cryptotis parvus	JF	FT640	New Mexico	Chaves	9/10/2005	Female
Cryptotis parvus	JF	FT641	New Mexico	Chaves	9/11/2005	Male
Cryptotis parvus	JF	FT642	New Mexico	Chaves	9/11/2005	Male

Cryptotis parvus	JF	FT643	New Mexico	Chaves	9/11/2005	Female
Cryptotis parvus	JF	FT644	New Mexico	Chaves	9/11/2005	Female
Cryptotis parvus	JF	FT645	New Mexico	Chaves	9/24/2005	Female
Cryptotis parvus	JF	FT646	New Mexico	Quay	9/24/2005	Male
Cryptotis parvus	JF	FT647	New Mexico	Quay	9/25/2005	Female
Cryptotis parvus	JF	FT648	New Mexico	Quay	9/25/2005	Male
Cryptotis parvus	JF	FT650	New Mexico	Quay	9/25/2005	Male
Cryptotis parvus	JF	FT652	New Mexico	Quay	9/25/2005	Female
Cryptotis parvus	JF	FT653	New Mexico	Roosevelt	9/30/2005	Male
Cryptotis parvus	JF	FT654	New Mexico	Chaves	9/12/2005	Female
Cryptotis parvus	JF	FT655	New Mexico	Chaves	9/13/2005	Female
Cryptotis parvus	JF	FT656	New Mexico	Chaves	9/13/2005	Female
Cryptotis parvus	JF	FT658	New Mexico	Chaves	9/13/2005	Male
Cryptotis parvus	JF	FT659	New Mexico	Chaves	9/13/2005	Female
Cryptotis parvus	JF	FT660	New Mexico	Chaves	9/2/2005	Unk.
Notiosorex crawfordi	NK	305193	New Mexico	Curry	1/27/2018	Unk.
Notiosorex crawfordi	NK	305203	New Mexico	Santa Fe	7/8/2018	Unk.