Biological Journal of the Linnean Society, 2021, **133**, 120–134. With 5 figures.

Signatures of north-eastern expansion and multiple refugia: genomic phylogeography of the Pine Barrens tree frog, *Hyla andersonii* (Anura: Hylidae)

ALEXA R. WARWICK[1,*], LISA N. BARROW[2], MEGAN L. SMITH[3], D. BRUCE MEANS[4], ALAN R. LEMMON[5] and EMILY MORIARTY LEMMON[6]

[1]*Department of Fisheries and Wildlife, Michigan State University, East Lansing, MI, USA*
[2]*Museum of Southwestern Biology and Department of Biology, University of New Mexico, Albuquerque, NM, USA*
[3]*Department of Biology and Department of Computer Science, Indiana University, Bloomington, IN, USA*
[4]*Coastal Plains Institute and Land Conservancy, Tallahassee, FL, USA*
[5]*Department of Scientific Computing, Florida State University, Tallahassee, FL, USA*
[6]*Department of Biological Science, Florida State University, Tallahassee, FL, USA*

Received 20 October 2020; revised 2 February 2021; accepted for publication 3 February 2021

Range fragmentation poses challenges for species persistence over time and can be caused by both historical and contemporary processes. We combined genomic data, phylogeographical model testing and palaeoclimatic niche modelling to infer the evolutionary history of the Pine Barrens tree frog (*Hyla andersonii*), a seepage bog specialist, in eastern North America to gain a better understanding of the historical context of its fragmented distribution. We sampled *H. andersonii* populations across the three disjunct regions of the species range: Alabama/Florida (AF), the Carolinas (CL) and New Jersey (NJ). Phylogenetic relationships within *H. andersonii* were consistent between the nuclear species tree and mitochondrial analyses, indicating divergence between AF and CL/NJ (Atlantic clade) ~0.9 Mya and divergence of the NJ clade ~0.15 Mya. Several predictions of north-eastern expansion along the Atlantic coast were supported by phylogeographical analyses. Model testing using genome-wide single nucleotide polymorphism data and species distribution models both provided evidence for multiple disjunct refugia. This comprehensive phylogeographical study of *H. andersonii* demonstrates a long history of range fragmentation within an endemic coastal plain species and highlights the influence of historical climate change on the current distribution of species and their genetic diversity.

ADDITIONAL KEYWORDS: anchored hybrid enrichment – divergence time – Last Glacial Maximum – model selection – sequence capture – species distribution models.

INTRODUCTION

Range fragmentation can be detrimental for species persistence by reducing dispersal and gene flow, population sizes and the genetic diversity that underlies adaptation to changing environments (Fahrig, 2003; Frankham, 2005). Given that fragmented distributions have arisen through a combination of historical and contemporary factors, species with disjunct ranges can be particularly intriguing, yet challenging to study (Hochheimer & Hoffmann, 2016). Genetic breaks within continuously distributed species can coincide with obvious biogeographical barriers (Soltis *et al.*, 2006; Bell *et al.*, 2012), but the processes that gave rise to highly fragmented distributions may no longer be readily apparent. Inferring the evolutionary history of fragmented species with genomic data and model-based approaches can reveal how long populations have been isolated, how they responded to past environmental changes and how diversity is currently distributed across the landscape.

The North American Coastal Plain (NACP) is recognized as a global biodiversity hotspot (Noss *et al.*, 2015) and has been a focal area in phylogeography for several decades (Avise *et al.*, 1987; Avise, 2000). Various taxa within the NACP have fragmented

*Corresponding author. E-mail: awarwick@msu.edu

distributions, probably influenced by glacial cycles and concomitant shifts in forests, sea levels and coastlines starting during the Pleistocene ~2.58 Mya (Haq *et al.*, 1987; Williams *et al.*, 2004). During glacial periods, many species ranges shifted southwards to refugia, or areas of climatic stability and expanded land mass, such as peninsular Florida and the western Gulf coast (Marshall *et al.* 2002; Swenson & Howard, 2005). As glaciers receded, species then tracked shifts in suitable habitat and climate while expanding northwards (prediction of north-eastern expansion). In some cases, species persisted into the present only in disjunct patches (Sorrie & Weakley, 2001). The degree of range fragmentation varies widely among species and was recently quantified for amphibians in the NACP (Newman & Austin, 2019). Amphibians may be particularly vulnerable to range fragmentation given their reliance on aquatic and terrestrial environments, continued habitat loss and major threats, such as disease (Stuart *et al.*, 2004; Lips, 2016), highlighting the need to gain a better understanding of the processes underlying their distributions and genetic diversity.

One amphibian species endemic to the NACP with a highly disjunct distribution is the Pine Barrens tree frog [*Hyla andersonii* (Baird, 1854); Fig. 1]. This seepage bog specialist occurs in three regions, each separated by > 470 km: Alabama and Florida (AF), North and South Carolina (CL) and New Jersey (NJ). Resolving the relationships among regions is necessary for management of *H. andersonii* because it is considered near threatened (Hammerson, 2017) and state-listed in all but Florida (Moler *et al.*, 2020). Disjunct distributions can be difficult to address in a conservation framework, particularly when it is unclear whether isolation was the result of natural or human causes (Reilly *et al.*, 2014). The strong similarities in morphology and behaviour among the three regions of *H. andersonii* (Warwick *et al.*, 2015) suggest hypotheses of either recent divergences, with little time for differences to accumulate, or older divergences, with consistent selective pressures across regions to maintain these similarities. Previous work indicated that NJ and CL populations were most closely related (Karlin *et al.*, 1982; Lemmon *et al.*, 2007; Warwick *et al.*, 2015; Oswald *et al.*, 2020), but the phylogeographical history of the species has not been investigated fully. Advances in genomic data collection over the last decade provide higher-resolution data that will enable rigorous hypothesis testing (Garrick *et al.*, 2015).

Owing to the habitat specialization of *H. andersonii* (Means, 2006), our hypotheses about its evolutionary history can be informed by wetland-associated plant species, many of which are distributed in a similar manner (*Chamaecyparis thyoides*, Little, 1971; *Drosera filiformis*, Rice, 2011; and part of the range of *Sarracenia purpurea*, Godt & Hamrick, 1998). In addition, Sorrie & Weakley (2001) identified 27 different geographical distribution patterns of vascular plants in the NACP, which included disjunct patterns similar to that of *H. andersonii*. One disjunct pattern was found for eight plant species across NJ/Delaware and the Carolinas, possibly explained by previously continuous (late Pleistocene) distributions that were divided when sea levels rose in the Holocene (Sorrie & Weakley, 2001). A second pattern was identified in 11 plant species across the Carolinas and eastern Gulf Coastal Plain. In this region, the Georgia Tifton/Vidalia uplands and Florida Central Highlands have been suggested as a significant phylogeographical barrier for some coastal lowland taxa because lowlands would have been subject to periodic oceanic flooding during Plio-Pleistocene interglacials (Liu *et al.*, 2006). Given this historical context, we addressed three main questions and their associated predictions for *H. andersonii*:

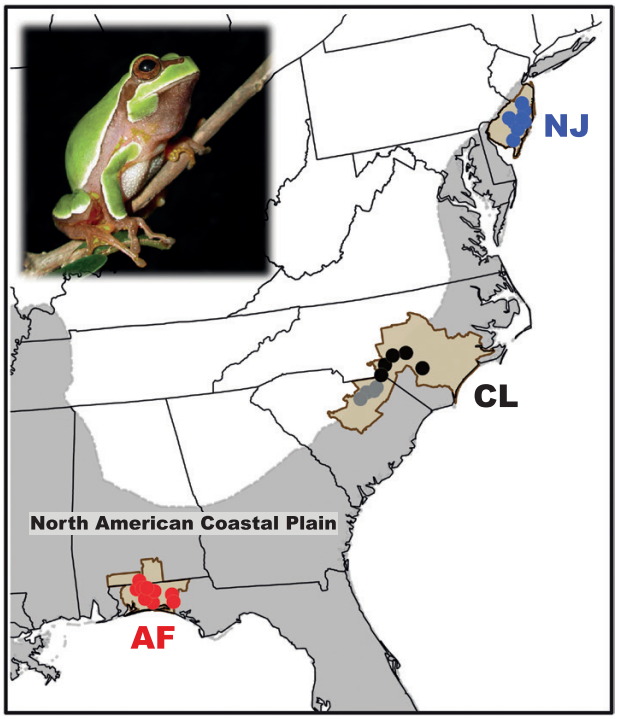

Figure 1. Sampling localities for *Hyla andersonii* genomic data. The North American Coastal Plain is shown in grey. The county-level species range map is shown in tan and is based on Hammerson (2017), with the addition of Marlboro County, South Carolina, from which historical localities are known. Abbreviations: AF, Alabama/Florida; CL, the Carolinas; NJ, New Jersey. Photograph credit: A. R. Warwick.

1. How long have populations in these regions been isolated? We predict either more recent (Holocene) or older (Pleistocene) divergences.

2. How many refugia were there? We predict multiple [two (AF and CL) or three (AF, CL, and NJ)] refugial areas.
3. What was the direction of colonization? We predict north-eastern expansion, with the lowest genetic diversity in NJ and greater dispersal distances in CL and NJ.

We combined new genome-scale mitochondrial DNA (mtDNA) and nuclear data with palaeoclimate species distribution modelling (SDM) to assess our predictions independently. Our objectives were as follows: (1) to infer the phylogenetic relationships and estimate divergence times within *H. andersonii*; (2) to test predictions of north-eastern expansion with a spatial diffusion approach; (3) to compare models with different numbers of refugia and colonization histories; and (4) to generate SDMs for current climate and three historical time points to identify potential refugia and evaluate the response of this species to historical climate change.

MATERIAL AND METHODS

Field and genomic data collection

Historical and known contemporary locations across the range of *H. andersonii* were visited from 2010 to 2014. Tissue samples were collected by capturing adult frogs by hand, of which 26 samples were selected for the present study (Fig. 1; Supporting Information, Table S1). Each sample represents a different population, and those population locations reflect the current known species distribution with equal representation per region. Two *Hyla femoralis* individuals, one each from North Carolina and Florida, were used as outgroups (Supporting Information, Table S1). Samples were immersed in tissue buffer (20% dimethyl sulfoxide and 0.25 M EDTA, salt saturated) in the field and subsequently stored at −80 °C at Florida State University (FSU). Total genomic DNA was purified from toe clips or liver using either an Omega Bio-Tek E.Z.N.A. Tissue DNA Kit or an alcohol precipitation method (for details, see Warwick *et al.*, 2015).

We processed all 28 samples at FSU's Center for Anchored Phylogenomics (www.anchoredphylogeny.com) using the hybrid enrichment protocol described by Lemmon *et al.* (2012). Briefly, each sample was sonicated to an average fragment size of ~300 bp using a Covaris E220 focused ultrasonicator. Library preparation and indexing were completed on a Beckman-Coulter Biomex FXp liquid-handling robot according to a protocol modified from the study by Meyer & Kircher (2010). Indexed samples were pooled in equal quantities (12–16 samples per pool) before hybrid enrichment using an Agilent Custom SureSelect kit. We used the *Pseudacris* v.1 probe set described by Banker *et al.* (2020), which consists of 366 deep-scale loci designed for amphibians (Barrow *et al.*, 2018; Hime *et al.*, 2021) and 1250 shallow-scale loci designed from the chorus frogs *Pseudacris nigrita* and *Pseudacris feriarum* (Family Hylidae). After enrichment for these 1616 anchored loci, samples were pooled into a group of ≤ 48 samples for sequencing on one PE150 lane of an Illumina HiSeq 2000 at the FSU College of Medicine.

Bioinformatic processing

Raw sequencing reads were quality filtered and demultiplexed using CASAVA v.1.8.2 (Illumina). Mitochondrial (mtDNA) genomes were assembled as sequencing 'bycatch' using SeqMan Pro NGEN v.12.3.1, following the protocol described by Barrow *et al.* (2017). Briefly, a reference *H. femoralis* mtDNA genome (ARW436) was constructed by mapping reads to *H. femoralis* gene sequences downloaded from GenBank (Supporting Information, Table S2). Raw reads for each individual were then mapped to that reference. We inspected assemblies manually, aligned consensus sequences in Geneious v.8.1.3 using the MAFFT v.7.222 (Katoh *et al.*, 2002) plugin, and annotated genomes based on the *H. japonica* genome (GenBank accession no. AB303949; Igawa *et al.*, 2008).

Anchored loci were assembled using the quasi-*de novo* approach described by Prum *et al.* (2015) and Hamilton *et al.* (2016). After merging paired-end reads as described by Rokyta *et al.* (2012), assemblies were generated using both reference-based and extension methods. Allele phasing and orthology assessment were completed following the methods of Pyron *et al.* (2016). We aligned consensus sequences across individuals using MAFFT v.7.023b (Katoh & Standley, 2013) and then trimmed and masked alignments to remove ambiguously aligned regions (Hamilton *et al.*, 2016).

We generated two nuclear datasets for downstream analyses. For phylogenetic analyses, we used the sequence alignments with 28 individuals including the outgroup samples (*H. femoralis*). For phylogeographical model testing, we extracted single nucleotide polymorphisms (SNPs) from only *H. andersonii* samples using custom python scripts (Barrow *et al.*, 2018; available on Dryad). To maximize the number of SNPs with no missing data, we retained the 21 individuals (seven per region) with the most loci recovered. We randomly selected one SNP per locus to avoid violating assumptions of independence. Genetic diversity metrics (expected heterozygosity and allelic richness) were estimated in R (R Core Team, 2019) using the packages '*adegenet*' (Jombart, 2008) and 'PopGenReport' (Adamack & Gruber, 2014).

Phylogenetic relationships and divergence time estimation

We inferred mtDNA relationships and divergence times within *H. andersonii* using BEAST v.1.10.1 (Suchard *et al.*, 2018). Additional taxa were downloaded from GenBank (Supporting Information, Table S1) and included as outgroups. Given that fossil calibrations were unavailable within *H. andersonii*, we used a substitution rate estimate for *ND2* (0.00957 mutations per lineage per million years; Macey *et al.*, 1998; Crawford, 2003). We partitioned the dataset by gene and linked tree models. After preliminary runs to test different prior settings, we chose a strict clock, HKY site models, and a coalescent constant tree prior for the full Markov chain Monte Carlo run of 50 million generations, sampling every 5000 generations. We verified that all parameters had reached adequate effective sample sizes (> 200) using Tracer v.1.7.1 (Rambaut *et al.*, 2018), discarded the first 10% of trees as burn-in, and summarized samples using TreeAnnotator.

For comparison with mtDNA relationships, we used anchored loci to infer nuclear phylogenetic relationships with two coalescent-based species tree methods, SVDQuartets (Chifman & Kubatko, 2014, 2015) and ASTRAL v.5.4.5 (Mirarab *et al.*, 2014; Mirarab & Warnow, 2015). SVDQuartets uses algebraic statistical techniques to infer quartet relationships from multi-locus sequence data, assuming each locus has its own genealogy under the multispecies coalescent model. We ran SVDQuartets in PAUP* v.4.0a157 (Swofford, 2002), with two alleles assigned to each individual, exhaustive quartet sampling and 100 bootstrap replicates. The second method, ASTRAL, summarizes over gene trees to infer an unrooted species tree that maximizes the number of quartets shared across the gene and species trees. We estimated gene trees for each nuclear locus using RAxML v.8 (Stamatakis, 2014), phased alleles, the GTR+G model, and rapid bootstrapping with 100 replicates. We then used the best trees as input for ASTRAL, mapping alleles to each individual, and conducting multi-locus bootstrapping with the RAxML bootstrap trees and 100 replicates. Branches with bootstrap support <70 were collapsed using TreeGraph v. 2.14.0 (Stöver & Müller, 2010).

Testing for north-eastern expansion and mitochondrial DNA ancestral location

To evaluate the direction of colonization within *H. andersonii*, we tested predictions using PhyloMapper v.1.0 (Lemmon & Lemmon, 2008). This approach uses a spatially explicit random walk model of migration to estimate the geographical location of an ancestor, dispersal rate and migration direction, given a set of georeferenced taxa and their phylogenetic relationships. PhyloMapper requires a resolved genealogy and clades with at least five tips to conduct tests; thus, we inferred the mtDNA history for well-sampled clades within *H. andersonii* (see Results). We tested the predictions that: (1) recently expanded lineages (NJ or CL + NJ, hereafter Atlantic) would have higher dispersal than lineages that have not expanded (AF); (2) the geographical centre of present-day samples for expanded lineages would be north-east relative to the maximum likelihood (ML) estimate of the ancestral centre of origin; and (3) the direction of migration for expanded lineages would be non-random and to the north-east.

Initially, we tested for phylogeographical association within each clade using randomization tests to compare the ML dispersal estimate with a null distribution. We generated null distributions by randomizing geographical coordinates across the tips of the phylogeny 10 000 times, estimated the dispersal parameters, and calculated *P*-values as the proportion of samples from the null that were lower than the original ML dispersal estimate. We then used a likelihood ratio test (LRT) to compare models with different numbers of dispersal rates: either a single dispersal rate for the whole species, or two dispersal rates (AF vs. Atlantic). A significantly higher dispersal rate for the Atlantic clade would support expansion (prediction 1).

After estimating the ML ancestral location for each clade, we conducted LRTs to determine whether these were significantly different from the centre of sampled locations. We calculated the mean coordinates of sampled individuals (Fig. 1) in QGIS v.2.18.2 and compared the likelihood when the central location was constrained to this point to the likelihood from the unconstrained (ML) ancestral location. We also assessed uncertainty in the ancestral location by estimating the location from 1000 random trees sampled from the posterior. A difference between the geographical centre and ancestral location would suggest expansion (prediction 2).

Finally, we tested whether the inferred direction of migration was non-random or in the predicted direction (north-east) for clades of interest (prediction 3). PhyloMapper includes a set of standard directions (east, north-east, west, south-west, etc.) as decimal values in radians between zero and 2π. North-east is 45° or π/4. We generated null distributions by randomizing the geographical coordinates of the tips 10 000 times, computed the expected net dispersal as the average change in location from ancestor to descendant (random direction test), and computed the average of all angles between ancestor–descendant arrows and the predicted direction (north-east; a priori direction test). We also generated a graph to visualize

directional tendency by repeating the a priori test with different predicted migration directions from zero to 2π radians. The *P*-value for each direction is depicted as a line in a circle, with longer lines representing smaller *P*-values. Additional details are provided by Lemmon & Lemmon (2008).

Testing models of multiple refugia and colonization history

We used the R package 'delimitR' (Smith & Carstens, 2020) to test phylogeographical models with the genomic SNP dataset. This approach conducts demographic model selection by summarizing SNP data using the site frequency spectrum (SFS) and comparing models with machine learning. We designed a custom model set to test our predictions for *H. andersonii*, specifically, whether populations survived in one, two or three refugia during the Pleistocene (Fig. 2). Model 1 included a single Pleistocene refugium (AF), with colonization of CL and NJ occurring during the Holocene. Model 2 included two refugia, AF and CL, with the colonization of NJ occurring during the Holocene. Models 3 and 4 included three Pleistocene refugia, with model 4 considering the potential for gene flow during the initial stages of divergence between the CL and NJ populations.

Population sizes were drawn from uniform priors (1000, 100 000 individuals) for all models. For models including post-Pleistocene colonization (models 1 and 2), colonization times were drawn from uniform priors (1000, 10 000 generations before the present), and the proportion of the population remaining during the bottleneck was drawn from a uniform prior (0.1, 1.0). Bottlenecks began 500 years before colonization (looking backward in time). Divergence times between refugial populations were drawn from broad uniform priors (20 000, 5 000 000 generations), and we restricted divergence and colonization times such that the topologies of the models did not change. For the model including secondary contact between the CL and NJ populations, migration rates were drawn from a uniform prior (5×10^{-6}, 5×10^{-5}), and migration began halfway between the present and the time of divergence between the two populations (looking backward in time).

We used custom python scripts (available at https://github.com/meganlsmith) to construct a multidimensional site frequency spectrum (mSFS) from our empirical data. We simulated 10 000 datasets under each model using the fastsimcoalsims function in 'delimitR', which uses FASTSIMCOAL v.2.6 (Excoffier *et al.*, 2013) to simulate SFS under the specified model. Next, we binned the mSFS using four classes, and constructed a random forest (RF) classifier from the simulated data using 500 decision trees. We calculated error rates based on out-of-bag errors, and then applied the RF classifier to our observed data and selected the best model according to the number of votes or decision trees out of 500 that agreed with that model. We then approximated the posterior probability of the best model using the RF_predict_abcrf function in 'delimitR', which follows the approach described by Pudlo *et al.* (2016).

Species distribution modelling

To estimate the historical distribution independently and identify potential refugia within *H. andersonii*, we generated SDMs in MAXENT v.3.3.3 (Phillips *et al.*, 2006). We downloaded 19 bioclimatic variables from WorldClim v.1.4 (Hijmans *et al.*, 2005), representing four time points: current (averaged from 1950 to 2000), Holocene (~0.006 Mya), Last Glacial Maximum (LGM; ~0.022 Mya) and Last Interglacial (LIG; ~0.12–0.14 Mya). Historical data included changes to the coastline. We removed significantly correlated climate layers ($r^2 > 0.8$; Sheppard, 2013) and retained nine variables describing aspects of temperature and precipitation (Supporting Information, Table S3).

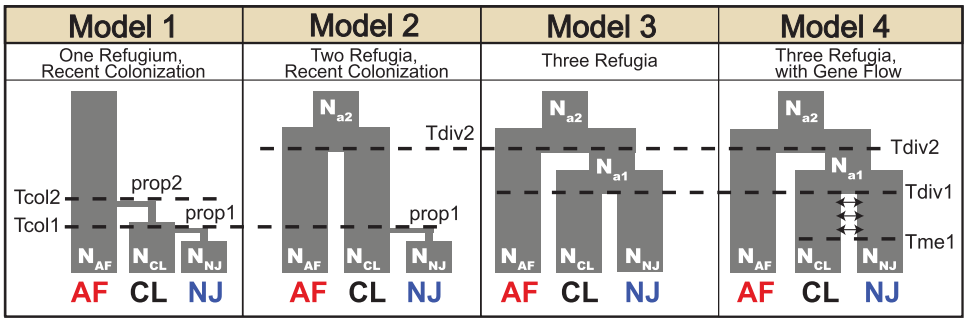

Figure 2. Models compared in delimitR. Abbreviations: a1, ancestral population size at Tdiv1; a2, ancestral population size at Tdiv2; AF, Alabama/Florida; CL, North and South Carolina; N, population size; NJ, New Jersey; prop, proportion of the population remaining after the bottleneck; Tcol, time of recent colonization with a bottleneck; Tdiv, divergence time between populations; Tme, time of migration ending.

We obtained locality records for *H. andersonii* from various sources (museums, state databases, personal and published field records; Supporting Information, Table S4; Fig. S1), excluding the only two records before 1950. To avoid sampling bias, we thinned localities for each region with a filtering distance of 5 km and 100 replicates using the R package 'spThin' (Aiello-Lammens *et al.*, 2015). To generate background points relevant to the disjunct distribution of *H. andersonii*, we selected random localities within a 400 km buffer of the sampling localities, clipped to land borders (Supporting Information, Fig. S2). In MAXENT, we tested multiple values of the regularization multiplier (R = 1, 2.5 and 5) and compared the resulting area under the curve (AUC) values. For the final models, we used R = 1, a random set of 25% testing data and 100 bootstrap replicates. We calculated the 95% lowest presence threshold (LPT95; Pearson *et al.*, 2007) using custom R scripts that averaged the probability at which 95% of testing localities were included across bootstrap replicates. Based on the LPT95 value, we visualized binary presence-only models and rescaled probability models.

RESULTS

DATA SUMMARY

We sequenced a total of 241 million reads across all 28 individuals (~8.6 million reads on average per individual), from which mtDNA genomes and anchored loci were assembled. The mtDNA genome alignment consisted of 15 457 bp with 2.14% missing data, 430 (2.8%) variable sites and 315 (2.0%) parsimony-informative sites. Reads were assembled into an average of 804 anchored loci (±57) per individual. After aligning and filtering, we retained 458 nuclear loci for species tree analyses, which included 2.3% missing data, a total of 628 954 bp, 12 444 (2.0%) variable sites and 9157 (1.5%) parsimony-informative sites. Loci were retained only when ≥ 24 of 28 individuals had data, even if the data were partial. Of the 458 loci retained, 319 loci had all 28 individuals, and the remaining 139 had varying numbers of missing individuals. The average anchored locus length was 1375 bp (range: 146–2617 bp). The SNP dataset for *H. andersonii* included 609 loci, with a total of 6428 SNPs. Genetic diversity estimates were highest for the CL region and lowest for NJ (Table 1).

Table 1. Genetic diversity estimates for each region based on single nucleotide polymorphism datasets including 21 individuals

Dataset	All SNPs, 609 loci (6428 SNPs)		Random SNPs, 609 loci (609 SNPs)	
Region	He	AR	He	AR
AF	0.1936	1.530	0.1790	1.482
CL	0.2455	1.690	0.2281	1.654
NJ	0.1413	1.395	0.1183	1.327

Abbreviations: AF, Alabama/Florida; AR, allelic richness; CL, North and South Carolina; He, expected heterozygosity; NJ, New Jersey; SNP, single nucleotide polymorphism.

PHYLOGENETIC RELATIONSHIPS AND DIVERGENCE TIMES

Phylogenetic relationships based on mtDNA genomes and nuclear anchored loci were similar and were highly congruent with geography (Fig. 3). The AF samples formed a clade and diverged from the rest of the range ~0.93 Mya [95% highest posterior density (HPD): 0.80–1.05 Mya]. The Atlantic clade comprised the remaining individuals, within which diversification began ~0.54 Mya (95% HPD: 0.45–0.62 Mya). The NJ samples formed a clade and were the most recently diverged, at ~0.15 Mya (95% HPD: 0.12–0.18 Mya). The relationships of the CL samples were less certain. In both species tree analyses, the South Carolina (SC) samples formed a clade, but with unclear placement relative to the North Carolina samples. In the mtDNA analysis, three SC samples formed a clade that diverged from the remaining Atlantic samples.

SIGNATURES OF NORTH-EASTERN EXPANSION

We found significant phylogeographical association within *H. andersonii* and each clade with at least five tips (Table 2). As predicted, the Atlantic clade had the highest dispersal distance, and this was significantly higher than the AF clade based on an LRT (Table 3). The dispersal distance for the Atlantic clade (318.5 m) was approximately four times larger than that for the AF clade (81.5 m). The geographical centre of present-day samples (central location) shifted in the expected direction (north-east) for the entire species and for the Atlantic clade, but neither was significantly different from the unconstrained ML estimates of each ancestral location (Supporting Information, Table S5; Fig. 4). The NJ and AF central locations overlapped with the ML estimates. Uncertainty in the ancestral estimates was relatively low for each clade, as indicated by the tight clustering of points around the ML estimate (Fig. 4). The direction of migration was significantly non-random for the AF and Atlantic clades, and the a priori direction test indicated north-eastern migration for the AF, Atlantic and NJ clades. Finally, the directional tendency of migration visualized for the Atlantic clade was towards the north-east, as predicted (Fig. 4).

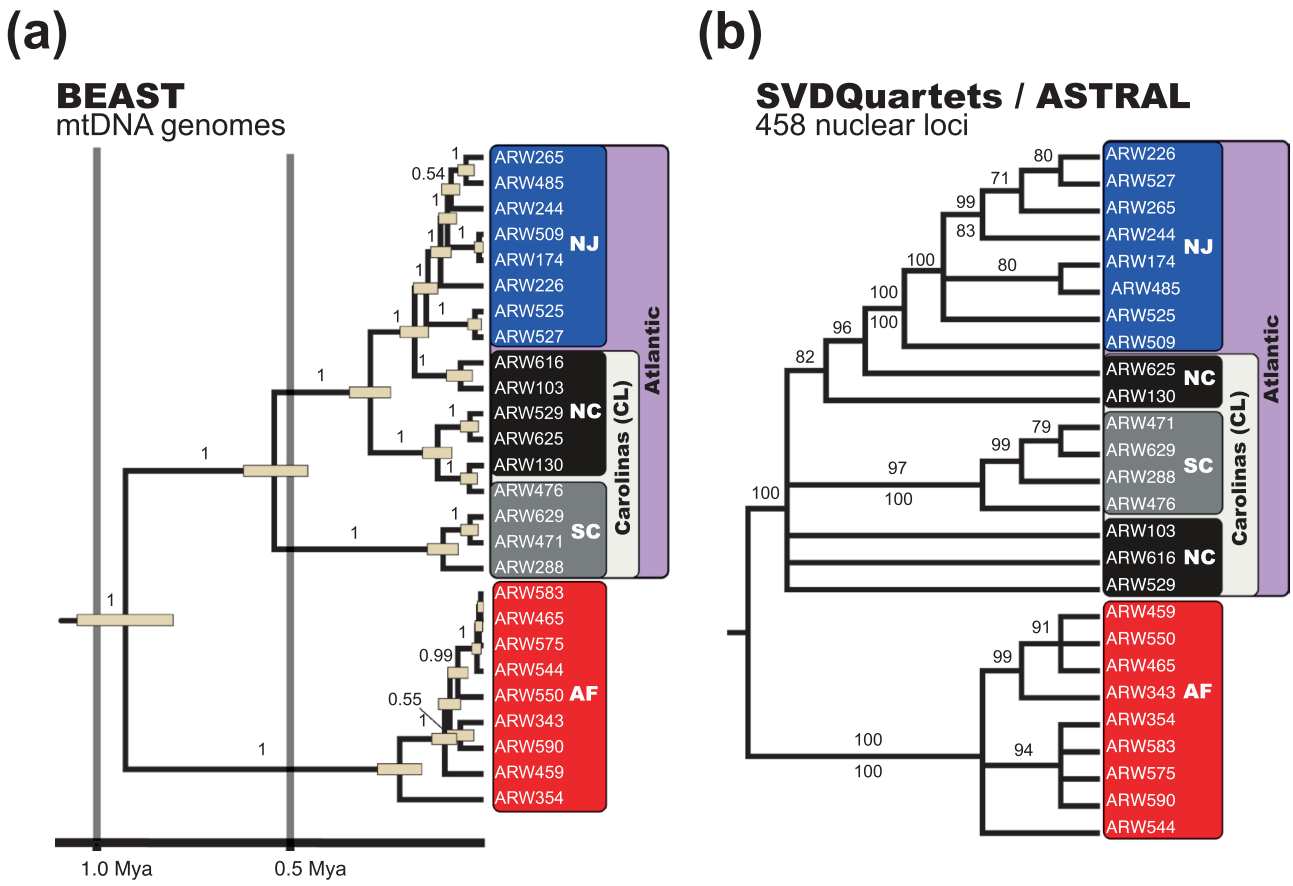

Figure 3. Phylogenetic relationships within *Hyla andersonii* inferred from mtDNA genomes (A) and 458 nuclear loci (B). A, divergence times were estimated in BEAST using a mutation rate for *ND2*. Bars represent the 95% highest posterior density interval for node age, with the *x*-axis representing millions of years ago. Branch labels are posterior probabilities. B, species trees from SVDQUARTETS, with bootstrap support (above branch) and ASTRAL bootstrap support values (below branch). Branches with support < 70 were collapsed using TREEGRAPH v.2.14.0 (Stöver & Müller, 2010). Abbreviations: AF, Alabama/Florida; NC, North Carolina; NJ, New Jersey; SC, South Carolina.

MULTIPLE LINES OF EVIDENCE FOR MULTIPLE REFUGIA

The best model (model 2; posterior probability = 0.836) using 'delimitR' and the SNP dataset supported the existence of two refugia (AF and CL), with post-Pleistocene colonization of NJ. Models 3 and 4 received 13 and nine of the 500 decision tree votes, respectively, suggesting very little support for a three-refugia model. Model 2 received 307 of 500 of the decision tree votes, and model 1 received 171 votes. The overall error rate in 'delimitR' was 14.21%, and all models except model 3 (three refugia, with no gene flow) had error rates < 6% (Table 4). Model 3 was often (~50% of the time) misclassified as model 4. We suspect that this misclassification is a function of difficulties in distinguishing between a model with more recent divergence times in comparison to a model with more ancient divergence times but including migration.

Species distribution models using climate data also provided support for multiple refugia (Fig. 5; Supporting Information, Figs S3, S4). MAXENT models were generated with a thinned set of 260 localities (61 AF, 73 CL and 126 NJ; Supporting Information, Fig. S1). The average AUC was 0.977, and the variable that contributed most to model fit was the mean temperature of the wettest quarter (Bio8; 45.2%; Supporting Information, Table S3). The current model included high suitability for the known range and indicated suitable climate in two areas within the NACP where the species does not occur at present, Louisiana (LA) and the Delmarva peninsula (Fig. 5A). During each of three historical time points, multiple disjunct regions were predicted to include suitable climate for *H. andersonii*. The Holocene, LGM and LIG models indicated that the CL region was a stable refugium for the species, and they also predicted suitable areas in LA, which is west of the current range on the

Table 2. Results from PHYLOMAPPER randomization tests of phylogeographical association and directionality with 10 000 randomizations

		Phylogeographical association		Non-random directionality		A priori direction (north-east)	
Clade	Number of taxa	Dispersal distance	*P*-value	Mean displacement	*P*-value	Mean angle difference (radians)	*P*-value
Hyla andersonii	26	217.56	< 0.0001*	NA	NA	NA	NA
AF	9	73.21	< 0.0001*	0.2376	< 0.0001*	1.622	< 0.0001*
Atlantic	17	232.61	< 0.0001*	0.4569	< 0.0001*	1.342	< 0.0001*
NJ	8	78.89	< 0.0001*	0.1941	1	1.453	< 0.0001*

For the a priori direction test, north-east was chosen as the predicted direction. Directional tendency was inferred only for clades within *Hyla andersonii* to test predictions, and not for the full species. Asterisks indicate significance at the $\alpha = 0.05$ level.
Abbreviations: AF, Alabama/Florida; NA, not assessed; NJ, New Jersey.

Table 3. Likelihood ratio test of dispersal distance variation

Clade	Subclade	Dispersal distance	Dispersal classes	Model	LnL	χ^2	*P*-value
Hyla andersonii		217.54	1	H_O	−231.22	–	–
	AF	81.51	2	H_A	−224.63	13.18	0.0003*
	Atlantic	318.50					

Dispersal distance is in metres per generation (assuming a generation time of 1 year). $\alpha = 0.05$, d.f. = 1. A model with two dispersal classes is significantly better than a model with a single class, and the dispersal distance of the Atlantic clade is much larger than that of the AF clade.
Abbreviations: AF, Alabama/Florida; ; H_A, alternative model; H_O, null model; LnL, log likelihood.

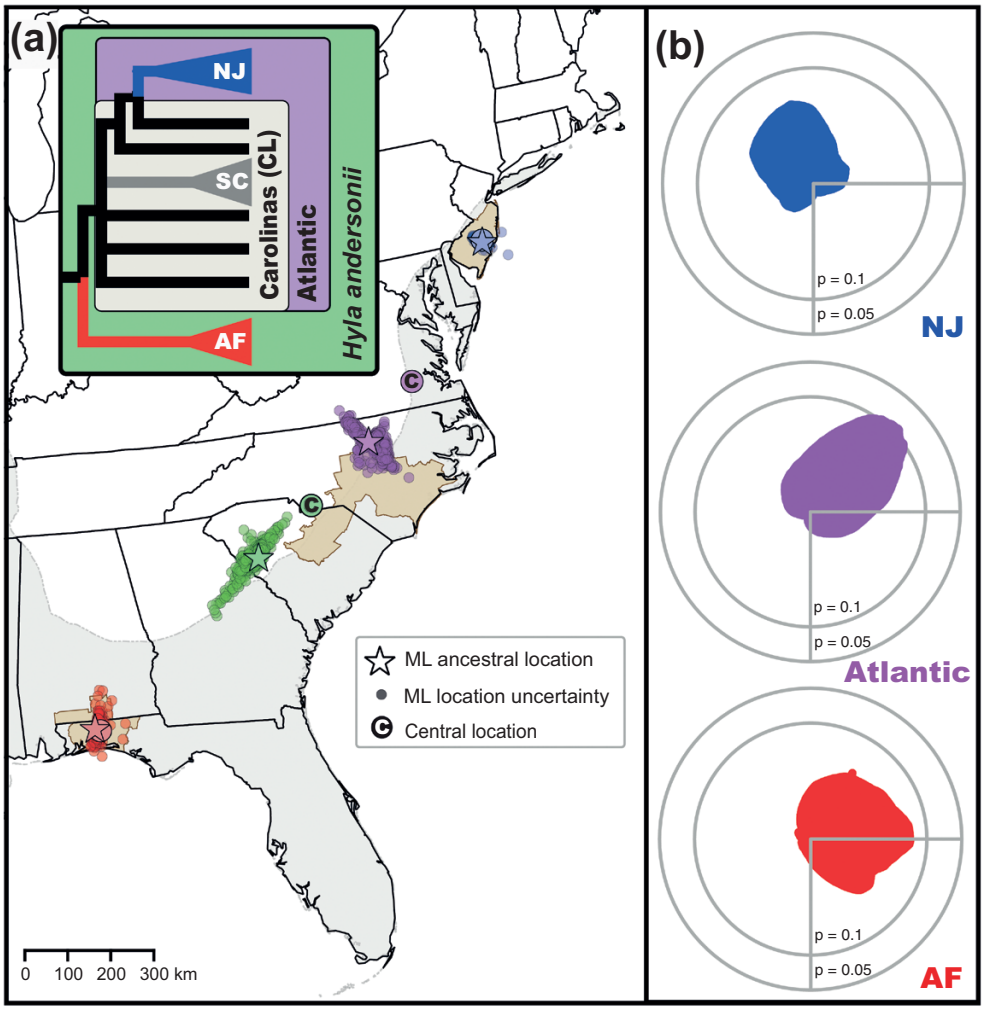

Figure 4. PHYLOMAPPER results, including the geographical locations of clade ancestors (A) and the directional tendency of migration (inkblot plots; B). The nuclear phylogeny is also shown in A. Abbreviations: AF, Alabama/Florida; CL, Carolinas; ML, maximum likelihood; NJ, New Jersey; SC, South Carolina. In A, the North American Coastal Plain is indicated on the map in grey. Stars represent the ML estimates of ancestral location for each clade. Small circles represent uncertainty in the ancestral location. Larger circles (labelled 'C') are the central locations. The NJ (blue) and AF (red) central locations overlap with the ML estimates. The inner circle around the inkblot is $P = 0.1$ and outer circle $P = 0.05$.

Gulf Coast (Fig. 5B–D). The Holocene and LGM models showed no suitable areas on the northern Atlantic coast (Fig. 5B, C), whereas the LIG model indicated suitable areas in New England, which is north of the current range (Fig. 5D).

DISCUSSION

Here, we tested predictions regarding the evolutionary history of *H. andersonii* using large-scale genomic and climate data. First, both nuclear species tree and mtDNA analyses confirmed the existence of two major clades, AF and Atlantic (containing CL and NJ individuals), with an estimated mtDNA divergence time between them of ~0.9 Mya. NJ also formed a clade, diverging ~0.15 Mya from the CL individuals. Second, our results are consistent with predictions of multiple refugia and north-eastern expansion along the Atlantic coast. We found the largest dispersal distance in the Atlantic clade, a significant north-eastern dispersal direction, the lowest genetic diversity in NJ, and support for a model with two refugia and recent colonization of NJ. Finally, the current SDM closely matched the known species range, and three historical models predicted disjunct areas of suitability. The

Table 4. Error rates for models tested in delimitR

Simulating model	Model 1 selected	Model 2 selected	Model 3 selected	Model 4 selected	Model error rate	Number of votes	Posterior probability
Model 1	0.9990	0.0009	0.0000	0.0001	0.0010	171/500	NA
Model 2*	0.0053	0.9920	0.0001	0.0026	0.0080	307/500	0.836
Model 3	0.0001	0.0088	0.4937	0.4974	0.5063	13/500	NA
Model 4	0.0001	0.0145	0.0385	0.9469	0.0531	9/500	NA

Each row corresponds to a model, and columns 2–5 report the proportion of the time each model was selected when data were simulated under that model for that row. Columns 6–8 report the overall error rate as a proportion for the model, the number of votes for each model, and the posterior probability of the best model.
Abbreviation: NA, not assessed.
*Best model.

CL region was probably a stable refugium, given its predicted suitability in all SDMs, and areas in the eastern Gulf Coast region outside the current range were also consistently predicted as suitable. Overall, this comprehensive phylogeographical study demonstrates a long history of range fragmentation within an endemic NACP species and highlights the influence of historical climate change on the current distribution of species and their genetic diversity.

Multiple refugia and north-eastern expansion

Using several lines of congruent evidence, we confirmed that *H. andersonii* populations in disjunct regions have probably been isolated for hundreds of thousands of years. Previous estimates from one or two localities per region left uncertainty in the mtDNA divergence times between the Atlantic and AF clades (~3 Mya, Lemmon *et al.*, 2007; ~0.46 Mya, Oswald *et al.*, 2020), which we clarified here (95% HPD: 0.80–1.05 Mya). Our SDM results support the proposal by Noss *et al.* (2015) that climatic variation and natural fragmentation generated by fluctuating sea levels have contributed to divergence and high endemism in the NACP. Historical SDMs ranging from ~6000 to 140 000 years ago predicted disjunct areas of suitability for *H. andersonii* across the Gulf and Atlantic coasts, and the best delimitR model supported two refugia. Plant species with similar distributions, such as Atlantic white cedar (*Chamaecyparis thyoides*) and *Sarracenia* pitcher plants, also exhibit genetic divergences between disjunct populations (Mylecraine *et al.*, 2004; Stephens *et al.*, 2015), the latter of which has divergence date estimates similar to *H. andersonii* (~1.1–1.3 Mya; Ellison *et al.*, 2012). Few vertebrate species in the NACP, however, have such a high degree of range disjunction as *H. andersonii* (Newman & Austin, 2019), perhaps as a result of these historical periods of fragmentation, in addition to its habitat specificity. In contrast, pine snakes (*Pituophis melanoleucus*), for example, have a similar disjunction between the Carolinas and New Jersey, yet are fairly continuous between Florida and the Carolinas (Means, 2006).

Quantitative tests in PhyloMapper supported the prediction of north-eastern expansion. Dispersal estimates within the northern clade were larger than AF, which is consistent with north-eastward invasion (Lemmon & Lemmon, 2008). The per-generation dispersal distance estimates ranged from ~70 to 320 m, which was reasonable compared with the single study that estimated dispersal in one NJ population (0–20 m daily movements, 40–100 m distance from breeding ponds; (Freda & Gonzalez, 1986). In addition, larger-bodied animals are expected to show greater dispersal abilities (Paz *et al.*, 2015).

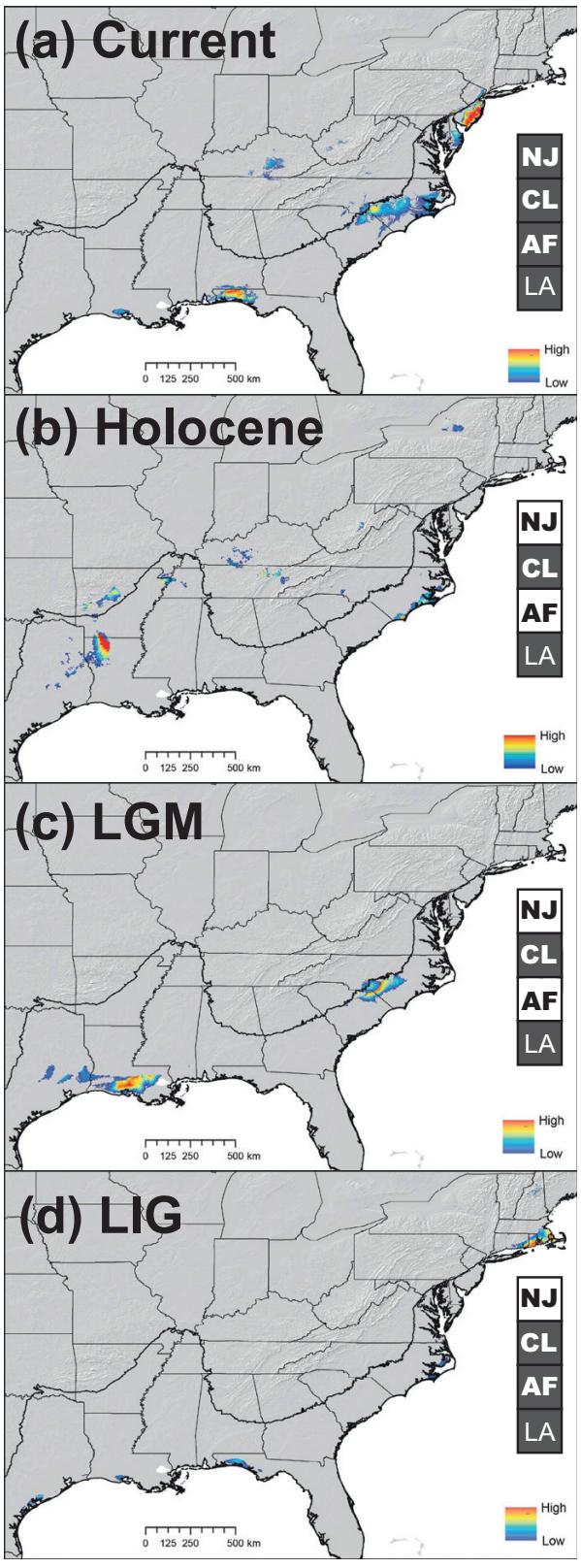

Figure 5. Predicted suitability at the 95% lowest presence threshold for each time period (LGM, Last Glacial Maximum; LIG, Last Interglacial), with predicted presence in each region summarized in boxes (AF, Alabama/Florida; CL, North and South Carolina; NJ, New Jersey). Louisiana (LA) is outside the current range but consistent across all models. Warm colours indicate higher suitability for *Hyla andersonii* occurrence. The Fall Line is indicated, which represents the northern boundary of the North American Coastal Plain physiographical region. Only the contemporary coastline is shown across all time periods, although climate data included relevant changes in coastline.

Consistent with this prediction, NJ individuals were significantly larger in snout–vent length than those in the other two regions (mean = 3.65 cm; ~0.1 cm larger on average; ANOVA, d.f. = 2, $F = 20.4995$, $P < 0.001$; Warwick *et al.*, 2015, with additional unpublished data from AR Warwick; Supporting Information, Fig. S5). Post-glacial northern expansion in this region has also been inferred for other anuran species: *Pseudacris crucifer* (Austin *et al.*, 2002), *Pseudacris feriarum* (Lemmon & Lemmon, 2008), *Hyla cinerea* and *Rana sphenocephala* (Barrow *et al.*, 2017), demonstrating a similar influence of climate on species ranges.

Future directions

Spatially explicit and model-based methods are important tools that have added rigour for inference in phylogeography, but some limitations should be addressed (Bradburd & Ralph, 2019; Mable, 2019). First, PHYLOMAPPER does not currently take into account landscape information when estimating movement direction and ancestral areas. This issue might be most problematic when generating null distributions, because random movement in all directions, such as towards the ocean, is not realistic. Second, our PHYLOMAPPER analyses only infer mtDNA genome history because resolved gene trees are needed; thus, our inferences might not reflect the entire species history. Although delimitR enables analysis of massive SNP datasets, we analysed ~600 SNPs to minimize missing data and avoid violating the assumption of independence among loci. We also designed models to test our specific predictions rather than exploring a large portion of model space that might not correspond to biologically realistic scenarios for the species under study. Future analyses using more SNPs could allow additional, more detailed models of gene flow and population size to provide further insights in this system.

In the current SDM, areas of suitability were generally consistent with the species distribution, with two discrepancies within the NACP, the Delmarva peninsula and Louisiana, where the

current absence of *H. andersonii* might be explained by human effects, such as land use. The Delmarva is located between CL and NJ, and thus it is plausible that *H. andersonii* might have inhabited the area previously and has since been extirpated, but we are unaware of any historical records from this region. Interestingly, the predicted species distribution of *H. andersonii* under the current climate matches closely with the known distribution of Atlantic white cedar (Supporting Information, Fig. S6; Little, 1971), including the same area within the Delmarva. Future modelling could test the influence of other environmental and landscape data layers (e.g. vegetation, soil, fire frequency; (Pekin *et al.*, 2012; Laliberté *et al.*, 2013), although hindcasting SDMs was only possible using climate in this study. The historical climate data were also limited to three time points; thus, we lacked finer resolution during climate conditions that aligned with mtDNA divergence estimates. It is possible that suitable habitat might have occurred in the intervening times and served as a corridor for movement. Finally, we acknowledge that when modelling any species distribution with palaeoclimatic data we assume the species climate envelope has not changed over time (Richards *et al.*, 2007). Given the habitat specialization of *H. andersonii*, this assumption might be reasonable.

Conclusions

We found evidence that *H. andersonii* has existed in disjunct regions for much of its evolutionary history. Given its habitat specialization and the continued loss and degradation of wetland habitat (Moler *et al.*, 2020; Oswald *et al.*, 2020), however, it might be more difficult for this species to track climate shifts and persist into the future. In addition, the relatively deep divergence between the AF and Atlantic clades indicates that these areas should be managed separately. Although we do not propose taxonomic changes to *H. andersonii*, particularly given the concordance of behavioural and morphometric data across regions (Warwick *et al.*, 2015), our results show that the AF and Atlantic clades represent distinct evolutionary histories, with little to no recent gene flow. Management efforts within regions would benefit from higher-resolution, population-level estimates of genetic diversity. In addition, the inability to detect the species from historical localities in recent surveys, especially in CL (Warwick *et al.*, 2015), is of concern, because CL shows the highest genetic diversity. Although the current populations appear stable in Florida (Moler *et al.*, 2020), the species is only found actively in four counties in AF (one in Alabama and three in Florida) and was recently delisted in Florida. Future management efforts of this species should be considered to secure the AF clade, given its distinct evolutionary history from the Atlantic populations. Taken together, our comprehensive dataset and inferences from genomic and climatic data have uncovered the complex history of a uniquely fragmented habitat specialist in this important area for biodiversity and phylogeographical discovery.

ACKNOWLEDGEMENTS

All procedures were approved under Florida State University (FSU) Animal Care and Use Committee protocols 1015 and 1624, and collection permits were obtained from each state and landowner. A.R.W. and L.N.B. contributed equally to the work. We greatly appreciate field assistance from many students and colleagues, in addition to species historical records. We also thank S. Banker, K. Birch, M. Conry, S. A. Emme, A. Hassinger, S. Holland, A. Jalal and M. Kortyna for laboratory and computational assistance. Financial support was provided by an FSU Dissertation Research Grant, a New Jersey Department of Environmental Protection Conserve Wildlife Grant, a Society for the Study of Evolution Rosemary Grant Award, an American Museum of Natural History Theodore Roosevelt Memorial Grant and a Sigma Xi Grants-in-Aid of Research to A.R.W. This material is based upon work supported by the National Science Foundation under grant no. 1120516 to E.M.L. We thank the Carstens laboratory group and three anonymous reviewers for their helpful comments.

REFERENCES

Adamack AT, **Gruber B. 2014.** PopGenReport: simplifying basic population genetic analyses in R. *Methods in Ecology and Evolution* **5:** 384–387.

Aiello-Lammens ME, **Boria RA**, **Radosavljevic A**, **Vilela B**, **Anderson RP. 2015.** spThin: an R package for spatial thinning of species occurrence records for use in ecological niche models. *Ecography* **38:** 541–545.

Austin JD, **Lougheed SC**, **Neidrauer L**, **Chek AA**, **Boag PT. 2002.** Cryptic lineages in a small frog: the post-glacial history of the spring peeper, *Pseudacris crucifer* (Anura: Hylidae). *Molecular Phylogenetics and Evolution* **25:** 316–329.

Avise JC. 2000. *Phylogeography: the history and formation of species*. Cambridge: Harvard University Press.

Avise JC, **Arnold J**, **Ball RM**, **Bermingham E**, **Lamb T**, **Neigel JE**, **Reeb CA**, **Saunders NC. 1987.** Intraspecific phylogeography: the mitochondrial DNA bridge between

population genetics and systematics. *Annual Review of Ecology and Systematics* **18:** 489–522.

Banker SE, **Lemmon AR**, **Hassinger AB**, **Dye M**, **Holland SD**, **Kortyna ML**, **Ospina OE**, **Ralicki H**, **Lemmon EM. 2020.** Hierarchical hybrid enrichment: multitiered genomic data collection across evolutionary scales, with application to chorus frogs (*Pseudacris*). *Systematic Biology* **69:** 756–773.

Barrow LN, **Lemmon AR**, **Lemmon EM. 2018.** Targeted sampling and target capture: assessing phylogeographic concordance with genome-wide data. *Systematic Biology* **67:** 979–996.

Barrow LN, **Soto-Centeno JA**, **Warwick AR**, **Lemmon AR**, **Lemmon EM. 2017.** Evaluating hypotheses of expansion from refugia through comparative phylogeography of southeastern Coastal Plain amphibians. *Journal of Biogeography* **44:** 2692–2705.

Bell RC, **MacKenzie JB**, **Hickerson MJ**, **Chavarria KL**, **Cunningham M**, **Williams S**, **Moritz C. 2012.** Comparative multi-locus phylogeography confirms multiple vicariance events in co-distributed rainforest frogs. *Proceedings of the Royal Society B: Biological Sciences* **279:** 991–999.

Bradburd GS, **Ralph PL. 2019.** Spatial population genetics: it's about time. *Annual Review of Ecology, Evolution, and Systematics* **50:** 427–449.

Chifman J, **Kubatko L. 2014.** Quartet inference from SNP data under the coalescent model. *Bioinformatics* **30:** 3317–3324.

Chifman J, **Kubatko L. 2015.** Identifiability of the unrooted species tree topology under the coalescent model with time-reversible substitution processes, site-specific rate variation, and invariable sites. *Journal of Theoretical Biology* **374:** 35–47.

Crawford AJ. 2003. Huge populations and old species of Costa Rican and Panamanian dirt frogs inferred from mitochondrial and nuclear gene sequences. *Molecular Ecology* **12:** 2525–2540.

Ellison AM, **Butler ED**, **Hicks EJ**, **Naczi RFC**, **Calie PJ**, **Bell CD**, **Davis CC. 2012.** Phylogeny and biogeography of the carnivorous plant family Sarraceniaceae. *PLoS One* **7:** e39291.

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

Fahrig L. 2003. Effects of habitat fragmentation on biodiversity. *Annual Review of Ecology, Evolution, and Systematics* **34:** 487–515.

Frankham R. 2005. Genetics and extinction. *Biological Conservation* **126:** 131–140.

Freda J, **Gonzalez RJ. 1986.** Daily movements of the Treefrog, *Hyla andersoni*. *Journal of Herpetology* **20:** 469–471.

Garrick RC, **Bonatelli IAS**, **Hyseni C**, **Morales A**, **Pelletier TA**, **Perez MF**, **Rice E**, **Satler JD**, **Symula RE**, **Thomé MTC**, **Carstens BC. 2015.** The evolution of phylogeographic datasets. *Molecular Ecology* **24:** 1164–1171.

Godt MJW, **Hamrick JL. 1998.** Genetic divergence among infraspecific taxa of *Sarracenia purpurea*. *Systematic Botany* **23:** 427–438.

Hamilton CA, **Lemmon AR**, **Lemmon EM**, **Bond JE. 2016.** Expanding anchored hybrid enrichment to resolve both deep and shallow relationships within the spider tree of life. *BMC Evolutionary Biology* **16:** 212–212.

Hammerson GA. **2017.** *Dryophytes andersonii* (amended version of 2004 assessment). The IUCN Red List of Threatened Species 2017: e.T10350A112711185. doi:10.2305/IUCN.UK.2017-1.RLTS.T10350A112711185.en (3 March 2021, date last accessed).

Haq BU, **Hardenbol J**, **Vail PR. 1987.** Chronology of fluctuating sea levels since the Triassic. *Science* **235:** 1156–1167.

Hijmans RJ, **Cameron SE**, **Parra JL**, **Jones PG**, **Jarvis A. 2005.** Very high resolution interpolated climate surfaces for global land areas. *International Journal of Climatology* **25:** 1965–1978.

Hime PM, **Lemmon AR**, **Lemmon ECM**, **Prendini E**, **Brown JM**, **Thomson RC**, **Kratovil JD**, **Noonan BP**, **Pyron RA**, **Peloso PLV**, **Kortyna ML**, **Keogh JS**, **Donnellan SC**, **Mueller RL**, **Raxworthy CJ**, **Kunte K**, **Ron SR**, **Das S**, **Gaitonde N**, **Green DM**, **Labisko J**, **Che J**, **Weisrock DW. 2021.** Phylogenomics reveals ancient gene tree discordance in the amphibian tree of life. *Systematic Biology* **70:** 49–66.

Hochheimer J, **Hoffmann MH. 2016.** Footprints of the Last Glacial: Past competition may have shaped strongly disjunct distribution ranges in *Artemisia rupestris* and *A. laciniata*. *Flora* **224:** 203–210.

Igawa T, **Kurabayashi A**, **Usuki C**, **Fujii T**, **Sumida M. 2008.** Complete mitochondrial genomes of three neobatrachian anurans: a case study of divergence time estimation using different data and calibration settings. *Gene* **407:** 116–129.

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

Karlin AA, **Means DB**, **Guttman SI**, **Lambright DD. 1982.** Systematics and the status of *Hyla andersonii* (Anura: Hylidae) in Florida. *Copeia* **1982:** 175–178.

Katoh K, **Misawa K**, **Kuma K**, **Miyata T. 2002.** MAFFT: a novel method for rapid multiple sequence alignment based on fast Fourier transform. *Nucleic Acids Research* **30:** 3059–3066.

Katoh K, **Standley DM. 2013.** MAFFT multiple sequence alignment software version 7: improvements in performance and usability. *Molecular Biology and Evolution* **30:** 772–780.

Laliberté E, **Grace JB**, **Huston MA**, **Lambers H**, **Teste FP**, **Turner BL**, **Wardle DA. 2013.** How does pedogenesis drive plant diversity? *Trends in Ecology & Evolution* **28:** 331–340.

Lemmon AR, **Emme SA**, **Lemmon EM. 2012.** Anchored hybrid enrichment for massively high-throughput phylogenomics. *Systematic Biology* **61:** 727–744.

Lemmon AR, **Lemmon EM. 2008.** A likelihood framework for estimating phylogeographic history on a continuous landscape. *Systematic Biology* **57:** 544–561.

Lemmon EM, **Lemmon AR**, **Cannatella DC. 2007.** Geological and climatic forces driving speciation in the continentally distributed trilling chorus frogs (*Pseudacris*). *Evolution; International Journal of Organic Evolution* **61:** 2086–103.

Lips KR. 2016. Overview of chytrid emergence and impacts on amphibians. *Philosophical Transactions of the Royal Society B: Biological Sciences* **371:** 20150465.

Little EL Jr. 1971. *Atlas of United States Trees: Volume 1. Conifers and Important Hardwoods*. United States Department of Agriculture, Forest Service, Miscellaneous Publication 1146. Washington, DC: United States Government Printing Office.

Liu F-GR, Moler PE, Miyamoto MM. 2006. Phylogeography of the salamander genus *Pseudobranchus* in the southeastern United States. *Molecular Phylogenetics and Evolution* **39:** 149–159.

Mable BK. 2019. Conservation of adaptive potential and functional diversity: integrating old and new approaches. *Conservation Genetics* **20:** 89–100.

Macey JR, Schulte JA, Larson A, Fang Z, Wang Y, Tuniyev BS, Papenfuss TJ. 1998. Phylogenetic relationships of toads in the *Bufo bufo* species group from the eastern escarpment of the Tibetan Plateau: a case of vicariance and dispersal. *Molecular Phylogenetics and Evolution* **9:** 80–87.

Marshall SJ, James TS, Clarke GKC. 2002. North American Ice Sheet reconstructions at the Last Glacial Maximum. *Quaternary Science Reviews* **21:** 175–192.

Means DB. 2006. Vertebrate faunal diversity of longleaf pine ecosystems. In: Jose S, Jokela EJ, Miller DL, eds. *The longleaf pine ecosystem: ecology, silviculture, and restoration*. New York: Springer, 157–213.

Meyer M, Kircher M. 2010. Illumina sequencing library preparation for highly multiplexed target capture and sequencing. *Cold Spring Harbor Protocols* **2010:** 10.1101/pdb.prot5448-10.1101/pdb.prot5448.

Mirarab S, Bayzid MS, Warnow T. 2014. Evaluating summary methods for multilocus species tree estimation in the presence of incomplete lineage sorting. *Systematic Biology* **65:** 366–380.

Mirarab S, Warnow T. 2015. ASTRAL-II: coalescent-based species tree estimation with many hundreds of taxa and thousands of genes. *Bioinformatics* **31:** i44–i52.

Moler PE, Enge KM, Tornwall B, Farmer AL, Harris BB. 2020. Status and current distribution of the Pine Barrens Treefrog (*Hyla andersonii*) in Florida. *Southeastern Naturalist* **19:** 380–394.

Mylecraine KA, Kuser JE, Smouse PE, Zimmermann GL. 2004. Geographic allozyme variation in Atlantic white-cedar, *Chamaecyparis thyoides* (Cupressaceae). *Canadian Journal of Forest Research* **34:** 2443–2454.

Newman CE, Austin CC. 2019. Quantifying amphibian range fragmentation in the southeastern United States. *Frontiers of Biogeography* **11:** 1. doi:10.21425/F5FBG37772

Noss RF, Platt WJ, Sorrie BA, Weakley AS, Means DB, Costanza J, Peet RK. 2015. How global biodiversity hotspots may go unrecognized: lessons from the North American Coastal Plain. *Diversity and Distributions* **21:** 236–244.

Oswald KJ, Roberts MA, Moler PE, Arndt RG, Camper JD, Quattro JM. 2020. Wetlands, evolution, and conservation of the Pine Barrens Treefrog (*Hyla andersonii*). *Journal of Herpetology* **54:** 206–215.

Paz A, Ibáñez R, Lips KR, Crawford AJ. 2015. Testing the role of ecology and life history in structuring genetic variation across a landscape: a trait-based phylogeographic approach. *Molecular Ecology* **24:** 3723–3737.

Pearson RG, Raxworthy CJ, Nakamura M, Townsend Peterson A. 2007. Predicting species distributions from small numbers of occurrence records: a test case using cryptic geckos in Madagascar. *Journal of Biogeography* **34:** 102–117.

Pekin BK, Boer MM, Wittkuhn RS, Macfarlane C, Grierson PF. 2012. Plant diversity is linked to nutrient limitation of dominant species in a world biodiversity hotspot. *Journal of Vegetation Science* **23:** 745–754.

Phillips S, Anderson R, Schapire R. 2006. Maximum entropy modeling of species geographic distributions. *Ecological Modelling* **190:** 231–259.

Prum RO, Berv JS, Dornburg A, Field DJ, Townsend JP, Lemmon EM, Lemmon AR. 2015. A comprehensive phylogeny of birds (Aves) using targeted next-generation DNA sequencing. *Nature* **526:** 569–573.

Pudlo P, Marin JM, Estoup A, Cornuet JM, Gautier M, Robert CP. 2016. Reliable ABC model choice via random forests. *Bioinformatics* **32:** 859–866.

Pyron RA, Hsieh FW, Lemmon AR, Lemmon EM, Hendry CR. 2016. Integrating phylogenomic and morphological data to assess candidate species-delimitation models in brown and red-bellied snakes (*Storeria*). *Zoological Journal of the Linnean Society* **177:** 937–949.

R Core Team. 2019. *R: a language and environment for statistical computing*. Vienna: R Foundation for Statistical Computing. Available at: https://www.R-project.org/

Rambaut A, Drummond AJ, Xie D, Baele G, Suchard MA. 2018. Posterior summarization in Bayesian phylogenetics using Tracer 1.7. *Systematic Biology* **67:** 901–904.

Reilly SB, Portik DM, Koo MS, Wake DB. 2014. Discovery of a new, disjunct population of a narrowly distributed salamander (*Taricha rivularis*) in California presents conservation challenges. *Journal of Herpetology* **48:** 371–379.

Rice BA. 2011. The thread-leaf sundews *Drosera filiformis* and *Drosera tracyi*. *Carnivorous Plant Newsletter* **40:** 4–16.

Richards CL, Carstens BC, Knowles LL. 2007. Distribution modelling and statistical phylogeography: an integrative framework for generating and testing alternative biogeographical hypotheses. *Journal of Biogeography* **34:** 1833–1845.

Rokyta DR, Lemmon AR, Margres MJ, Aronow K. 2012. The venom-gland transcriptome of the eastern diamondback rattlesnake (*Crotalus adamanteus*). *BMC Genomics* **13:** 312–312.

Sheppard CS. 2013. How does selection of climate variables affect predictions of species distributions? A case study of three new weeds in New Zealand. *Weed Research* **53:** 259–268.

Smith ML, Carstens BC. 2020. Process-based species delimitation leads to identification of more biologically relevant species. *Evolution; International Journal of Organic Evolution* **74:** 216–229.

Soltis DE, **Morris AB**, **McLachlan JS**, **Manos PS**, **Soltis PS. 2006.** Comparative phylogeography of unglaciated eastern North America. *Molecular Ecology* **15:** 4261–4293.

Sorrie BA, **Weakley AS. 2001.** Coastal plain vascular plant endemics: phytogeographic patterns. *Castanea* **66:** 50–82.

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

Stephens JD, **Rogers WL**, **Heyduk K**, **Cruse-Sanders JM**, **Determann RO**, **Glenn TC**, **Malmberg RL. 2015.** Resolving phylogenetic relationships of the recently radiated carnivorous plant genus *Sarracenia* using target enrichment. *Molecular Phylogenetics and Evolution* **85:** 76–87.

Stöver BC, **Müller KF. 2010.** TreeGraph 2: combining and visualizing evidence from different phylogenetic analyses. *BMC Bioinformatics* **11:** 7.

Stuart SN, **Chanson JS**, **Cox NA**, **Young BE**, **Rodrigues ASL**, **Fischman DL**, **Waller RW. 2004.** Status and trends of amphibian declines and extinctions worldwide. *Science* **306:** 1783–1786.

Suchard MA, **Lemey P**, **Baele G**, **Ayres DL**, **Drummond AJ**, **Rambaut A. 2018.** Bayesian phylogenetic and phylodynamic data integration using BEAST 1.10. *Virus Evolution* **4:** vey016.

Swenson NG, **Howard DJ. 2005.** Clustering of contact zones, hybrid zones, and phylogeographic breaks in North America. *The American Naturalist* **166:** 581–591.

Swofford DL. 2002. *PAUP*. Phylogenetic analysis using parsimony (*and other methods)*. Sunderland: Sinauer Associates.

Warwick AR, **Barrow LN**, **Smith ML**, **Means DB**, **Lemmon AR**, **Lemmon EM. 2021.** Data from: Signatures of north-eastern expansion and multiple refugia: genomic phylogeography of the Pine Barrens tree frog, *Hyla andersonii* (Anura: Hylidae). *Dryad, Dataset*. https://doi.org/10.5061/dryad.59zw3r26j

Warwick AR, **Travis J**, **Lemmon EM. 2015.** Geographic variation in the Pine Barrens Treefrog (*Hyla andersonii*): concordance of genetic, morphometric and acoustic signal data. *Molecular Ecology* **24:** 3281–3298.

Williams JW, **Shuman BN**, **Thompson W**, **Bartlein PJ**, **Leduc PL. 2004.** Late-Quaternary vegetation dynamics in North America: scaling from taxa to biomes. **74:** 309–334.

SUPPORTING INFORMATION

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

Table S1. Specimens used for genomic data collection.
Table S2. GenBank accession numbers for *Hyla femoralis* genes used to construct a reference mitochondrial genome for assembly in SEQMAN NGEN.
Table S3. All 19 bioclimatic layers used for species distribution models.
Table S4. All *Hyla andersonii* locality sources by state used as input for species distribution models.
Table S5. Likelihood ratio tests for geographical location.
Figure S1. All *Hyla andersonii* occurrence localities that were thinned to 5 km nearest-neighbour distances and used as input for MAXENT species distribution models.
Figure S2. The distribution of background localities used in MAXENT, which was generated using a 400 km radius around all occurrence records.
Figure S3. Species distribution models for each time point using all 19 climatic variables. The raw output from MAXENT is shown.
Figure S4. Species distribution models for all four time points as binary outputs.
Figure S5. Range of snout–vent lengths (in centimetres) for males only from each of the three regions.
Figure S6. The species distribution model for *Hyla andersonii* in current climatic conditions compared with the known distribution of Atlantic white cedar (*Chamaecyparis thyoides*).

SHARED DATA

Sequences are available on GenBank (MW002696–MW002723), additional data files on Dryad (Warwick *et al.*, 2021), and remaining tissues and extracts are available at the Museum of Southwestern Biology, University of New Mexico.