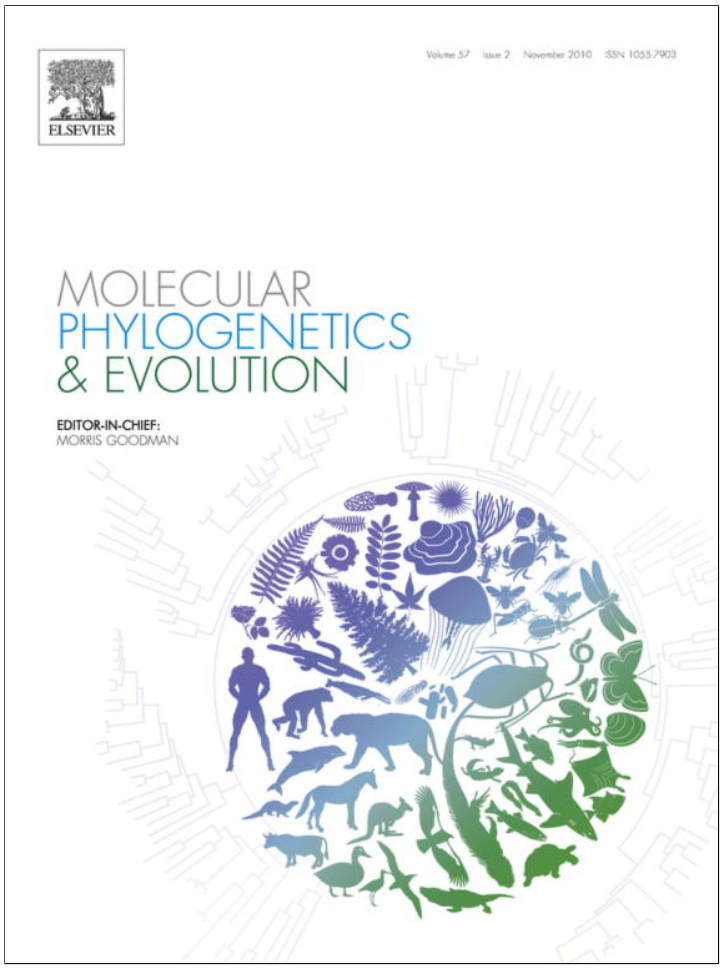

Molecular Phylogenetics and Evolution 57 (2010) 552–560

Contents lists available at ScienceDirect

Molecular Phylogenetics and Evolution

journal homepage: www.elsevier.com/locate/ympev

# Does dispersal across an aquatic geographic barrier obscure phylogeographic structure in the diamond-backed watersnake (*Nerodia rhombifer*)?

Matthew C. Brandley [a,*,1], Tim J. Guiher [b,c], R. Alexander Pyron [b,c,2], Christopher T. Winne [d], Frank T. Burbrink [b,c,1]

[a] University of California, Museum of Vertebrate Zoology, Department of Integrative Biology, Berkeley, CA, USA
[b] College of Staten Island, CUNY, Biology Department, Staten Island, NY, USA
[c] City University of New York, The Graduate School, University Center, New York, NY, USA
[d] University of Georgia, Savannah River Ecology Laboratory, Aiken, SC, USA

ARTICLE INFO

*Article history:*
Received 6 December 2009
Revised 22 July 2010
Accepted 22 July 2010
Available online 3 August 2010

*Keywords:*
Coalescence
Mississippi River
Molecular divergence dating
Natricinae
Nerodia
Phylogeography

ABSTRACT

The impact of barriers to dispersal and gene flow is often inferred to be the primary cause of lineage divergence and phylogeographic structure in terrestrial organisms. In particular, the Mississippi River has been implicated as a barrier to gene flow in many species, including aquatic taxa. However, if barriers are permeable to organisms, then phylogeographic structure may be difficult to detect due to gene flow between lineages. Using time-calibrated Bayesian phylogenetic analyses of mtDNA, and phylogeographic coalescent simulations, we determine if the Mississippi River operates as a barrier to gene flow in the aquatic diamond-backed watersnake (*Nerodia rhombifer*). The phylogenetic analyses support a basal division within *N. rhombifer* mtDNA lineages that coincides with populations generally east and west of the Mississippi River. These results, and that of the divergence dating analyses, therefore suggest that the river was a significant barrier to gene flow in the Pleistocene ~1.4 million years ago, presumably during an interglacial period when the river was much wider. However, we also detect western haplotypes in the eastern clade, and vice versa, thereby indicating that this barrier has not been complete. Nonetheless, the coalescent simulations that account for limited migration suggest that the Mississippi River was an important feature that shaped the phylogeographic history of this aquatic snake in the USA despite limited gene flow.



## 1. Introduction

The strength of phylogeographic barriers in shaping the genetic structure of populations may be due, in part, to the ability of organisms to cross these impediments to dispersal, and certain physical barriers present a dispersal challenge to a wide variety of organisms, regardless of lability. For example, numerous studies have implicated the Mississippi River as one of the most important phylogeographic barriers for both terrestrial and aquatic North American vertebrates (Avise, 2000; Blair, 1958; Brant and Ortí, 2003; Burbrink et al., 2000; Leaché and Reeder, 2002; Near et al., 2001; Soltis et al., 2006; Burbrink and Castoe, 2009; Pyron and Burbrink, in press). The Mississippi River is one of the oldest and largest drainage basins in the world. Originating in the Cretaceous, it forms a large, wide valley referred to as the Mississippi Embayment that extends from southern Illinois to the Gulf of Mexico (Cushing et al., 1964). The embayment represents a physical, hard allopatric barrier limiting dispersal in many terrestrial populations (e.g., Burbrink et al., 2000, 2008; Howes et al., 2006; Soltis et al., 2006; Pyron and Burbrink, 2009, 2010), but it is unclear whether it is consistently a strong barrier in aquatic or semi-aquatic taxa. Furthermore, the Mississippi River is presumably a stronger barrier to terrestrial populations during interglacial periods when it receives massive volumes of meltwater from melting glaciers. Two recent studies of semi-aquatic snakes, the water moccasin (*Agkistrodon piscivorus*) and the plain-bellied watersnake (*Nerodia erythrogaster*) produced no evidence for lineage divergence at the Mississippi River (Guiher and Burbrink, 2008; Makowsky et al., 2010). This lack of evidence implies that extensive gene flow across the Mississippi River embayment occurred during the Pleistocene or that there has been extensive recent colonization. Therefore, the available evidence suggests that large river barriers may impede dispersal of terrestrial vertebrates to a far greater extent than it does semi-aquatic species.

* Corresponding author. Address: University of Sydney, School of Biological Sciences, Heydon-Laurence Building (A08), NSW 2006, Australia. Fax: +61 2 9351 4119.
*E-mail address:* mbrandley@gmail.com (M.C. Brandley).
[1] These authors contributed equally.
[2] Present address: Stony Brook University, Ecology and Evolution Department, Stony Brook, NY, USA.

1055-7903/$ - see front matter © 2010 Elsevier Inc. All rights reserved.
doi:10.1016/j.ympev.2010.07.015

One example of such a species may be the diamond-backed watersnake, *Nerodia rhombifer*. This snake occurs throughout the south-central and southeastern USA and inhabits permanent bodies of water, including rivers and lakes (Gibbons and Dorcas, 2004; Fig. 1). *Nerodia rhombifer* is one of the most aquatic North American watersnakes, spending a vast majority of its time in or near freshwater environments and consuming fish almost exclusively (97% of its diet; Gibbons and Dorcas, 2004; Vincent et al., 2009).

We hypothesize that, given the aquatic lifestyle of this species, the Mississippi River may not serve as a strong barrier to gene flow currently or historically as it has with other, more terrestrial vertebrates. We use a time-calibrated phylogenetic analysis of mitochondrial DNA to investigate the phylogeographic structure of *Nerodia rhombifer*. We then used coalescent statistics that account for divergence given population size, generation time, and limited migration to test the hypothesis that the Mississippi River promoted population genetic structure of this aquatic snake. Combining these methods provides much a clearer picture of the dynamics of aquatic barriers for aquatic organisms throughout time. We conclude that the Mississippi River has not completely precluded recent gene flow, but were nonetheless an important barrier in this aquatic snake during the Pleistocene, presumably during periods when the river was much wider than at present.

## 2. Methods and materials

### 2.1. DNA sequencing

We obtained tissue samples from 388 individual *Nerodia rhombifer rhombifer* collected throughout their current range in the USA, and from one individual from Mexico (*N. r. werleri*; Fig. 1; Appendix A). Representatives of the other Mexican subspe-

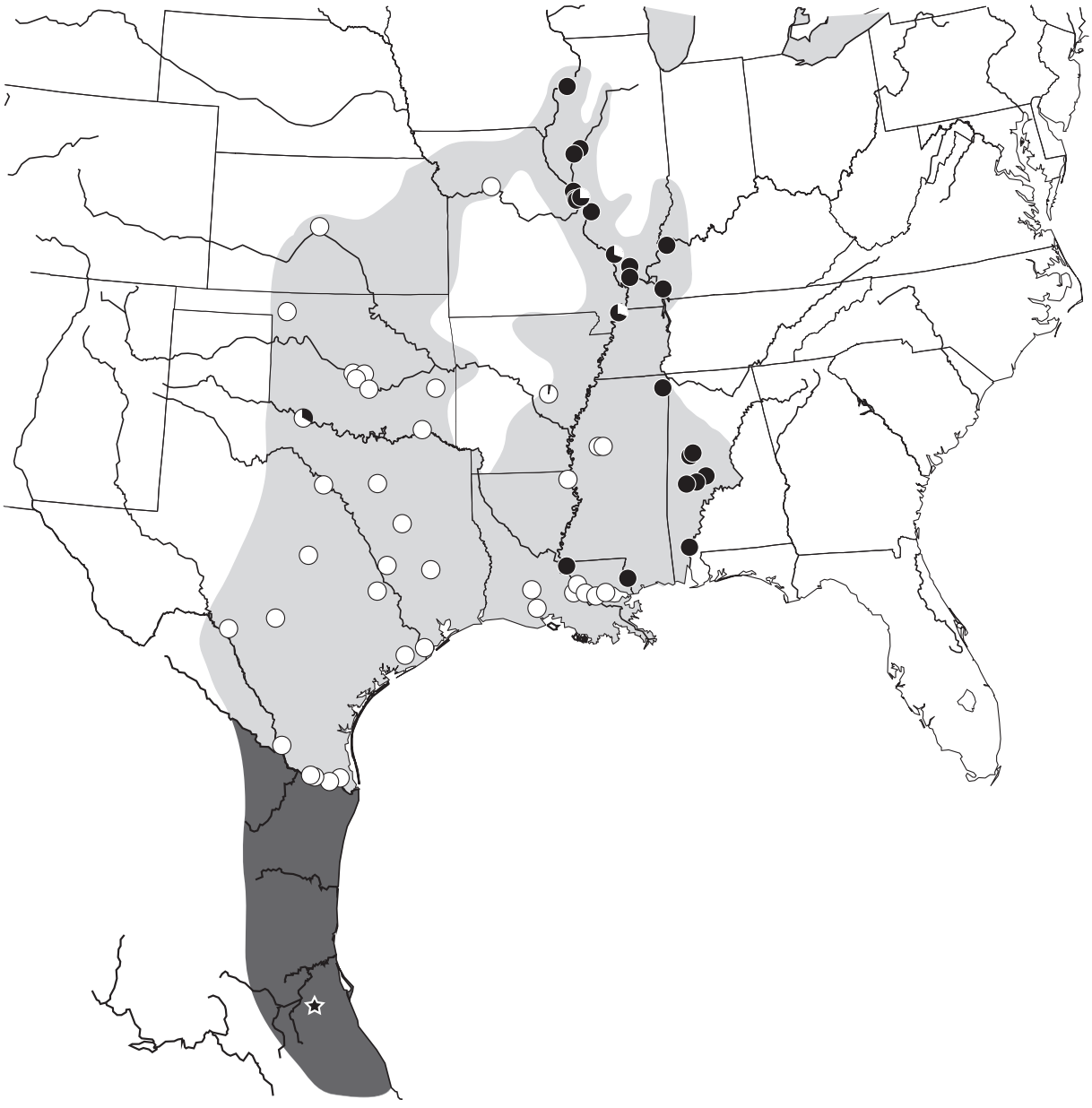

**Fig. 1.** Sampling localities of *Nerodia rhombifer*. Light grey shading represents the geographic range of *N. r. rhombifer* and dark shading represents the range of *N. r. blanchardi* and *N. r. werleri*. Open circles represent West clade haplotypes, and black circles, East clade haplotypes. The sampled population of *N. r. werleri* is indicated by the black star. Populations with mixed haplotypes are represented by pie charts indicating the relative proportion of each haplotype.

cies, *N. r. blanchardi*, were unavailable. Our sampling reflects broad geographic breadth, and in several cases, dense intrapopulational sampling (>15 individuals per population; Appendix A). We sampled several closely related species to use as outgroups in our phylogenetic analyses, including *N. erythrogaster* ($n = 6$), *N. fasciata* ($n = 3$), *N. sipedon* ($n = 3$), *N. taxispilota* ($n = 3$), and *Regina grahami* ($n = 1$) to root the tree (Lawson, 1987; Alfaro and Arnold, 2001; de Queiroz et al., 2002).

We isolated DNA from tissue using Qiagen DNeasy™ columns. The mitochondrial NADH dehydrogenase 1 (*ND1*) gene has been shown to be a useful genetic marker for elucidating population genetic structure in snakes given its relatively rapid rate of evolution (Wood et al., 2008). Thus, we amplified an 886 base pair segment of *ND1* using the following primers: ND1F-Colubrid (5′-AAA CCW AGA TAA GGT TAA TTA AGG AC-3′), ND1R-Nerodia (5′-CAT AGT GCG GAT GTA GAG GAA AT-3′). We used standard PCR techniques (95 °C for 2 min followed by 40 cycles of 95 °C for 30 s, 50 °C for 30 s and 72 °C for 60 s) and cleaned PCR products using ExoSap-it (USB Corp.). Purified templates were dye-labeled using BigDye™ (ABI) and sequenced on an ABI 377™ or 3077™ automated DNA sequencer. Samples were sequenced in both directions. In some cases, samples were re-sequenced if the phylogenetic position of the sample was anomalous (e.g., the Jersey County, Illinois sample). Although not every singleton was re-sequenced, we did re-inspect the chromatograms and verified that different haplotypes were not the result of errors in interpreting the chromatograms. Nucleotide sequences were examined and aligned by eye, and an open reading frame was verified using MacClade v4.08 (Maddison and Maddison, 2005). No indels were present in any individuals. Sequences were accessioned into GenBank (HQ121593–HQ121995).

We acknowledge that, although we have sampled extensively throughout the North American range of *N. rhombifer* and likely captured much of the mtDNA diversity within this species, our phylogeographic and coalescent simulation results (see below) are nonetheless dependent on a single locus (see Beerli and Felsenstein (1999, 2001) for review). Surveys of nuclear loci among New World natricines have yet to reveal any loci suitable for phylogeographic analysis of *Nerodia* (Brandley, unpublished data).

### 2.2. Phylogenetic and molecular dating analyses

To reconstruct the phylogenetic relationships among extant haplotypes and to estimate an explicit temporal scale for our coalescent simulations (below), we employed a partitioned Bayesian phylogenetic analysis and relaxed-clock dating using BEAST v1.4.8 (Drummond et al., 2006; Drummond and Rambaut, 2007). We used Bayesian relaxed-clock phylogenetic analyses because they enabled us to estimate divergence times while simultaneously incorporating rate heterogeneity and uncertainty in the substitution parameters, tree topology, and calibration ages (Drummond et al., 2006).

Dating molecular divergences requires *a priori* assumptions of the ages of one or more clades to calibrate the relationship between age and molecular rate (Drummond et al., 2006). This is most commonly achieved by using fossil calibrations as age constraints. However, directly using fossil calibration is not possible within *Nerodia rhombifer* because the morphology of the known fossils of *N. rhombifer* does not permit us to determine their phylogenetic placement within this species. Therefore, it is necessary to expand the dataset to include most New World natricines (the subfamily that includes *N. rhombifer*) to use older fossils that can be assigned to clades outside of *N. rhombifer*. For this purpose, we used the existing extensive natricine *cytochrome-b* (*cyt-b*) mtDNA data from Genbank (Appendix B). Additionally, we sequenced *cyt-b* for six populations that span the root of the *N. r. rhombifer* phylogeny using primers from Burbrink et al. (2000) and Lawson et al. (2005) in order to combine and date these clades with the larger natricine dataset. This strategy takes advantage of the extensive natricine snake *cyt-b* data set to infer the ages of the deep divergences in *N. rhombifer* and closely related species, and then we subsequently use this information for coalescence simulations and secondary age constraints for BEAST analyses of our more extensive *N. rhombifer* population *ND1* data set.

For the analyses of the *cyt-b* dataset in BEAST, we used the earliest New World natricine fossil, *Neonatrix elongata*, from the Hemingfordian in South Dakota (Holman, 1973), to calibrate the divergence between the New World (NW) natricines and their closest living Old World (OW) relative (*Amphiesma sauteri*). For the divergence age priors constraints, we employed a lognormal distribution (Drummond et al., 2006; Ho, 2007) with an offset of 15.5 Ma (the earliest date for the Hemingfordian; Ma = million years ago) and standard deviation so that 95% of the distribution falls within the Hemingfordian (15.5–19 Ma), with 5% sampling older ages (SD = 0.762). Our second fossil constraint is the earliest known specimen of the monophyletic genus *Thamnophis*, originating in the late Barstovian of Nebraska (Holman, 2000). The basal relationships within NW natricines are ambiguous (Alfaro and Arnold, 2001), and thus, the closest living NW relative to *Thamnophis* is unknown. However, *Thamnophis* is certainly one of the earliest divergences within NW natricines (Alfaro and Arnold, 2001). We therefore used this fossil to date the root of the NW natricine tree using a lognormal distribution with an offset of 11.8 Ma (the earliest date for the Barstovian) and a standard deviation so that 95% of the distribution falls within the Barstovian (11.8–15.5 Ma), with 5% sampling older ages (SD = 0.796).

We used data partition-specific evolutionary models to improve phylogenetic inference by accounting for the heterogenous characteristics of sequence evolution among different codon positions (Castoe et al., 2004; Brandley et al., 2005; Nylander et al., 2004). The most appropriate evolutionary model for each partition was determined using the hierarchical likelihood-ratio test implemented by MrModeltest (Nylander, 2004) using an initial neighbor-joining tree calculated from Jukes-Cantor-corrected genetic distances calculated by PAUP* 4b10 (Swofford, 2003). BEAST analyses consisted of a single MCMC chain (current versions of BEAST do not allow Metropolis-coupled MCMC) run for $1 \times 10^7$ generations, sampled every 1000 generations, using a starting tree estimated under the coalescent process, a Yule process tree prior, and uncorrelated lognormal distribution of rates. We used the program's default prior distributions of model parameters with the exception of GTR substitution rates in which we used a uniform (0,100) distribution, and the date distributions of the most recent common ancestor of the two clades used for calibration (see above). To determine convergence, cumulative posterior probability plots for each clade were constructed for four separate BEAST analyses using the *cumulative* function in AWTY (Nylander et al., 2008). Stationarity was assumed when the cumulative posterior probabilities of all clades stabilized. These plots indicated that excluding the first 2 million generations as burn-in was sufficient, and the frequency of inferred relationships in the remaining trees represent estimated clade posterior probabilities. To decrease the chance of reaching apparent stationarity on local optima, four separate analyses were performed. Posterior probability estimates for each clade were then compared among all analyses using a scatterplot created by the *compare* command in AWTY. If posterior probability estimates for clades were similar in all analyses, the results were combined. Posterior probabilities $\geqslant 0.95$ are considered statistically significant clade support (Huelsenbeck and Rannala, 2004).

We used the resulting age distributions of the *cyt-b* analyses from BEAST as log-normally distributed secondary calibration constraints (Table 1) for additional partitioned BEAST analyses of the

**Table 1**
Calibration age constraints used in the Bayesian relaxed-clock phylogenetic analyses of the *ND1* mtDNA data set. These are secondary constraints calculated from Bayesian relaxed-clock phylogenetic analyses of a more extensive *cytochrome-b* data set of natricine snakes employing two fossil age constraints (see text).

| Calibrated clade | Ages of secondary calibrations | | | Parameter values used in BEAST | |
|---|---|---|---|---|---|
| | Mean age (Ma) | Lower 95% HPD (Ma) | Upper 95% HPD (Ma) | Ln mean age (Ma) | SD |
| *Nerodia* | 7.10 | 5.47 | 8.84 | 1.96 | 0.133 |
| *N. fasciata* + *N. sipedon* | 5.36 | 3.63 | 7.11 | 1.68 | 0.200 |
| *N. rhombifer* + *N. taxispilota* | 3.90 | 2.54 | 5.37 | 1.36 | 0.218 |
| US *N. rhombifer* | 0.86 | 0.44 | 1.34 | −0.15 | 0.342 |

unique *ND1* unique haplotypes. The BEAST analytical conditions were the same as those used in the previous natricine *cyt-b* analyses. We acknowledge that the use of secondary calibrations has received deserved criticism when used as errorless point calibrations (see Graur and Martin, 2004). However, unlike previous divergence dating methods, Bayesian age estimation permits the explicit incorporation of this error by permitting age calibration constraints (rather than point estimates) in the form of statistical distributions, thus eliminating at least one negative feature of secondary calibrations.

### 2.3. Coalescent simulations

Our phylogeographic analyses revealed a deep division in *Nerodia rhombifer* between populations that mostly reside on the east and west of the Mississippi River (see Section 3; Figs. 2 and 3), but there exist several examples of "east" haplotypes west of the river and vice versa, implying at least limited recent migration or failure for lineages to sort. To determine if the Mississippi River could serve as a strong barrier to gene flow to affect population genetic structure despite a limited amount of migration, we performed simulations using a method that accounts for uncertainty in the coalescent by including estimates of population size and generation times to the most recent common ancestor (MRCA; see Carstens and Knowles, 2007; Knowles and Maddison, 2002; Steele and Storfer, 2006; Shepard and Burbrink, 2008, 2009). These simulations required a single "best" hypothesis of *N. rhombifer* intraspecific relationships and branch lengths (as opposed to consensus summary of the posterior distribution). We therefore conducted an additional maximum likelihood (ML) analysis of the entire *ND1* dataset with outgroups ($n = 404$) using the GTR + Γ model with the program RAxML v.7.0.4 (Stamatakis, 2006). We further augmented our coalescent simulations to incorporate the probability of migration along with isolation and divergence to reflect the likelihood of dispersal across the Mississippi River by this aquatic snake. Using Mesquite v2.5 (Maddison and Maddison, 2008), we simulated 1000 trees of 388 taxa (equal to our entire *ND1* dataset for USA *N. rhombifer* populations). These trees correspond to the same sample size and localities as our real dataset and were simulated on the ML tree reflecting strict east and west divergence at the Mississippi River with branch lengths equivalent to the number of generations prior to coalescence and branch widths equivalent to the effective population size ($N_e$) for each lineage. Since most of the discordance between individuals was found on the border of the Mississippi river, and the discordance appears at the tips of the trees, we compared the number of migration events using the *S* statistic developed by Slatkin and Maddison (1989) in our *ND1* phylogeny to the number of migration events in our simulations (Knowles and Maddison, 2002). We would conclude that divergence occurred at the Mississippi River in *N. rhombifer* if the number of migration events is the same or smaller in our *ND1* tree when compared to the simulated trees accounting for migration.

Simulating trees for a null model of divergence at the Mississippi River with migration requires estimates of coalescent times (in number of generations), $N_e$ of each lineage, and the individual probability of migration for each individual per generation. Our Bayesian divergence dating analyses indicate that the two lineages diverged ~1.4 Ma (Fig. 2), and it is known that females require 2.2 years to attain sexual maturity (Gibbons and Dorcas, 2004). Thus, the two lineages would have coalesced in 636,000 generations. Assuming a substitution rate/site/generation ($m$) of $1.52 \times 10^{-8}$ estimated from the *ND1* relaxed-clock divergence dating analyses (see above), we used the program Migrate 2.1.3 (Beerli, 2008) to calculate $N_e$, $M$ (the number of migrants/$m$), and $\Theta$ ($N_e * m$). Specifically, we ran four separate Migrate runs (under the Bayesian criteria) for 5 million generations, each with eight chains and four with adaptive heating schemes and each run starting with random seed values. We used uniform priors and slice sampling for both $M$ (0–1000, mean = 100) and $\Theta$ (0.0–0.10, mean = 0.01). The first 10,000 samples were discarded as burn-in. The mean values for $\Theta$ and M of each run were compared to the 95% CI of all other to runs to ensure that they fit within with the interval and were similar to the other mean values. For use in our simulations, we took the mean of $\Theta$ and $M$ for all four runs. With these values, the probability of migration/population/generation was calculated for a single lineage using the equation: $(\Theta * M)/N_e$. The probabilities for each lineage were summed and used as the overall probability of migration for each individual/generation. Additionally, we compared our *ND1* tree to a simulation that did not include migration, using the same parameters for $N_e$ and tree depth in generations described above, to determine the value of *S* due only to the failure to coalesce.

Additionally, we used MDIV to examine models of isolation that include gene flow among the two *Nerodia rhombifer* populations situated east and west of the Mississippi River (Nielsen and Wakeley, 2001). This program uses Metropolis Hastings MCMC method to obtain credible joint estimates for population divergence times ($T = t_{div}/2N_e$), migration rates ($M = 2N_e m$) and $\Theta$ ($2N_e\mu$) for two populations using a sequence data from a single locus. To convert these values into estimates of divergence time and $N_e$, we used a rate of $6.9 \times 10^{-9}$ substitutions/site/year (discussed above) multiplied by the size of gene (886 bp), yielding a value of $6.11 \times 10^{-6}$ substitutions/gene/year. Following the approach in Nielsen and Wakeley (2001) and Spellman et al. (2007), we sampled 50,00,000 generations (with the first 500,000 discarded as burn-in) using the finite site model (HKY) to estimate the posterior density surfaces for all parameters. Initial parameter estimates were obtained using default priors to inform later searches using better priors that produce credible posterior distributions similar to Nielsen and Wakeley (2001). Priors for the final replicate searchers were $T_{max} = 10$, $\Theta = 100$ and $M_{max} = 10$.

## 3. Results

### 3.1. Phylogeography and divergence dating

The estimated rates of evolution for the *cyt-b* and *ND1* genes are similar; the mean and 95% credible interval of the rate for both loci is 0.011 (0.0091–0.013) and 0.0079 (0.0059–0.01) substitutions/

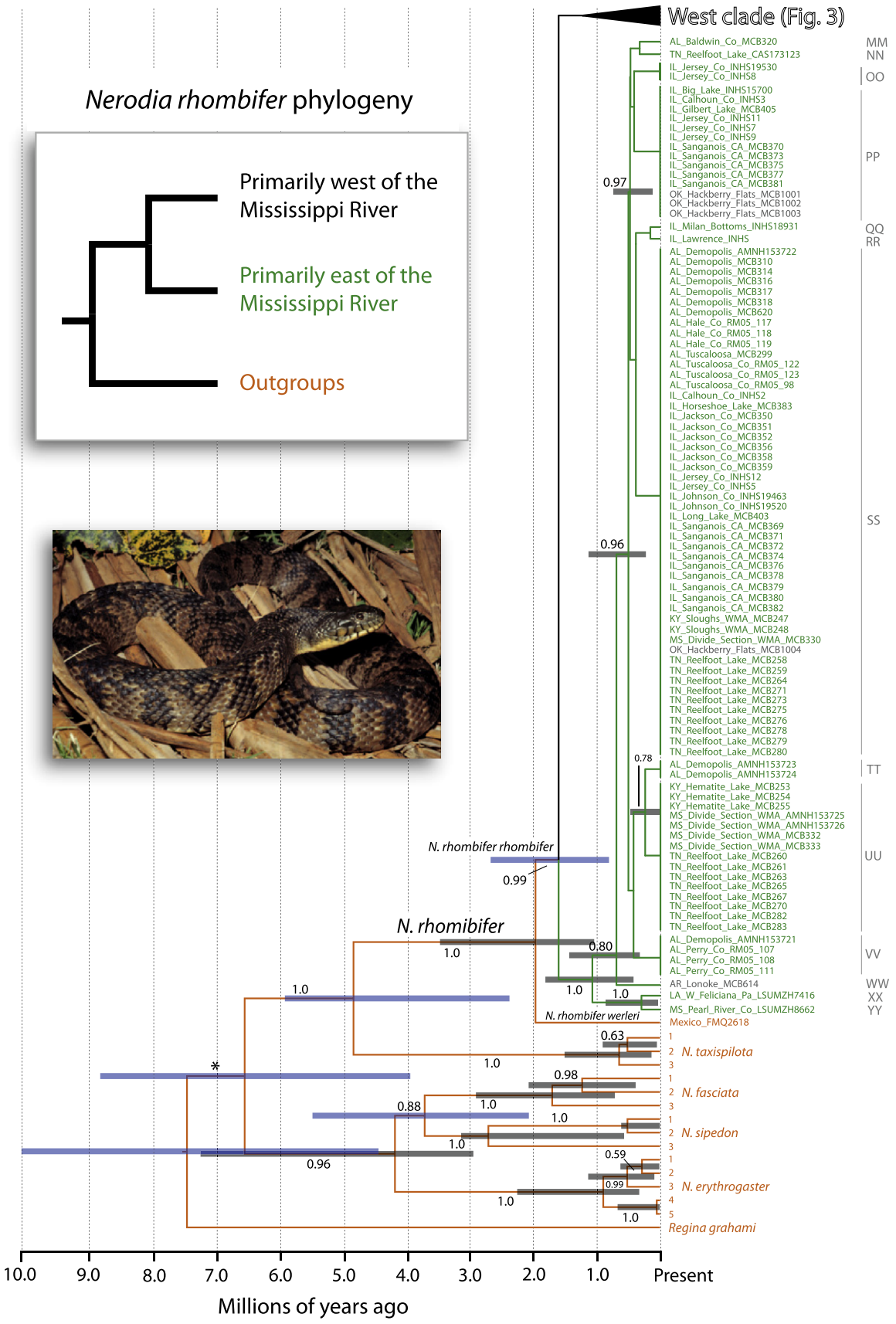


**Fig. 2.** *ND1* mitochondrial DNA phylogenetic relationships of *Nerodia rhombifer* with emphasis on the East clade (in green). West clade haplotypes are indicated in grey. Numbers above or below the nodes are posterior probabilities. Grey bars indicate the 95% credible interval of estimated divergence ages. Blue bars indicate the 95% credible interval of prior age calibrations. The asterisk indicates that this clade was constrained to be monophyletic. Grey letters in the margin indicate the haplotype names.

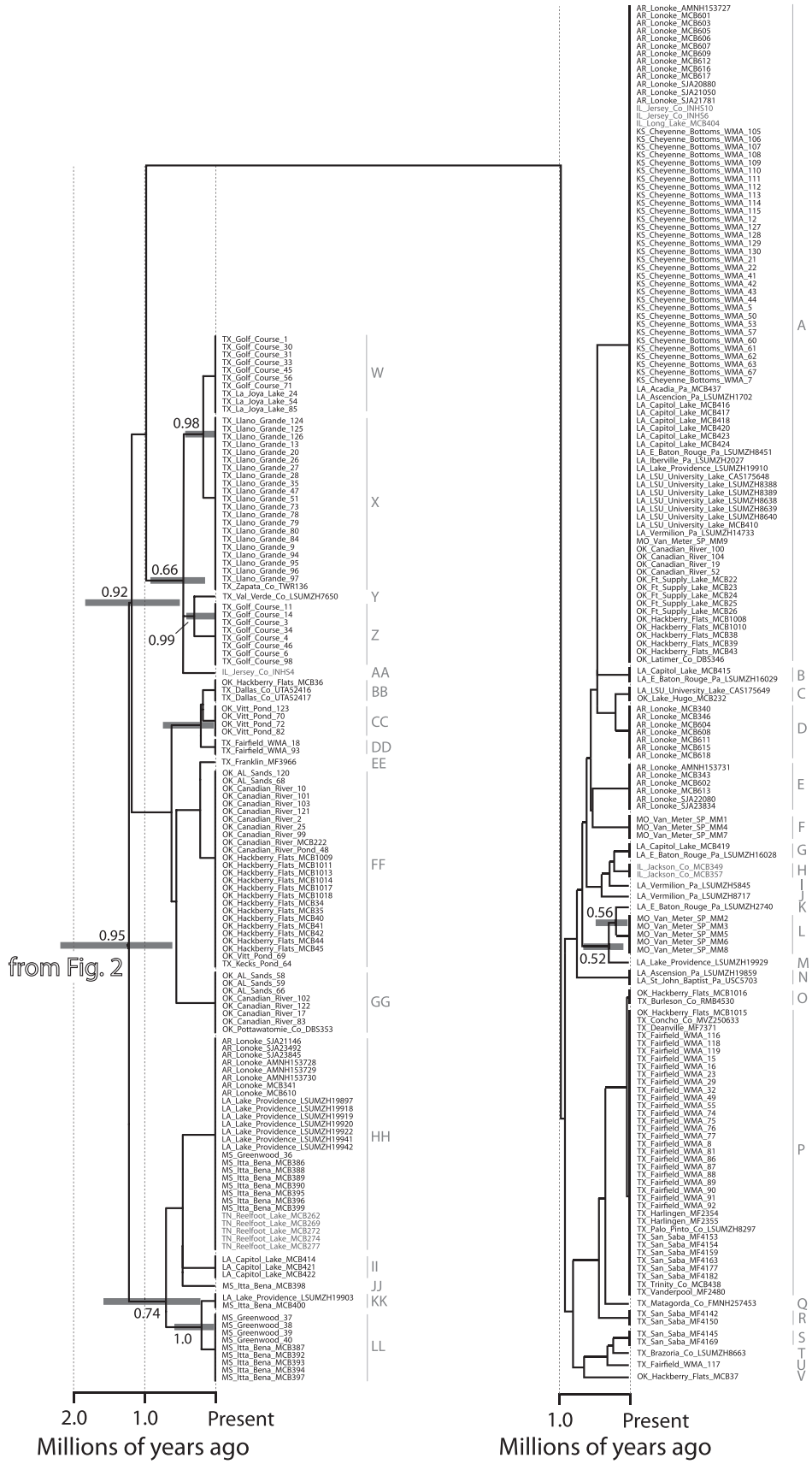


**Fig. 3.** A continuation of Fig. 2. *ND1* mitochondrial DNA phylogenetic relationships of *Nerodia rhombifer* with emphasis on the West clade. East clade haplotypes are indicated in grey. Numbers above or below the nodes are posterior probabilities. Grey bars indicate the 95% credible interval of estimated divergence ages. Grey letters in the margin indicate the haplotype names.

million years, respectively. However, we note that rate heterogeneity among loci should not negatively affect our estimates of divergence time as both analyses use time estimates as common calibration prior constraints. Of all 388 samples from the USA, there were 51 unique haplotypes; of these, 13 were located east of the Mississippi River, 38 to the west. For all USA *N. rhombifer* populations sampled, the average nucleotide diversity ($\pi$) was 0.00673, and the maximum uncorrected % genetic difference ($\pi$ distance) was 0.0181. The likelihood-ratio test indicated that the HKY + Γ, HKY, and GTR + Γ nucleotide evolution models were the most appropriate for the first, second, and third codon positions, respectively. All four BEAST analyses of the *Nerodia rhombifer ND1* haplotypes using these models and secondary calibration age constraints estimated from the natricine *cyt-b* analysis (below) achieved stationarity quickly, and only the first 2 million generations of each analysis were discarded as burn-in. The compare function of AWTY indicated that all four analyses converged on the same posterior distribution. Thus, we pooled the post-burn-in trees to calculate all subsequent clade posterior probabilities and divergence date estimates. The consensus tree of the remaining 32,000 trees from the four combined analyses is shown in Figs 2 and 3. Monophyly of *Nerodia rhombifer* is significantly supported (PP = 1.0), as is the split between the Mexican *N. r. werleri* and the samples of *N. r. rhombifer* from the USA (PP = 0.99). Within the USA, there exist two lineages of *N. rhombifer* roughly oriented east (PP = 1.0) and west (PP = 0.95) of the Mississippi River (Fig. 2). However, the geographic range of these clades overlaps (Figs. 1–3).

Our simultaneous estimation of molecular divergence dates for the *ND1* tree, using secondary calibrations estimated from analyses of *cyt-b* (Table 1) reveals that the split between *N. rhombifer* and *N. taxispilota* occurred in the Pliocene, approximately 4.00 Ma (95% CI = 2.86–5.23 Ma). *Nerodia rhombifer* in the USA diverged from the Mexican subspecies 1.86 Ma (95% CI = 1.08–2.77 Ma), and the tMRCA of the two extant lineages of *N. rhombifer* found within the USA is 1.39 Ma (95% CI = 0.84–1.99 Ma). Our estimated date for the origin of the USA population is similar to the date of the oldest identified fossil of *N. rhombifer* (the late Pliocene, ~1.8 Ma) found in Nebraska (Rogers, 1984). The tMRCA of the West clade is 1.13 Ma (95% CI = 0.64–1.69 Ma) and the tMRCA of the East clade is 0.90 Ma (95% CI = 0.45–1.39).

### 3.2. The Mississippi River and lineage divergence

If the Mississippi River were a barrier to migration between East and West clades of *N. rhombifer*, then the *S* statistic of the ML tree should be less than or equal to that of the coalescent simulations of 1000 trees (Slatkin and Maddison, 1989). Using mean values from the four replicated analyses in Migrate we obtained the following estimates of mode and 95% CI in parentheses: for the East clade we estimated $N_e$ = 590,953 (279,605–912,828). $\Theta$ = 0.009 (0.004–0.0139), and $M_{east-west}$ = 335.4 (48.75–692.5); for the West clade we estimated $N_e$ = 581,907 (279,605–879,934), $\Theta$ = 0.009 (0.004–0.014), and $M_{west-east}$ = 227 (6.25–573), and a tree depth of 636,000 generations. We predicted a total probability of per-individual migrations per generation of $8.44 \times 10^{-6}$ ($7.67 \times 10^{-7}$ – $1.9 \times 10^{-5}$) between East and West clades. Using the mean and upper 95% CI of $N_e$ and *M*, our result is consistent with a divergence and migration scenario. Our ML tree value for *S* was 35 and significantly lower (mean *S* = 65.7; $P = 1.48 \times 10^{-3}$) than the 1000 trees simulated reflecting a Mississippi River divergence with migration using the upper 95% CI estimates of $N_e$ and *M* and not rejected when using the mean estimates of $N_e$ and *M* (mean *S* = 37.141, *P* = 0.31). In contrast, any combination of lower, mean or upper $N_e$ used with the lower 95% CI estimate of migration ($7.67 \times 10^{-7}$) always produced significantly lower values for *S* than did our ML tree (mean *S* = 8, $P = 1.0 \times 10^{-28}$). Additionally, simulations without migration estimate a mean *S* statistic of 1, further supporting the hypothesis of divergence occurring concomitantly with migration at the Mississippi in *N. rhombifer*. Finally, the program Migrate also predicts migration across the Mississippi River.

Testing migration and divergence time using MDIV produced similar results to our coalescent estimates. Four runs using the estimated priors all produced similar modes with overlapping credible intervals. Values (with 95% CI in parentheses) for the parameters of interest were: *T* = 0.1 (0.05–0.38), *M* = 1.8 (0.3–4.78) and $\Theta$ = 70.6 (49.2–83.8). Posterior probability estimates for migration were non-zero and the mode of the divergence time estimate is 1155,482 (402,618–5,211,783) years before present. This supports a model of ancient divergence with gene flow. Confidence intervals for the MDIV divergence dates overlap with those estimates provided by BEAST. Both estimates place the divergence times to the early Pleistocene.

## 4. Discussion

The Mississippi River has been implicated as a major barrier for gene flow in many terrestrial and aquatic vertebrates (Burbrink et al., 2000, 2008; Howes et al., 2006; Soltis et al., 2006), presumably because these organisms are unable to disperse across the river. In contrast, one would expect that aquatic vertebrates would be able to cross the Mississippi River more frequently – enough to permit at least limited gene flow among populations (Guiher and Burbrink, 2008; Makowsky et al., 2010) and potentially to erase signatures of this barrier. We tested the role of the Mississippi River in shaping the genetic history of the aquatic snake, *N. rhombifer*. Phylogenetic analyses (Figs. 2 and 3) strongly suggest that the river is a genetic barrier, and thus has limited gene flow between populations residing east and west of the river. However, the presence of shared haplotypes on both sides of the Mississippi River may be cases of mitochondrial introgression or migration and provide evidence that the Mississippi River may not promote long-term separation of semi-aquatic lineages. Moreover, the results of our coalescent simulations testing for divergence while incorporating gene flow demonstrate that indeed, the Mississippi River promotes lineage divergence despite limited gene flow.

Although the Mississippi River is permeable to the aquatic snake, *N. rhombifer*, our divergence dating analyses using BEAST suggests that the introgression was recent and that the river was likely a stronger barrier in the past. The split between USA populations of *Nerodia rhombifer* occurred in the early Pleistocene. Our coalescent simulations accounting for migration indicate that the ML tree cannot reject the Mississippi River as the vicariant barrier responsible for forming the East and West lineages. Thus, divergence initially occurred between the East and West clades in the early Pleistocene at the Mississippi River and was followed by increased migration in the last part of the Pleistocene and the Holocene. Diversification in the early Pleistocene and low levels of migration were also predicted independently using the model of isolation and migration implemented in MDIV. Given that we detect recent mitochondrial introgression, our results suggest that that this divergence coincided with an interglacial event when the Mississippi River was much wider than the present. However, we acknowledge that little is known about potential expansions and contractions of the Mississippi River that do not coincide with maximum glacial and interglacial periods.

Our results are similar to studies of several terrestrial snakes (*Pantherophis obsoletus* complex, *P. guttatus* complex, *Coluber constrictor*, *Lampropeltis getula*) that also have a distinct phylogeographic break at the Mississippi River (Burbrink, 2001; Burbrink et al., 2000, 2008; Pyron and Burbrink, 2009). However, they contrast greatly with the studies of other semi-aquatic North

American snakes studied to date (*Agkistrodon piscivorus*; *Nerodia erythrogaster*) which do not exhibit patterns of vicariance at the Mississippi River (Guiher and Burbrink, 2008; Makowsky et al., 2010). However, it is unclear whether the river is simply not an impediment to migration for these snakes, mtDNA markers cannot yet detect the signal of vicariance, or if recent colonization of habitats on the western side of the Mississippi River precludes detection of phylogeographic structure using mtDNA. Nonetheless, the contrasting phylogeographic patterns between *N. rhombifer* and *A. piscivorus*/*N. erythrogaster* demonstrate the potentially different effects of aquatic barriers on the population genetic structure of an aquatic organism.

The impact of barriers to dispersal and gene flow is often inferred to be the primary cause of lineage divergence and phylogeographic structure in terrestrial organisms. Our phylogeographic analysis of the diamond-backed water snake, *N. rhombifer* has demonstrated that this divergence may exist despite an organism's ability to traverse it. Although lineage divergence with gene flow seems counterintuitive, this result supports a growing body of research that has inferred the same phenomenon (e.g., Fitzpatrick et al., 2008; Niemiller et al., 2008; Nosil, 2008). By incorporating statistical phylogeographic techniques that test for lineage divergence despite gene flow, it is quite likely that future phylogeographic studies will reveal that this phenomenon is far more common than currently documented.

## Acknowledgments

Our gratitude goes to the many landowners and fisheries who allowed us access to their facilities, and to the following individuals and institutions for generous tissue loans, without which, this project would not have been possible: S. Arnold, S. Ballard, R. Brown, J. Boundy, B. Crother, J. Dixon, T. Esker, O. Flores, M. Keck, S. Lewis, R. Makowsky, F. Mendoza Quijano, M. Mills, A. Nieto Montes de Oca, T. Papenfuss, J. Petzing, M. Pfrender, S. Richter, T. Reeder, S. Secor, D. Shepard, J. Tucker, P. Ustach, J. Willson, American Museum of Natural History (D. Kizirian), California Academy of Sciences (J. Vindum), Field Museum of Natural History, Illinois Natural History Survey (C. Phillips), Louisiana State University Museum of Natural History (D. Dittmann), Museum of Vertebrate Zoology (Carol Spencer), and Southwest Texas State University tissue collections (M. Forstner). We also thank J. Boundy, W. Fair, A. Seago, and B. Smith and Harvey (USA Army Corps of Engineers, Demopolis, Alabama) for fieldwork assistance, L. Seago and H. Voris for access to field equipment, and M. Keck for photos. The J. McGuire lab group, A. Seago, D. Warren, and two anonymous reviewers made helpful comments on the manuscript. Portions of this study were funded by the American Museum of Natural History Theodore Roosevelt grant, Museum of Vertebrate Zoology Koford grant, and UC Berkeley Integrative Biology summer research grant awarded to MCB.

## Appendix A. Supplementary data

Supplementary data associated with this article can be found, in the online version, at doi:10.1016/j.ympev.2010.07.015.

## References

Alfaro, M.E., Arnold, S.J., 2001. Molecular systematics and evolution of *Regina* and the thamnophiine snakes. Mol. Phylogenet. Evol. 16, 408–423.

Avise, J.C., 2000. Phylogeography: The History and Formation of Species. Harvard University Press, Cambridge.

Beerli, P., 2008. Migrate-N, Version 2.4. Available from: <http://popgen.scs.fsu.edu/Migrate-n.html>.

Beerli, P., Felsenstein, J., 1999. Maximum likelihood estimation of migration rates and effective population numbers in two populations. Genetics 152, 763–773.

Beerli, P., Felsenstein, J., 2001. Maximum likelihood estimation of a migration matrix and effective population sizes in n subpopulations using a coalescent approach. Proc. Natl. Acad. Sci. USA 98, 4563–4568.

Blair, W.F., 1958. Distributional patterns of vertebrates in the southern United States in relation to past environments. In: Hubbs, C.L. (Ed.), Zoogeography American Association for the Advancement of Science Publication No. 51, pp. 433–463.

Brandley, M.C., Schmitz, A., Reeder, T.W., 2005. Partitioned Bayesian analyses, partition choice, and the phylogenetic relationships of scincid lizards. Syst. Biol. 54, 373–390.

Brant, S.V., Ortí, G., 2003. Phylogeography of the Northern short-tailed shrew, *Blarina brevicauda* (Insectivora: Soricidae): past fragmentation and postglacial recolonization. Mol. Ecol. 12, 1435–1449.

Burbrink, F.T., 2001. Systematics of the North American rat snake complex (*Elaphe obsoleta*). Herp. Monogr. 15, 1–53.

Burbrink, F.T., Castoe, T.A., 2009. Molecular snake phylogeography. In: Mullin, Seigel (Eds.), Snakes: Applied Ecology and Conservation. Cornell University Press.

Burbrink, F.T., Lawson, R., Slowinski, J.B., 2000. Mitochondrial DNA phylogeography of the polytypic North American rat snake (*Elaphe obsoleta*): a critique of the subspecies concept. Evolution 54, 2107–2118.

Burbrink, F.T., Fontanella, F.M., Pyron, R.A., Guiher, T.J., Jimenez, C., 2008. Phylogeography across a continent: the evolutionary and demographic history of the North American Racer (Serpentes: Colubridae: *Coluber constrictor*). Mol. Phylogenet. Evol. 47, 274–288.

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

Castoe, T.A., Doan, T.M., Parkinson, C.L., 2004. Data partitions and complex models in Bayesian analysis: the phylogeny of gymnophthalmid lizards. Syst. Biol. 53, 448–469.

Cushing, E.M., Boswell, E.H., Hosman, R.L., 1964. General Geology of the Mississippi Embayment. USGS Professional Paper 448-B.

De Queiroz, A., Lawson, R., Lemos-Espinal, J.A., 2002. Phylogenetic relationships of North American garter snakes (*Thamnophis*) based on four mitochondrial genes: how much DNA sequence is enough? Mol. Phylogenet. Evol. 22, 315–329.

Drummond, A.J., Rambaut, A., 2007. BEAST: Bayesian evolutionary analysis by sampling trees. BMC Evol. Biol. 7, 214.

Drummond, A.J., Ho, S.Y.W., Phillips, M.J., Rambaut, A., 2006. Relaxed phylogenetics and dating with confidence. PloS Biol. 4, e88.

Fitzpatrick, B.M., Placyk, J.S., Niemiller, M.L., Casper, G.S., Burghardt, G.M., 2008. Distinctiveness in the face of gene flow: hybridization between specialist and generalist gartersnakes. Mol. Ecol. 17, 4107–4117.

Gibbons, J.W., Dorcas, M.E., 2004. North American Watersnakes: A Natural History. University of Oklahoma Press, Norman.

Graur, D., Martin, W., 2004. Reading the entrails of chickens: molecular timescales of evolution and the illusion of precision. Trends Genet. 20, 80–86.

Guiher, T.J., Burbrink, F.T., 2008. Demographic and phylogeographic histories of two venomous North American snakes of the genus *Agkistrodon*. Mol. Phylogenet. Evol. 48, 543–553.

Ho, S.Y.W., 2007. Calibrating molecular estimates of substitution rates and divergence times in birds. J. Avian Biol. 38, 409–414.

Holman, J.A., 1973. Reptiles of the Egelhoff local fauna (Upper Miocene) of Nebraska. Contrib. Mus. Paleon. Univ. Michigan 24, 125–134.

Holman, J.A., 2000. Fossil Snakes of North America. Origin, Evolution, Distribution, Paleoecology. Indiana University Press, Indianapolis, IN.

Howes, B.J., Lindsay, B., Lougheed, S.C., 2006. Range-wide phylogeography of a temperate lizard, the five-lined skink (*Eumeces fasciatus*). Mol. Phylogenet. Evol. 40, 183–194.

Huelsenbeck, J.P., Rannala, B., 2004. Frequentist properties of Bayesian posterior probabilities of phylogenetic trees under simple and complex substitution models. Syst. Biol. 53, 904–913.

Knowles, L.L., Maddison, W.P., 2002. Statistical phylogeography. Mol. Ecol. 11, 2623–2635.

Lawson, R., 1987. Molecular studies of thamnophiine snakes. 1. The phylogeny of the genus *Nerodia*. J. Herpetol. 21, 140–157.

Lawson, R., Slowinski, J.B., Crother, B.I., Burbrink, F.T., 2005. Phylogeny of the Colubroidea (Serpentes): new evidence from mitochondrial and nuclear genes. Mol. Phylogenet. Evol. 37, 581–601.

Leaché, A.D., Reeder, T.W., 2002. Molecular systematics of the eastern fence lizard (*Sceloporus undulatus*): a comparison of parsimony, likelihood, and Bayesian approaches. Syst. Biol. 51, 44–68.

Maddison, D., Maddison, W., 2005. MacClade Version, 4.08, Sinauer Associates, Inc., Sunderland, MA.

Maddison, W.P., Maddison, D.R., 2008. Mesquite: A Modular System for Evolutionary Analysis, Version 2.5. Available from: <http://mesquiteproject.org>.

Makowsky, R., Marshall Jr., J.C., McVay, J., Chippindale, P.T., Rissler, L.J., 2010. Phylogeographic analysis and environmental niche modeling of the plain-bellied watersnake (*Nerodia erythrogaster*) reveals low levels of genetic and ecological differentiation. Mol. Phylogenet. Evol. 55, 985–995.

Near, T.J., Page, L.M., Mayden, R.L., 2001. Intraspecific phylogeography of *Percina evides* (Percidae: Etheostomatinae): an additional test of the Central Highlands pre-Pleistocene vicariance hypothesis. Mol. Ecol. 10, 2235–2240.

Nielsen, R., Wakeley, R., 2001. Distinguishing migration from isolation: a Markov chain Monte Carlo approach. Genetics 158, 885–896.

Niemiller, M.L., Fitzpatrick, B.M., Miller, B.T., 2008. Recent divergence with gene flow in Tennessee cave salamanders (Plethodontidae: *Gyrinophilus*) inferred from gene genealogies. Mol. Ecol. 17, 2258–2275.

Nosil, P., 2008. Speciation with gene flow could be common. Mol. Ecol. 17, 2103–2106.
Nylander, J.A.A., 2004. MrModeltest v2. Programme distributed by the author. Evolutionary Biology Centre, Uppsala University.
Nylander, J.A.A., Ronquist, F., Huelsenbeck, J.P., Nieves-Aldrey, J.L., 2004. Bayesian phylogenetic analysis of combined data. Syst. Biol. 53, 47–67.
Nylander, J.A.A., Wilgenbusch, J.C., Warren, D.L., Swofford, D.L., 2008. AWTY (are we there yet?): a system for graphical exploration of MCMC convergence in Bayesian phylogenetics. Bioinformatics 24, 581–583.
Pyron, R.A., Burbrink, F.T., 2009. Lineage diversification in a widespread species: roles for niche divergence and conservatism in the Common Kingsnake, *Lampropeltis getula*. Mol. Ecol. 18, 3443–3457.
Pyron, R.A., Burbrink, F.T., 2010. Hard and soft allopatry: physically and ecologically mediated modes of geographic speciation. J. Biogeogr. doi:10.1111/j.1365-2699.2010.02336.x.
Rogers, K.L., 1984. Herpetofauna of the Big Springs and Hornets Nest Quarries (Northeastern Nebraska, Pleistocene: Late Blancan). Trans. Nebraska Acad. Sci. 12, 81–94.
Shepard, D.B., Burbrink, F.T., 2008. Lineage diversification and historical demography of a sky island salamander, *Plethodon ouachitae*, from the Interior Highlands. Mol. Ecol. 17, 5315–5335.
Shepard, D.B., Burbrink, F.T., 2009. Phylogeographic and demographic effects of Pleistocene climatic fluctuations in a montane salamander, *Plethodon fourchensis*. Mol. Ecol. 18, 2243–2262.
Slatkin, M., Maddison, W.P., 1989. A cladistic measure of gene flow inferred from the phylogenies of alleles. Genetics 123, 603–613.
Soltis, D.E., Morris, A.B., Mclachlan, J.S., Manos, P.S., Soltis, P.S., 2006. Comparative phylogeography of unglaciated eastern North America. Mol. Ecol. 15, 4261–4293.
Spellman, G.M., Riddle, B., Klicka, J., 2007. Phylogeography of the mountain chickadee (*Poecile gambeli*): diversification, introgression, and expansion in response to quaternary climate change. Mol. Ecol. 16, 1055–1068.
Stamatakis, A., 2006. RAxML-VI-HPC: maximum likelihood-based phylogenetic analyses with thousands of taxa and mixed models. Bioinformatics 22, 2688–2690.
Steele, C.A., Storfer, A., 2006. Coalescent-based hypothesis testing supports multiple Pleistocene refugia in the Pacific Northwest for the Pacific Giant Salamander (*Dicamptodon tenebrosus*). Mol. Ecol. 15, 2477–2487.
Swofford, D.L., 2003. PAUP*. Phylogenetic Analysis Using Parsimony (*and Other Methods). Version 4. Sinauer Associates, Sunderland, Massachusetts.
Vincent, S.E., Brandley, M.C., Herrel, A., Alfaro, M.E., 2009. Convergence in trophic morphology and feeding performance among piscivorous natricine snakes. J. Evol. Biol. 22, 1203–1211.
Wood, D.A., Fisher, R.N., Reeder, T.W., 2008. Novel patterns of historical isolation, dispersal, and secondary contact across Baja California in the Rosy Boa (*Lichanura trivirgata*). Mol. Phylogenet. Evol. 46, 484–502.