North American phylogeography of the white-footed mouse (Peromyscus leucopus) reveals strong differentiation in the desert Southwest and rapid expansion

Sydney Walters[1], Marketa Zimova[2], Jenifer Mallinoff[3], and Elizabeth Kierepka[4]

[1]North Carolina State University at Raleigh
[2]Ohio Unviersity
[3]Appalachian State University
[4]North Carolina Museum of Natural Sciences

July 16, 2024

Abstract

Pleistocene glaciation events had a dramatic impact on temperate taxa by displacing animal and plant populations south of ice sheets into glacial refugia. Genetic variation often reflects these histories of isolation within glacial refugia and subsequent recolonization. The highly speciose rodent genus Peromyscus, in particular, is well known for its rapid diversification during the Pleistocene. Peromyscus are also significant reservoirs for a myriad of zoonoses, and many cosmopolitan species are undergoing range expansions due to human land use and climate change. This study focused on the range-wide phylogeography of the white-footed mouse (Peromyscus leucopus), a common species found in eastern North America that is one the primary reservoirs for Lyme Disease (Borrelia burgdorferi). We used two mitochondrial genes, cytochrome b and control region, to identify evolutionary lineages of white-footed mice and characterize patterns of expansion of each lineage across their geographic range. Overall, we found evidence for four evolutionary lineages with a Southwest lineage largely restricted to grassland and desert habitats. Time since recent common ancestors placed all lineages diverging within the Last Glacial Maximum (˜19-25k years ago). All lineages exhibited signatures of expansion, particularly the two northern lineages known to host Lyme Disease. Overall, white-footed mice underwent rapid diversification similar to other Peromyscus species and potentially exhibit habitat-based divergence within the Southwest lineage. Signatures of expansion also indicate that white-footed mice will continue to facilitate increased spread of zoonoses like Lyme Disease, but further study is needed to clarify how these evolutionary dynamics interact with other factors associated with human disease incidence.

Running title : Phylogeography of white-footed mice

Title: North American phylogeography of the white-footed mouse (*Peromyscus leucopus*) reveals strong differentiation in the desert Southwest and rapid expansion

SYDNEY P. WALTERS[1], JENIFER MALLINOFF[2], MARKETA ZIMOVA[2], and ELIZABETH M. KIEREPKA[1]

Affiliation:

[1]North Carolina Museum of Natural Sciences, Department of Forestry and Environmental Resources, North Carolina State University, Raleigh, NC, USA

[2]Department of Biological Sciences, Ohio University, Athens, OH, USA

Correspondence:

Elizabeth Kierepka

North Carolina Museum of Natural Sciences

11 W Jones Street

Raleigh, NC 27601

EMAIL: emkierep@ncsu.edu

Word Count: 5,589

ABSTRACT

Pleistocene glaciation events had a dramatic impact on temperate taxa by displacing animal and plant populations south of ice sheets into glacial refugia. Genetic variation often reflects these histories of isolation within glacial refugia and subsequent recolonization. The highly speciose rodent genus *Peromyscus* , in particular, is well known for its rapid diversification during the Pleistocene. *Peromyscus* are also significant reservoirs for a myriad of zoonoses, and many cosmopolitan species are undergoing range expansions due to human land use and climate change. This study focused on the range-wide phylogeography of the white-footed mouse (*Peromyscus leucopus*), a common species found in eastern North America that is one the primary reservoirs for Lyme Disease (*Borrelia burgdorferi*). We used two mitochondrial genes, cytochrome b and control region, to identify evolutionary lineages of white-footed mice and characterize patterns of expansion of each lineage across their geographic range. Overall, we found evidence for four evolutionary lineages with a Southwest lineage largely restricted to grassland and desert habitats. Time since recent common ancestors placed all lineages diverging within the Last Glacial Maximum (˜19-25k years ago). All lineages exhibited signatures of expansion, particularly the two northern lineages known to host Lyme Disease. Overall, white-footed mice underwent rapid diversification similar to other *Peromyscus* species and potentially exhibit habitat-based divergence within the Southwest lineage. Signatures of expansion also indicate that white-footed mice will continue to facilitate increased spread of zoonoses like Lyme Disease, but further study is needed to clarify how these evolutionary dynamics interact with other factors associated with human disease incidence.

Keywords: phylogeography, white-footed mouse, Pleistocene,*Peromyscus leucopus* , Lyme Disease

INTRODUCTION

Quaternary climatic cycles greatly influenced numerous species across the northern hemisphere by forcing populations into refugia south of glaciated areas. Contemporary populations often retain the genetic signatures of isolation within separate refugia and subsequent recolonization after glacial recession (Hewitt 2004). Considerable diversity exists among responses to glaciation, owing to an interplay between species' life history (e.g., dispersal capability, degree of ecological specialization, geographic distribution; Shafer et al. 2010, Kierepka and Latch 2016) and landscape factors (e.g., presence of contemporary barriers and number of occupied refugia). This interaction is particularly evident in wide-ranging, generalist species where they potentially could have occupied multiple refugia and utilized distinct colonization routes, making it difficult to infer how such species responded to Pleistocene glaciation events. Indeed, wide-ranging generalist species exhibit diverse responses to glaciation, ranging from a single lineage (Kierepka and Latch 2016) and simple eastern and western lineages (Reding et al. 2012, Kierepka et al. 2023, Klicka et al. 2023) to multiple refugial lineages (Puckett et al. 2015) and formation of cryptic species (Burbrink et al. 2022, McDonough et al. 2022).

The murid genus *Peromyscus* exemplifies the difficulty in inferring responses to Pleistocene glaciation due to their complex evolutionary histories. *Peromyscus* is a highly speciose group that occupies every terrestrial biome in North America, resulting in complex patterns of speciation across numerous refugia and geographic barriers. Despite these evolutionary dynamics, *Peromyscus* mice exhibit very similar morphologies, which makes it difficult to identify individual species and evolutionary lineages. Indeed, many taxa require genetic identification to conclusively differentiate species when they occur in sympatry. With genetic data, multiple

cryptic species and thirteen species complexes have been found within previously thought cosmopolitan species (e.g., Bradley et al. 2007, Castaneda-Rico et al. 2014). Many of these newly described species have relatively narrow geographic ranges within western biogeographic hotspots like Mexico (Bradley et al. 2022, Leon-Tapia et al. 2022), Pacific Northwest (Boria and Blois 2023), and California (Riddle et al. 2000, Boria and Blois 2023) while several species are still considered generalists with large geographic ranges.

While many *Peromyscus* species have small geographic ranges, others are highly cosmopolitan and have experienced range expansion due to human land use and climate change. These expansions are of particular concern in eastern North America because *Peromyscus* are the primary reservoirs of multiple tick-borne illnesses, principally*Borrelia burgdorferi* , the most common causative agent of Lyme disease. The most competent reservoir for *B. burgdorferi* is the white-footed mouse, *Peromyscus leucopus* , an abundant rodent throughout central and eastern North America. *P. leucopus* is found in many habitats including highly urban habitats with a distribution range stretching from central Mexico and Arizona in the west to southern coastal Canada in the east. Numerous studies have documented the northward expansion of *P. leucopus* and associated*B. burgdorferi* into Canada and the Great Lakes region (Roy-Dufresne et al. 2013, Fiset et al. 2015, Garcia-Elfring et al. 2017).

Previous work with *P. leucopus* in eastern North America has identified two monophyletic evolutionary lineages in eastern North America: a broadly distributed eastern lineage (East) and another lineage in the Upper Midwest and north of the St. Lawrence River (Midwest; Rowe et al. 2006, Fiset et al. 2015). These lineages contain 8 morphological subspecies that correspond to five islands and three mainland groups (Hall 1981). Strong genetic signatures of expansion have been detected in both lineages along their northern border in Michigan and Canada (Fiset et al. 2015, Prado et al. 2022). These lineages appear to have formed in allopatry during the Pleistocene (Rowe et al. 2006, Moscarella et al. 2019, Prado et al. 2022). Several studies have recorded additional lineages west of these lineages, but had low support in phylogenetic analysis or were based on a few individuals (e.g., Shipp-Pennock et al. 2005, Rowe et al. 2006, Herrera 2021).

Western *P. leucopus* populations occur across multiple grassland and desert habitats unlike the largely forest-associated East and Midwest lineages. The westernmost extent of *P. leucopus* occurs in the desert southwest, a well-recognized glacial refugium for many species (Riddle and Hafner 2006, McDonnough et al. 2022). Western*P. leucopus* also contain three chromosomal inversions that separate them from eastern populations, and thus, have been considered chromosomal "races" (Baker et al. 1983). The three inversion difference between western and eastern chromosomal races also represents a higher divergence than other *Peromyscus* species pairs (Baker et al. 1983). However, these two chromosomal races meet in Oklahoma and appear to inbreed extensively (Stangl 1986), making it unclear if these races correspond to evolutionary lineages like East and Midwest. Taken together, it is likely there is further divergence found in *P. leucopus* beyond the well-recognized East and Midwest lineages with at least one more lineage in the west. Thus, this study sampled across the range of *Peromyscus leucopus* to examine the species-wide effects of Pleistocene glaciation on observed genetic variation.

Collectively, Pleistocene glaciation and large water barriers (e.g., Atlantic Ocean, Great Lakes, and St. Lawrence River) appear to have had the largest impact on *P. leucopus* within the eastern portion of their range (e.g., Shipp-Pennock et al. 2005, Rowe et al. 2006, Fiset et al. 2015, Moscarella 2019). Much of the habitat within these areas are forested, which allows high gene flow. More habitat heterogeneity and geographic barriers exist in the western portion of *P. leucopus* ' range, increasing the potential for higher genetic differentiation. Indeed, western *P. leucopus* is separated into 11 morphological subspecies (9 mainland, 2 island; Hall 1981). *Peromyscus spp* . often speciate due to habitat heterogeneity in biodiversity hotspots (e.g., León-Tapia et al. 2022), and western *P. leucopus* occur within areas known for high endemism (e.g., desert southwest and Mexico). We predict that Pleistocene glaciation influenced the western portion of *P. leucopus* range and that the western populations form a separate evolutionary lineage from East and Midwest. Furthermore, given the known species' ability to colonize variety of habitats and subsequently expand its range, we predict strong signatures of population expansion in all evolutionary lineages.

METHODS

Sample Collection

We collected putative *P. leucopus* tissues (n = 157) from museums, NSF NEON tissue collection, and field sampled animals from across their range (Table S1). Collection emphasized animals caught west of the Mississippi River including Kansas, Mexico, Nebraska, New Mexico, North Dakota, Oklahoma, and Texas (n = 85; Fig 1). The remaining individuals originated from areas east of the Mississippi River including previously glaciated areas like Canada, New England, and the Great Lakes regions. All samples were georeferenced, and stored in a -20°C until DNA extraction.

Laboratory Methods

All tissues were extracted via Zymo Miniprep tissue kits following the manufacturer guidelines. We amplified two mitochondrial genes: 750 bp of cytochrome b and the entire control region (920 bp) using published mammal (MTCB-F: 5'-CCHCCATAAATAGGNGAAGG-3', MTCB-R: (5'-WAGAAYTTCAGCTTTGGG-3'; Naidu et al. 2012) and custom (F: 5'-CCAAAGCTGATATTCTATTTAAAC-3', R: 5'-ATAAGGCTAGGACCAAACCT-3', Internal: 5'-ACATATCTGCGTTATCTTACATAC-3') primer sets. We chose these two regions as they are the most commonly used in phylogeographic literature on*P. leucopus* , allowing for inclusion of Genbank sequences. Polymerase chain reactions (10 µL total volume) contained 5-10 ng of extracted DNA, 10 nM of forward and reverse primer, 2.0 mM of $MgCl_2$, 5.0 µL of 1:5 mix of Takara Taq Polymerase and Promega GoTaq MasterMix, and 2.9 µL nuclease-free water. PCRs began with initiation at 95 °C for 2 min then 35 cycles of denaturing at 95 °C for 15 seconds, annealing at 52 (cyt b) or 54°C (control region) for 30 seconds, and extension at 72 °C for 1 min, completed by an extension at 72 °C for 10 min. We sequenced the forward primer for cytochrome b and the forward and internal primers for control region to obtain the final dataset. Sequencing reactions initiated at 96 °C for 3 min then underwent 30 cycles at 96 °C for 10 seconds, 50 °C for 5 seconds, and 60 °C for 2.5 min. Sequencing reactions were then cleaned via an ethanol precipitation method (Latch and Rhodes 2005) and sequenced on an ABI 3500. All sequences were aligned and concatenated in Geneious Prime (Kearse et al. 2012).

Species Identification

Peromyscus spp. can be extremely difficult to distinguish based on morphology, and *P. leucopus* ' range overlaps with multiple other species including both wide-ranging species (e.g., *P. maniculatus* species group) and range-restricted species (*Peromyscus eremicus* species group). Therefore, we confirmed species identification via BLAST within Geneious Prime (Kearase et al. 2012). Of the 157 tissue samples, 27 were genetically identified as different species (*P. maniculatus, pectoralis, attwateri, gossypinus,* and *eremicus*) and 12 failed to amplify at both mitochondrial regions, leaving a total of 118 *P. leucopus* samples for phylogeographic analysis.

Combined Mitochondrial Dataset Analysis

To reconstruct phylogenetic relationships among *P. leucopus* , we used Bayesian coalescent models in BEAST v 2.6.6 (Bouckaert et al. 2019). Model parameters were set in BEAUti and included the following priors: coalescent constant population size, clock rate of 5% and 20% per million years for cytochrome b and control region respectively, and a strict molecular clock. With these priors, we performed Bayesian Markov chain Monte Carlo searches to estimate the most recent common ancestor for each major lineage. We sampled trees and divergence dates every 10,000 iterations for 10,000,000 generations, and checked model convergence in Tracer v.1.7.1 (Rambaut et al. 2018). Final consensus trees were visualized in FigTree v.1.4.4 (Rambaut 2018).

To complement the tree-building analysis, we constructed a median joining haplotype network within PopArt version 1.7 (Leigh and Bryant 2015). Large gaps of mutational steps are expected to separate Pleistocene lineages, allowing comparison to the BEAST analysis. Additionally, we performed a principal coordinates analysis (PCoA) with individual-based Kimura's 2-parameters distance (Kimura 1980). These distances were calculated in the R package ape via the "dist.dna" function (Paradis et al. 2004). We tested other evolutionary models for genetic distance between individuals, and the results were similar across all distances.

We expected lineages to be most distant from one another within the PCoA and correspond to the BEAST analysis and haplotype network.

For each lineage and the total dataset, we calculated measures of diversity within DNAsp v. 6 (Rozas et al. 2017). These metrics included number of haplotypes (h), haplotype diversity (h_D), nucleotide diversity (π), and mean number of pairwise nucleotide differences (k). We also calculated F_{ST} between all lineages in Arlequin v. 3.5.2.2 (Excoffier and Lischer 2010). Finally, we used Arlequin to estimate two metrics, Fu's Fs (Fu 1997) and Tajima's D (Tajima 1989), as both are expected to be significantly negative if a lineage underwent rapid expansion.

Reduced Control Region Dataset Analysis

We also downloaded 216 Genbank sequences of the control region to increase our geographic coverage. These sequences were georeferenced based on their original studies. Our control region sequences were trimmed to 636 bp to maximize power while allowing for additional sequences. In total, this reduced control region dataset included 334 sequences. We repeated the analyses described above (BEAST, haplotype network, PCoA, diversity, Fst, Fu's Fs, and Tajima's D) for the reduced dataset.

In addition to Fu's Fs and Tajima's D, we used Arlequin to calculate raggedness index and sum of square differences (SSD) under both demographic and spatial expansion. These metrics are derived from mismatch distributions where the distribution is expected to be unimodal if populations experienced recent demographic or spatial expansion. Multimodal or ragged distributions indicate a stable population over time (Rogers and Harpending 1992). Raggedness indices, therefore, describe the shape of the mismatch distributions and sum of SSD serves as a goodness of fit. Non-significant tests indicate that lineages either underwent demographic and/or spatial expansion. These metrics were not possible in the combined dataset as the markers evolve at different rates.

RESULTS

Combined Mitochondrial Dataset

For the combined dataset, we recovered 107 unique haplotypes among the 118 individuals. We trimmed the sequences to 1652 bp, which included 240 polymorphic sites (7 indels and 233 substitutions). Of the 240 polymorphic sites, 95 occurred once while 145 were parsimony informative.

For the BEAST analysis, the consensus tree highly supported three lineages (all BPP = 1.0; East, Midwest, and Southwest; Fig. 1). A fourth group was sister to East and was found to be distinct in all other analyses (see below), so we defined this group as another lineage denoted Central (Fig 1). The first split occurred between the Southwest lineage and the remaining three (time to most recent common ancestor [TMRCA] = 0.2511 ma, 95% HPD = 0.2100 – 0.2907 ma). BEAST estimated the next split was between the Midwest and East lineages to be 0.2026 ma (95% HPD = 0.1705 – 0.2364 ma). Although not statistically supported in the BEAST analysis (BPP = 0.652), the final split between East and Central was the newest divergence estimated at 0.1942 ma (95% HPD = 0.1627 – 0.2267 ma).

All lineages grouped geographically across North America with multiple areas of overlap (Fig 2a). East was broadly distributed east of the Mississippi River while the Midwest lineage occurred in the upper Midwest including the Upper Peninsula of Michigan, Wisconsin, Minnesota, and North Dakota as recorded in numerous other studies (e.g., Rowe et al. 2006, Fiset et al. 2015, Perno et al. 2022). Central occurred primarily in the southern Midwest and the eastern Great Plains while Southwest was found across Texas, Mexico, and New Mexico (Fig 2a). Significant overlap exists between the Central and Southwest lineages. The PCoA separated the four lineages with Axis 1 (variance explained = 25.78%) separating Southwest from the remaining three lineages. East and Midwest were most distant on Axis 2 (variance explained = 21.71%) with Central in the middle of all three lineages (Fig 2b). Similarly, all lineages are separated by at least 10 mutational steps within the haplotype network, and Central remains in the middle of all three lineages as observed in the PCoA (Fig 2c).

We found high diversity in all lineages (Table 1), with the majority of individuals in each lineage constituting a unique haplotype (81.57-96.88%). Haplotype diversity was similar across lineages (0.980-0.998) and nucleotide diversity ranged from 0.0073-0.0136 across lineages (all lineages combined = 0.0202). Pair-wise differences ranged from 12.106 (Midwest) to 22.595 (Central). Fu's Fs were significant in three (East, Midwest, and Southwest: -11.619-18.248, all $p < 0.001$) of the four lineages (Central = -1.998, $p = 0.421$) while no Tajima's D was significant (D = -1.207 - 1.657, all $p > 0.05$; Table 1). Fst values ranged from 0.424 (Central vs. Southwest) to 0.647 (Midwest vs. Southwest), and all were significant.

Reduced Control Region Dataset Analysis

Our reduced control region dataset consisted of 634 bp for a total of 276 haplotypes across 334 individuals. These haplotypes contained 167 polymorphic sites and 10 indels. Of the 167 polymorphic sites, 45 were singletons and 122 informative sites.

The reduced control region analyses largely agreed with the combined dataset despite lower power. With increased geographic coverage, it is apparent that there are areas of overlap between the lineages. East, Central, and Midwest individuals occur in Illinois and Indiana whereas Central and Southwest individuals occur throughout Texas and New Mexico (Fig 3a). The PCoA is largely similar to the combined dataset (Fig 3b), but Axis 1 (variance explained = 29.67%) separates East and Midwest first while Southwest, Midwest, and Central are spread along Axis 2 (variance explained = 13.42%). Lineages were less distinct in the haplotype network as expected with reduced power, but each lineage is separated by at least 4 mutations (Fig 3c).

All four lineages exhibited high diversity metrics as observed in the combined dataset (Table 2). Both haplotype (0.993-0.995) and nucleotide diversities (0.0144-0.0218) were similar across lineages whereas pairwise differences were highest in the Central (13.801) and Southwest (13.004) lineages compared to East (9.068) and Midwest (7.898). Sample sizes were much higher in East and Midwest (n = 166 and 93 respectively) compared to the other lineages (Central n = 44 and Southwest n = 32), which likely explains these differences. Fst values ranged from 0.448 (Central vs. Southwest) to 0.620 (East vs. Southwest) with all Fsts being significantly different from 0 (all $p < 0.001$).

All Fu's Fs metrics were significant (Fs = -13.132- -191.893, all $p < 0.001$) while all Tajima's D were not (D = -1.009- -1.905, all $p > 0.05$; Table 2). Tests for expansion found evidence for demographic expansion in all lineages (all Raggedness indices < 0.008, $p > 0.064$, SSD < 0.008, $p > 0.078$). No other tests for expansion were significant (Table 3).

DISCUSSION

Based on the combined and reduced control region datasets, we found strong evidence for three evolutionary lineages of white-footed mouse (East, Midwest, and Southwest) with a likely fourth lineage (Central) that received less support. These lineages disagree with the recognized morphological subspecies (Hall 1981). East, Midwest, and Southwest formed strongly supported monophyletic lineages within the BEAST analysis and were the most strongly differentiated groups in the haplotype networks, PCoA, and Fst analyses. Central lineage was monophyletic, but did not receive strong support in the phylogeny. However, it was differentiated from the other three lineages in the haplotype network, PCoA, and Fst analyses. Three of the four lineages crossed the Mississippi River, providing little evidence for the Mississippi River acting as a barrier for white-footed mice. Collectively, these results are consistent with additional Pleistocene lineages beyond the well characterized East and Midwest lineages.

Four Lineages in Peromyscus leucopus

In all analyses, the most divergent lineage was the Southwest lineage, a group found in Mexico, New Mexico, Texas, and southern Oklahoma. Multiple lines of evidence support divergence of the Southwest lineage, namely the highest estimated divergence time in BEAST (˜25kya) and Fst values among all between lineage comparisons. Furthermore, the Southwest lineage roughly correspond to the previously defined western chromosomal race as identified in Baker et al. (1983). Baker et al. (1983) found that the chromosomal races meet

at an ecotone transition between the Great Plains and eastern temperate forests (Strangl 1986) in Oklahoma, Kansas, and northern Texas. Our analysis extends the contact zone beyond Oklahoma to include eastern Texas and New Mexico, boundaries largely coincident with habitat transitions between grassland/desert and forests. A similar pattern of divergence is observed in the more cosmopolitan *Peromyscus maniculatus* where habitat is thought to reinforce divergence between putative species and lineages (Kalvik et al. 2012). Indeed, most lineages associated with forest and grassland habitats were once allopatric during the Pleistocene (Kalvik et al. 2012). Contact zones also occurred at habitat transitions, much like the breaks between the Southwest lineage and the remaining groups.

We believe the contact zone between the Southwest and the other lineages identified here exists in the southern Great Plains as observed in Strangl et al. (1986) where eastern deciduous forests transition into grassland habitats (i.e., eastern Oklahoma, Kansas, and Texas). This habitat transition acts as a contact zone between eastern and western lineages in both birds (Rising 1983) and mammals that were isolated during the Pleistocene (Reding et al. 2012, Kierepka et al. 2023), but levels of differentiation vary greatly across taxa. While our analysis presents evidence for fairly strong differentiation of southwestern *P. leucopus* , further study is needed to clarify this relationship. Mitochondrial DNA does not provide the full picture of evolutionary history due to its maternal inheritance and differing evolutionary rates. Reclassification has occurred in other *Peromyscus* species including *P. maniculatus* , which was separated into six species based on numerous genetic, ecological, and morphological studies (e.g., Dragoo et al. 2006, Kalvik et al. 2012, Greenbaum et al. 2019). The Southwest lineage could be a cryptic species as observed in the *P. maniculatus* species group (*P. maniculatus* and. *P. sonoriensis*), but nuclear DNA should confirm differentiation and patterns of hybridization within the contact zone.

Unlike the Southwest lineage, the East and Midwest lineages are primarily found in forested habitats. Splits between these lineages were aged at ˜20,000 years ago during the Last Glacial Maximum (LGM) and received strong support in the haplotype network, PCoA, and Fst analyses. We acknowledge these estimates may be underestimations based on previous estimates (e.g. Moscarella et al. 2019, Prado et al. 2019). However, our TMRCA remained in the Pleistocene, which supports previous findings of two Pleistocene lineages separated by the Great Lakes and the St. Lawrence River (Rowe et al. 2006, Fiset et al. 2015, Baumgartner and Hoffman 2019, Moscarella et al. 2019, Prado et al. 2022). Beyond the Great Lakes and northeastern United States/Canada, East has the largest geographic distribution of *P. leucopus'* lineages, extending along the entire eastern coast and Appalachians into western Tennessee, Alabama, and Louisiana. We also recorded East individuals in Wisconsin and Illinois whereas the Midwest lineage was largely confined to northern Midwest states including Minnesota, Illinois, the Upper Peninsula of Michigan, Wisconsin, Nebraska, and North Dakota. These areas largely correspond with the different hypothesized colonization routes taken following the LGM in the East and Midwest lineages and large water barriers that existed prior to the LGM (Rowe et al. 2006, Fiset et al. 2015, Moscarella et al. 2019, Prado et al. 2022). With the greater coverage of the reduced dataset, we found that East and Midwest co-occur in Indiana and Illinois along with the Central lineage. All three lineages crossed the Mississippi River, so we did not find evidence for it as a barrier.

While Southwest, East, and Midwest were highly supported in every analysis, the Central lineage is the least distinct. Central did not receive high support in the BEAST analysis, had the youngest divergence time, and was in the middle of the haplotype network and PCoA. Fst metrics were also smallest in Central comparisons. Interestingly, Central was placed sister to East in the phylogenetic tree, but was the least differentiated from Southwest in the PCoA and haplotype network. Rowe et al. (2006)'s dataset placed Central sister to the Midwest lineage, but also did not receive high support in the phylogeny. Fst values were the smallest between Central and Southwest, which appears to support that Central is more closely related to Southwest. Central also does not occur east of Indiana and primarily occurs in Missouri, east Texas, Louisiana, Illinois, Indiana, and northern New Mexico. All these areas are largely forested, which may indicate habitat-based differentiation between Southwest and Central or a separate refugium like the Ozarks. The hypothesized refugium for the Midwest lineage occurred in Illinois (Rowe et al. 2006), so another isolated refugium within a southern area is most likely for Central. We cannot discount that these results could be due to low sample sizes of Central or homoplasy. Much of our eastern Central individuals were

from Genbank, and therefore, only had control region sequences, a marker more susceptible to homoplasy than slower evolving regions. This lineage was also detected in Herrera (2021), which used cytochrome b, so homoplasy in two markers is unlikely for the Central lineage. Consequently, more data is needed to elucidate the evolutionary relationship of Central to the other lineages.

All Lineages Underwent Expansion

All expansion tests supported expansion for all lineages despite Central and Southwest occurring in unglaciated areas within North America. This result was somewhat surprising as climatic projections indicated suitable habitat persisted within these areas during the LGM (Herrera 2021). Refugial populations are expected to harbor the most genetic diversity versus those on the expansion front, which could help with identifying refugia for the Southwest and Central lineages. For the Southwest lineage, repeated haplotypes only occurred in northern Texas and Kansas, which suggests a possible northward expansion front. Within the reduced dataset, the Central lineage exhibited higher genetic diversity in the southern range with shared haplotypes only occurring within Indiana, Illinois, and Kansas. We caution that the high frequency of unique haplotypes within these lineages may obscure patterns of expansion, but both appear to have expanded north as greater diversity occurs in the southern areas within each lineage. Furthermore, expansion from southern refugia has been found in multiple lineages within the southwestern United States and Mexico (e.g., Barton and Wisely 2012, Myers et al. 2020, McDonough et al. 2022).

While expansion patterns in Central and Southwest are somewhat unclear, our analysis confirmed strong signatures of expansion within the East and Midwest lineages. Both lineages occupy areas once covered by glaciers, so both spatial and demographic expansion is expected following glacial recession. Furthermore, both lineages are expanding north into Canada due to climate change (e.g., Fiset et al. 2015, Garcia-Elfring et al. 2017). Previously glaciated areas like the northeastern United States and Michigan had the lowest number of haplotypes compared to more southern populations as expected for refugial and recolonized populations (Petit et al. 2003). In addition to northward expansions, the East lineage, in particular, appears to be expanding west given that haplotypes were found in Wisconsin, Louisiana, Illinois, and western Tennessee. The East and Midwest lineages are both known to carry *B. burgdorferi* , so this westward movement likely will facilitate the continued expansion of Lyme Disease when black-legged ticks (*Ixodes scapularis*) are present. Indeed,*Ixodes scapularis* populations are becoming established and increasing throughout their range including in western areas (Maestas et al. 2016, Gardner et al. 2020, Pasternak and Palli 2022), and typically, Lyme Disease reports quickly follow population establishment. Expansion of East and Midwest lineage white-footed mice may certainly contribute to Lyme Disease, but many other factors contribute to this spread including climate, landscape heterogeneity, questing behavior of tick nymphs, and movement of other hosts (e.g., Eisen et al. 2016, Halsey et al. 2018, Gardner et al. 2020). It is currently unknown if Central and Southwest lineages carry *B. burgdorferi* like East and Midwest, but Lyme Disease remains rare in warmer climates despite both *P. leucopus* and *I. scapularis* being present (Ginsberg et al. 2021).

Conclusions

Overall, this study found four evolutionary lineages of white-footed mice in North America based on two mitochondrial regions, all of which diverged during Pleistocene glaciation. The most divergent lineage, Southwest, was largely limited to grassland and desert habitats whereas the other three mainly occurred in deciduous forest. Thus, Southwest could be adapted to such habitats and exhibit genomic change that allows them to persist outside the typical forested habitats of *P. leucopus* . This lineage, therefore, may represent incipient speciation as observed in other *Peromyscus spp* , but further clarification is needed, especially within contact zones to measure patterns in hybridization. Additionally, we recorded signatures of expansion in all four lineages, but these signals were particularly evident in the East and Midwest lineages. These lineages are known reservoirs for tick-borne illnesses, but current work has focused on regions where only East and Midwest occur (i.e., northern latitudes). Comparative little is known about the lineages of Central and Southwest, and their relationship to*B. burgdorferi* and other potential human pathogens. Collectively, these results highlight the complex evolutionary histories of cosmopolitan species like white-footed mice, and the importance of understanding these dynamics in reservoir species that are actively expanding.

AUTHOR CONTRIBUTIONS

Sydney Walters was involved in Conceptualization (Supporting), Data Curation (Lead), Formal Analysis (Supporting), Investigation (Lead), Methodology (Lead), Validation (Lead), and Writing-original draft (Supporting). Both **Marketa Zimova** and **Jenifer Mallinoff** contributed through Investigation (Supporting), Methodology (Supporting), Resources (Supporting), and Writing-review and editing (Equal). Elizabeth Kierepka participated through Conceptualization (Lead), Data Curation (Equal), Formal Analysis (Lead), Funding Acquisition (Lead), Investigation (Equal), Methodology (Equal), Project Administration (Lead), Supervision (Lead), Validation (Supporting), Visualization (Lead), Writing – original draft (Equal), and Writing – review and editing (Equal).

ACKNOWLEDGEMENTS

We would like to thank the mammal collections staff of the Museum of Southwestern Biology, Texas A&M University's Biodiversity Research and Teaching Collections, Museum of Texas Tech University, Angelo State Natural History Collection, and Arizona State University (NEON Biorepository) for providing tissue samples of *Peromyscus leucopus* .

FUNDING INFORMATION

Funding for this project was provided by the James R. and Judy Allen Pick Award at North Carolina State University and was instrumental in providing research experience for undergraduate scholars.

CONFLICT OF INTEREST STATEMENT

The authors declare no conflict of interest.

DATA AVAILABILITY STATEMENT

All sequence data will be uploaded to Dryad and Genbank upon acceptance.

LITERATURE CITED

Baker, R.J., Robbins, L.W., Stangl Jr, F.B., Birney, E.C. 1983. Chromosomal evidence for a major subdivision in *Peromyscus leucopus* . Journal of Mammalogy 64(2): 356-359.

Barton, H.D., Wisely S.M. 2012. Phylogeography of striped skunks (*Mephitis mephitis*) in North America: Pleistocene dispersal and contemporary population structure. Journal of Mammalogy 93(1): 38-51.

Baumgartner, J.M., Hoffman, S.M. 2019. Comparison of the responses of two Great Lakes lineages of *Peromyscus leucopus* to climate change. Journal of Mammalogy 100(2): 354-364.

Boria, R.A., Blois, J.L. 2023. Phylogeography within the*Peromyscus maniculatus* species group: understanding past distribution of genetic diversity and areas of refugia in western North America. Molecular Phylogenetics and Evolution 180: 107701.

Bouckaert, R., Vaughan, T.G., Barido-Sottani, J., Duchene, S., Fourment, M., Gavryushkina, A., Heled, J., Jones, G., Kühnert, D., De Maio, N., Matschiner, M., Mendes, F.K., Müller, N.F., Ogilvie, H.A., du Plessis, L., Popinga, A., Rambaut, A., Rasmussen, D., Siveroni, I., Suchard, M.A., Wu, C., Xie, D., Zhang, C., Stadler, T., Drummond, A.J. 2019. BEAST 2.5: An advanced software platform for Bayesian evolutionary analysis. PloS Computational Biology 15(4): e1006650.

Bradley, R.D., Durish, N.D., Rogers, D.S., Miller, J.R., Engstrom, M.D., Kilpatric, C.W. 2007. Toward a molecular phylogeny for*Peromyscus* : evidence from mitochondrial cytochrome-b sequences. Journal of Mammalogy 88(5): 1146-1159.

Bradley, R.D., Ordóñez-Garza, N., Thompson, C.W., Wright, E.A., Ceballos, G., Kilpatrick, C.W., Schmidly, D.J. 2022. Two new species of*Peromyscus* (Cricetidae: Neotominae) from the Transverse Volcanic Belt of Mexico. Journal of Mammalogy 103(2): 255-274.

Burbrink, F.T., Bernstein, J.M., Kuhn, A., Gehara, M., Ruane, S. 2022. Ecological divergence and the history of gene flow in the Nearctic milksnakes (*Lampropeltis triangulum* complex). Systematic Biology 71(4): 839-858.

Castañeda-Rico, S., León-Paniagua, L., Vázquez-Domínguez, E., Navarro-Sigüenza, A.G. 2014. Evolutionary diversification and speciation in rodents of the Mexican lowlands: the *Peromyscus melanophrys* species group. Molecular Phylogenetics and Evolution 70: 454-463.

Dragoo, J.W., Lackey, J.A., Moore, K.E., Lessa, E.P., Cook, J.A., Yates, T.L. 2006. Phylogeography of the deer mouse (*Peromyscus maniculatus*) provides a predictive framework for research on hantaviruses. Journal of General Virology 87(7): 1997-2003.

Eisen, R.J., Eisen, L. Ogden, N.H., Beard, C.B. 2016. Linkages of weather and climate with *Ixodes scapularis* and *Ixodes pacificus* (Acari: Ixodidae), enzootic transmission of *Borrelia burgorferi* , and Lyme Disease in North America. Journal of Medical Entomology 53(2): 250-261.

Excoffier, L., Lischer, H.E.L. 2010. Arlequin suite ver 3.5: a new series of programs to perform population genetics analyses under Linux and Windows. Molecular Ecology Resources 10(3): 564–567

Fiset, J., Tessier, N. Millien, V., Lapointe, F.J. 2015. Phylogeographic structure of the white-footed mouse and the deer mouse, two Lyme Disease reservoir hosts in Quebec. PloS One 10(12): e0144112.

Fu, Y. 1997. Statistical tests of neutrality of mutations against population growth, hitchhiking and background selection. Genetics 147(2): 915-925.

Garcia-Elfring, A., Barrett, R.D.H., Combs, M., Davies, T.J., Munshi-South, J., Millien, V. 2017. Admixture on the northern front: population genomics of range expansion in the white-footed mouse (*Peromyscus leucopus*) and secondary contract with the deer mouse (*Peromyscus maniculatus*). Heredity 119: 447-458.

Gardner, A.M., Pawlikowski, N.C., Hamer, S.A., Hickling, G.J., Miller, J.R., Schotthoefer, A.M. Tsao, J.I., Allan, B.F. 2020. Landscape features predict the current and forecast the future geographic spread of Lyme Disease. Proceedings of the Royal Society B Biological Sciences 287(1941): 20202278.

Ginsberg, H.S., Hickling, G.J., Burke, R.L., Ogden, N.H., Beati, L., LeBrun, R.A., Arsnoe, I.M., Gerhold, R., Han, S., Jackson, K., Maestas, L., Moody, T., Pang, G., Ross, B., Rulison, E.L., Tsao, J.I. 2021. Why Lyme Disease is common in the northern US, but rare in the south: the roles of host choice, host-seeking behavior, and tick density. PloS Biology 19(9): e3001396.

Greenbaum, I.F., Honeycutt, R.L., Chirhart, S.E. 2019. Taxonomy and phylogenetics of the *Peromyscus maniculatus* species group. In From Field to Laboratory: A Memorial Volumer in Honor of Robert J. Baker. R.D. Genoways, H.H., Schmidley, D.J., Bradley, L.C. Eds. Special Publications, Museum of Texas Tech University: Lubbock, Texas, United States of America. Volume 71, pp. 559-575.

Hahn, E.N., Fore, S.A. 2019. Changes in abundance of the *Ixodes scapularis* say (Blacklegged tick) in Adair County, Missouri, from 2006 to 2015. Northeastern Naturalist 26: 137-140.

Hall, E.R. 1981. The mammals of North America. 2ndedition. Volume 2. John Wiley and Sons, Inc., New York, United States of America.

Halsey, S.J., Allan, B.F., Miller, J.R. 2018. The role of *Ixodes scapularis* , *Borrelia burgdorferi* , and wildlife hosts in Lyme disease prevalence: a quantitative review. Ticks and Tick-borne Diseases 9(5): 1103-1114.

Herrera, T.M. 2021. Comparative phylogeography of small mammals across the Great Plains Suture Zone highlights repeated processes of speciation and community assembly coincident with the 100thmeridian. MS Thesis, Kansas State University.

Hewitt, G.M. 2004. Genetic consequences of climatic oscillations in the quaternary. Philosophical Transactions of the Royal Society B Biological Sciences 359 (1442): 183–195.

Kalvik, H.M., Stout, I.J., Doonan, T.J, Parkinson, C.L. 2012. Investigating niche and lineage diversification in widely distributed taxa: phylogeography and ecological niche modeling of the*Peromyscus maniculatus* species group. Ecography 35(1): 54-64.

Kearse, M., Moir, R., Wilson, A., Stones-Havas, S., Cheung, M. Sturrock, S., Buxton, S., Cooper, A., Markowitz, S., Duran, C., Their, T., Ashton, B., Meintjes, P., 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.

Kierepka, E.M., Latch, E.K. 2016. High gene flow in the American badger overrides habitat preferences and limits broadscale genetic structure. Molecular Ecology 25(24): 6055-6076.

Kierepka, E.M., Preckler-Quisqater, S., Reding, D.M., Piaggio, A.J., Riley, S.P.D., Sacks, B.N. 2023. Genomic analyses of gray fox lineages suggest ancient divergence and secondary contact in the southern Great Plains. Journal of Heredity 114(2): 110-119.

Kimura, M. 1980. A simple method for estimating evolutionary rates of base substitutions through comparative studies of nucleotide sequences. Journal of Molecular Evolution 16: 111-120.

Klicka, J., Epperly, K., Smith, B.T., Spellman G.M., Chaves, J.A., Escalante, P., Witt, C.C., Canales-del-Castillo, R., Zink, R.M. 2023. Lineage diversity in a widely distributed New World passerine bird, the house wren. Ornithology 140(11): ukad018.

Latch, E.K., Rhodes Jr, O.E. 2005. The effects of gene flow and population isolation on the genetic structure of reintroduced wild turkey populations: are genetic signatures of source populations retained? Conservation Genetics 6: 981-997.

Leigh, J.W., Bryant, D. 2015. POPART: full-feature software for haplotype network construction. Methods in Ecology and Evolution 6(9): 1110–1116.

León-Tapia, M.A., Rico, Y., Fernández, J.A., Espinosa de los Monteros, A. 2022. Molecular, morphometric, and spatial data analyses provide new insights into the evolutionary history of the *Peromyscus boylii* species complex (Rodentia: Cricetidae) in the mountains of Mexico. Systematics and Biodiversity 20(1): 1-19.

Maestas, L.P., Adams, S.L., Britten, H.B. 2016. First evidence of an established population of *Ixodes scapularis* (Acari: Ixodidae) in South Dakota. Journal of Medical Entomology 53(4): 965-966.

McDonough, M.M., Ferguson, A.W., Dowler, R.C., Gompper, M.E., Maldonado, J.E. 2022. Phylogenomic systematics of the spotted skunks (Carnivora: Mephitidae, *Spilogale*): additional species diversity and Pleistocene climate change as a major driver of diversification. Molecular Phylogenetics and Evolution 167: 107266.

Moscarella, R.A., Hoffman, S.M.G., Myers, P., Yahnke, C.J., Lundrigan, B.L. 2019. Genetic and demographic analysis of invasive *Peromyscus leucopus* in the northern Great Lakes region. Journal of Mammalogy 100(2): 345-353.

Myers, E.A., McKelvy, A.D., Burbrink, F.T. 2020. Biogeographic barriers, Pleistocene refugia, and climatic gradients in the southeastern Nearctic drive diversification in cornsnakes (*Pantherophis guttatus* complex). Molecular Ecology 29: 797-811.

Naidu, A., Fitak, R.R., Munguia-Vega, A., Culver, M. 2012. Novel primers for complete mitochondrial cytochrome b gene sequencing in mammals. Molecular Ecology Resources 12(2): 191-196.

Paradis, E., Claude, J., Strimmer, K. 2004. APE: analyses of phylogenetics and evolution in R language. Bioinformatics 20(2): 289-290.

Pasternak, A.R., Palli, S.R. 2022. Mapping distributions of the Lyme disease vector, *Ixodes scapularis* , and spirochete, *Borrelia burgdorferi* , in Kentucky using passive and active surveillance. Ticks and Tick-borne Diseases 13(2): 101885.

Petit, R.J., Aguinagalde, I., De Beaulieu, J., Bittkau, C., Brewer, S., Cheddadi, R., Ennos, R., Fineschi, S., Grivet, D., Lascoux, M., Mohanty, A., Müller-Stark, G., Demesure-Musch, B., Palmé, A., Martín, J.P., Rendell, S., Vendramin, G.G. 2003. Glacial refugia: hotspots but not melting pots of genetic diversity. Science 300(5625): 1563-1565.

Prado, J.R., Rubi, T.L., Baumgartner, J., Hoffman, S.M., Dantzer, B., Knowles, L.L. 2022. Postglacial colonization in the Great Lakes Region by the white-footed mouse (*Peromyscus leucopus*): conflicts between genomic and field data. Journal of Mammalogy 103(2): 243-254.

Puckett, E.E., Etter, P.D., Johnson, E.A., Eggert, L.S. 2015. Phylogeographic analyses of American black bears (*Ursus americanus*) suggest four glacial refugia and complex patterns of postglacial admixture. Molecular Biology and Evolution 32(9): 2338-2350.

Rambaut, A. 2018. FigTree, version 1.4.4 [online]. Available from http://tree.bio.ed.ac.uk/software-/figtree/ [accessed 2 January 2022].

Rambaut, A., Drummond, A.J., Xie, D., Baele, G., Suchard, M.A. 2018. Posterior summarization in Bayesian phylogenetics using Tracer 1.7. Systematic Biology 67(5): 901-904.

Reding, D.M., Bronikowski, A.M., Johnson, W.E., Clark, W.R. 2012. Pleistocene and ecological effects on continental-scale genetic differentiation in the bobcat (*Lynx rufus*). Molecular Ecology 21(12): 3078-3093.

Riddle, B.R., Hafner, D.J. 2006. A step-wise approach to integrating phylogeographic and phylogenetic biogeographic perspectives on the history of a core North American warm deserts biotas. Journal of Arid Environments 66: 435-461.

Riddle, B.R., Hafner, D.J., Alexander, L.F. 2000. Phylogeography and systematics of the *Peromyscus eremicus* species group and the historical biogeography of North American warm regional deserts. Molecular Phylogenetics and Evolution 17(2): 145-160.

Rising, J.D. 1983. "The Great Plains hybrid zones" in *Current ornithology.* Plenum Press, New York, United States of America. pp. 131–157

Rogers, A.R., Harpending, H. 1992. Population growth makes waves in the distribution of pairwise genetic differences. Molecular Biology and Evolution 9(3): 552–569.

Rowe, K.C., Heske, E.J., Paige, K.N. 2006. Comparative phylogeography of eastern chipmunks and white-footed mice in relation to the individualistic nature of species. Molecular Ecology 15(13): 4003-4020.

Roy-Dufresne, E., Logan, T., Simon, J.A., Chmura, G.L., Millien, V. 2013. Poleward expansion of the white-footed mouse (*Peromyscus leucopus*) under climate change: implications for the spread of Lyme Disease. PloS One 8(11): e80724.

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 datasets. *Molecular Biology and Evolution* **34(12):** 3299-3302.

Shafer, A.B.A., Cullingham, C.I., Cote, S. D., Coltman, D.W. 2010. Of glaciers and refugia: a decade of study sheds new light on the phylogeography of northwestern North America. Molecular Ecology 19(21): 4589-4621.

Shipp-Pennock, M.A., Webster, D., Freshwater, D.W. 2005. Systematics of the white-footed mouse (*Peromyscus leucopus*) in the mid-Atlantic region. Journal of Mammalogy 86(4): 803-813.

Stangl Jr, F.B. 1986. Aspects of a contact zone between two chromosomal races of *Peromyscus leucopus* (Rodentia: Cricetidae). Journal of Mammalogy 67(3): 465-473.

Tajima, F. 1989. Statistical method for testing the neutral mutation hypothesis by DNA polymorphism. Genetics 123(3): 585–595.

Table 1. Diversity and expansion metrics for the full dataset of cytochrome b and control region (732 bp cytochrome b, 920 bp control region; 1652 bp total) for four lineages (Central, East, Midwest, and Southwest) and the total dataset (Total). Metrics include number of individuals (n), number of haplotypes (h), haplotype diversity (h_D), nucleotide diversity (π), average pair-wise nucleotide differences (k), Fu's Fs (Fs), and Tajima's D (D). All Fu's Fs values except in Central were significant (bold) while no Tajima's D tests were significant.

Lineage	n	h	h_D	π	k	Fs	D
Central	18	16	0.980	0.0136	22.595	-1.998	-1.207
East	38	31	0.986	0.0082	13.644	**-11.619**	-1.657
Midwest	30	29	0.998	0.0073	12.106	**-18.248**	-1.549
Southwest	32	31	0.998	0.0122	20.204	**-13.839**	-1.386
Total	118	107	0.998	0.0202	33.424	**-33.126**	-1.099

Table 2. Diversity and expansion metrics for the reduced control region dataset (634 bp control region) for inferred lineages (Central, East, Midwest, and Southwest) and the total dataset (Total). We report the same metrics as the total dataset: number of individuals (n), number of haplotypes (h), haplotype diversity (h_D), nucleotide diversity (π), average pair-wise nucleotide differences (k), Fu's Fs (Fs), and Tajima's D (D). Both Fu's Fs and Tajima's D were negative, with only Fs being significant (bold).

Lineage	n	h	h_D	π	k	Fs	D
Central	44	39	0.993	0.0218	13.801	**-21.949**	-1.488
East	166	131	0.995	0.0144	9.068	**-191.893**	-1.676
Midwest	93	78	0.995	0.0125	7.898	**-33.792**	-1.905
Southwest	32	28	0.994	0.0206	13.004	**-13.132**	-1.009
Total	334	276	0.998	0.0289	18.051	**-431.882**	-1.310

Table 3: Tests of demographic and spatial expansion for the reduced control region dataset. Both tests calculate two statistics (raggedness index [Ragg] and sum of square differences [SSD]), and non-significant values indicate the lineage underwent either demographic (Demo) or spatial (Spatial) expansion. All lineages exhibited signatures of both demographic and spatial expansion (i.e., all Ragg and SSD values were significant).

Lineage	Ragg_Demo	SSD_Demo	Ragg_Spatial	SSD_Spatial
Central	0.003	0.004	0.003	0.005
East	0.006	0.006	0.005	0.002
Midwest	0.008	0.018	0.008	0.004
Southwest	0.006	0.006	0.006	0.003

FIGURE CAPTIONS

Fig 1. Bayesian tree based on 107 unique haplotypes in the combined cytochrome b and control region dataset (1652 bp) calculated in BEAST. Black dots indicate nodes with a posterior probability (BPP) of 1.0. Three lineages (East, Midwest, and Southwest) received high support whereas Central did not (BPP = 0.652). Time since most recent common ancestor (TMRCA) analysis estimated the oldest split as Southwest vs. the other lineages. Each TMRCA estimate is positioned above its corresponding node along with its 95% confidence interval.

Fig 2. Geographic distribution (A), principal coordinates analysis (PCoA, B), and median joining haplotype network (C) for the combined cytochrome b and control region (1652 bp, n = 118). The Central (light

blue, squares), East (dark blue, diamonds), Midwest (purple, circles), and Southwest (orange, triangles) showed differentiation in both the PCoA and haplotype network with Central in the middle of the three other lineages. The first axis of the PCoA (variance explained = 25.78%) separated Southwest and Central (negative values) from Midwest and East (positive values) whereas the second axis (variance explained = 21.71%) separated East and Midwest. Correspondingly, the median joining haplotype network exhibited four main groups separated by a minimum of 15 mutational steps. The most divergent groups were Southwest and Midwest (52 total mutational steps).

Fig 3. Results of the reduced control region dataset (636 bp) in 336 white-footed mice. Greater geographic sampling clarified the distribution of each lineage (A) where Central (light blue, squares) occurred in the southern Midwest and parts of the southwest, East (dark blue, diamonds) in most of the eastern United States and Canada, Midwest in the upper Midwest (purple, circles), and Southwest within Texas, Mexico, Oklahoma, and New Mexico (orange, triangles). Despite decreased power in the reduced dataset, all four lineages were apparent in the principal coordinate analysis (PCoA; B), and median joining haplotype network (C). Axis 1 of the PCoA separated Central, Midwest, and Southwest (negative values) from East (positive values; variance explained = 29.67%). The second axis revealed differentiation between Midwest, Central, and Southwest (variance explained = 13.42%). In the haplotype network, all lineages were separated by at least 6 mutational steps, and the most divergent groups were Southwest and East (17 total mutation steps).