Journal of Mammalogy, 103(2):243–254, 2022
https://doi.org/10.1093/jmammal/gyab158

Postglacial colonization in the Great Lakes Region by the white-footed mouse (*Peromyscus leucopus*): conflicts between genomic and field data

Joyce R. Prado[1,*], Tricia L. Rubi[2], Joseph Baumgartner[3], Susan M. G. Hoffman[4], Ben Dantzer[2,5], and L. Lacey Knowles[5]

[1]*Departamento de Ciências Biológicas, Escola Superior de Agricultura "Luiz de Queiroz", Universidade de São Paulo, Piracicaba, SP, Brazil*
[2]*Department of Psychology, The University of Michigan, Ann Arbor, MI, USA*
[3]*Department of Biology, UC Blue Ash College, Blue Ash, Ohio, USA*
[4]*Department of Biology, Miami University, Oxford, Ohio, USA*
[5]*Department of Ecology and Evolutionary Biology, The University of Michigan, Ann Arbor, MI, USA*

**To whom correspondence should be addressed: joycepra@gmail.com*

For regions that were covered by ice during the Pleistocene glaciations, species must have emigrated from unglaciated regions. However, it can be difficult to discern when and from what ancestral source populations such expansions took place, especially since warming climates introduce the possibility of very recent expansions. For example, in the Great Lakes region, pronounced climatic change includes past glaciations as well as recent, rapid warming. Here we evaluate different expansion hypotheses with a genomic study of the white-footed mouse (*Peromyscus leucopus noveboracensis*), which is one of the most common mammals throughout the Great Lakes region. Ecological surveys coupled with historical museum records suggest a recent range expansion of *P. leucopus* associated with the warming climate over the last decades. These detailed records have yet to be complemented by genomic data that provide the requisite resolution for detecting recent expansion, although some mitochondrial DNA (mtDNA) sequences have suggested possible hypotheses about the geography of expansion. With more than 7,000 loci generated using RADseq, we evaluate support for multiple hypotheses of a geographic expansion in the Upper Peninsula of Michigan (UP). Analysis of a single random single-nucleotide polymorphism per locus revealed a fine-scale population structure separating the Lower Peninsula (LP) population from all other populations in the UP. We also detected a genetic structure that reflects an evolutionary history of postglacial colonization from two different origins into the UP, one coming from the LP and one coming from the west. Instead of supporting a climate-driven range expansion, as suggested by field surveys, our results support more ancient postglacial colonization of the UP from two different ancestral sources. With these results, we offer new insights about *P. leucopus* geographic expansion history, as well as a more general phylogeographic framework for testing range shifts in the Great Lakes region.

Key words: climatic change, geographic barriers, phylogeography, pleistocene, rodentia

Understanding the factors that shape the genetic structure of species by separating contemporary drivers (e.g., recent warming of temperate regions) from past drivers (e.g., the formation of geographical barriers or historic climatic shifts such as those associated with the Pleistocene glaciations) is critical, as historical processes may confound demographic analyses focused on the recent past (He et al. 2013; Lanier et al. 2015). In northern glaciated landscapes, processes linked to the Pleistocene climatic shifts are posited to be some of the most important in structuring genetic variation (Knowles and Massatti 2017; Baumgartner and Hoffman 2019), especially when the ranges of species were displaced to southern refugia during glacial periods (Rowe et al. 2004; Bemmels et al. 2019).

The Great Lakes region offers an interesting opportunity to understand the effects of contemporary and past events on the genetic structure of terrestrial species. The lakes are a substantial water barrier limiting the dispersal of taxa (Rowe et al. 2006; Moscarella et al. 2019). For example, there are approximately 80 native mammal species in the Great Lakes region, over half of which have southern or northern distributional limits in the area (Myers et al. 2009; Kurta 2017). The region has also been ice free for only about 12,000 years (Holman 2001; Moscarella et al. 2019). Moreover, like other northern latitudes, the Great Lake region has experienced substantial warming during the last century (Handler et al. 2014; Baumgartner and Hoffman 2019).

The white-footed mouse (*Peromyscus leucopus noveboracensis*; hereafter *P. leucopus*) is the most common mammal species throughout the region (Moscarella et al. 2019). This generalist species is found in virtually all deciduous forest stands in the midwestern and eastern United States (Lackey et al. 1985). It is small (~20–30 g), nocturnal, semi-arboreal, and a common host for the major vector of Lyme disease in North America (*Ixodes scapularis*), as well as serving as the primary natural reservoir for the bacterium that causes Lyme disease (Lackey et al. 1985; Cook and Barbour 2015; Pardiñas et al. 2017). Recent expansion of *P. leucopus* has been linked to changes in the climate (Rowe et al. 2006; Baumgartner and Hoffman 2019; Moscarella et al. 2019). For example, multiple studies have traced the demographic history of *P. leucopus* in the northern Midwest, which includes an expansion during the postglacial colonization of the region, as well as a recent contemporary northward expansion (e.g., from the southern Great Lakes deciduous forests to the Upper Peninsula of Michigan (hereafter, the UP), northern Wisconsin and Minnesota, and central Quebec) due to the rapid warming of northern latitudes (Rowe et al. 2006; Jannett et al. 2007; Myers et al. 2009; Roy-Dufresne et al. 2013; Baumgartner and Hoffman 2019; Moscarella et al. 2019). Within the Great Lakes region specifically, historical museum and literature records from the early to mid-1900´s indicate that the northern Lower Peninsula of Michigan (hereafter, the LP) was the limit of *P. leucopus*'s range in the north (Osgood 1909; Myers et al. 2009). However, records of occurrence began in 1939 in Michigan's UP (Myers et al. 2009), but only in a restricted area in the southernmost part of the peninsula (in Menominee County). Starting in 1981, *P. leucopus* began to appear in other parts of the UP, and they were first captured in the easternmost part of the peninsula in 2004 (Myers et al. 2009).

Such detailed tracing through museum specimens and historical records can provide information about the distributional shifts of *P. leucopus* in response to recent climate change. However, there are limits to the insights they provide. In particular, they cannot test hypotheses about gene flow associated with shifting distributions or population bottlenecks that impact genetic structure and diversity. Such inferences about the demography of colonization of previously glaciated areas are important beyond a focus on the history of expansion itself. For example, gene flow and/or reduced genetic diversity have consequences for the ongoing response of species to climatic changes (e.g., constraining genetic variation, and/or limiting adaptive responses). Such information can only be gained from genomic studies.

There have been a few genetic analyses in *P. leucopus*, but given they were based on single locus studies (i.e., mitochondrial DNA, mtDNA) or a relatively small set of microsatellites, they are limited in their resolution. For example, no genetic differences between *P. leucopus* populations in the eastern UP and northern LP were detected based on a set of eight microsatellite markers (Baumgartner and Hoffman 2019). In contrast, analysis of sequence data of mtDNA detected regional differences in Great Lakes *P. leucopus*. Specifically, one mtDNA lineage found in individuals from UP populations (i.e., areas previously covered by glacial ice) is also found in Wisconsin, while another mtDNA lineage found in the LP of Michigan was estimated to have diverged from a shared common ancestor about 34,500 BP (Rowe et al. 2006; Moscarella et al. 2019). This divergence predates the formation of Lake Michigan (13,000 BP; Holman 2001). Moreover, the geographic distribution of the divergent mtDNA lineages suggests that the postglacial colonization of the previously glaciated area of UP proceeded from the western mtDNA lineage, which had spread from Illinois to Wisconsin, Minnesota, and the southernmost UP, as opposed to the eastern mtDNA lineage, which spread from Ohio into the LP of Michigan (Rowe et al. 2006; Moscarella et al. 2019). The ancestral source population for the recent expansion of *P. leucopus* into the UP was therefore identified as part of the western lineage that followed a colonization route from populations in Wisconsin. A few individuals of the eastern mtDNA lineage were also detected in the eastern UP during sampling of the species in 2007 (Moscarella et al. 2019), as well as at additional sites in the eastern UP in samples collected in 2017 (Baumgartner 2017), which suggests a possible contact zone between two expanding fronts (i.e., one lineage from Wisconsin and another from the LP that collide in the UP).

Here we investigate the expansion history of *P. leucopus* using genomic data, testing hypotheses about its geographic expansion into the UP, including the previous hypotheses suggested by mtDNA analyses. Specifically, capitalizing on the resolution provided by random genomic markers from RADSeq (Vendrami et al. 2019), we (i) test for genetic structure among populations, (ii) infer the time of genetic divergence, and (iii) evaluate different expansion scenarios in *P. leucopus*. Specifically, we test whether the divergence times estimated from parameterized models correspond better with postglacial expansion in the more distant past (e.g., following the last glacial maximum —LGM), versus recent climatic changes (i.e., the warming of the Northern Hemisphere) as suggested by historical museum records, since this was not addressed by previous studies based on mtDNA. In addition, based on the information presented above, we also test hypotheses about two different putative ancestral sources for the colonization of the UP. A first scenario where the UP populations were colonized by two separate events, i.e., the common ancestry of the western UP populations, is distinct from the common ancestry of the eastern UP and LP populations (see Rowe et al. 2006; Moscarella et al. 2019), and a second scenario with a single colonization of the UP populations from the western part of the species range (see Myers et al. 2009; Baumgartner and Hoffman 2019). Based on the best fit model identified statistically (using AIC criteria), we discuss the implications of a correspondence (or lack thereof) between genomic tests and field surveys over the last several decades.

Materials and Methods

Sampling, genomic data generation and processing.—Genomic data were generated for 62 individual white-footed mice (*P. leucopus*) vouchered at the University of Michigan Museum of Zoology or collected by the Hoffman Laboratory at Miami University, following the ASM guidelines (Sikes et al. 2016). All individuals were sampled from four sites in northern Michigan: three sites in the UP (Menominee, Schoolcraft, and Chippewa Counties) and one in the LP (Cheboygan County; Fig. 1). These populations span the geographic distribution (west to east) of *P. leucopus* in the UP, and they should represent different periods of occupancy, with UP populations established after 1981 in the east (Myers et al. 2009) versus historic populations in the west that have been assumed to represent post-Pleistocene ancestral population sources for the expansion. Hereafter these populations are referred to as western historic (W_H; Menominee County), western expansion (W_E; Schoolcraft County), eastern historic (E_H; Cheboygan County), and eastern expansion (E_E; Chippewa County; Fig. 1, see Supplementary Data SD1 for details about specimens).

Genomic DNA was extracted from bone tissue using a DNeasy Blood and Tissue Kit (Qiagen, Valencia, California), with modifications for working with museum specimens (see Rubi et al. 2020 for full protocol). To minimize damage to the skulls, we sampled the microturbinates (small nasal bones; Wisely et al. 2004; Taylor and Hoffman 2010). Prior to DNA extraction, the bone fragments were placed into a thick-walled

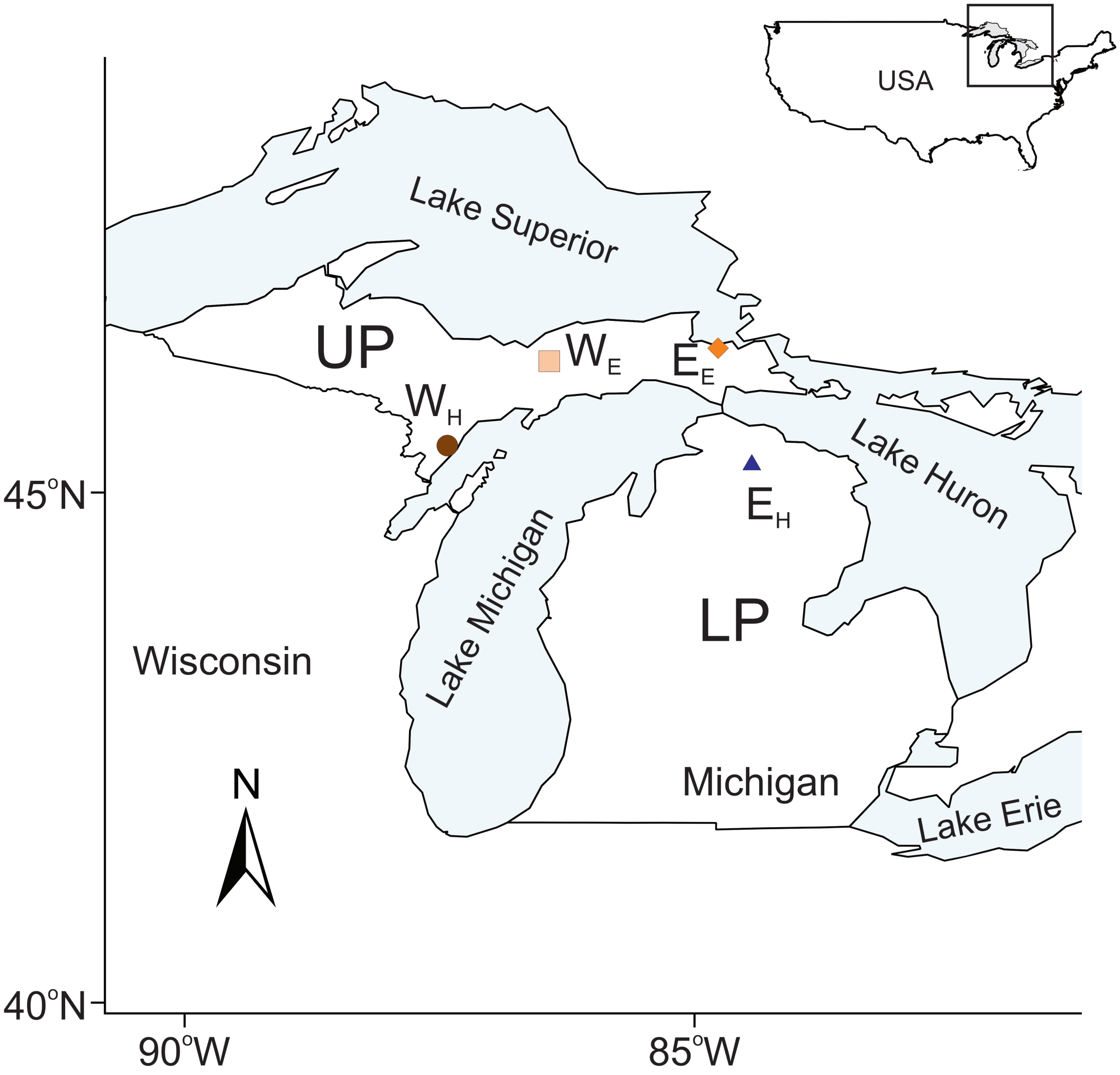

Fig. 1.—Map of the four studied populations sampled in either the LP (shown in blue) and UP populations (shown in shades of brown-orange) from the western or eastern part of UP. Differences in the periods of occupancy of the populations are denoted by the subscripts (i.e., expansion versus historic regions). Western historic (W_H, Menominee county), western expansion (W_E, Schoolcraft county), eastern historic (E_H, Cheboygan county), and eastern expansion (E_E, Chippewa county).

2 ml microcentrifuge tube with four 2.4 mm stainless steel beads and processed in a FastPrep tissue homogenizer (MP Biomedicals) for 1 min at 6.0 m/s. All pre-amplification steps were performed in a laboratory for processing low-quality specimens and followed stringent anti-contamination protocols.

One double-digest reduced representation library was constructed following the protocol for ddRADSeq from Peterson et al. (2012). Briefly, the DNA was double digested with two restriction enzymes (SphI-HF and MluCI), followed by a ligation step and amplification by PCR, where unique barcodes (custom Illumina adapters (Sigma-Aldrich)) were added to the digested DNA. PCR products were cleaned and 376 to 412 bp fragments were size-selected using a Pippin Prep (Sage Biosciences). The library was sequenced in one lane of an Illumina HiSeq 2500 (San Diego, California), producing 54 million 100 bp paired-end reads.

The pipeline STACKS version 2.41 (Catchen et al. 2013) was used to check for adapter contamination, demultiplex, and process the genomic sequences. One mismatch in the adapter sequence (–adapter_mm) and a barcode distance of one was used in *process radtags* (–barcode_dist). All reads were trimmed to a length of 96 bp, and reads with unambiguous barcodes and Phred score >10 retained. Individuals with less than 200,000 raw reads were excluded, resulting in a dataset with 32 individuals. The reads for each individual were aligned to the *Peromyscus leucopus* genome (NCBI assembly ID: GCA_004664715.1) using Bowtie2 (Langmead and Salzberg 2012) with default parameters (i.e., maximum and minimum mismatch penalties were 6 and 2, respectively). Because of limited overlap of the paired-end sequences, we restricted our analysis to the first reads only.

Consensus sequences for each assembled locus in the data were extracted using *gstacks*, with default options, and these files were read in *populations* to identify loci across populations (with –p 2). The resulting output was processed in R version 3.3.2 (R Core Team, 2017) to eliminate single-nucleotide polymorphisms (SNPs) from the 16 last base pairs in the 3′-end of each locus, as well as loci with exceedingly high genetic diversity, as such high values are suggestive of sequencing and assembly errors. One random SNP per locus was selected using *populations* for loci with a maximum of 20% missing data (selected using PLINK 1.07; Purcell et al. 2007). The final dataset contained 7,088 loci from 32 individuals (16 individuals from the E_H, 8 from the E_E, 5 from the W_H, and 3 from the W_E population). All STACKS modules were run under parallel execution on the University of Michigan ARC cluster, and specimens level library information can be found in Supplementary Data SD2.

Genetic diversity statistics.—Genetic diversity was summarized for the 7,088 loci (as opposed to a single SNP per locus) by nucleotide diversity (π), expected heterozygosity (*Exp_Het*), and Wright's inbreeding coefficient (*FIS*) at each locus. A paired *t*-test (conducted in R, BSDA package; Arnholt 2017) was used to test for significant differences in genetic diversity between all pair-wise comparisons of populations. A Bonferroni correction for multiple comparisons was applied using the *p.adjust* function in R version 3.3.2 (R Core Team, 2017).

Genetic structure.—A principal component analysis (PCA) on the genetic covariance matrix was performed to visualize the major axes of population genetic variation with the *prcomp* function in R version 3.3.2 (R Core Team, 2017). Specifically, we computed the PCA with the two historical populations, projecting the individuals from the two expanded populations onto the axes of the historic populations (following Lipson et al. 2018).

Potential genomic structure was estimated using STRUCTURE 2.3.4 (Pritchard et al. 2000), for a range of different genetic clusters (i.e., *K*-values ranging from 1 to 5). Ten independent runs were performed for each *K*-value, with 300,000 MCMC iterations and a burn-in of 100,000. STRUCTURE HARVESTER (Earl and vonHoldt 2012) was used to identify the *K*-value that best fit the data based on the ΔK (Evanno et al. 2005), and each individual genomic makeup (i.e., the proportional contribution of each of the different putative ancestors) was visualized using CLUMPAK (Kopelman et al. 2015).

F_{ST} values and significance were calculated with the R package dartR (function *gl.fst.pop*(); Gruber et al. 2018). An association between all pairwise F_{ST} values and the Euclidean geographic distances among populations (i.e., isolation-by-distance; IBD) was tested with a Mantel test using the R package VEGAN version 2.5-6 (Oksanen et al. 2013). We also applied a sequential population dropout procedure, in which the test was repeated excluding one population at time, in order to confirm that the results were robust.

Divergence time and expansion scenarios.—To test the hypothesis of a range expansion, we calculated the directionality index ψ (Peter and Slatkin 2013). This statistic detects the allele frequency clines created by successive founder events, where the further away a population is from the origin of the range expansion, the higher the probability that a SNP increases in allele frequency or becomes fixed. We estimated the statistic between all pairwise populations pairs using the scripts available in the *X-Origin* pipeline (He et al. 2017) to calculate the pairwise ψ values.

For a formal statistic evaluation of the demographic past, including the timing of divergence, we parameterized several different models of the potential colonization history. These models differed with respect to the ancestry of the different historical and expansion populations. Specifically, we tested two alternative models (which were outlined in the Introduction): namely, (i) a model of two colonizations of the UP populations (i.e., the W_E and E_E) that posits a common ancestry of the western UP populations (i.e., W_H and W_E), which is distinct from the historical population from the LP and the eastern population from UP (i.e., E_H and E_E) (see Rowe et al. 2006; Moscarella et al. 2019), and (ii) a model of a single colonization of the UP populations (i.e., the W_E and E_E) from the western part of the species range (represented by W_H; see Myers et al. 2009; Baumgartner and Hoffman 2019). These two models were tested with and without the migration parameter. For each model, parameters were estimated using a composite-likelihood simulation-based approach as implemented in FASTSIMCOAL2 (Excoffier and Foll 2011;

Excoffier et al. 2013) using the site frequency spectrum (SFS). To improve performance, one population parameter (the effective population size, N_1) was fixed and calculated directly from the empirical data (based on nucleotide diversity (π) of variant and invariant sites; see Excoffier et al. 2013). Other parameters such as *N2* and *N3*, ancestral population sizes N_{ANC}, and divergence times T_{DIV}, were estimated from the SFS using uniform priors, with a genomic mutation rate, μ, of 3.67×10^{-8} per site per generation, and a 2.5-year generation, based on information from the close taxon *P. maniculatus* (Pfeifer et al. 2018). A total of 40 FASTSIMCOAL2 runs were conducted per model, with 100,000 simulations per likelihood estimation based upon a stopping criterion of 0.001, and 40 expectation-conditional maximization (ECM) cycles. The Akaike information criterion (AIC) was used to select the best-fit model, based on the single run with the highest composite likelihood. A parametric bootstrap was used to estimate 95% confidence intervals for the parameter estimates by estimating the parameters from 100 simulated datasets under the parameter-values estimated for the empirical data (Excoffier et al. 2013). Note that separate datasets were generated for each of the two models using the *populations* program (STACKS; Catchen et al. 2013) to maximize the amount of data; calculation of the SFS requires no missing data (Excoffier and Foll 2011). We used a Python script to remove all missing data and calculate the joint SFS between each population pair (available on Github/KnowlesLab; Knowles et al. 2016). Five individuals from each population were used to calculate the SFS; however, this required pooling of individuals from the western populations (W_H and W_E, 5 and 3 individuals, respectively), and to avoid an overparameterization model (i.e., a model with four population lineages) given the limited amount of data (i.e., a simpler model was used, which assumes a single ancestry for individuals from W_H and W_E,; see results from STRUCTURE analyses to justify the simplification of pooling keep the W_H and W_E, populations).

RESULTS

Genetic diversity did not differ substantially among the UP populations (W_E, W_H, and E_E), irrespective of the diversity statistic considered; the genetic diversities of those populations were lower than that of the LP population (E_H), although the difference was not statistically significant (Fig. 2; Supplementary Data SD3). Note that the lower genetic diversity of the W_H and W_E populations when compared with the eastern populations is not an artifact of the smaller number of individuals sampled (5 and 3 individuals, respectively) given that the same results were obtained when subsampling three individuals for all populations (see Supplementary Data SD4).

Projection of the populations from the UP (W_E, W_H, and E_E) upon the primary axes of variation (i.e., PC1 and PC2) characterized by a principal component analysis of the genetic variation of the eastern historic population in the LP (E_H) shows the genetic separation of the UP populations (W_E, W_H, and E_E) from the LP population (E_H) along PC1 (Fig. 3). There are different degrees of distinctiveness among the UP populations, with some overlapping more than others (e.g., W_E; Fig. 3). The genetic makeup of the eastern expansion (E_E) is more diverged from the individuals from the LP population (E_H), compared with the western historic (W_H) individuals, which show a stronger affinity with the LP population (E_H) along the axis of genetic variation defined by PC1, despite being the most geographically distant from the LP population (E_H).

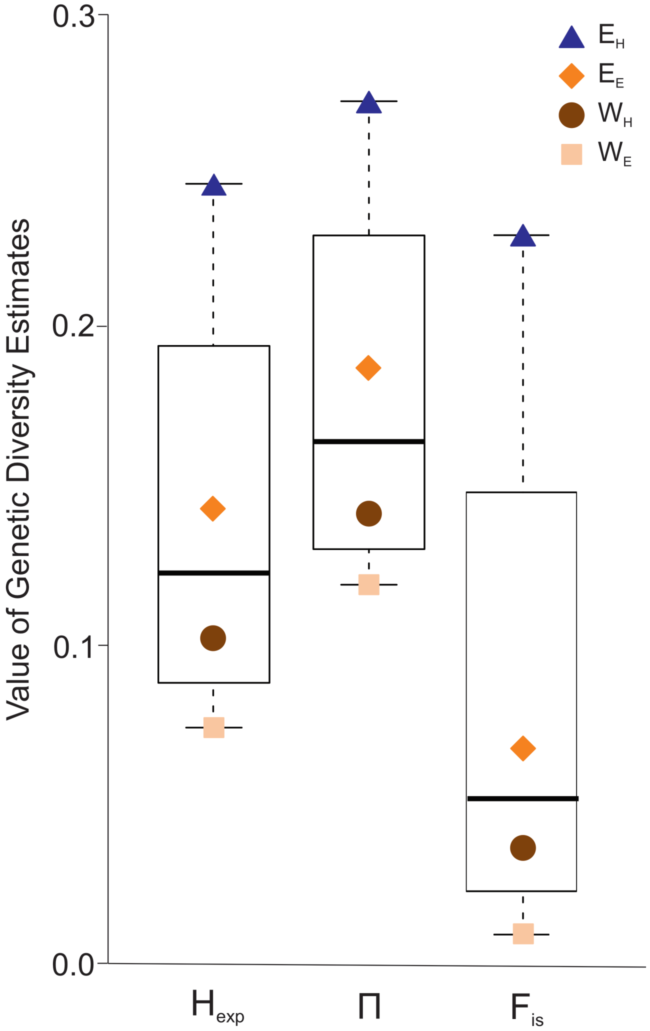

Fig. 2.—Comparison of genetic diversity of the four different populations shown as box plots (calculated with *boxplot* function in R) for each summary statistic, the LP population (E_H) shown in blue and the three UP populations (W_H, W_E, and E_E) in the brown-orange tones.

Genetic differentiation (F_{ST}) was low among all populations, as expected given their recent shared ancestry (Supplementary Data SD5). However, the degree of differentiation was not correlated with geographic distance between populations (Supplementary Data SD6, Mantel test $r = 0.61$, $P = 0.12$). For example, the lowest level of F_{ST} (0.002) was between the geographically distant western expansion (W_E) and eastern historic (E_H) populations (Supplementary Data SD5). This lack of a correspondence between pairwise F_{ST} values and the geographic distances among populations did not change when one population was excluded from the analysis (i.e., the lack of a correlation between the degree of genetic differentiation

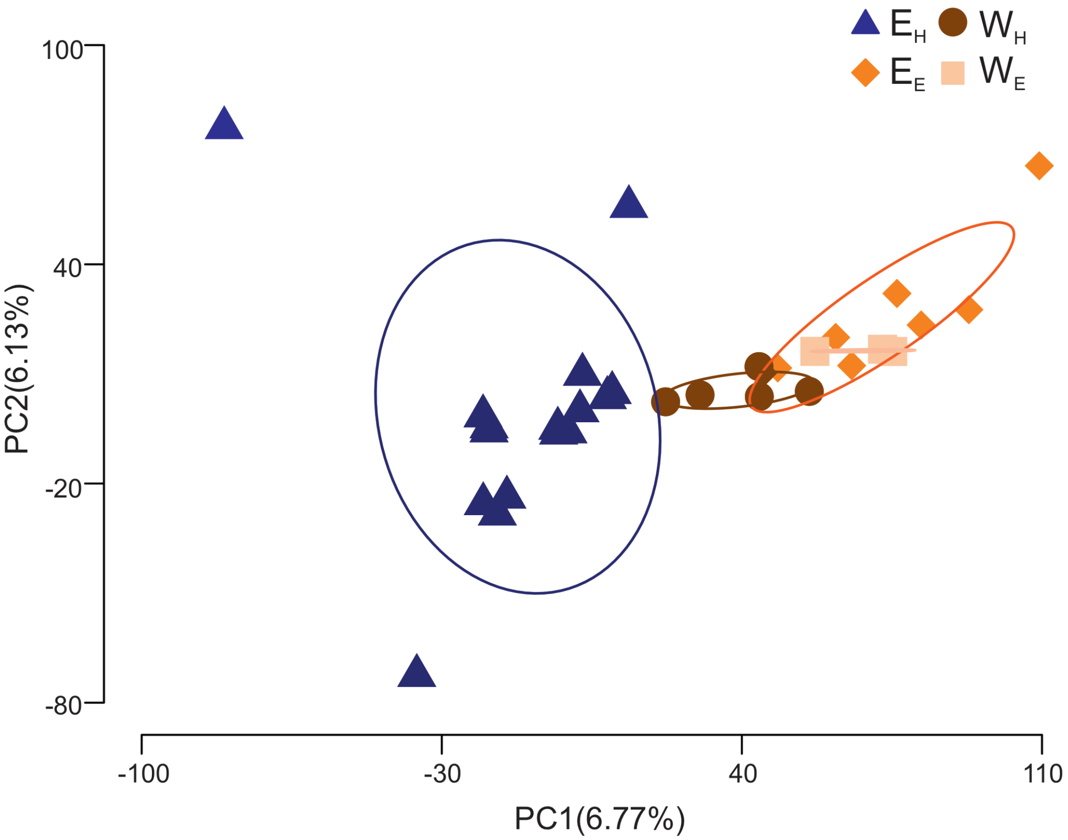

Fig. 3.—Distribution of genetic variation based on a principle components analysis of the individuals from the historical populations (E_H and W_H), with individuals from the expansion populations (E_E and W_E) projected onto the PC axes. The E_H population was sampled in the LP, and the W_H, W_E, and E_E populations were collected in the UP.

and geographic distance was not due to one outlier population; Supplementary Data SD6). We also did not find a conspicuous positive relationship between the genetic distance of individuals (rather than populations) as a function of geographic distance.

The STRUCTURE analyses identified $K = 3$ as the most likely number of genetic clusters. However, the individuals all have a large component of their genetic makeup that reflects a shared ancestral source that is represented in similar proportions across all individuals (Fig. 4). Likewise, all populations showed a similar amount of admixture, albeit with the two western populations sharing a proportion of their ancestry from an ancestral population that is distinct from the eastern populations (for $K = 3$).

The directionality index showed a weak signal of a range expansion scenario that was marginally nonsignificant ($r^2 = 0.50$, $P = 0.06$). The strongest differences in allele frequencies were observed in pairwise comparisons between populations from eastern (E_H and E_E) versus western (W_H and W_E) regions (e.g., $\psi > 0.02$), whereas within each of the regions (e.g., between the E_H and E_E populations), pairwise comparisons were close to 0.01 (Supplementary Data SD7).

The best model to describe the divergence history was the model with two independent colonizations of the UP, that is, populations from the east (E_E) and west (W_E and W_H) in the UP had distinct shared ancestors (Fig. 5, Tables 1 and 2) based on the parameterization of the models using the SFS in FASTSIMCOAL2. Divergence time estimates suggest that within the eastern region, the E_H population in the LP diverged from the E_E population in the UP around 4,800 years ago, which was more recent than the estimated divergence of the eastern and western regions from their last common ancestor at around 91,000 years ago (Fig. 5); note that we did not estimate a time of divergence between the western populations, given the limited sample sizes of the W_E population (see Methods for details). In all cases, estimates of the ancestral population sizes were larger than estimates for the current populations (Table 1). Also note that although a model with migration fit the data better than one without migration, estimated migration rates were very low (e.g., ≤0.0003).

Discussion

As with other taxa in northern latitudes (Rowe et al. 2006; Garcia-Elfring et al. 2017; Klütsch et al. 2017; Bi et al. 2019), climatic changes structure patterns of genetic variation in *P. leucopus*. Analyses of genomic data show that such effects are not limited to a single event, but rather are consistent with divergences that trace to the retreat of ice during glacial cycles in the more distant past, as well as to the last glacial maximum (Fig. 5). These findings support some of the previous studies based on analyses of mtDNA or microsatellites (Baumgartner and Hoffman 2019; Moscarella et al. 2019). However, genomic

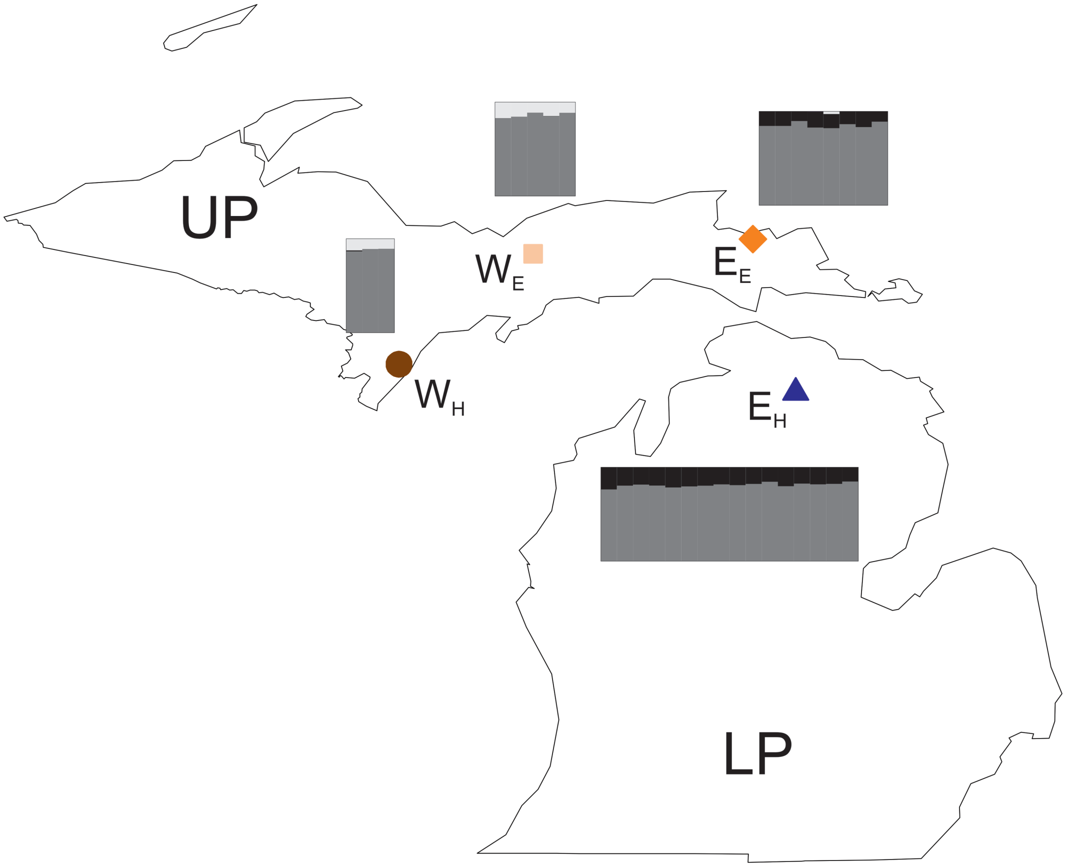

Fig. 4.—Results from STRUCTURE analyses depicting the genetic structure with $k = 3$ groups, and the ancestry of the three groups represented by three different colors. Individual admixture proportions show reduced differentiation among populations.

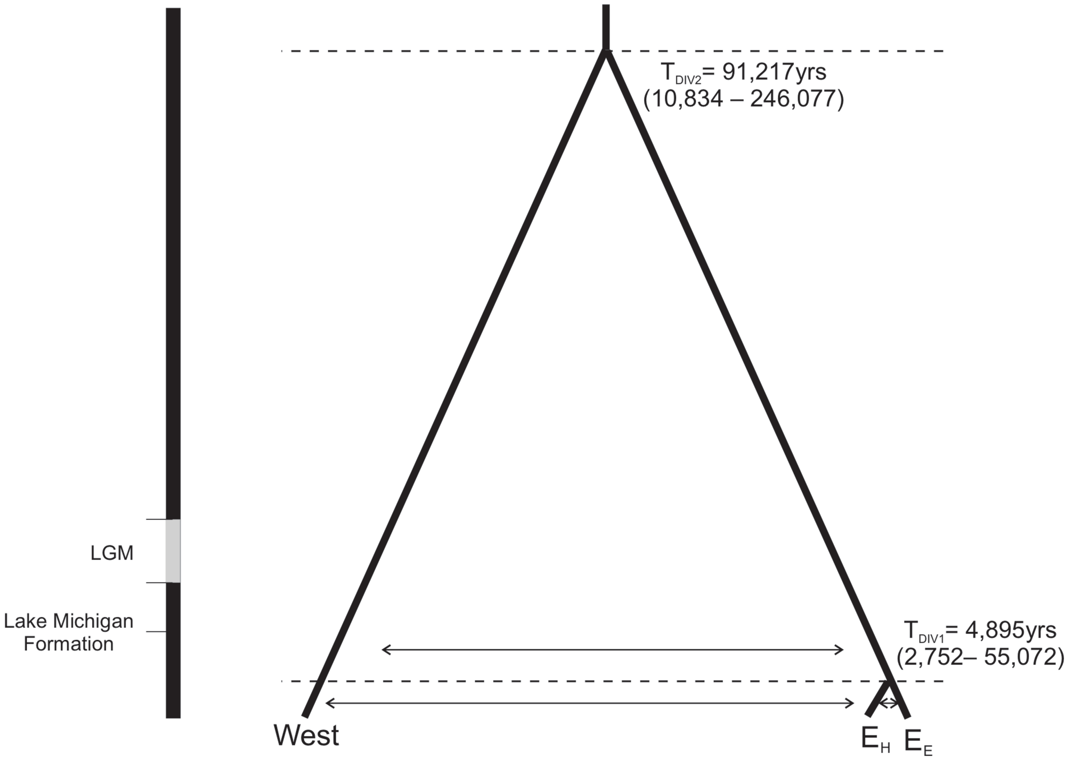

Fig. 5.—The best model of divergence estimated using FASTSIMCOAL2, in which expansion populations in the east shared a more common ancestor with the E_H population (i.e., the LP population), compared with a more distant ancestry with western populations (as opposed to a single colonization route from the W_H population). Arrows indicate the presence of migration among populations.

Table 1.—ML estimates of model parameters for the best fit model (see Fig. 5) estimated with migration across 40 runs in FASTSIMCOAL2. Specifically, the model that posits two colonization of the UP from an eastern and western source that shared a common ancestry over 91,000 years (T_{DIV2}), compared with the more recent divergence, T_{DIV1}, between the historical population from the LP (N_1) in the east with the E_{E} population (N_2); also shown are 95% confidence intervals in parentheses. Note that we did not estimate a time of divergence of the western populations (N_1) given the limited sample sizes of the W_{E} population (see methods for details). Divergence times are shown in years, assuming, 2.5 years per generation.

Loci	N_1 (fixed)	T_{DIV1}	T_{DIV2}	N_{ANc2}	N_{ANC1}	N_2	N_3
1180	48,501	4,895 (2,752–55,072)	91,217 (10,834–246,077)	260,4676	123,981	1,976 (1,104–23,402)	36,227 (19,367–92,129)

Table 2.—Composite likelihood [max ln(L)], Akaike information criterion (AIC), and ΔAIC for each of the tested demographic models for *Peromyscus leucopus*; the best model is shown in bold (see Table 1 for parameter estimates). The histories were chosen based on the demographic results previously showed.

Model	Maxln(L)	n.param	AIC	ΔAIC
Two colonization sources	−3,438.554	6	6,889.108	673.954
Two colonization sources with migration	**−3,095.577**	**12**	**6,215.154**	**0**
Single colonization source	−3,497.217	6	7,006.434	791.28
Single colonization source with migration	−3,143.563	12	6,311.126	95.972

analyses suggest a more nuanced history for the white-footed mouse than previously recognized. Specifically, with the resolution provided by thousands of genomic loci and a formal statistical evaluation of alternative hypotheses, the proposal of a single ancestral source for UP populations derived from the west, as was suggested by previous studies based on mtDNA (Rowe et al. 2006; Moscarella et al. 2019), is not supported; nor do the genomic data indicate a lack of any population differentiation among UP populations, as inferred from analyses of microsatellites (Baumgartner and Hoffman 2019). Instead, the genomic data support a postglacial colonization of the UP from two different ancestral sources. This includes a shared ancestry of the LP population (E_{H}) and the eastern populations that expanded into the UP (E_{E}), as well as a shared ancestry between the western populations (i.e., W_{H} and W_{E}). The genomic makeup of individuals also highlights their shared common ancestry (Fig. 4), as well as a distinct ancestry of western versus eastern populations.

The divergence separating the eastern population in the LP (E_{H}) and western populations in the UP (W_{H} and W_{E}) is older than the divergence between the LP (E_{H}) and E_{E} populations; however, it is still recent (occurring within the last 100,000 years; Fig. 5). These recent divergence estimates are consistent with the large amounts of shared ancestry among all the populations (Fig. 4); if such divergences occurred in the distant past, it is unlikely that the ancestral genetic variation would still be shared among individuals from different populations (Edwards and Beerli 2000). These divergence times are also considerably older than the expansion history documented by historical museum collections and faunal surveys (Myers et al. 2009). As such, the genomic data do not support the hypothesis that the UP populations were founded recently by *P. leucopus*, in contrast with the results of recent trapping efforts. Likewise, although the genetic diversity among the UP populations (W_{E}, W_{H}, and E_{E}) was lower than the diversity within the LP population (E_{H}; see Fig. 2), evidence for an extreme bottleneck associated with the founding of the UP populations is lacking. Nevertheless, the genetic makeup of the LP population (E_{H}) is more similar to the geographically more distant W_{H} population than to the E_{E} and W_{E} expansion populations (Fig. 3). This suggests that there may have been a shift in the genetic variants (i.e., drift-induced genetic changes) associated with the expanded populations. However, this shift is not sufficient to erase a pattern of common ancestry of each expanded population with its respective historical population, since a small fraction of their genetic makeup reflects the contribution of a common regional ancestral source (Fig. 4).

It may be tempting to interpret the patterns of similar genetic makeup among populations in the UP and LP as reflecting large amounts of gene flow. As an ecological generalist, *P. leucopus* appears to have a considerable dispersal capacity relative to its body size. For example, the distance of natal dispersal in *P. leucopus* is known to vary from a few hundred meters to almost one kilometer (Krohne et al. 1984; Gaitan and Millien 2016), and the shift in distribution to the UP represents a range expansion on the order of 250 km. Moreover, unlike the expanding distribution of *P. leucopus*, field surveys suggest that the distribution of its close relative *P. maniculatus gracilis* has remained historically stable (at least based on faunal surveys) since its postglacial colonization of the UP (Myers et al. 2009). However, there are several aspects of the analysis that make it more likely that the shared genetic makeup of individuals reflects the sorting of ancestral variation from common ancestors, rather than dispersal per se. First, there is no support for isolation by distance (Supplementary Data SD6). Second, the distribution of individuals from the UP expansion populations ($\mathrm{W_E}$ and $\mathrm{E_E}$) does not overlap with the genomic space of the historical LP population (E_{H}), as defined by the PCA of genetic variation (Fig. 3). This does not mean the populations are entirely isolated. For example, the best fit model to the data includes migration among the populations (Table 2), and the greater genetic diversity of the E_E relative to the W_H and W_E populations (Fig. 2) may reflect the contribution of genetic variants from these populations, in addition to variation from E_H. Moreover, aside from the shared ancestral variation common to all individuals (Fig. 4), there are portions of the genetic makeup of eastern (E_E) versus western (W_E) expansion populations that are shared with the W_H and E_H populations, respectively, but are nonetheless distinct between the two expansion populations (i.e., $K = 3$, Fig. 4). Such a genetic signature is inconsistent with a hypothesis of widespread and high levels of gene flow, which would have homogenized genetic variation across individuals, erasing the genetic distinctions among populations (Fig. 3).

An open question regarding the expanding range of *P. leucopus* in the UP of Michigan is whether it reflects environmental tracking versus rapid adaptive change. The same populations studied here genetically are morphologically variable (based on analyses of the skull and mandible); however, that morphological variation is not geographically structured (Baumgartner and Hoffman 2019). Instead, there is considerable overlap in morphological variation between populations, suggesting a plastic response to local environmental conditions. In addition, analysis of epigenetic variation in the same populations studied here is not consistent with a compensatory hypothesis in which there are genetic constraints imposed on expanding populations that might be alleviated by epigenetic variation (unpublished data), which suggests that the genetic variants available in the genome of the ancestral population may be sufficient to accommodate any differences in the environment in the UP. Together, the tangential evidence from analyses of morphology and epigenetics suggests that the expansion of *P. leucopus* reflects environmental tracking. Formal testing of this hypothesis in *P. leucopus* will require additional sampling of populations across both the ancestral and expanded distribution in order to identify direct evidence of environmental tracking (see Massatti and Knowles 2020).

Finally, the divergence time of the ancestors giving rise to the western populations (Wisconsin) and eastern (LP of Michigan) populations (≅91,000 years ago) is older than previous estimates for the split between these western and eastern lineages based on mitochondrial data (≅34,000 years ago; Moscarella et al. 2019). The difference in estimated dates likely reflects the fact that divergence times were estimated from the genomic data, but not from the mtDNA, using a model to accommodate the possibility of migration. Irrespective of the different estimated divergence times, both mtDNA and genomic data sets indicate that there was a historical substructure in *P. leucopus* that predates the formation of Lake Michigan about 13,000 years ago (Holman 2001).

In contrast, the genetic differentiation detected here between the eastern LP (E_H) and UP (E_E) populations that dates to 4,800 years ago is consistent with the retreat of glacial ice and the closing of the land connection between the LP and UP, due to the formation of the Straits of Mackinac, which connect Lake Huron and Lake Michigan (Holman 2001). The current physical separation of the UP and LP populations by Lake Michigan and Lake Huron appears to be playing an ongoing role in the genetic differentiation between the UP and LP populations, since genetic variation between these populations has not been homogenized by gene flow (Fig. 3). The possibility of human-mediated dispersal suggested by Moscarella et al. (2019) is contradicted by the genetic dissimilarity between geographically proximate populations in the northern LP and the eastern UP (Fig. 3).

Several scenarios may reconcile the historical museum records and field surveys of expansion (and/or human mediated dispersal) with what appears to be a longer history of occupancy of the eastern UP by *P. leucopus*, as suggested by the genomic data. Survey data admittedly only provide a minimum estimate for expansion, so it is theoretically possible, even though extremely unlikely given a history of intensive surveying, that long-term populations of *P. leucopus* in the UP simply went undetected. Furthermore, it is also possible that the expansion into the UP was not homogeneous, such that the areas previously surveyed (Myers et al. 2009) did lack *P. leucopus* until very recently, and these areas were colonized by other populations in the UP that had been more stable. Alternatively, the difference in the time of colonization might not reflect the limitations of the survey data. For example, if populations within the LP have been historically differentiated, and if the eastern UP population studied here (E_E) was founded from a different LP ancestral population than the one sampled here (i.e., E_H), the older divergence between the LP and eastern UP population E_E (Fig. 4) could reflect this ancestral substructure. In this scenario, it is possible that the E_E may have indeed been colonized recently, as suggested by the faunal surveys. However, without additional evidence of substructure in the LP, which would require the sampling of additional populations (and is beyond the scope of this study), this hypothesis cannot be tested.

Irrespective of which of these alternative explanations is correct, this study is another example of how genetic studies raise alternative hypotheses that are beyond the scope of contemporary data based on field surveys (see Biek et al. 2006). Given that there is geological evidence to support routes of colonization from the LP to the UP, and the support of the genomic data for hypothesized historical postglacial expansion, the most pertinent question may be: why do not the genomic data correspond to what faunal surveys suggest about the expansion history of *P. leucopus?* Our findings are consistent with the many examples of shifts in distribution coinciding with climatic shifts that are associated with the last glacial maximum (Carstens and Knowles 2007; Shafer et al. 2010; Fasanella et al. 2013; Ortego et al. 2015; González-Wevar et al. 2016). However, there are certainly cases where the shifts in distribution of species documented in field surveys are supported by genetic analyses (e.g., the Grinnell surveys of elevational shifts in California mammal communities; see Moritz et al. 2008). Given the association of *P. leucopus* with specific forest habitats (which includes data from other parts of its range not studied here; Baumgartner 2017) and the regional environment of the UP, a hypothesis of historical postglacial expansion into the UP was not simply overlooked. Such a hypothesis is inconsistent with ecological information about this species. At this point, given that we do not have evidence of a general correspondence between historical faunal surveys in *P. leucopus* and genomic estimates of divergence, there is a clear conundrum that remains to be resolved. Only with future investigations that include more populations can we test whether the divergence between the LP and the UP (and in particular, E_E) estimated here represents a different geographic configuration.

In conclusion, the expansion of *P. leucopus* into the UP has indeed occurred recently. However, genomic data identifies genetic divergences between the LP and UP that coincide with the retreat of ice during the last glacial maximum and the formation of the Great Lakes, rather than with recent climatic warming, as was suggested from previous studies based upon historical museum records and field surveys. Genomic data from additional populations will be needed to evaluate whether these differences between ecological and genomic data might be reconciled by a different model of colonization: specifically, could the divergence observed between the LP (E_H) and eastern UP (E_E) actually predate the colonization of the eastern UP? Within the UP itself, we find evidence of separate ancestral sources (i.e., eastern and western sources) that trace their common ancestry to a population that predates the formation of the Great Lakes, corroborating evidence based on mtDNA (Moscarella et al. 2019). These results highlight the need for integration of genomic data with field surveys, in order to test hypotheses about ecological range expansions due to recent contemporary climate change versus alternative hypotheses such as historical postglacial range expansion.

Acknowledgements

We are grateful to museum curators and staff who provided access to their collections. Computational resources and services were provided by Advanced Research Computing at the University of Michigan. This study was funded and supported by a National Science Foundation Postdoctoral Research Fellowship in Biology (Award No.: 1612143), as well as the Dantzer and Knowles Labs at the University of Michigan and the Hoffman Lab at Miami University.

Conflict of Interest

None declared.

Supplementary Data

Supplementary Data SD1.—Information about the specimens sequenced.

Supplementary Data SD2.—Summaries of genomic data.

Supplementary Data SD3.—Summaries of genetic diversity.

Supplementary Data SD4.—Summaries of genetic diversity with standardized sample size.

Supplementary Data SD5.—Pairwise F_{ST}-values.

Supplementary Data SD6.—Isolation by distance plot.

Supplementary Data SD7.—Directionality index results.

Literature Cited

Arnholt A.T. 2017. BSDA: basic statistics and data analysis. R package version 1.2.0. http://CRAN.R-project.org/package=BSDA.

Baumgartner J.M. 2017. The genetic and morphometric responses of *Peromyscus leucopus* populations to the changing environment of the Great Lakes region, Dissertation, Miami University.

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

Bemmels J.B., Knowles L.L., Dick C.W. 2019. Genomic evidence of survival near ice sheet margins for some, but not all, North American trees. Proceedings of the National Academy of Sciences of the United States of America 116:8431–8436.

Bi K., Linderoth T., Singhal S., Vanderpool D., Patton J.L., Nielsen R., Moritz C., Good J.M. 2019. Temporal genomic contrasts reveal rapid evolutionary responses in an alpine mammal during recent climate change. PLoS Genetics 15:e1008119.

Biek R., Drummond A.J., Poss M. 2006. A virus reveals population structure and recent demographic history of its carnivore host. Science (New York, N.Y.) 311:538–541.

Carstens B.C., Knowles L.L. 2007. Shifting distributions and speciation: species divergence during rapid climate change. Molecular Ecology 16:619–627.

Catchen J., Hohenlohe P.A., Bassham S., Amores A., Cresko W.A. 2013. Stacks: an analysis tool set for population genomics. Molecular Ecology 22:3124–3140.

Cook V., Barbour A.G. 2015. Broad diversity of host responses of the white-footed mouse *Peromyscus leucopus* to Borrelia infection and antigens. Ticks and Tick-Borne Diseases 6:549–558.

Earl D.A., vonHoldt B.M. 2012. STRUCTURE HARVESTER: a website and program for visualizing STRUCTURE output and implementing the Evanno method. Conservation Genetics Resources 4:359–361.

Edwards S.V., Beerli P. 2000. Perspective: gene divergence, population divergence, and the variance in coalescence time in phylogeographic studies. Evolution; International Journal of Organic Evolution 54:1839–1854.

Evanno G., Regnaut S., Goudet J. 2005. Detecting the number of clusters of individuals using the software STRUCTURE: a simulation study. Molecular Ecology 14:2611–2620.

Excoffier L., Dupanloup I., Huerta-Sánchez E., Sousa V.C., Foll M. 2013. Robust demographic inference from genomic and SNP data. PLoS Genetics 9:e1003905.

Excoffier L., Foll M. 2011. fastsimcoal: a continuous-time coalescent simulator of genomic diversity under arbitrarily complex evolutionary scenarios. Bioinformatics (Oxford, England) 27:1332–1334.

Fasanella M., Bruno C., Cardoso Y., Lizarralde M. 2013. Historical demography and spatial genetic structure of the subterranean rodent *Ctenomys magellanicus* in Tierra del Fuego (Argentina). Zoological Journal of the Linnean Society 169:697–710.

Gaitan J., Millien V. 2016. Stress level, parasite load, and movement pattern in a small-mammal reservoir host for lyme disease. Canadian Journal of Zoology 94:565–573.

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 contact with the deer mouse (*Peromyscus maniculatus*). Heredity 119:447–458.

González-Wevar C.A., Rosenfeld S., Segovia N.I., Hüne M., Gérard K., Ojeda J., Mansilla A., Brickle P., Díaz A., Poulin E.

2016. Genetics, gene flow, and glaciation: the case of the South American Limpet *Nacella mytilina*. PLoS One 11:e0161963.

Gruber B., Unmack P.J., Berry O.F., Georges A. 2018. dartr: an r package to facilitate analysis of SNP data generated from reduced representation genome sequencing. Molecular Ecology Resources 18:691–699.

Handler S., et al. 2014. Michigan forest ecosystem vulnerability assessment and synthesis: a report from the Northwoods Climate Change Response Framework project. The United States Department of Agriculture, Forest Service, Newtown Square, Pennsylvania.

He Q., Edwards D.L., Knowles L.L. 2013. Integrative testing of how environments from the past to the present shape genetic structure across landscapes. Evolution; International Journal of Organic Evolution 67:3386–3402.

He Q., Prado J.R., Knowles L.L. 2017. Inferring the geographic origin of a range expansion: Latitudinal and longitudinal coordinates inferred from genomic data in an ABC framework with the program x-origin. Molecular Ecology 26:6908–6920.

Holman A.J. 2001. In quest of Great Lakes Ice Age vertebrates. Michigan State University Press, East Lansing, Michigan, USA.

Jannett F.J.J., Broschart M.R., Grim L.H., Schaberl J.P. 2007. Northerly range extensions of mammalian species in Minnesota. The American Midland Naturalist 158:168–176.

Klütsch C.F.C., Manseau M., Anderson M., Sinkins P., Wilson P.J. 2017. Evolutionary reconstruction supports the presence of a Pleistocene Arctic refugium for a large mammal species. Journal of Biogeography 44:2729–2739.

Knowles L.L., Massatti R. 2017. Distributional shifts – not geographic isolation – as a probable driver of montane species divergence. Ecography 40:1475–1485.

Knowles L.L., Massatti R., He Q., Olson L.E., Lanier H.C. 2016. Quantifying the similarity between genes and geography across Alaska's alpine small mammals. Journal of Biogeography 43:1464–1476.

Kopelman N.M., Mayzel J., Jakobsson M., Rosenberg N.A., Mayrose I. 2015. Clumpak: a program for identifying clustering modes and packaging population structure inferences across K. Molecular Ecology Resources 15:1179–1191.

Krohne D.T., Dubbs B.A., Baccus R. 1984. An analysis of dispersal in an unmanipulated population of *Peromyscus leucopus*. The American Midland Naturalist 112:146–156.

Kurta A. 2017. Mammals of the Great Lakes region. 3rd ed. University of Michigan Press, Ann Arbor, Michigan, USA.

Lackey J.A., Huckaby D.G., Ormiston B.G. 1985. *Peromyscus leucopus*. Mammalian Species 247:1–10.

Langmead B., Salzberg S.L. 2012. Fast gapped-read alignment with Bowtie 2. Nature Methods 9:357–359.

Lanier H.C., Massatti R., He Q., Olson L.E., Knowles L.L. 2015. Colonization from divergent ancestors: glaciation signatures on contemporary patterns of genomic variation in Collared Pikas (*Ochotona collaris*). Molecular Ecology 24:3688–3705.

Lipson M., et al. 2018. Ancient genomes document multiple waves of migration in Southeast Asian prehistory. Science 361:92–95.

Massatti R., Knowles L.L. 2020. The historical context of contemporary climatic adaptation: a case study in the climatically dynamic and environmentally complex southwestern United States. Ecography 43:735–746.

Moritz C., Patton J.L., Conroy C.J., Parra J.L., White G.C., Beissinger S.R. 2008. Impact of a century of climate change on small-mammal communities in Yosemite National Park, USA. Science (New York, NY) 322:261–264.

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 100:345–353.

Myers P., Lundrigan B.L., Hoffman S.M.G., Haraminac A.P., Seto S.H. 2009. Climate-induced changes in the small mammal communities of the Northern Great Lakes Region. Global Change Biology 15:1434–1454.

Oksanen A.J., et al. 2013. Package Vegan. Community ecology package. R package version 2.0-10. https://CRAN.R-project.org/package=vegan.

Ortego J., Noguerales V., Gugger P.F., Sork V.L. 2015. Evolutionary and demographic history of the Californian scrub white oak species complex: an integrative approach. Molecular Ecology 24:6188–6208.

Osgood W.H. 1909. Revision of mice of the American genus *Peromyscus*. North American Fauna 28:1–285.

Pardiñas U.F.J., et al. 2017. Family Cricetidae (true hamsters, voles, lemmings and new world rats and mice). In: Wilson D.E., Lacher T.E., Mittermeier R.A., editors. Handbook of the Mammals of the World. Vol. 7, Rodents 2, Lynx Editions, Barcelona, Spain; p. 204–279.

Peter B.M., Slatkin M. 2013. Detecting range expansions from genetic data. Evolution; International Journal of Organic Evolution 67:3274–3289.

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

Pfeifer S.P., Laurent S., Sousa V.C., Linnen C.R., Foll M., Excoffier L., Hoekstra H.E., Jensen J.D. 2018. The evolutionary history of Nebraska deer mice: local adaptation in the face of strong gene flow. Molecular Biology and Evolution 35:792–806.

Pritchard J.K., Stephens M., Donnelly P. 2000. Inference of population structure using multilocus genotype data. Genetics 155:945–959.

Purcell S., Neale B., Todd-Brown K., Thomas L., Ferreira M.A., Bender D., Maller J., Sklar P., de Bakker P.I., Daly M.J., et al., 2007. PLINK: a tool set for whole-genome association and population-based linkage analyses. American Journal of Human Genetics 81:559–575.

R Development Core Team. 2017. R: a language and environment for statistical computing. R Foundation for Statistical Computing, Vienna, Austria. www.R-project.org/.

Rowe K.C., Heske E.J., Brown P.W., Paige K.N. 2004. Surviving the ice: Northern refugia and postglacial colonization. Proceedings of the National Academy of Sciences of the United States of America 101:10355–10359.

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 15:4003–4020.

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: e80724.

Rubi T.L., Knowles L.L., Dantzer B. 2020. Museum epigenomics: characterizing cytosine methylation in historic museum specimens. Molecular Ecology Resources 20:1161–1170.

Shafer A.B., Cullingham C.I., Côté 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 19:4589–4621.

Sikes R.S.; Animal Care and Use Committee of the American Society of Mammalogists. 2016. 2016 Guidelines of the American Society of Mammalogists for the use of wild mammals in research and education. Journal of Mammalogy 97:663–688.

Taylor Z.S., Hoffman S.M.G. 2010. Mitochondrial DNA genetic structure transcends natural boundaries in great lakes populations of woodland deer mice (*Peromyscus maniculatus gracilis*). Canadian Journal of Zoology 88:404–415.

Vendrami D.L.J., De Noia M., Telesca L., Handal W., Charrier G., Boudry P., Eberhart-Phillips L., Hoffman J.I. 2019. RAD sequencing sheds new light on the genetic structure and local adaptation of European scallops and resolves their demographic histories. Scientific Reports 9:1–13.

Wisely S.M., Maldonado J.E., Fleischer R.C. 2004. A technique for sampling ancient DNA that minimizes damage to museum specimens. Conservation Genetics 5:105–107.

Submitted 11 January 2021. Accepted 30 November 2021.

Associate Editor was Amy Baird.