doi:10.1111/evo.12967

Genetic surfing, not allopatric divergence, explains spatial sorting of mitochondrial haplotypes in venomous coralsnakes

Jeffrey W. Streicher,[1,2,3] Jay P. McEntee,[2,4] Laura C. Drzich,[3] Daren C. Card,[3] Drew R. Schield,[3] Utpal Smart,[3] Christopher L. Parkinson,[5] Tereza Jezkova,[2] Eric N. Smith,[3] and Todd A. Castoe[3,6]

[1]*Department of Life Sciences, The Natural History Museum, London, United Kingdom*

[2]*Department of Ecology and Evolutionary Biology, University of Arizona, Tucson, Arizona*

[3]*Department of Biology, University of Texas at Arlington, Arlington, Texas*

[4]*Department of Biology, University of Florida, Gainesville, Florida*

[5]*Department of Biology, University of Central Florida, Orlando, Florida*

[6]*E-mail: todd.castoe@uta.edu*

Received October 6, 2015
Accepted May 16, 2016

Strong spatial sorting of genetic variation in contiguous populations is often explained by local adaptation or secondary contact following allopatric divergence. A third explanation, spatial sorting by stochastic effects of range expansion, has been considered less often though theoretical models suggest it should be widespread, if ephemeral. In a study designed to delimit species within a clade of venomous coralsnakes, we identified an unusual pattern within the Texas coral snake (*Micrurus tener*): strong spatial sorting of divergent mitochondrial (mtDNA) lineages over a portion of its range, but weak sorting of these lineages elsewhere. We tested three alternative hypotheses to explain this pattern—local adaptation, secondary contact following allopatric divergence, and range expansion. Collectively, near panmixia of nuclear DNA, the signal of range expansion associated sampling drift, expansion origins in the Gulf Coast of Mexico, and species distribution modeling suggest that the spatial sorting of divergent mtDNA lineages within *M. tener* has resulted from genetic surfing of standing mtDNA variation—not local adaptation or allopatric divergence. Our findings highlight the potential for the stochastic effects of recent range expansion to mislead estimations of population divergence made from mtDNA, which may be exacerbated in systems with low vagility, ancestral mtDNA polymorphism, and male-biased dispersal.

KEY WORDS: **Directionality index psi, Elapidae, heterozygosity, mitonuclear discordance, private alleles, RADseq, serial founder effect.**

Spatial patterns of genetic variation have played an important role in our understanding of evolutionary processes (Dobzhansky 1937; Mayr 1942; Avise 2000). Both neutral and selective processes can result in the spatial sorting of genetic variation. One classic mechanism for spatial sorting of genetic variation is population divergence in allopatry due to genetic drift (Mayr 1970). In the absence of regular gene flow, isolated populations can accrue frequency differences for alleles, with some alleles becoming fixed in particular populations (Slatkin 1985). Spatial sorting of genetic variation may also be explained by local adaptation, where selective pressures vary across space (Williams 1966; Slatkin 1973; Endler 1977; Kawecki and Ebert 2004). Spatial sorting of genetic variation may be maintained by selection in the face of ongoing gene flow from populations occupying different selective environments. As a consequence, strong spatial variation can arise and persist in contiguous populations without any period of isolation (Bulmer 1972; Lenormand 2002).

A third mechanism for the spatial segregation of genetic variation—range expansion associated drift—does not require allopatric isolation or adaptation. Range expansion events may

EVOLUTION
INTERNATIONAL JOURNAL OF ORGANIC EVOLUTION

directly influence the spatial distribution of genotypes and phenotypes within a species in predictable ways (Austerlitz et al. 1997; Excoffier et al. 2009; Shine et al. 2011). During range expansion, some alleles, even deleterious alleles (Klopfstein et al. 2006; Travis et al. 2007; Excoffier and Ray 2008; Hallatschek and Nelson 2010; Hallatschek 2011), may reach high frequency because of repeated sampling effects (Edmonds et al. 2004). This process is called genetic surfing and it often results in spatially segregated sectors of "surfed" alleles (Excoffier et al. 2009; Graciá et al. 2013).

Although all three mechanisms result in the spatial segregation of genetic variation, the extent of their influence in producing variation across different genomic regions varies, as does their ability to create discordance between nuclear (nucDNA) and organelle genomes. Random genetic drift during long-term isolation should produce spatially correlated variation within the nuclear genomes, and between nuclear and organelle genomes, because the loss of ancestral diversity and the accumulation of novel alleles are related to a common history of isolation. The effects of local adaptation, meanwhile, should be limited to those genomic regions linked to, or including variation that confers a fitness benefit. Local adaptation has sometimes been detected where strong spatial structuring in organelle genomes is not matched by similar spatial structuring in nuclear markers (Cheviron and Brumfield 2009; Toews and Brelsford 2012; Pavlova et al. 2013). Genetic surfing during range expansion may leave a similar signature—strong spatial structuring in organelle genomes without concordant structure of nuclear variation—under some conditions. Because the capacity of genetic surfing to generate spatial sectors of variation should be stronger for loci with lower effective population sizes, maternally inherited organelle genomes should be prime candidates for experiencing surfing (Petit et al. 1997; Excoffier and Ray 2008). Similarly, the frequency of surfing-associated genetic sectors should be greater in loci with weak effective dispersal. Thus, opportunities for the surfing of maternally inherited genomes into spatial sectors should occur more frequently in species with male-biased dispersal.

In a species of extra-tropical venomous snake, the Texas coral snake (*Micrurus tener*), we discovered two deeply divergent mitochondrial (mtDNA) haplotypes that are sympatric in the southern part of the range, but spatially segregated in northern regions. To understand the processes that led to this pattern, we tested for evidence of three different plausible mechanisms that could have caused this discordance: allopatric divergence followed by secondary contact, local adaption, and genetic surfing. Specifically, we tested for (1) divergence in allopatry by inferring population structure and phylogenetic relationships from up to 22,547 nuclear single nucleotide polymorphisms (SNPs), (2) evidence of molecular and environmental selection on mitochondrial types, and (3) recent northern range expansion using genetic data and species distribution models for the last glacial maximum (LGM). Ultimately, we found no support for selection or allopatric divergence in northern regions, but did identify patterns consistent with recent, rapid range expansion. Owing to the combined evidence for recent range expansion, including a strong signal of sequential founder effects across the nuclear genome, we hypothesize that the two divergent mtDNA haplotypes simultaneously "surfed" out of southern refugia into fixed spatial sectors on the expanding northern range front, a process that was likely accentuated by low vagility and/or male-biased dispersal. Thus, we present empirical evidence for a scenario of "mitochondrial surfing"—a largely undescribed scenario in which genetic markers inherited from a single parent can become spatially segregated without the action of selection or divergence in allopatry.

Methods

STUDY SYSTEM AND SAMPLING DESIGN

Coralsnakes belong to the family Elapidae, a group of highly venomous snakes that also includes sea snakes, mambas, cobras, and kraits (Campbell and Lamar 2004). The genus *Micrurus* contains a large radiation of colorful coralsnakes that is restricted to the New World, with the vast majority of species being restricted to the tropics. One exception is the *Micrurus fulvius* complex (sensu Castoe et al. 2012), which ranges as far north as the southeastern United States. This complex includes four species: *M. bernadi* (Bernad's coral snake), *M. fulvius* (eastern coral snake), *M. tener* (Texas coral snake), and *M. tamaulipensis* (Tamaulipan coral snake). Although most species have been described on the basis of color pattern, a recent microsatellite study did not find segregating genotypic variation for the species *M. bernadi*, *M. tener*, and *M. tamaulipensis*, or the four subspecies of *M. tener* (Castoe et al. 2012). Our sampling included representatives of all subspecies of *M. tener*: *M. t. tener* ($N = 60$), *M. t. fitzingeri* ($N = 3$), *M. t. maculatus* ($N = 2$), and *M. t. microgalbineus* ($N = 4$). These subspecies are differentiated from one another by a combination of geographic range, scale counts, number of black body rings, and the size of the nuchal ring (Roze 1996; Campbell and Lamar 2004; Lavin-Murcio and Dixon 2004), however, few morphological characteristics hold up to scrutiny when compared across multiple individuals (see Appendix S1). We also sampled multiple individuals of *M. bernadi* ($N = 4$) and *M. tamaulipensis* ($N = 1$). We found that genome-wide SNP and mtDNA analyses did not support any of these species or subspecies as being clearly differentiated from the nominate form of *M. tener*. As such, in Appendix S1 we recommend that *M. bernadi*, *M. tamaulipensis*, and all subspecies of *M. tener* be considered junior synonyms of *M. tener* and hereafter refer to them as this single taxon.

Field collections from 1998 to 2012 in the United States and Mexico resulted in samples from 90 coralsnakes (16 *M. fulvius*,

74 *M. tener*; Table S1). Genomic DNA was isolated from tissue samples (liver, heart, shed skins, or scale fragments preserved in 95% ethanol or stored frozen at −80°C) using the Qiagen DNeasy extraction kit and protocol (Qiagen, Inc., Hilden, Germany) or by proteinase K digestion followed by phenol-chloroform-isoamyl alcohol organic extraction.

MITOCHONDRIAL DNA SEQUENCING AND ANALYSIS

We sequenced two mtDNA gene fragments: (1) a fragment of the NADH dehydrogenase subunit 4 (ND4; 665 bp) and (2) cytochrome-b (cyt-b; 711 bp) gene from 90 individuals of the *M. fulvius* complex, and six outgroups (Table S1). The ND4 and cyt-b regions were amplified following Castoe et al. (2007). Amplicons were purified using the GeneCleanIII kit (MP Biomedicals) or with ExoSap-IT (USB), and sequenced using either the CEQ D Dye Terminator Cycle Sequencing Quick Start Kit on a Beckman CEQ2000 sequencer (Beckman Coulter) or using BigDye® cycle sequencing mix (Applied Biosystems) run on an Applied Biosystems 3130 genetic analyzer. We edited and aligned mitochondrial sequence data using Geneious 8.0. All sequences were deposited to NCBI's GenBank.

We analyzed mtDNA data using paired Markov Chain Monte Carlo Bayesian analysis in the program MrBayes 3.2.1 (Ronquist and Huelsenbeck 2003). We performed model selection and selected appropriate models using Partition Finder 1.1 (Lanfear et al. 2012; Table S2). We used default setting in MrBayes with the following exceptions; we ran analyses for 10 million generations and sampled trees and parameters every 1000 generations and a variable rate prior. We assessed topological convergence of paired runs using AWTY ("are we there yet?"; Nylander et al. 2008) and the potential scale reduction factor between runs.

GENOME-WIDE NUCLEAR SNP ANALYSIS

We used the double-digest restriction site associated DNA sequence (ddRADseq) protocol of Peterson et al. (2012) to collect nuclear SNPs from the *M. fulvius* complex. Of 90 individuals, 45 had extractions with sufficient amounts of DNA to proceed with the ddRADseq library creation (including seven *M. fulvius*, 38 *M. tener*); detailed laboratory methods are given in Appendix S2. Unprocessed RADseq data (i.e., fastq files) were deposited in the NCBI Sequence Read Archive (Table S1).

We used STACKS version 1.20 (Catchen et al. 2011, 2013) and the FASTX-toolkit (available at http://hannonlab.cshl.edu/fastx_toolkit/index.html) to demultiplex and discard reads lacking restriction sites and high-quality base calls. To process ddRADseq loci and to produce SNP datasets, we used the workflow described by Catchen et al. (2011) to identify RAD-tags containing SNPs (i.e., process_radtags, ustacks, cstacks, and sstacks modules) and used custom Perl scripts to remove any reads that matched two or more stacks. For subsequent data analysis, we modified three parameters to allow for varying levels of missing data (up to 50% missing data per SNP, and up to 25% missing data per SNP). We also experimented with using only one of the paired reads (forward or reverse) or both paired reads (Tables S3 and S4).

We inferred the genetic structure of nuclear SNPs via four methods: (1) a Bayesian clustering method (STRUCTURE, Pritchard et al. 2000), (2) principal components analysis (PCA) using the program adegenet (Jombart 2008), (3) UPGMA cluster analysis implemented in Geneious 8.0 (Biomatters Ltd.), and (4) maximum-likelihood (ML) phylogenetic analysis using the program RAxML 8.0 (Stamatakis 2014). For Bayesian clustering and multivariate analyses (STRUCTURE and PCA), we used heterozygous SNPs (hereafter "biallelic"). We used adegenet 1.3–7 (Jombart 2008) to identify the optimal number of genetic clusters for describing the data using *k*-means clustering, which ranks clustering patterns using BIC scores from axes derived from a PCA (run using the ade4 package [Dray and Bufour 2007] in R version 2.15.2 [R Development Core Team 2014]). We evaluated *k*-means clustering schemes up to $K = 40$. In STRUCTURE analyses, we used default settings and ran each chain for 1–10 million generations, sampling every 1000 generations. In PCA analyses, we dealt with missing data using the "na.replace" command in two ways: (1) missing data were replaced with a value of zero and (2) missing data were replaced with the mean of a given allele, which was computed using all available allelic variants independently from population structure. We found that in all analyses, the first principal component (PC) was strongly correlated with the amount of data per sample ($N = 45$, $r^2 = 0.96$, $P = 0.000$; Fig. S1). Thus, we removed this PC from our analysis and used subsequent axes to interpret genomic variation. In tree-based analyses (UPGMA and RAxML), we only used SNPs that were homozygous within, but variable among individuals. We used default program settings with the following exception: in RAxML 8.0 analysis, we used nonparametric bootstrapping (100 pseudoreplicates), a GTR + Γ model of nucleotide evolution, and ascertainment bias correction (because SNP alignments contain no invariant sites).

TESTING FOR RANGE EXPANSION IN NUCLEAR DNA

A predicted consequence of range expansion is the steady reduction of genetic diversity (e.g., heterozygosity) with increasing distance from the ancestral population (i.e., the serial-founder model, Mayr 1942; Austerlitz et al. 1997; DeGiorgio et al. 2009; Slatkin and Excoffier 2012), which has been well documented in humans (Prugnolle et al. 2005; Ramachandran et al. 2005; Handley et al. 2007; Li et al. 2008; DeGiorgio et al. 2009; Deshpande et al. 2009) and other species (e.g., Pierce et al. 2014; Jezkova et al. 2015). Thus, theoretical expectations of range expansion include that (1) heterozygosity and (2) the number of private alleles will decrease with increasing distance from the ancestral population. We modeled heterozygosity and the number of

private alleles for both *M. tener* and *M. fulvius* using linear models (heterozygosity with latitude) and linear models with a covariate (number of private alleles with latitude, with nearest neighbor distance as a covariate to control for the clustering of samples). We calculated heterozygosity and number of private alleles for each individual using the "populations" module in STACKS (by treating each individual as a population).

We further examined the hypothesis that a south–north range expansion recently occurred in *M. tener* by calculating pairwise directionality indices (the ψ statistic; Peter and Slatkin 2013) for a subset of 32 individuals that were selected to minimize missing data across SNPs from both reads with a total coverage of at least $10\times$. We inferred the area of origin for the supported range expansion using the time distance of arrival (TDOA) method (Peter and Slatkin 2013).

The directionality index, calculated for two populations of interest, is based on range expansion theory involving serial founder effects (Slatkin and Excoffier 2012). Populations further from the origin of an expansion have incurred greater numbers of recent founder events, and consequently more sampling drift, than populations close to the origin of expansion. For derived alleles, which are expected to occur at low frequency in the population of origin, serial founder effects during range expansion should result in one of two general outcomes. In the first outcome, the derived allele is not sampled in one of a series of founder events, resulting in its local extinction and failure to surf along the wave of advance. Derived alleles that are locally extinct (or happen not to be in the molecular sample) are discarded when calculating directionality indices by excluding loci for which either of the two focal populations has private derived alleles. In the second outcome, the derived allele survives repeated founder events. For this circumstance, we must consider the history of the derived allele both at the wave of advance and near the population of origin. When a derived allele survives serial founder events in an advancing wave (i.e., survives sampling drift and random genetic drift) and survives the passage of time near the population of origin (i.e., survives random genetic drift within a stable population), the probability that it has attained high frequency is higher in the advancing wave than the population of origin. The directionality index makes use of loci whose history matches the second outcome to infer range expansion by examining frequencies of the derived allele in both populations.

In our analysis, we calculate pairwise ψ statistics among all individuals. Each individual is treated as a population. For individuals S_1 and S_2,

$$\psi = f_{21} - f_{12},$$

where f_{21} is the proportion of SNPs that are homozygous for the derived allele in S_2, but heterozygous in S_1, and f_{12} is the proportion of SNPs that are homozygous for the derived allele in S_1 but heterozygous in S_2. Positive values indicate that S_2 has been subjected to more sampling drift from serial founder events than S_1, and consequently is further from the origin of expansion than S_1. Negative values indicate the opposite.

Given evidence for northward range expansion in *M. tener*, we expected that subsets of nuclear diversity should be distributed in spatial sectors similar to those observed in the mtDNA dataset. To test for the presence of these fixed sectors of nuclear variation at the putative range edge of *M. tener*, we began by using a multivariate approach, spatial PCA (sPCA; Jombart 2008), to identify groups of spatially correlated SNPs. We then examined the spatial distribution of sPC scores to determine if any nuclear variation was grouped in sectors perpendicular to the putative axis of range expansion via Spearman's rho analysis (performed using SYSTAT 11) with longitude and latitude, respectively. This method ranks PCs by their degree of spatial autocorrelation which is informed by a spatial weighting matrix of geographic coordinates that are measured by Moran's I (Moran 1948, 1950). Following sPCA, eigenvalues are used to identify those PCs with significant and nonsignificant spatial structure. Eigenvalues of the sPCA are composite in nature such that they measure both genetic diversity and spatial structure. We performed sPCA using adegenet with (1) all samples of *M* tener ($N = 36$) and (2) individuals from the United States that should belong to more recently expanded populations ($N = 30$), and we retained all axes with spatial autocorrelation (positive eigenvalues).

TESTING FOR RANGE EXPANSION AND SELECTION IN MITOCHONDRIAL DNA

To examine our mtDNA data for evidence of range/population expansion, we used ARLEQUIN 3.5 (Excoffier and Lischer 2010) to conduct mismatch distribution analyses and BEAST version 1.8.0 (Drummond et al. 2012) to estimate historical effective population sizes using the extended Bayesian skyline coalescent method (EBSP; Heled and Drummond 2008). For mismatch distributions, we simulated distributions under a spatial expansion model, and compared empirically estimated distributions to these simulations using a raggedness index. Empirical mismatch distributions were calculated separately for each of the three major mtDNA lineages. For the EBSP analysis, cyt-b and ND4 sequences were concatenated, and analyzed using an HKY (Hasegawa Kishino and Yano; Hasegawa et al. 1985) substitution model, a strict molecular clock, and 7% $\times$ lineage^{-1} millions $\times$ years^{-1} mutation rate (Castoe et al. 2007). We conducted parallel MCMC (Markov chain Monte Carlo) runs of 20 million generations, with parameters logged every 1000 generations. Proper chain mixing was assessed in Tracer version 1.6 (Rambaut et al. 2014) by verifying that the ESSs were greater than 200, and we used AWTY to confirm topological convergence between EBSP runs. After discarding 10%

of the generations as burn-in, posterior samples were combined using LogCombiner (Drummond et al. 2012). This analysis was conducted on four groups (defined in results): (1) tener A ($N = 26$), (2) tener B ($N = 18$), (3) all tener ($N = 44$), and (iv) fulvius ($N = 11$).

We also estimated the marginal posterior probability curves for parameters of the isolation migration (IM) model for mtDNA using IMa2 (Hey and Neilson 2007). In particular, we were interested in the t_0 parameter (time to most recent common ancestor; TMRCA) and the q parameters (effective female population sizes) associated with major mitochondrial lineages of *M. tener* and *M. fulvius*. We performed two separate IMa2 analyses in parallel, each consisting of a comparison of one *M. tener* (tener A, $N = 26$; or B, $N = 18$) mitochondrial haplogroup and the *M. fulvius* ($N = 11$) haplogroup. We concatenated mtDNA data to form an alignment with a length of 1376 bases, and used the HKY model for analyses. We used an estimated generation time of 1.53 (the average of generation time for *M. tener* and *M. fulvius* according to Roze 1996) and an average mutation rate estimate for snake mitochondrial DNA to rescale parameter estimates into approximate demographic units (1.4×10^{-6} mutations per year; Castoe et al. 2007). For each of our parallel analyses, we performed four independent runs with a 5×10^6 generation burn-in period, followed by a 1.5×10^7 postburn-in sampling period. We determined the burn-in length based on trial runs, and for final runs, we confirmed mixing based on ESS values > 1000 for all parameters per run, and confirmed convergence by comparing independent runs.

Because selection on mitochondrial haplotypes may drive haplotype extinction and limit haplotype polymorphism, and evidence for spatially varying selection would represent a deterministic alternative to the surfing hypothesis, we investigated whether selection may have driven patterns of mitochondrial genetic diversity. We tested for evidence of positive selection in mitochondrial protein-coding sequences using codon-based selection tests, McDonald–Kreitman tests (McDonald and Kreitman 1991), and Tajima's *D* calculation (Tajima 1989) using *Micrurus laticollaris* as an outgroup. These tests were made on reduced alignments (to minimize missing data) of 328 bp and 493 bp in cyt-b and ND4, respectively, and were performed using the program MEGA 5.1 (Tamura et al. 2011) and R package PopGenome (Pfeifer et al. 2014).

To test for evidence that exogenous selection is related to the geographic distribution of *M. tener* mtDNA haplotypes (sensu Pavolva et al. 2013), we performed simple and partial Mantel tests (Mantel 1967; Legendre and Legendre 1998) between bio-climatic, genetic, and geographic distances using the R package vegan (Oksanan et al. 2012). We used point locality data to extract data corresponding to the full set of 19 BIOCLIM variables (2.5 min resolution) in the program DIVA-GIS (Hijmans et al. 2001). Distance matrices were generated using R, MEGA 5.1, and GenAlEx 6.4 (Peakall and Smouse 2006) for Bioclim (log-transformed), mtDNA (concatenated cyt-b + ND4), and geographic distance (in km), respectively. For each mtDNA haplogroup of *M. tener* (*M. tener* A [$N = 32$]; *M. tener* B [$N = 22$]), we performed three simple Mantel tests to test for correlation between dissimilarity matrices and one partial Mantel test, where geographic distance was used as a control matrix. We also conducted these tests on data matrices that combined all *M. tener* from the haplogroup-specific tests. We hypothesized that under a scenario of climate-induced selection, there would be a significant relationship between mtDNA divergence and climate dissimilarity after controlling for the effects of geographic distance.

ESTIMATING THE HISTORICAL DISTRIBUTION OF CORAL SNAKE SPECIES

To complement our studies of range expansion patterns inferred using genetic markers, we examined the current and past distribution of the *M. fulvius* complex by modeling the climatic niche under current and LGM conditions. Present-day models were projected from current occurrence records, and LGM distribution were inferred assuming that the climatic niche of a species has remained conserved between the LGM and present (Elith et al. 2010). We used methods described in Jezkova et al. (2015) to remove highly correlated variables, set MAXENT parameters (Phillips et al. 2006), and to evaluate the performance of various models. We first analyzed each species separately, but because we observed substantial overlap in distributional predictions, we ultimately pooled all species together and ran a single model on the entire *M. fulvius* complex. See Appendix S2 for additional information regarding niche modeling and model choice.

Results

THREE MITOCHONDRIAL HAPLOGROUPS WITH ENIGMATIC DISTRIBUTIONS

Phylogenetic analysis of cyt-b and ND4 revealed that there are three haplogroups within the *M. fulvius* complex (Figs. 1A and S2)—two of these haplogroups contain individuals referable to *M. tener*. The two haplogroups containing individuals of *M. tener*, however, were not monophyletic with regard to *M. fulvius*. Interestingly, the two divergent haplogroups that are contained within *M. tener* are codistributed in Mexico and southern Texas, but spatially segregated from one another in northern Texas and Louisiana (Fig. 1B). We refer to these haplogroups as *M. tener* A and B, respectively (see Fig. 1A and B).

NUCLEAR SNPs SUPPORT TWO SPECIES OF CORAL SNAKE IN THE *Micrurus fulvius* COMPLEX

We found that our results were qualitatively similar across all examined datasets, and thus we only describe the results obtained

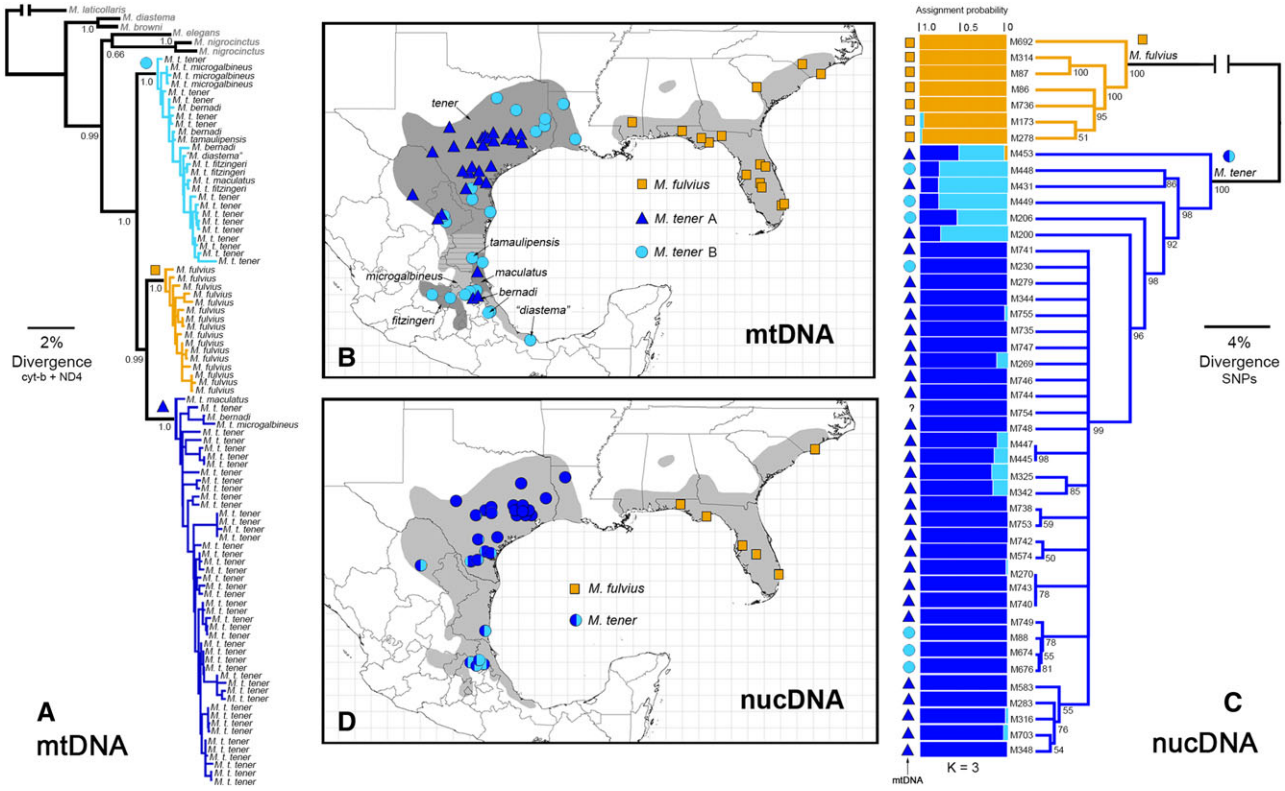

Figure 1. **Bayesian phylogram based on concatenated alignment of mitochondrial (mtDNA) ND4 and cyt-b sequences (1376 base pairs) from the *Micrurus fulvius* complex. Posterior probabilities are listed on relevant nodes and symbols (triangle, square, and circle) correspond to samples in(B) map (A). Spatial distribution of individuals sampled for mtDNA analysis. The geographic distribution of subspecies and junior synonyms of *M. tener* are indicated with putative area of intergradation (Campbell and Lamar 2004) indicated by barred lines (B). UPGMA cluster analysis of 3466 homozygous nuclear (nucDNA) SNPs (support values correspond to 2000 bootstrap pseudoreplicates) and STRUCTURE plot ($K = 3$) of 22,547 biallelic SNPs with corresponding mtDNA haplogroup designations (symbols to left of plot). A question mark indicates a single sample (M754) that we did not acquire mtDNA from, despite multiple attempts. Sample numbers correspond to Table S1 and symbols correspond to samples in (D) map (C). Spatial distribution of individuals sampled for nucDNA analysis. Proportion of color in each symbol related to the posterior assignment probabilities from (C) STRUCTURE analysis (D).**

from the dataset including 50% missing data and at least 10× coverage per locus using both paired reads, which resulted in 3466 phylogenetically informative SNPs (variable between individuals but homozygous within an individual) and 22,547 biallelic SNPs. The total number of RAD loci obtained from the *M. fulvius* complex using STACKS under different data combination and filtering schemes is summarized in Tables S3 and S4. Phylogenetic clustering (RAxML and UPGMA) performed using homozygous SNPs resulted in two distinct and well-supported clades, one corresponding to *M. fulvius* and another to *M. tener* (Fig. 1C), which are geographically distributed according to known ranges for these two species (Fig. 1D). The *k*-means clustering analyses suggested that between two and four clusters were useful in describing the observed SNP variation (Fig. S3). STRUCTURE analysis using a *K*-value of 3 shows three clusters of individuals: (1) all individuals of *M. fulvius*, (2) individuals of *M. tener* from Mexico, and (3) individuals of *M. tener* from the United States; individuals of *M. tener* from south Texas had mixed assignment to both clusters of *M. tener* (Fig. 1C and D). Similar mixed assignments were also observed when all individuals of *M. fulvius* were excluded from STRUCTURE analysis (Fig. S4). It is notable that there was little correspondence between the posterior assignments of nuclear clusters from STRUCTURE analyses and mtDNA haplotypes observed in *M. tener* individuals (Fig. 1B and D). Our PCAs of SNP variation produced patterns that correspond well with the geographic distribution of samples, but not with the geographic distribution of mtDNA haplotypes (Fig. 2A and B).

NUCLEAR SNPs REVEAL EVIDENCE OF RECENT NORTHERN RANGE EXPANSION

Within *M. fulvius*, weak associations of heterozygosity and private alleles were observed with latitude (heterozygosity, $N = 6$, $R^2 = 0.36$, $P = 0.14$; number of private alleles, $N = 6$, $R^2 = 0.10$, $P = 0.28$; Fig. 2C and D) and with longitude (heterozygosity, $N = 6$, $R^2 = 0.000$, $P = 0.73$; number of private alleles, $N = 6$, $R^2 = 0.000$, $P = 0.74$). Within *M. tener*, we found negative

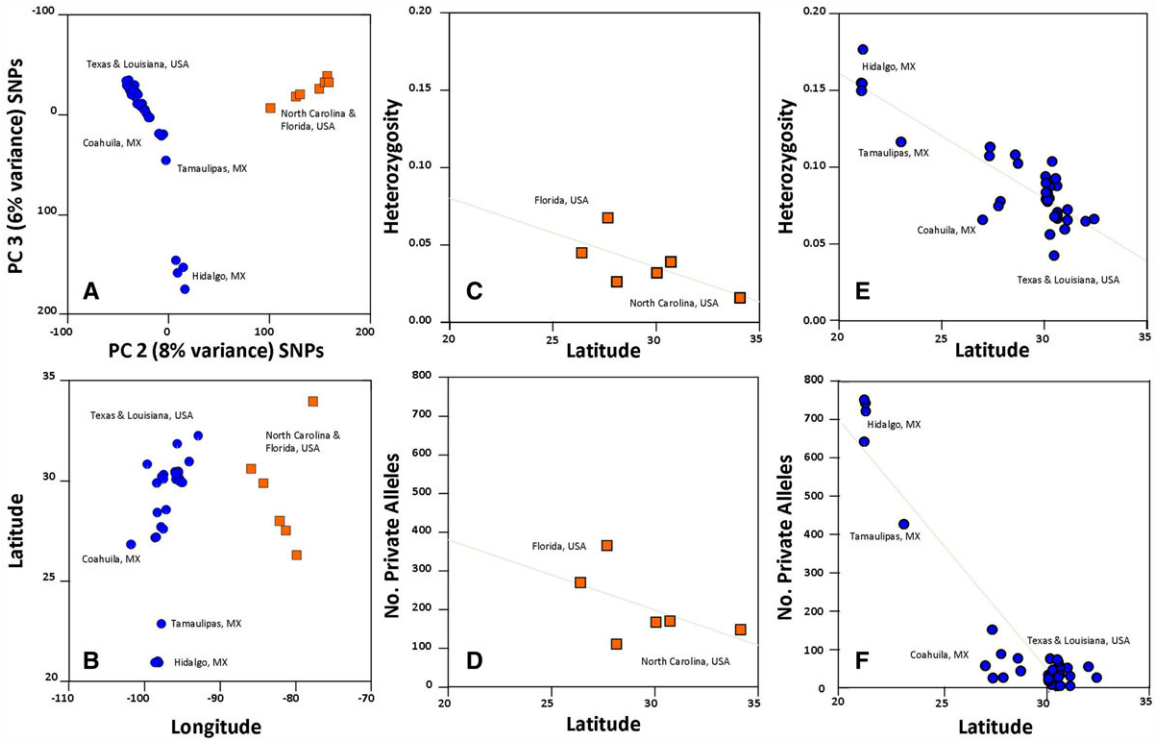

Figure 2. **Principal components (PCs) 2 and 3 from a multivariate analysis of 22,547 SNPs derived from the *Micrurus fulvius* complex samples used in this study (A). PC 1 corresponded to the amount of missing data (see Fig. S1). Geographic distribution of samples used in the PCs analysis (B). Locality names have been added and PC 3 has been inverted to demonstrate the similarity of the patterns. Relationship between heterozygosity and latitude in *M. fulvius* (C; orange squares) and *M. tener* (E; blue circles) sampled throughout their range. Relationship between the number of private alleles and latitude in *M. fulvius* (D; orange squares) and *M. tener* (F; blue circles).**

relationships between heterozygosity and latitude, as well as the number of private alleles and latitude (heterozygosity, $N = 37$, $R^2 = 0.75$, $P < 0.001$; number of private alleles, $N = 37$, $R^2 = 0.86$, $P < 0.001$; Fig. 2E and F). Neither heterozygosity nor number of private alleles showed strong relationships with longitude in *M. tener* (heterozygosity, $N = 37$, $R^2 = 0.15$, $P = 0.01$; number of private alleles, $N = 37$, $R^2 = 0.13$, $P = 0.02$). Although we did not find a significant relationship between any variables when examining *M. fulvius*, our sample size was small and the general pattern for latitude (in terms of R^2 values and negative slope) was similar to *M. tener*. We interpret these findings as evidence that northern range expansion events occurred in *M. tener*, and possibly *M. fulvius*.

We sampled 4028 nucDNA SNPs to calculate directionality indices in *M. tener*. Of these, 1135 were excluded because they were heterozygous in our *M. fulvius* sample (i.e., they could not be used to identify derived alleles in *M. tener*). From the remaining 2893 polarized SNPs, we calculated pairwise ψ statistics for all 496 possible individual pairwise combinations across 32 individuals. The mean number of polarized SNPs for which a derived allele was present in both individuals in pairwise comparisons was 451.7 ($\pm$ 2.8 SE; range 295–574).

Our inference of the area of origin for the range expansion using the TDOA method supports an origin of *M. tener* in central Mexico (Fig 3A). Under a hypothesis of northward range expansion, directionality indices should indicate relatively greater effects of sampling drift with higher latitude. We calculated the directionality index for individual pairwise individual comparisons for all *M. tener* samples. Among these comparisons, directionality indices should have greater magnitude for pairwise comparisons from more disparate latitudes. Additionally, values should be positive when the S_2 sample (see Methods) is at higher latitudes than the S_1 sample (and negative when S_1 is at higher latitudes). Both of these predictions were met consistently across the dataset (Fig. 3B), and therefore indicate recent range expansion from south to north.

MITOCHONDRIAL DATA CONSISTENT WITH NEUTRAL EVOLUTION AND POPULATION EXPANSION

Mismatch distributions and EBSP analyses provided evidence for recent expansion within each of the mtDNA haplogroups of the *M. fulvius* complex (Figs. 4A–D and S5). Interestingly, the BSP analysis estimated the timing of population expansion initiation as similar for all three haplogroups (ca. 100,000 years before present;

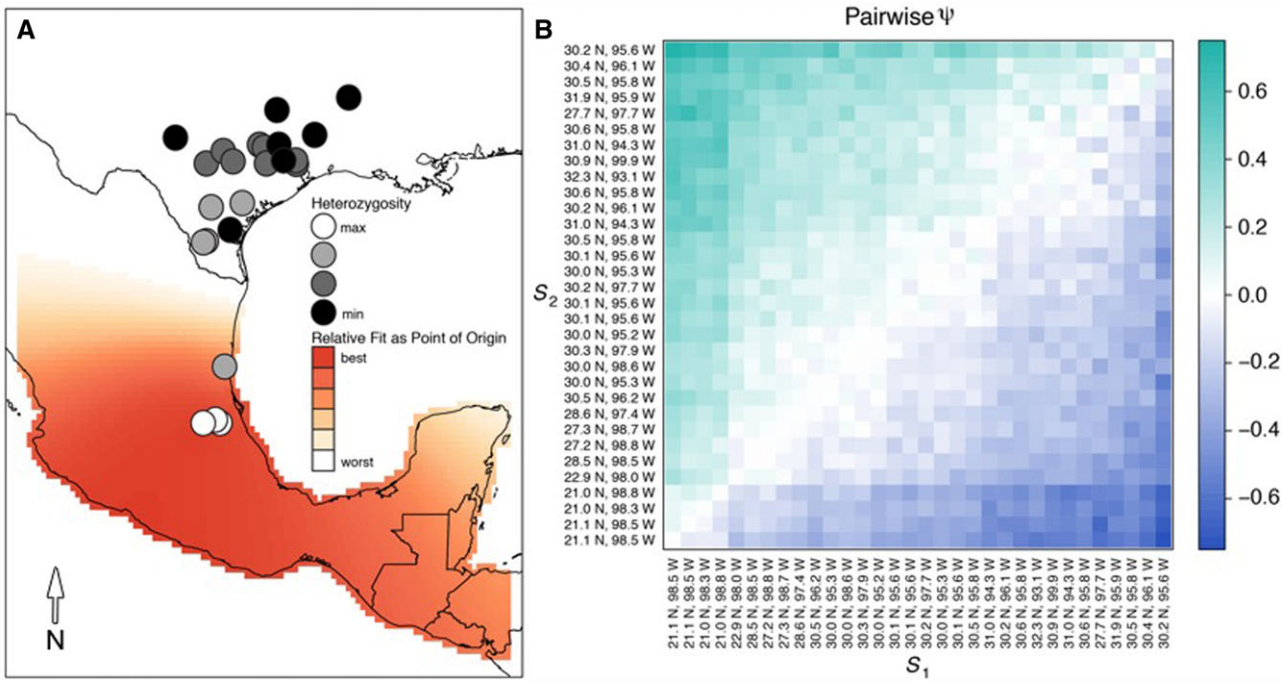

Figure 3. **Genomic evidence (nuclear DNA) for northern range expansion in *Micrurus tener*: Inference of the area of origin for the range expansion using the TDOA method supports an origin in Mexico (A). Redder colors indicate better fit as a point of origin for a given grid cell, under a least squares criterion. Sampling points are colored across a grayscale spectrum relative to the maximum and minimum heterozygosity values for the 32 samples (see also Fig. 2). Directionality index values for all pairwise comparisons. If northern range expansion occurred, we should observe positive values when the S_2 sample is more northerly than the S_1 sample (and negative when S_1 is more northerly). These predictions are met consistently across the dataset, indicative of recent northward range expansion (B).**

Fig. 4). Additional evidence that these haplogroups began to diversify at similar points in time, and have similar demographic histories, was observed in our IMa2 results when comparing estimates of effective population size and TMRCA (time to the most recent common ancestor) values for each haplogroup (Fig. 4E and F).

REJECTING SELECTION AS A DRIVER FOR MITOCHONDRIAL DIVERSITY PATTERNS

Selections tests (Tajima's *D* and codon based tests; Tables 1 and 2) were consistent with either neutral evolution or purifying selection (i.e., no evidence for positive or balancing selection). McDonald–Kreitman tests (conducted individually on cyt-b and ND4) revealed no significant differences in d*N*/d*S* ratios across haplogroups (Table 2). Our tests for exogenous selection using Mantel tests with the *M. tener* A and B haplogroup individuals revealed significant relationships between all matrices for simple comparisons, but nonsignificant relationships between climate and mitochondrial divergence when correcting for geographic distance in both haplogroups (Table 3). Thus, our expectations for selection-driven discordance were not met for either endogenous or exogenous analyses.

COMPARISONS BETWEEN SPATIALLY CORRELATED NUCLEAR SNP VARIATION AND LATITUDE/LONGITUDE

To further evaluate evidence for northward range expansion in *M. tener*, we tested if subsets of nuclear diversity were distributed in spatial sectors similar to those observed in the mtDNA dataset. We used sPCA to test this hypothesis by identifying groups of spatially correlated SNPs. We ranked PCs by their degree of spatial autocorrelation, using eigenvalues to identify PCs with significant spatial structure, and estimated if any nuclear variation was grouped in sectors perpendicular (roughly east–west) to the putative axis of range expansion (roughly north–south) by examining Spearman's rho correlation with longitude and latitude.

An sPCA performed on individuals of *M. tener* from Mexico and the United States resulted in 51 spatially correlated axes, with the first three sPCs explaining the most variation (Fig. S6 A and B). An sPCA performed on only populations of *M. tener* from the United States resulted in 11 spatially correlated axes (Fig. S6 C and D). We found that as sPCs explained progressively less of the cumulative variation, the correlation with latitude and longitude diminished. When analyzing individuals of *M. tener* from Mexico and the United States, we observed that the most explanatory axes (sPCs 1 and 2) were similarly correlated with latitude and

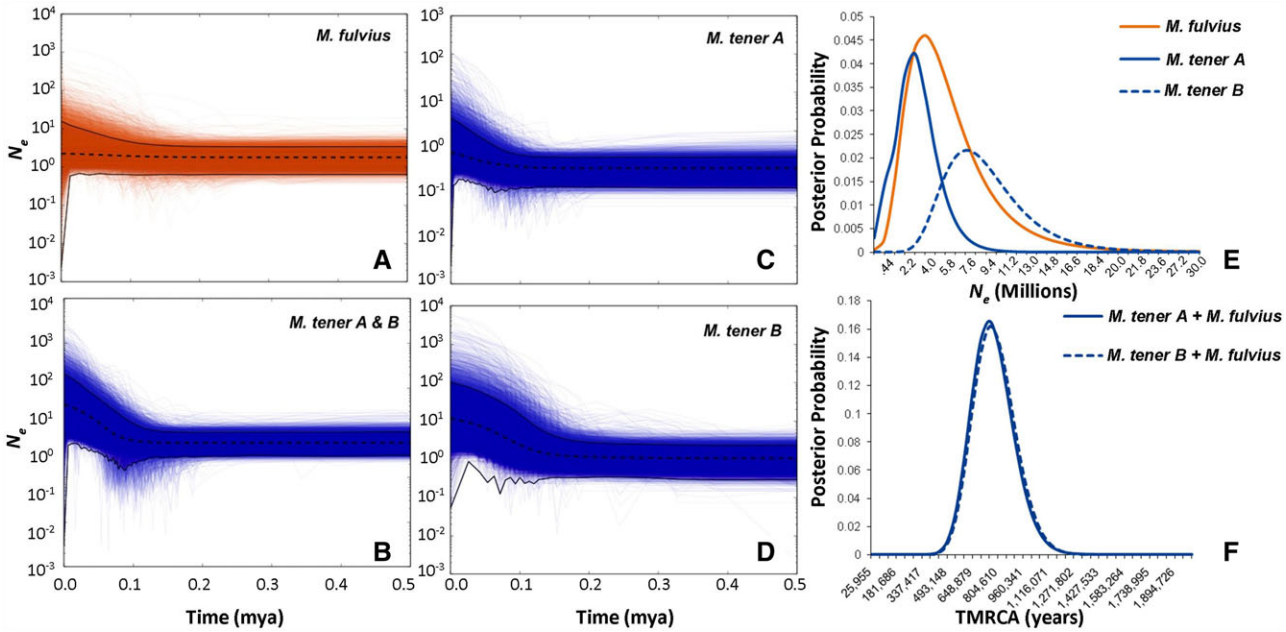

Figure 4. Evidence for population expansion in each mitochondrial DNA haplogroup: Extended Bayesian skyline plots of effective population size in (A) *Micrurus fulvius*, (B) both mtDNA haplotypes of *M. tener*, (C) *M. tener* mtDNA haplotype A, and (D) *M. tener* mtDNA haplotype B over the past 500,000 years. Dotted lines represent the median effective population sizes inferred using the MCMC analysis and solid lines represent the 95% highest posterior density. All Bayesian inferences are represented with orange (*M. fulvius*) and blue (*M. tener*) "densi-clouds." Marginal posterior probability density curves for isolation migration model parameters of *M. tener* and *M. fulvius* from representative MCMC runs in IMa2 (E). Density curve for the *q* (population size) parameter (F). Density curve for the time since splitting of an ancestral population into *M. tener* and *M. fulvius* populations. Each parameter was converted to demographic units using the average mutation rate for the ND4 and cyt-b mitochondrial genes for snakes and the average generation time for *M. tener* and *M. fulvius* (Roze 1996).

Table 1. Codon-based tests for selection for mitochondrial genes sequenced from the *Micrurus fulvius* complex.

			Alternative hypothesis tested		
mtDNA lineage	Locus	N	Positive $d_N > d_S$	Purifying $d_N < d_S$	Neutral $d_N = d_S$
All	cyt-b(109 codons)	73	−3.54 ($P = 1.000$)	3.62 ($P = 0.000$)	−3.62 ($P = 0.000$)
M. tener-A/triangle	cyt-b (109 codons)	37	−3.25 ($P = 1.000$)	3.26 ($P = 0.001$)	−3.23 ($P = 0.002$)
M. tener-B/circle	cyt-b (109 codons)	24	−2.22 ($P = 1.000$)	2.28 ($P = 0.012$)	−2.22 ($P = 0.028$)
M. fulvius	cyt-b (109 codons)	12	−0.92 ($P = 1.000$)	0.93 ($P = 0.177$)	−0.87 ($P = 0.386$)
All	ND4 (141 codons)	55	−4.10 ($P = 1.000$)	3.91 ($P = 0.000$)	−4.07 ($P = 0.000$)
M. tener-A/triangle	ND4 (141 codons)	27	−1.92 ($P = 1.000$)	1.86 ($P = 0.033$)	−1.82 ($P = 0.071$)
M. tener-B/circle	ND4 (141 codons)	17	−1.91 ($P = 1.000$)	1.87 ($P = 0.032$)	−1.91 ($P = 0.059$)
M. fulvius	ND4 (141 codons)	11	−1.52 ($P = 1.000$)	1.56 ($P = 0.061$)	−1.52 ($P = 0.131$)

Variance was estimated using 1,000 bootstrap pseudoreplicates. Significant statistics are indicated in bold.

longitude (Fig. S6A vs. B), and when analyzing individuals from more recently expanded populations (U.S. samples only), we observed that the most explanatory axes had stronger correlations with longitude (sPC1; Fig. S6C vs. D). Correlation coefficients (rho) for all tests are provided in Table S5. Visualizing lagged sPC scores (as recommended by Jombart et al. 2008) in relation to geography also revealed clustering patterns that are perpendicular to the putative axis of expansion, similar to the mtDNA dataset (e.g., sPCs 1 and 2 vs. mtDNA; Fig. S7). Collectively, sPCA analyses provide evidence that multiple axes of nucDNA variation are correlated with longitude. Importantly, more nucDNA variation is structured this way than would be expected to result from isolation by distance. This finding is consistent with the stochastic spatial sorting of alleles perpendicular to the axis of range expansion that is predicted under the mechanism of genetic surfing.

Table 2. Selection tests (McDonald–Kreitman and Tajima's *D*) for mitochondrial genes as inferred by PopGenome (Pfeifer et al. 2014). Tajima's D estimates that were significant at P < 0.05% are shown in bold.

	Dataset	Pn	Ps	d*N*	d*S*	*P*-value	*D*
M. tener-all	cyt-b	5	8	25	48	0.76	0.58
M. tener-A	cyt-b	7	4	26	49	0.10	**−1.94**
M. tener-B	cyt-b	3	8	26	47	0.74	**−0.94**
M. fulvius	cyt-b	12	15	23	43	0.48	**−1.72**
M. tener-all	ND4	20	26	36	47	1.00	0.01
M. tener-A	ND4	6	7	39	54	0.77	**−2.07**
M. tener-B	ND4	10	11	40	53	0.81	**−1.91**
M. fulvius	ND4	5	3	37	55	0.27	**−1.71**

For McDonald–Kreitman tests, *M. laticollaris* (M33) was used as an outgroup in all comparisons.

HISTORICAL SPECIES DISTRIBUTION MODELING IS CONSISTENT WITH NORTHWARD EXPANSION

Our inferences of past and present distributions using niche-modeling approaches independently support the hypothesis of a Gulf Coast of Mexico origin and recent circumcoastal (post-LGM) range expansion in *M. tener* (Fig. 5). The resulting models provide evidence that during the LGM both *M. fulvius* and *M. tener* were restricted to more southern geographic regions than they are at present (Figs. 5 and S8). Furthermore, based on conservative estimates, the range of *M. tener* was likely restricted south to the Mexican state of Tamaulipas, where we observe both mtDNA haplogroups of *M. tener* in sympatry.

Discussion

TESTING AMONG MECHANISMS THAT MAY EXPLAIN SPATIAL SORTING OF MITOCHONDRIAL DNA

In this study, we conducted analyses to discern which of three plausible mechanisms best explained the spatial distribution of mtDNA lineages observed in *M. tener*: allopatric divergence, local selection, or genetic surfing. Two lines of evidence reject allopatric divergence as a potential explanation. First, evidence from Bayesian clustering, phylogenetic reconstruction, and PCA of nucDNA indicates that individuals of *M. tener* from the northern regions (where mtDNA lineages are spatially segregated) comprise a single, closely related genomic cluster (Figs. 1C and Fig. 2A and B). Second, our LGM modeling suggests that the region of Texas and Louisiana where the spatial segregation of mtDNA lineages occurs was unsuitable habitat for *M. tener* during the LGM (Fig. 5). We also were able to reject selection as an explanation for the spatial segregation of haplotypes in *M. tener* because our tests for both endogenous and exogenous selection indicated that observed patterns were not significantly different than neutral expectations (Tables 1–3).

Genetic surfing occurs through the process of range expansion, and we found multiple lines of evidence for recent and rapid range expansion. First, we found significant negative relationships between latitude and both nuclear heterozygosity and number of private alleles, indicative of northern range expansion (Fig. 2E and F). Second, our directionality index analysis of nuclear diversity conformed to expectations for northern range expansion (Fig. 3B). Third, we found that mitochondrial haplogroups display patterns consistent with recent population expansion at approximately the same time (Fig. 4), with phylogenetic reconstructions placing most individuals from Mexico as basally diverging (Fig. 1B). Finally, using sPCA we revealed that multiple axes of nuclear genome-wide variation are structured perpendicular to the axis of expansion but not in a consistent manner (Figs. S6 and S7); this pattern matches the expectation for how genetic surfing would stochastically generate spatial sectors during an expansion event (Excoffier et al. 2009).

Evidence of recent range expansion from our analyses of mitochondrial and nuclear genetic data is also independently supported by LGM niche modeling predictions, suggesting *M. tener* arrived in the northern portions of its current range recently and (given the presence of both mtDNA haplogroups near the putative ancestral population) suggests they brought the observed mtDNA variation with them during colonization. If divergent mtDNA haplogroups co-occurred in the ancestral population of *M. tener* as standing variation, why did they spatially segregate near the most recently expanded range edge? We believe that the mechanism of genetic surfing provides the best explanation for this pattern—alleles (in this case mitochondrial haplogroups) that occur with similar frequency in the ancestral population should fix in clines perpendicular to the axis of expansion (Hallatschek et al. 2007)—the same process we describe for a subset of our nucDNA data.

EVIDENCE FOR MITOCHONDRIAL SURFING

It is an open question why mitochondrial surfing has not been reported more often in animals, and we are motivated to understand what conditions have made this phenomenon detectable in coralsnakes. First, Excoffier et al. (2009) note that the distinct lineages that arise as a consequence of range expansion are transient in nature, but might last longer in organisms with limited dispersal ability, such as snails (Cain and Currey 1963; Goodhart 1963; Gould and Woodruff 1996), trees (Petit et al. 1997), and snakes (this study). Therefore, limited overall dispersal capability in the coral snake system may have contributed to patterns consistent with mitochondrial surfing. Second, many elapid snakes, possibly including coralsnakes, have male-biased sex dispersal (e.g., Keogh et al. 2007). Such sex-biased dispersal should result in a decrease in the ratio of effective population sizes (Ne) of mtDNA

Table 3. Test for whether climatic variation is associated with mitochondrial divergence via simple and partial Mantel tests applied to two divergent mitochondrial haplogroups of *Micrurus tener* (*M. tener* A, $N = 32$; *M. tener* B, $N = 22$).

Haplogroup (test)	Dissimilarity matrices	Mantel statistic r	*P*-value
M. tener A (simple)	GEOG × CLIM	0.758	0.001
M. tener A (simple)	GEOG × GENE	0.754	0.001
M. tener A (simple)	GENE × CLIM	0.437	0.017
M. tener A (partial)	GENE × CLIM × GEOG	−0.271	**0.995 NS**
M. tener B (simple)	GEOG × CLIM	0.983	0.001
M. tener B (simple)	GEOG × GENE	0.610	0.001
M. tener B (simple)	GENE × CLIM	0.283	0.002
M. tener B (partial)	GENE × CLIM × GEOG	−0.008	**0.537 NS**
M. tener-all (simple)	GEOG × CLIM	0.717	0.001
M. tener-all (simple)	GEOG × GENE	0.336	0.001
M. tener-all (simple)	GENE × CLIM	0.245	0.001
M. tener-all (partial)	GENE × CLIM × GEOG	0.007	**0.404 NS**

Three matrices were used: climatic distance (CLIM), geographic distance (GEOG), and mitochondrial genetic distance (GENE). For partial Mantel tests, the third matrix, GEOG, was the control matrix.

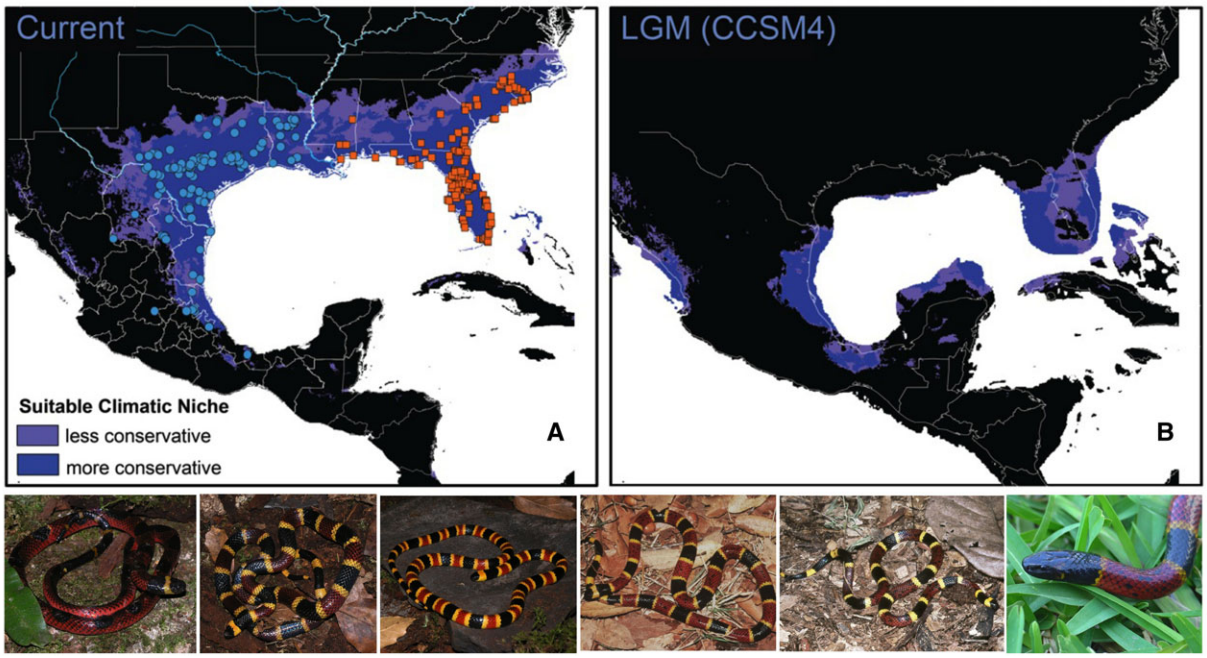

Figure 5. Top: Species distribution models for *Micrurus fulvius* and *M. tener* predicting current distribution (A) and distribution during the LGM (B). Occurrence records for *M. fulvius* and *M. tener* used for distribution prediction are shown as orange squares and blue circles on the current model, respectively. See Figure S8 for additional LGM model approximations. Bottom: Color pattern variation within *M. tener* (from left to right): Huejutla de Reyes, Hidalgo, Mexico; Guanajuato, Mexico; Tampico, Tamaulipas, Mexico; Gómez Farías, Tamaulipas, Mexico; Smith County, Texas, United States; and Sierra de Tamaulipas, Tamaulipas, Mexico. Photography credits: W. W. Lamar, E. A. Liner, E. N. Smith, and P. A. Lavin-Murcio.

versus nucDNA at the range front, relatively higher probability of mitochondrial surfing, and longer temporal persistence of sectors affected by surfing. Interestingly, this differs notably from the way that male-biased dispersal is typically hypothesized to contribute to mitonuclear discordance, such as through the immigration of foreign males to relatively isolated populations of distinct mitochondrial types (e.g., Peters et al. 2014; birds). Thus, unique characteristics of the coral snake system, including male-biased dispersal, relatively low vagility, and a source population that apparently contained two divergent mitochondrial haplotypes, may collectively explain why this coral snake system experienced mitochondrial surfing on a scale that is obvious and detectable.

AN INTRIGUING SCENARIO FOR MITOCHONDRIAL DIVERSIFICATION

A key component in our hypothesis of mitochondrial surfing in coralsnakes is that ancestral populations of *M. tener* contained two divergent (about 5% sequence divergence) yet sympatric mtDNA lineages prior to northern range expansion. Given our estimates of range limits during the LGM, it is likely that populations possessing these distinct mitochondrial haplotypes were confined to one or more southern refugia at this time. It is therefore relevant to consider how a panmictic population could be expected to maintain multiple divergent mtDNA haplotypes through time, as population and range expansion proceed northward. One might assume that both stochastic sampling and mitonuclear coevolution would tend to drive one of these haplotypes to extinction (similar to the hypothesis of Hill and Johnson [2013]). Empirical evidence, however, for the persistence of divergent mitochondrial haplotypes is available in multiple natural populations, including in ring species and in examples of ancient insular hybridization (Irwin et al. 2001; Larsen et al. 2010). The ability for each of these haplotypes to persist and expand northward suggests that each is compatible against a more or less common nuclear background, which is supported by our inability to detect any type of molecular selection other than general purifying selection (Table 1).

Simulation studies suggest that mtDNA lineage extinction is dramatically slowed in expanding populations (Avise et al. 1984). N_e is likely the key determinant of a population's likelihood of maintaining divergent mtDNA haplotypes, assuming compatibility is maintained between the nuclear genome and multiple mitochondrial variants. Specifically, the ability to maintain large amounts of mtDNA variation would require relatively stable and large N_e through time. Nucleotide diversity (π) should be directly proportional to N_e for sites evolving neutrally (Hartl and Clark 1997). We estimated π from our nucDNA SNPs using DNAsp 4.9 (Rozas et al. 2003) and found, not surprisingly based on our other analyses, that these estimates were higher in southern populations of *M. tener* than either northern populations of *M. tener* or all sampled populations of *M. fulvius* (Table 4).

Interestingly, fossil evidence is consistent with large ancestral population sizes, as well as inferences from genetic data and range model predictions of recent range expansion in *M. tener*. Two quaternary fossils from the *M. fulvius* complex have been described from Texas (Holman 2000; Lundelius 2003). The youngest fossil is from 10,900 years before present and should correspond to the post-LGM expansion of *M. tener*. The older fossil is known from 830,000 years before present and, based on our TMRCA estimates (Fig. 2F), may represent a common ancestor of *M. fulvius* and *M. tener*. Importantly, there are also Middle Pleistocene fossils of the *M. fulvius* complex known from Florida (Holman 2000). This would indicate that the ancestor of the *M. fulvius* complex possessed a large distribution, including what is now central Texas and Florida before experiencing substantial range reduction and the divergence of *M. fulvius* and *M. tener* (sensu Fig. 5B). Collectively, fossil, genetic, and niche modeling evidence are consistent with a particular historical scenario that could account for many unique features of the system, including ancestral mitochondrial polymorphism. This scenario includes the existence of a widespread *M. tener–M. fulvius* ancestral lineage that was subdivided initially into separate northern and southern populations (possibly near the current U.S.–Mexico border), and then further subdivided into a northeast and a northwest population (across the Mississippi River); this pattern would explain the mitochondrial phylogeny. If glacial cycles then drove both western populations south into one or multiple refugia, subsequent northward expansion, accompanied by gene flow, would have mixed nuclear genomes, while mitochondrial surfing led to spatial structuring of divergent haplotypes. If this was indeed the case, it is intriguing that while mitochondrial data mislead inferences of species limits and divergence, they do retain signatures of complex ancient events largely obscured by recent nuclear genomic admixture.

Table 4. Nucleotide diversity levels of mitochondrial genes (cyt-b and ND4) and nuclear genome-wide SNPs (gwSNPs) obtained from the *Micrurus fulvius* complex.

	Dataset	*N*	*S*	Ps	θ	π
mtDNA						
All	cyt-b	73	27	0.08	0.017	0.019
M. tener-A	cyt-b	37	21	0.06	0.015	0.007
M. tener-B	cyt-b	24	9	0.03	0.007	0.004
M. fulvius	cyt-b	12	4	0.01	0.004	0.002
All	ND4	55	32	0.08	0.016	0.020
M. tener-A	ND4	27	9	0.02	0.005	0.003
M. tener-B	ND4	17	9	0.02	0.006	0.004
M. fulvius	ND4	11	7	0.02	0.006	0.003
nucDNA						
M. tener-MX	gwSNPs	5	3326	0.96	0.462	0.546
M. tener-USA	gwSNPs	35	3466	1.00	0.242	0.467
M. tener-All	gwSNPs	40	3466	1.00	0.235	0.477
M. fulvius	gwSNPs	7	2760	0.84	0.342	0.415

MITOCHONDRIAL SURFING AND IMPLICATIONS FOR OTHER SYSTEMS

Genetic surfing may be an underappreciated mechanism by which mitochondrial structure and diversity can be strongly impacted by stochastic demographic processes, although it remains unclear how common this phenomenon is in nature. For example, is it reasonable to expect genetic surfing to have similarly generated sectors of organelle diversity for other species that have experienced post-Pleistocene range expansion? We suspect the answer is both yes, and no. Yes, it is likely that mtDNA haplotype

diversity and structure have been affected by mitochondrial surfing in organisms that have experienced range expansion. However, such effects may be either so ephemeral that they are rarely encountered, or they may generate structure of minimally differentiated mtDNA haplotypes. Indeed, the circumstances under which the coral snake example arose are rare: few species maintain two highly divergent mitochondrial haplotypes in sympatry. Thus, while it is necessary to consider the potential for mitochondrial surfing to generate spatial sectors of genetic variation in taxa from temperate regions, it is yet unclear how widespread it is. Although it may be limited to taxa with male-biased or strongly limited dispersal, we suspect species with histories of range fragmentation followed by secondary admixture may be more prone to exaggerated cases of mitochondrial surfing due to the increased probability of multiple divergent mitochondrial haplotypes existing in a single expanding population.

ACKNOWLEDGMENTS

We thank J. Castoe and R. Wostl for assistance with laboratory work and B. Peter for assistance with directionality index analyses. M. Fujita and J. Wiens provided helpful suggestions on early ideas for the study. We thank the Whiteman Laboratory at the University of Arizona for comments on preliminary analyses. We are grateful to the following researchers, snake enthusiasts, curators, and collection managers for providing tissues for this work: A. C. Saucedo, T. LaDuc, T. Sinclair, T. Cole, R. A. Cortes, O. F. Villela, D. L. Villareal, E. S., C. Franklin, U. O. G. Vasquez, J. Vindum, F. M. Quijano, F. R. M. Paz, J. R. Velasco, K. Krysko, J. Dixon, T. Hibbitts, T. Hibbitts, M. Hibbitts, M. Varela, R. Lawson, B. Stuart, W. V. Devender, R. Murphy, M. Sasa, C. Guirola, A. N. M. de Oca, J. Boundy, R. Mora, K. Wray, L. C. Márquez, P. Lavin-Murcio, C. Devine, G. Pauly, E. M. Salazar, S. Hernandez G, A. R. Tiburcio, R. H. Arciga, B. Bradford, M. T., E. P. Ramos, J. Perez, M. Benard, D. Hillis, D. Cannatella, R. Brumfield, F. Sheldon, D. Dittmann. We thank P. Lavin-Murcio and J. Dixon for providing a photograph of *M. tamaulipensis*. We sincerely thank J. McGuire, M. Alfaro, and an anonymous reviewer for comments that greatly improved the manuscript. Funding was provided by a National Science Foundation collaborative grant to CLP and ENS (DEB-0416000, 0416160), a grant from Instituto Bioclon to ENS, and through faculty startup funds from the University of Texas at Arlington to TAC.

DATA ARCHIVING

The doi for our data is doi:http://dx.doi.org/10.5061/dryad.p6m94.

LITERATURE CITED

Austerlitz, F., B. Jung-Muller, B. Godelle, and P. -H. Gouyon. 1997. Evolution of coalescence times, genetic diversity and structure during colonization. Theor. Popul. Biol. 51:148–164.

Avise, J. C. 2000. Phylogeography: the history and formation of species. 447 pp. Harvard Univ. Press, Cambridge, MA.

Avise, J., J. Neigel, and J. Arnold. 1984. Demographic influences on mitochondrial DNA lineage survivorship in animal populations. J. Mol. Evol. 20:99–105.

Bulmer, M. G. 1972. Multiple niche polymorphism. Am. Nat. 106:254–257.

Cain, A. J., and J. D. Currey. 1963. Area effects in Cepaea on the Larkhill Artillery ranges, Salisbury plain. Zool. J. Linn. Soc. 45:1–15.

Campbell, J. A., and W. W. Lamar. 2004. The venomous reptiles of the Western hemisphere. Vol. 1, 475 pp. Cornell Univ. Press, Ithaca, NY.

Castoe, T. A., C. L. Spencer, and C. L. Parkinson. 2007. Phylogeographic structure and historical demography of the western diamondback rattlesnake (*Crotalus atrox*): a perspective on North American desert biogeography. Mol. Phylogenet. Evol. 42:193–212.

Castoe, T. A., J. W. Streicher, J. M. Meik, M. J. Ingrasci, A. W. Poole, A. P. J. de Koning, J. A. Campbell, C. L. Parkinson, E. N. Smith, and D. D. Pollock. 2012. Thousands of microsatellite loci from the venomous coralsnake *Micrurus fulvius* and variability of select loci across populations and related species. Mol. Ecol. Resour. 12:1105–1113.

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

Catchen, J. M., A. Amores, P. Hohenlohe, W. Cresko, and J. H. Postlethwait. 2011. Stacks: building and genotyping loci de novo from short-read sequences. G3 Genes Genom. Genet. 1:171–182.

Cheviron, Z., and R. T. Brumfield. 2009. Migration-selection balance and local adapation of mitochondrial haplotypes in rofous-collared sparrows (Zonotrichia capensis) along an elevational gradient. Evolution 63:1593–1605.

DeGiorgio, M., M. Jakobsson, and N. A. Rosenberg. 2009. Explaining worldwide patterns of human genetic variation using a coalescent-based serial founder model of migration outward from Africa. Proc. Natl. Acad. Sci. USA 106:16057–16062.

Deshpande, O., S. Batzoglou, M. W. Feldman, and L. Luca Cavalli-Sforza. 2009. A serial founder effect model for human settlement out of Africa. Proc. R. Soc. Lond. B Biol. Sci. 276:291–300.

Dobzhansky, T. 1937. Genetics and the origin of species. Vol. 11, 364 pp. Columbia Univ. Press, New York, NY.

Dray, S., and A. B. Bufour. 2007. The ade4 package: implementing the duality diagram for ecologists. J. Stat. Softw. 22:1–20.

Drummond, A. J., M. A. Suchard, D. Xie, and A. Rambaut. 2012. Bayesian phylogenetics with BEAUti and the BEAST 1.7. Mol. Biol. Evol. 29:1969–1973.

Edmonds, C. A., A. S. Lillie, and L. L. Cavalli-Sforza. 2004. Mutations arising in the wave front of an expanding population. Proc. Natl. Acad. Sci. USA. 101:975–979.

Elith, J., M. Kearney, and S. Phillips. 2010. The art of modelling range-shifting species. Methods Ecol. Evol. 1:330–342.

Endler, J. A. 1977. Geographic variation, speciation and clines. 262 pp. Princeton Univ. Press, Princeton, NJ.

Excoffier, L., and H. E. L. Lischer. 2010. Arlequin suite ver 3.5: a new series of programs to perform population genetics analyses under Linux and Windows. Mol. Ecol. Res. 10:564–567.

Excoffier, L., and N. Ray. 2008. Surfing during population expansions promotes genetic revolutions and structuration. Trends Ecol. Evol. 23:347–351.

Excoffier, L., M. Foll, and R. J. Petit. 2009. Genetic consequences of range expansions. Annu. Rev. Ecol. Evol. Syst. 40:481–501.

Goodhart, C. B. 1963. "Area effects" and non-adaptive variation between populations of Cepaea. Heredity 18:459–465.

Gould, S. J., and D. S. Woodruff. 1996. History as a cause of area effects: an illustration from Cerion on Great Inagua, Bahamas. Biol. J. Linn. Soc. 40:67–98.

Graciá, E., F. Botella, J. D. Anadón, P. Edelaar, D. J. Harris, and A. Giménez. 2013. Surfing in tortoises? Empirical signs of genetic structuring owing to range expansion. Biol. Lett. 9:20121091.

Hallatschek, O. 2011. The noisy edge of traveling waves. Proc. Natl. Acad. Sci. USA 108:1783–1787.

Hallatschek, O., and D. R. Nelson. 2010. Life at the front of an expanding population. Evolution 64:193–206.

Hallatschek, O., P. Hersen, S. Ramanathan, and D. R. Nelson. 2007. Genetic drift at expanding frontiers promotes gene segregation. Proc. Natl. Acad. Sci. USA 104:19926–19930.

Handley, L. J. L., A. Manica, J. Goudet, and F. Balloux. 2007. Going the distance: human population genetics in a clinal world. Trends Genet. 23:432–439.

Hartl, D. L., and A. G. Clark. 1997. Principles of population genetics. Sinauer Associates, Sunderland.

Hasegawa, M., H. Kishino, and T. Yano. 1985. Dating of the human-ape splitting by a molecular clock of mitochondrial DNA. J. Mol. Evol. 22:160–174.

Heled, J., and A. J. Drummond. 2008. Bayesian inference of population size history from multiple loci. BMC Evol. Biol. 8:289.

Hey, J., and R. Nielson. 2007. Integration within the Felsenstein equation for improved Marchov chain Monte Carlo methods in population genetics. Proc. Natl. Acad. Sci. USA 104:2785–2790.

Hijmans, R., L. Guarino, M. Cruz, and E. Rojas. 2001. Computer tools for spatial analysis of plant genetic resources data: 1. DIVA-GIS. Plant Genet. Resour. Newsl. 127:15–19.

Hill, G. E., and J. D. Johnson. 2013. The mitonuclear compatability hypothesis of sexual selection. Proc. R. Soc. Lond. B Biol. Sci. 280:20131314.

Holman, J. A. 2000. Fossil snakes of North America: origin, evolution, distribution, paleoecology. Indiana University Press, Bloomington, IN.

Irwin, D. E., J. H. Irwin, and T. D. Price. 2001. Ring species as bridges between microevolution and speciation. Genetica 112–113:223–243.

Jezkova, T., B. R. Riddle, D. C. Card, D. R. Schield, M. E. Eckstut, and T. A. Castoe. 2015. Genetic consequences of postglacial range expansion in two rodents (genus Dipodomys) depend on ecology and genetic locus. Mol. Ecol. 24:83–97.

Jombart, T. 2008. adegenet: a R package for the multivariate analysis of genetic markers. Bioinformatics 24:1403–1405.

Kawecki, T. J., and D. Ebert. 2004. Conceptual issues in local adaptation. Ecol. Lett. 7:1225–1241.

Keogh, J. S., J. K. Webb, and R. Shine. 2007. Spatial genetic analysis and long-term mark-recapture data demonstrate male-biased dispersal in a snake. Biol. Lett. 3:33–35.

Klopfstein, S., M. Currat, and L. Excoffier. 2006. The fate of mutations surfing on the wave of a range expansion. Mol. Biol. Evol. 23:482–490.

Lanfaer, R., B. Calcott, S. Y. W. Ho, and S. Guindon. 2012. Partitionfinder: combined selection of partitioning schemes and substitution models for phylogenetic analyses. Mol. Biol. Evol. 29:1695–1701.

Larsen, P. A., M. R. Marchán-Rivadeneira, and R. J. Baker. 2010. Natural hybridization generates mammalian lineage with species characteristics. Proc. Natl. Acad. Sci. USA 107:11447–11452.

Lavin-Murcio, P. A., and J. R. Dixon. 2004. A new species of coral snake (Serpentes, Elapidae) from the Sierra de Tamaulipas, Mexico. Phyllomedusa 3:3–8.

Legendre, P., and L. Legendre. 1998. Numerical ecology. 2nd ed. (English). Elsevier, Amsterdam.

Lenormand, T. 2002. Gene flow and the limits to natural selection. Trends Ecol. Evol. 117:183–189.

Li, J. Z., D. M. Absher, H. Tang, A. M. Southwick, A. M. Casto, S. Ramachandran, H. M. Cann, G. S. Barsh, M. Feldman, L. L. Cavalli-Sforza, et al. 2008. Worldwide human relationships inferred from genome-wide patterns of variation. Science 319:1100–1104.

Lundelius, E. L. 2003. A history of paleontological investigations of quaternary cave deposits on the Edawrds Plateau, central Texas. Pp. 201-214. In B. W. Schubert, J. I. Mead, R. W. Graham, eds. Ice age cave faunas of North America. Indiana University Press, Bloomington, IN.

Mantel, N. 1967. The detection of disease clustering and a generalized regression approach. Cancer Res. 27:209–220.

Mayr, E. 1942. Systematics and the origin of species. Columbia Univ. Press, New York.

———. 1970. Populations, species, and evolution. Belknap Press of Harvard Univ. Press, Cambridge.

McDonald, J. H., and M. Kreitman. 1991. Adaptive protein evolution at the ADH locus in Drosophila. Nature 351:652–654.

Moran, P. A. P. 1948. The interpretation of statistical maps. J. R. Stat. Soc. Ser. B 10:243–251.

———. 1950. Notes on continuous stochastic phenomena. Biometrika 37:17–23.

Nylander, J. A. A., J. C. Wilgenbusch, D. L. Warren, and D. L. Swofford. 2008. AWTY (are we there yet?): a system for graphical exploration of MCMC convergence in Bayesian phylogenetics. Bioinformatics 24:581–583.

Oksanen, J., G. Blanchet, R. Kindt, P. Legendre, P. R. Minchin, R. B. O' Hara, G. L. Simpson, P. Solymos, M. Henry, H. Stevens et al. 2012. vegan: community ecology pacakge. R-package version 2.0-5.

Pavlova, A., J. Nevil Amos, L. Joseph, K. Loynes, J. J. Austin, J. S. Keogh, G. N. Stone, J. A. Nicholls, and P. Sunnucks. 2013. Perched at the mito-nuclear crossroads: divergent mitochondrial lineages correlate with environment in the face of ongoing nuclear gene flow in an Australian bird. Evolution 67:3412–3428.

Peakall, R., and P. E. Smouse. 2006. GenAlEx 6: genetic analysis in Excel. Population genetic software for teaching and research. Mol. Ecol. Resour. 6:288–295.

Peter, B. M., and M. Slatkin. 2013. Detecting range expansions from genetic data. Evolution 67:3274–3289.

Peters, J. L., K. Winker, K. C. Millam, P. Lavretsky, I. Kulikova, R. E. Wilson, Y. N. Zhuravlev, and K. G. McCracken. 2014. Mito-nuclear discord in six congeneric lineages of Holarctic duck (genus *Anas*). Mol. Ecol. 23: 2961–2974.

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

Petit, R. J, E. Pineau, B. Demesure, R. Bacilieri, A. Ducousso, and A. Kremer. 1997. Chloroplast DNA footprints of postglacial recolonization by oaks. Proc. Natl. Acad. Sci. USA 94:9996–10001.

Pfeifer, B., U. Wittelsbuerger, S. E. Ramos-Onsins, and M. J. Lercher. 2014. PopGenome: an efficient Swiss army knife for population genomic analysis in R. Mol. Biol. Evol. 31:1929–1936.

Phillips, S. J., R. P. Anderson, and R. E. Schapire. 2006. Maximum entropy modeling of species geographic distributions. Ecol. Model. 190:231–259.

Pierce, A. A., M. P. Zalucki, M. Bangura, M. Udawatta, M. R. Kronforst, S. Altizer, J. F. Haeger, and J. C. de Roode. 2014. Serial founder effects and genetic differntiation during worldwide range expansion of monarch butterflies. Proc. R. Soc. Lond. B Biol. Sci. 281:20142230.

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

Prugnolle, F., A. Manica, and F. Balloux. 2005. Geography predicts neutral genetic diversity of human populations. Curr. Biol. 15:R159–R160.

R Development Core Team. 2014. R: a language and environment for statistical computing. R Foundation for Statistical Computing, Vienna, Austria.

Ramachandran, S., O. Deshpande, C. C. Roseman, N. A. Rosenberg, M. W. Feldman, and L. L. Cavalli-Sforza. 2005. Support from the relationship of genetic and geographic distance in human populations for a

serial founder effect originating in Africa. Proc. Natl. Acad. Sci. USA 102:15942–15947.

Rambaut, A., M. A. Suchard, D. Xie, and A. J. Drummond. 2014. Tracer v1.6. Available at http://beast.bio.ed.ac.uk/Tracer.

Ronquist, F., and J. P. Huelsenbeck. 2003. MrBayes 3: Bayesian phylogenetic inference under mixed models. Bioinformatics 19:1572–1574.

Rozas, J., J. C. Sánchez-DelBarrio, X. Messeguer, and R. Rozas. 2003. DNAsp, DNA polymorphism analyses by the coalescent and other methods. Bioinformatics 19:2496–2497.

Roze, J. A. 1996. Coral snakes of the Americas: biology, identification, and venoms. 340 pp. Krieger Publishing, Malabar, FL.

Shine, R., G. P. Brown, and B. L. Phillips. 2011. An evolutionary process that assembles phenotypes through space rather than through time. Proc. Natl. Acad. Sci. USA 108:5708–5711.

Slatkin, M. 1973. Gene flow and selection in a cline. Genetics 75:733–756.

———. 1985. Gene flow in natural populations. Annu. Rev. Ecol. Syst. 16:393–430.

Slatkin, M., and L. Excoffier. 2012. Serial founder effects during range expansion: a spatial analog of genetic drift. Genetics 191:171–181.

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

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

Tamura, K., D. Peterson, N. Peterson, G. Stecher, M. Nei, and S. Kumar. 2011. MEGA5: molecular evolutionary genetics analysis using maximum likelihood, evolutionary distance, and maximum parsimony methods. Mol. Biol. Evol. 28:2731–2739.

Toews, D. P. L., and A. Brelsford. 2012. The biogeography of mitochondrial and nuclear discordance in animals. Mol. Ecol. 21:3907–3930.

Travis, J. M. J., T. Münkemüller, O. J. Burton, A. Best, C. Dytham, and K. Johst. 2007. Deleterious mutations can surf to high densities on the wave front of an expanding population. Mol. Biol. Evol. 24:2334–2343.

Williams, G. C. 1966. Adaptation and natural selection. 328 pp. Princeton Univ. Press, Princeton, NJ.

Associate Editor: M. Alfaro
Handling Editor: R. Shaw

Supporting Information

Additional Supporting Information may be found in the online version of this article at the publisher's website:

Appendix S1. Systematics of the *Micrurus fulvius* complex and taxonomic revision of Micrurus tener.
Appendix S2. Additional information regarding laboratory and analytical methods performed in this study.
Table S1. Taxonomic and geographic sampling of the *Micrurus fulvius* complex.
Table S2. Best nucleotide model scheme identified by Partition Finder.
Table S3. Number of RAD tags (unique restriction digestion loci) obtained from Illumina PE100 sequencing of the *Micrurus fulvius* complex allowing for different levels of percent missing data.
Table S4. Number of nuclear genome-wide SNPs obtained from the *Micrurus fulvius* complex.
Table S5. Correlation coefficients (Spearman's rho) for comparisons between latitude, longitude, and spatial PCs from *Micrurus tener*.
Figure S1. Relationships between percent missing data and PC scores from the most explanatory axis of the multivariate nuclear DNA SNP analysis.
Figure S2. Bayesian phylograms of cytochrome-b (cyt-b) and NADH dehydrogenase subunit 4 (ND4) generated using MrBayes.
Figure S3. Results of k-means clustering in adegenet (Jombart 2008) using a maximum of 40 clusters (30 retained for figure) on the 22,547 biallelic nuclear SNP dataset from 45 individuals of the Micrurus fulvius complex.
Figure S4. STRUCTURE analysis ($K = 3$) of Micrurus tener.
Figure S5. Mismatch distributions of three mitochondrial haplogroups observed in the *Micrurus fulvius* complex (solid lines).
Figure S6. Spearman's rho correlation coefficients for comparisons between latitude, longitude, and nuclear DNA spatial PCs in *Micrurus tener*.
Figure S7. Lagged scores from the three most explanatory nuclear DNA (nucDNA) spatial principal components (SPCs) in *Micrurus tener* (A–C).
Figure S8. Alternative species distribution models for last glacial maximum predictions using the MIROC-ESM and MPI-ESM-P models.