

# UC Merced

## Frontiers of Biogeography

### Title

Does habitat stability structure intraspecific genetic diversity? It's complicated...

### Permalink

<https://escholarship.org/uc/item/4jq256x0>

### Journal

Frontiers of Biogeography, 12(2)

### Authors

Wieringa, Jamin G  
Boot, Matthew R.  
Dantas-Queiroz, Marcos V.  
[et al.](#)

### Publication Date

2020

### DOI

10.21425/F5FBG45377

### Supplemental Material

<https://escholarship.org/uc/item/4jq256x0#supplemental>

### Copyright Information

Copyright 2020 by the author(s). This work is made available under the terms of a Creative Commons Attribution License, available at <https://creativecommons.org/licenses/by/4.0/>

Peer reviewed



## Does habitat stability structure intraspecific genetic diversity? It's complicated...

Jamin G. Wieringa<sup>1</sup>, Matthew R. Boot<sup>1,2</sup>, Marcos V. Dantas-Queiroz<sup>3</sup>, Drew Duckett<sup>1,2</sup>, Emanuel M. Fonseca<sup>1,2</sup>, Heather Glon<sup>1,2</sup>, Natalie Hamilton<sup>1,2</sup>, Sungsik Kong<sup>1</sup>, Flavia M. Lanna<sup>1,2</sup>, Kali Z. Mattingly<sup>1</sup>, Danielle J. Parsons<sup>1,2</sup>, Megan L. Smith<sup>1,2</sup>, Benjamin W. Stone<sup>1</sup>, Coleen Thompson<sup>1,2</sup>, Lei Zuo<sup>1,2</sup> and Bryan C. Carstens<sup>1,2\*</sup>

<sup>1</sup>Department of Evolution, Ecology and Organismal Biology, The Ohio State University, 318 W. 12<sup>th</sup> Ave., Columbus, OH, USA. 43210; <sup>2</sup>Museum of Biological Diversity, The Ohio State University, 1315 Kinnear Rd., Columbus, OH, USA, 43212; <sup>3</sup>Departamento de Ecologia, Programa de Pós-graduação em Biologia Vegetal, Universidade Estadual Paulista, Av. 24-A, 1515, Bela Vista, Rio Claro, SP, Brazil. 13506–900. \*correspondence: carstens.12@osu.edu; <https://carstenslab.osu.edu/>

### Abstract

Regional phylogeographic studies have long been conducted in the southeastern United States for a variety of species. With some exceptions, many of these studies focus on single species or single clades of organisms, and those considering multiple species tend to focus on deep historical breaks causing differentiation. However, in many species more recent factors may be influencing genetic diversity. To understand the roles of historic and contemporary processes in structuring genetic diversity, we reanalyzed existing genetic data from Southeast of North America using approaches gleaned from phylogeographic and landscape genetic literature that were implemented across species including AMOVAs, PCoAs, Species Distribution Modelling, and tests of isolation by distance, environment, and habitat instability. Genetic variance was significantly partitioned by ecoregions, watersheds, and across phylogeographic breaks in the majority of species. Similarly, genetic variation was significantly associated with some combination of geographic or environmental distance or habitat instability in most species. Patterns of genetic variation were largely idiosyncratic across species. While habitat instability over time is significantly correlated with genetic diversity in some species, it appears generally less important than isolation by geographic or environmental distance. Our results suggest that many factors, both historical and contemporary, impact genetic diversity within a species, and more so, that these patterns aren't always similar in closely related species. This supports the importance of species-specific factors and cautions against assumptions that closely related species will respond to historical and contemporary forces in similar ways.

### Highlights

- We used a variety of approaches to quantify factors that potentially influence genetic structure across multiple species in the Southeastern United States
- Our analyses indicate that both contemporary and historical factors promote genetic divergence across multiple taxa
- Machine learning analysis indicated that taxa are largely idiosyncratic in terms of which factors are most important, suggesting that specific factors regulate how particular taxa respond to external factors
- Our results highlight the importance of considering both landscape and phylogeographic methodology when investigating genetic structure

**Keywords:** Phylogeography, data repurposing, Pleistocene glaciation, comparative phylogeography, mtDNA, cpDNA, barriers

## Introduction

Phylogeography aims to understand how climatic and geologic events have structured genetic diversity; while comparative phylogeography seeks this understanding in multiple species. Traditionally, this discipline interprets similar patterns (e.g., genealogies, STRUCTURE plots, etc.) across species as evidence for a shared species response to climate and/or geologic events (Avice 1987, Sullivan et al. 2000). One region that has served as the setting for many early phylogeographic studies (e.g., Avice 2000; Soltis 2006) and has continued to capture the interest of researchers (e.g., Satler and Carstens 2016; Barrow et al. 2017) is the Southeastern United States. The geologic and climatic history of this region is complex, involving diverse processes acting over millions of years, including mountain orogeny, changing river basins, and climate cycles, and species may have been affected in diverse ways by these factors. Soltis et al. (2006) reviewed phylogeographic patterns from the Southeastern US and suggested that these patterns could be categorized into three categories: populations structured by 1) river basins, 2) mountain ranges, and 3) the locations of glacial refugia. However, these categories are not independent or exclusive; for example, species with population structure across a given river or the Appalachian Mountains may have been isolated across these barriers in separate refugia during Pleistocene glaciation.

Studies of congruence in phylogeography have often aimed to identify identical breakpoints and population structure (e.g. Avice 2000, Satler and Carstens 2016), a task that is complicated by differences in species-specific attributes, such as dispersal ability and ecological niche (Papadopoulou and Knowles 2016). Another strategy for identifying similarity in pattern across species is to assess the relative importance of different factors in structuring genetic diversity. This approach is popular in landscape genetics (Rissler 2016), where many studies aim to evaluate the relative effects of isolation-by-distance (IBD; Wright 1943) and isolation-by-environment (IBE). By considering habitat stability in such an approach we can potentially disentangle the relative contribution of both contemporary and historical factors structuring genetic diversity (e.g., Vasconcellos et al. 2019).

Comparative phylogeographic investigations have long used climate modeling to identify regions of stable habitat that putatively served as climate refugia (e.g., Hugall et al. 2002, Carstens & Richards 2007, Carnaval et al. 2009). The general interpretation of such results is that similarity in genetic structure across species is a product of species persistence during periods of climatic change in these regions of stable habitat. The relationship between habitat and genetic diversity is a feature of landscape genetic investigations (e.g., Orsini et al. 2013), particularly in combination with species distribution modeling of habitat suitability (e.g., Geber 2011). Therefore, analytical methods that explicitly consider the suitability of the environment in the context of the genetic structure of species enable integration between phylogeography and landscape genetics (Rissler 2016). One promising analysis was proposed by Vasconcellos et al. (2019),

who investigated the effects of climatic shifts on the genetic structure of a savanna frog from Brazil. Their strategy used an evaluation of the relative contributions of isolation by distance and isolation by habitat suitability modeled over evolutionary time (habitat instability). Vasconcellos et al. (2019) argue that the second measure, their 'isolation by instability (IBI)' model, is a special case of the isolation by resistance model developed by McRae (2006). Like McRae's model, which posits that genetic diversity is structured by the separation of individuals using corridors of suitable habitat rather than a straight line over geographic space, Vasconcellos et al.'s model predicts that genetic diversity is structured by the connectedness of individuals over evolutionary time. Use of a multiple matrix regression with randomization (MMRR) approach (Wang 2013) allowed Vasconcellos et al. to evaluate the relative contribution of habitat instability in comparison to other factors.

While habitat instability is one important component of historical effects on genetic diversity, it is not the only factor to consider. In addition to the effects of Pleistocene glacial cycles, the effects of river formation and mountain orogeny are evident in phylogeographic datasets from eastern North America (Soltis 2000, Avice 2006, Satler and Carstens 2016). To evaluate the effects of these and other phylogeographic breaks on structuring genetic diversity, we conduct analyses of molecular variance (AMOVAs; Excoffier 1992) and habitat stability in order to evaluate whether genetic diversity is significantly structured across river basins, ecoregions, or proposed phylogeographic breaks. In combination, these approaches allow us to evaluate the relative contributions of contemporary and various historic factors across species from eastern North America. Because they provide easily comparable metrics across species, the degree to which species have similar phylogeographic patterns and responses to the factors that lead to incongruence across species can be assessed. Here, we use repurposed data collected from a variety of sources to test the factors influencing genetic variation in a comparative framework in the Southeastern United States.

## Methods

### Data Processing

Fifty-seven species were identified from published phylogeographic studies in the Southeastern United States, with most of the taxa included by Soltis et al. (2006) supplemented by data from several more recent investigations (Table 1). Overall guidelines for processing are presented in Fig. S1. To start, genetic data were downloaded from GenBank for each species. Genetic data were then aligned using CLUSTAL (Thompson et al. 1994) before selecting appropriate models of sequence evolution using the Akaike Information Criterion implemented in jModeltest ver 2.1.10 (Darriba et al. 2012, Guindon and Gascuel 2003) and/or MrModeltest ver 2.4 (Nylander 2004). Models implemented (Table S1) included K80 (Kimura

**Table 1.** Species used in analysis. For each species, the scientific name, type of organism, type of data, number of sequences, and reference of original publication is shown.

Species	Broad Taxon	Type of Data	# sequences	Original publication
<i>Bryopsis</i> sp.	Green Algae	cpDNA	66	Krellwitz et al. (2001)
<i>Gracilaria tikvahiae</i>	Red Algae	cpDNA	20	Gurgel et al.(2004)
<i>Xerula furfuracea</i>	Fungi	nuDNA	41	Yang et al.(2009) & Petersen and Hughes (2010) & Hao et al.(2016)
<i>Sphagnum bartlettianum</i>	Bryophyta	cpDNA + nuDNA	12	Shaw et al.(2005)
<i>Acer rubrum</i>	Angiosperm	cpDNA	38	McLachlan et al.(2005)
<i>Apios americana</i>	Angiosperm	nuDNA	18	Joly & Bruneau (2004)
<i>Dicerandra spp</i>	Angiosperm	cpDNA	30	Oliveira et al.(2007)
<i>Fagus grandifolia</i>	Angiosperm	cpDNA	23	McLachlan et al.(2005)
<i>Liquidambar styraciflua</i>	Angiosperm	cpDNA	109	Morris et al.(2008)
<i>Prunus spp</i>	Angiosperm	cpDNA	226	Shaw & Small (2005)
<i>Tilia americana</i>	Angiosperm	cpDNA	297	McCarthy and Mason-Gamer (2016)
<i>Trillium cuneatum</i>	Angiosperm	cpDNA	281	Gonzales et al.(2008)
<i>Uniola paniculata</i>	Angiosperm	cpDNA	131	Hodel & Gonzales (2013)
<i>Bugula neritina</i>	Bryozoa	mtDNA	30	McGovern & Hellberg (2003)
<i>Daphnia obtusa</i>	Crustacean	mtDNA	36	Penton et al.(2004)
<i>Emerita talpoida</i>	Crustacean	mtDNA	4	Tam et al.(1996)
<i>Farfantepenaeus aztecus</i>	Crustacean	mtDNA	76	McMillen-Jackson and Bert (2003)
<i>Litopenaeus setiferus</i>	Crustacean	mtDNA	92	McMillen-Jackson and Bert (2003) & Maggioni et al. (2001) & Vazquez-Bader et al.(2004) & Bremer et al.(2010)
<i>Pagarus longicarpus</i>	Crustacean	mtDNA	67	Young et al.(2002)
<i>Pagarus pollicaris</i>	Crustacean	mtDNA	13	Young et al.(2002)
<i>Busycon sinistrum</i>	Gastropod	mtDNA	31	Wise et al.(2004)
<i>Lampsilis altilis</i>	Mollusk	mtDNA	5	Roe et al.(2001)
<i>Lampsilis australis</i>	Mollusk	mtDNA	5	Roe et al.(2001)
<i>Lampsilis ovata</i>	Mollusk	mtDNA	2	Roe et al.(2001) & Campbell et al.(2005)
<i>Lampsilis perovalis</i>	Mollusk	mtDNA	5	Roe et al.(2001)
<i>Lampsilis teres</i>	Mollusk	mtDNA	2	Roe et al.(2001) & Lydeard et al.(2000)
<i>Spisula solidissima</i>	Mollusk	mtDNA	52	Hare and Weinberg (2005)
<i>Ambystoma tigrinum</i>	Amphibian	mtDNA	56	Church et al.(2003)
<i>Desmognathus wrightii</i>	Amphibian	mtDNA	29	Crespi et al.(2003)
<i>Eumeces fasciatus</i>	Amphibian	mtDNA	82	Howes et al.(2006)
<i>Eurycea bislineata</i>	Amphibian	mtDNA	56	Kozak et al.(2006)
<i>Eurycea cirrigera</i>	Amphibian	mtDNA	251	Kozak et al.(2006)
<i>Eurycea junaluska</i>	Amphibian	mtDNA	6	Kozak et al.(2006)
<i>Eurycea multiplicata</i>	Amphibian	mtDNA	46	Bonett & Chippindale (2004)
<i>Eurycea tymerensis</i>	Amphibian	mtDNA	16	Bonett & Chippindale (2004)
<i>Eurycea wilderae</i>	Amphibian	mtDNA	129	Kozak et al.(2006)

Table 1. Continued...

Species	Broad Taxon	Type of Data	# sequences	Original publication
<i>Pseudacris brachyphona</i>	Amphibian	mtDNA	25	Lemmon et al.(2007)
<i>Pseudacris brimleyi</i>	Amphibian	mtDNA	8	Lemmon et al.(2007)
<i>Pseudacris clarkii</i>	Amphibian	mtDNA	9	Lemmon et al.(2007)
<i>Pseudacris crucifer</i>	Amphibian	mtDNA	62	Austin et al.(2004)
<i>Pseudacris feriarium</i>	Amphibian	mtDNA	61	Lemmon et al.(2007)
<i>Pseudacris nigrita</i>	Amphibian	mtDNA	19	Lemmon et al.(2007)
<i>Pseudacris spnov</i>	Amphibian	mtDNA	27	Lemmon et al.(2007)
<i>Rana catesbeiana</i>	Amphibian	mtDNA	213	Austin et al.(2004)
<i>Tryphlotriton spelaeus</i>	Amphibian	mtDNA	14	Bonett & Chippindale (2004)
<i>Acipenser oxyrinchus</i>	Fish	mtDNA	615	Grunwald et al.(2008)
<i>Carcharhinus limbatus</i>	Fish	mtDNA	10	Keeney et al.(2005)
<i>Hypentelium nigricans</i>	Fish	mtDNA	85	Berendzen et al.(2003)
<i>Percina caprodes</i>	Fish	mtDNA	15	Near (2008)
<i>Percina nasuta</i>	Fish	mtDNA	6	Robison et al.(2014)
<i>Percina phoxocephala</i>	Fish	mtDNA	34	Robison et al.(2014)
<i>Seriola dumerli</i>	Fish	mtDNA	16	Gold & Richardson (1998)
<i>Blarina brevicauda</i>	Mammal	mtDNA	74	Brant et al.(2003)
<i>Apalone ferox</i>	Reptile	mtDNA	9	Weisrock and Jenzen (2000)
<i>Apalone mutica</i>	Reptile	mtDNA	9	Weisrock and Jenzen (2000)
<i>Apalone spinifera</i>	Reptile	mtDNA	30	Weisrock and Jenzen (2000)
<i>Pantherophis obsoletus</i>	Reptile	mtDNA	39	Burbrink et al. 2000

1980), K81 (Kimura 1981), F81 (Felsenstein 1981), or TN93 (Tamura and Nei 1993), with gamma correction for rate variation in some species. For species with multiple genes, we chose the model that applied to the greatest number of base pairs and concatenated all cases of multiple genes. FASTA formatted data alignments were used to analyze sequence diversity and polymorphism using DnaSP ver 6 software (Rozas et al. 2017). Prior to calculating genetic distance matrices, we removed outgroups and dealt with missing data as follows. For 44 / 57 species, we deleted sites with at least one missing base across all sequences. Because this procedure deleted most sites for 13 species with many missing data, for these species we instead performed pairwise deletion of missing bases, and, for 10 of these, also replaced remaining undefined pairwise distances with the mean distance across all other sequence comparisons for that site. Adopting procedures for addressing missing data (Jombart 2008, Jombart and Ahmed 2011) allowed us to define comparable distance matrices for all species.

Geographic locality information was collected from source publications where possible. In cases where specific coordinates were not provided, we used sample accession numbers to check GenBank for sample-associated metadata. If specific coordinates could not be determined from either source, we used GEOlocate, an online platform for georeferencing data, to generate coordinates based on the most specific locality available (Rios and Bart 2010). Once the data was compiled, all coordinates, including those from original

sources, were double-checked using GEOlocate and then visually inspected using Google Earth. Generation length and substitution rate were obtained for each species from the literature (Supplementary Table S2).

### Nucleotide Mapping

In order to develop a qualitative understanding of how nucleotide diversity is distributed across the landscape, we created a heat map of nucleotide diversity across all species. To calculate nucleotide diversity, we divided North America into a 1° grid and, using geographic coordinates for each species, assigned individuals to a cell in this grid. As nucleotide diversity requires multiple individuals for comparison, grid cells with fewer than two individuals of a species were removed. In order to avoid sampling bias, species with fewer than three occupied cells were also removed from subsequent analyses. Nucleotide diversity ( $\pi$ ) was calculated within each grid cell using the R package “pegas” (Emmanuel et al. 2018). Given the challenges associated with using different molecular markers, we normalized each individual raster to ensure that species richness did not exert an undue influence on nucleotide diversity ( $\pi$ ) as follows:

$$\frac{\pi - \min(\pi)}{\max(\pi) - \min(\pi)} \quad (1)$$

Subsequently, GIS layers for each species were summed to create a final heat map that included all species to indicate the overall level of nucleotide



diversity across the region. While this conservative approach provides a simplistic impression of nucleotide diversity, it prevents overinterpretation that may result from differences in species richness or the use of genetic markers across species that differ in their rates of evolution and as such may exhibit different levels of haplotype diversity. In addition, Moran's I was calculated using the R package "ape" (Paradis et al. 2004) to assess autocorrelation under the assumption that significant positive autocorrelation in nucleotide diversity may be indicative of shared refugia across species.

### Analysis of molecular variation

We used analysis of molecular variance (AMOVA; Excoffier et al. 1992) to compare genetic variation across geographic features thought to structure populations. For each georeferenced species, we matched GPS data for individuals to geographic features delineated by shapefiles (sp::over, Pebesma and Bivand 2005). We used tools in ArcMap 10.3 (ESRI 2014) to create shapefiles from preexisting spatial data for two current landscape features, Ecoregion (CECWG 1997; Spalding et al. 2007; USEPA 2012; AAFC 2017) and Watershed (Spalding et al. 2007; NRC, USGS, and INEG 2010a), and four phylogeographic breaks described in Soltis et al. (2006): the Atlantic-Gulf Coast discontinuity (Spalding et al. 2007), the Apalachicola River-Tombigbee River discontinuity (which separates rivers flowing into the Gulf of Mexico from those flowing into the Atlantic Ocean; Spalding et al. 2007; NRC, USGS, and INEG 2010a), the Appalachian Mountains (UNEP-WCMC 2002), and the Mississippi River (NRC, USGS, and INEG 2010b). We implemented AMOVA in Arlequin using *arlecore* with 1000 permutations to test for significance (Excoffier et al. 2005). Outgroups and species lacking geographic data were omitted from these analyses.

### Population assignment

Intraspecific genetic variation is often structured by populations, and the characterization of population genetic structure was prominently featured by many of the original publications that described the data we use here (Table 1). This type of information is also often necessary for other types of analysis, as was the case here. To facilitate comparison across species, we inferred the number of genetic populations using an assignment protocol implemented in the adegenet R package (Jombart 2008, Jombart and Ahmed 2011). This protocol utilizes principal components (PCs) to calculate successive K-means (adegenet::find.clusters, Jombart and Collins 2015) which are then used to infer the number of population groups. We modified this method such that it utilized principal coordinates analysis (PCoA, adegenet::dudi.pco) to analyze a distance matrix calculated according to a model of sequence evolution. Note that species with < 4 individuals (*Emerita talpoida*, *Lampsilis ovata*, *Lampsilis teres*) were not evaluated using this algorithm, but were assigned a value of K = 1 for other analyses.

We explored how well population assignments explained the genetic variation by performing a permutational multivariate analysis of variance (PERMANOVA, adonis::vegan, Oksanen et al. 2018). The dispersion of clusters relative to each other (vegan::betadisper) was also assessed. Here, a significant beta-dispersion value indicates violation of PERMANOVA's assumption of homogeneity of variance (Anderson 2006).

### Species Distribution Modelling

Species distribution modelling (SDM) was used to characterize habitat suitability in the present by associating species locality data with maps of environmental variables (limited to climate and ocean depth), and to estimate habitat stability over time. We downloaded the GBIF Records in November 2018 using the R package "dismo" (Hijmans et al. 2017) and retained records that passed the following filters: preserved specimen, living specimen, material sample from 1900 to present, present in the United States and Canada. We combined the downloaded GBIF records with the specimen localities associated with the genetic data to create a complete presence record dataset. These were reduced to one occurrence per grid cell within the boundary of [-130, -60, 23, 50] at 5 arcmin resolution using the R package "raster" (Hijmans et al. 2019); species that had less than 25 occurrences were removed from the analysis following results from van Proosdij et al. (2016). Three sets of environmental variables were used: WorldClim 2.0 bioclimatic variables for the current (Fick and Hijmans 2017) and WorldClim 1.4 for past climate at the Mid-Holocene and LGM (Hijmans et al. 2005) under the CCSM4 model for all terrestrial and freshwater species, and MARSPEC (Sbrocco and Barber 2013), including bathymetry, also at 5 arcmin resolution and under the current and LGM conditions (all models) for the marine species. All bioclimatic variables were downloaded at a resolution of 5 arcmin and trimmed to the same constraints as the occurrence records using the R packages "raster", "maptools" (Bivand et al. 2019a), "rgdal" (Bivand et al. 2019b), and "rgeos" (Bivand et al. 2019c). We performed a correlation analysis on the raster layers under current conditions using the "layerStats" function and subsequently removed highly correlated variables ( $r > 0.8$ ; Mateo et al. 2013).

SDM analyses were conducted in R using the packages "randomForest" (Liaw and Wiener 2018), "raster", "rgeos", "maptools", "dismo", "sp" (Pevesma et al. 2018), "ecospat" (Cola et al. 2017), and "rJava" (Urbanek 2019). Pseudo absence points were drawn randomly from the background layers, and we partitioned the model into testing and training data using the "kfold" function. Distributional models were estimated for each species using the present environmental layers and cast onto the last glacial maximum layers using the MaxEnt, Random Forest, and GLM approaches (Phillips et al. 2017 and Hijmans 2017). In order to combine model results for an ensemble model, we evaluated the AUC of each present-day model to weight the contribution of each to the final output. We calculated a separate binary output using a

threshold determined using 'ecospat.mpa' (Cola et al. 2017), resulting in a total of six raster layers for each species: present-day continuous and binary rasters, mid-Holocene continuous and binary rasters, and last glacial maximum continuous and binary rasters.

As different species have different life history traits, species were combined into functional groups (Reptiles, Amphibians, Plants, Mammals, Freshwater, and Marine) in order to build cumulative maps of suitable habitat for the present day and the LGM using raster addition. Additionally, we determined potential refugia for each species through the identification of the areas of overlap between the present day and LGM binary models for each species. Using the functional groups defined as above, we took the overlapped species models and compared areas of interspecific overlap within each major grouping to determine areas of multi-species refugia, and we added these maps together for cumulative Terrestrial, Freshwater, and Marine groupings.

### *Multiple Matrix Regression with Randomization*

Three factors are often highlighted in phylogenetic and landscape genomic studies: Isolation by Distance (IBD), Isolation by Environment (IBE), and Isolation by Instability (IBI). To examine the impact of each force as independent predictors on genetic distance (calculated above in the AMOVA), we used multiple matrix regression with randomization (MMRR; Wang 2013). We performed the MMRR as outlined in Legendre et al. (1994) and implemented by Wang (2013) using the MRM command in the 'ecodist' R package with 1000 permutations (Goslee and Urban 2007). One advantage of MMRR is the ability to account for small amounts of covariation (less than 0.5) in the data through inclusion of randomization (Wang 2013). For use with the MRM command, each independent factor (IBD, IBE, and IBI) was transformed to distance matrices using the 'dist' function present in the stats R package. Values for IBD were determined from geographic coordinates collected with the genetic data, IBE and IBI values were extracted at sampling geographic points from species distribution models in the present, and stability maps, respectively. To generate stability maps we averaged the species distribution models from the present, Mid-Holocene, and LGM for each species. Using the averaged model gives us the relative stability of the environment over this time period, and using the 'extract' function, we sampled this stability surface at each sampling point.

The rationale for each metrics inclusion are as follows: IBD is the isolation by Euclidean distance between two geographic points (Wright 1943); two individuals that are closer together should be more similar genetically than those that are not as close, as there is a greater chance of gene flow between two close individuals. IBE implies that large differences in the environment prevent movement across the landscape and therefore gene flow between individuals (Nosil 2004). Under IBI, movement across the landscape is affected by the relative stability of different regions. Where there is high instability, movement is restricted, preventing

gene flow between individuals separated by unsuitable areas (Vasconcellos et al. 2019). Due to the possibility of strong correlation (greater than 0.5) between some predictors, correlations were checked for all variables. Those that were found to be significant using MMRR and were strongly correlated were checked using partial mantel tests to more accurately determine which predictors were important. MMRR also gives a value for the correlation coefficient ( $R^2$ ) for each species. This value gives the amount of genetic variation explained by IBD, IBE, and IBI.

### *Random Forest*

To evaluate whether some factors were predictive of species-specific responses to geographic and environmental factors across species, we constructed a Random Forest (RF) classifier. RF is a machine learning approach that uses multiple decision trees to predict the response based on many predictor variables (Liaw and Wiener 2002, Biau 2012). Random forest samples the data for each decision tree with replacement and uses the unsampled data to test the model. This is used to build a confusion matrix for the prediction and calculate the out of bag error rates (OOB). Each variable's importance is determined by measuring the mean decrease in accuracy of the prediction after the removal of a given variable.

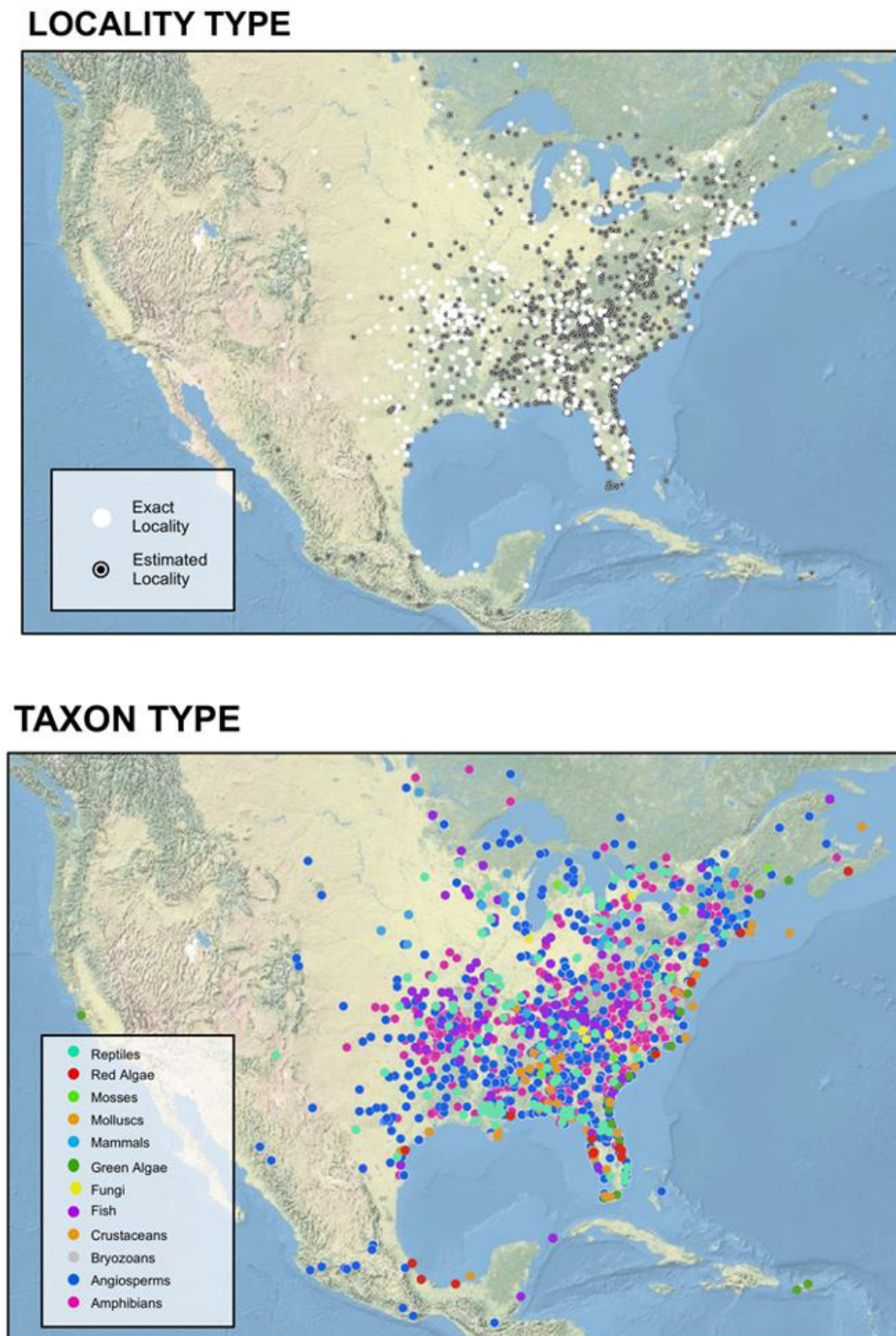
We designed different models that attempted to predict whether or not a species would have significant genetic structuring based on the categories used in the AMOVA and MMRR tests (Appalachian, Apalachicola, Coastal splits and Mississippi phylogeographic breaks; Ecoregion or Watershed; IBD, IBE, and IBI) as response variables. For predictor variables, we included organismal traits of body size, aquatic vs terrestrial, and generation length, and we included broad taxonomic groups (Table 1) in an attempt to identify results being driven by phylogeny. Additionally, we included the mean climatic variables for each species across their sampling range. Finally, we included the proportion of North America determined via SDM analysis to be suitable for both current and LGM timeframes, and the change in proportional area between LGM and the present.

## **Results**

### *Data processing*

The number of sequences per species ranged from three to 615 and the length ranged from 255 to 2395 base pairs (Table S2). The best fitted models of sequence evolution identified for each sequence alignment in either jModeltest or MrModeltest as well as the diversity and polymorphism calculations (i.e., Tajima  $D$ ,  $\Pi$ ,  $\theta$ ) are recorded in Table S2 for lowest AIC by jmodeltest or mrmmodeltest, and a narrowed list in Table S1 for analyses that couldn't use all possible models of substitution. Geographic data for individuals with associated genetic data showed widely distributed sampling across the eastern US, with the highest density in the south east (Fig. 1), suggesting our dataset has coverage adequate for testing our hypotheses. This





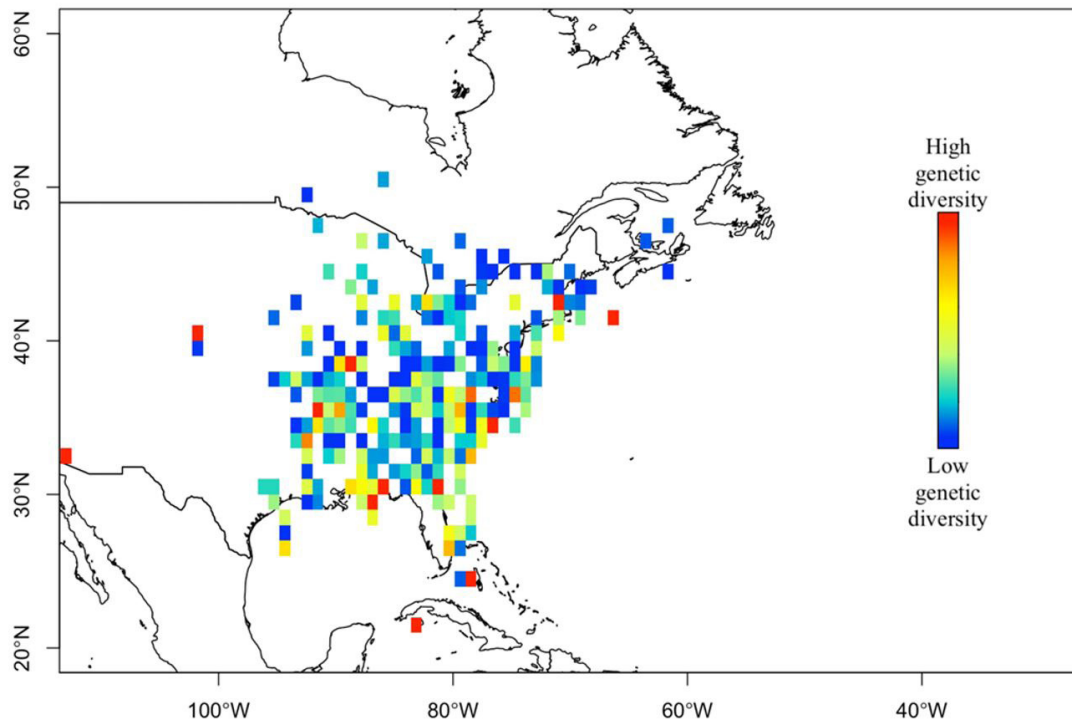
**Figure 1.** Summary of samples analyzed here. The upper figure shows the location of all samples, color coded by how the sample location was determined. The lower figure shows the location of all samples, color coded by the type of organism.

was regardless of sampling method (exact location vs estimated location), although there is a higher density of estimated localities in northern reaches. There is also no apparent effect of sampling based on taxon type (Fig. 1). While angiosperms clearly had the widest range, the data do not suggest that taxa are unevenly sampled across regions.

### *Nucleotide Mapping*

Nucleotide diversity for the taxa analyzed in this study is heterogeneously distributed across southeastern North America. While some regions contain high levels of nucleotide diversity, there is no evidence for a geographical trend in our final diversity map (Fig. 2). Moran's I was





**Figure 2.** Nucleotide diversity mapping across the eastern United States. Estimates of nucleotide diversity across species were visualized by summarizing values within a 1° grid cell for cells with two or more individuals and then normalized.

found to be close to 0 (0.037,  $p=0.32$ ), indicating little positive spatial autocorrelation. This indicates that grid cells that are near one another in geographic space are not any more similar in their nucleotide diversity than would be expected by chance. An out of refugia hypothesis would predict that areas of refugia would have higher nucleotide diversity. We also compared this to rasters based on predictions of trends (glacial refugia, null and random), and none were found to be significant.

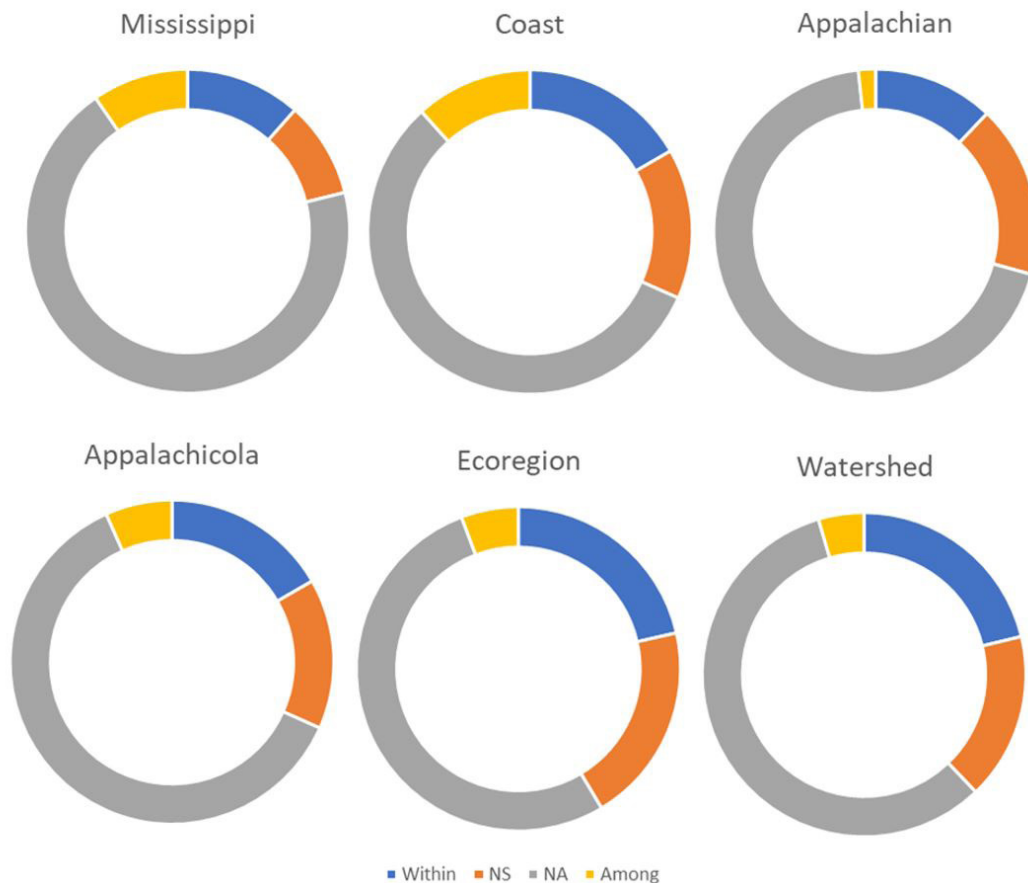
### AMOVA

We used analysis of molecular variance to explore the influence of ecological regions and putative phylogeographic breaks on intraspecific diversity. Samples were partitioned according to regions (i.e., Ecoregion, Watershed) or by putative phylogeographic breaks (i.e., Coast, Appalachian Mountains, Apalachicola, Mississippi River). For example, if population genetic structuring results from historical breaks, we expect significant differentiation across phylogeographic breaks, while if genetic structure results from contemporary ecological barriers, we expect that ecoregions would be an important promotor of genetic variance. Ecoregion AMOVA analyses indicate that 15 (37%) species had significant proportion of the observed variation occurring within groups and that 4 (10%) species had significant proportion of the observed variation occurring among groups (Fig. 3; Table S3). The results of our watershed AMOVA analyses for groups assigned by watersheds indicate that 14 (34%) species had significant proportion of the variation within group and three species (7%) had significant proportion of the observed variation

occurring among groups (Table S4). An AMOVA that tested the impact of phylogeographic breaks on genetic structure was conducted in 4 analyses that were named after the putative physical barrier: Coast, Appalachian Mountains, Apalachicola, and Mississippi River. For our Coast AMOVA analyses for groups either along the coast of the Gulf of Mexico or the coast of the Atlantic Ocean indicate that 10 species (34%) had significant portion of the variation within groups and 7 species (17%) had significant proportion of the observed variation among groups (Table S5). Results of our Appalachian AMOVA analyses for groups either on the west or east side of the Appalachian Mountains indicate that 7 species (17%) had significant portion of the variation within groups and one species (2%) had significant proportion of the observed variation among groups (Table S6). The AMOVA analyses for groups either on the west or east side of the Apalachicola River indicate that 10 species (24%) had significant portion of the variation within groups and four species (10%) had significant proportion of the observed variation among groups (Table S7). Lastly, the Mississippi River AMOVA split show that six species (15%) had significant portion of the variation within groups and five species (12%) had significant proportion of the observed variation among groups (Table S8). Results of all AMOVA analyses are summarized in Table 2.

### Population assignment

The K-means procedure typically produced population assignments that agreed with visual examinations of BIC vs. K plots (Fig. S2). For seven species, we chose a K that differed from that identified via the automated



**Figure 3.** Pie charts summarizing AMOVA results with a variety of different potential breaks/factors. Grey indicates species that could not be tested due to not being in multiple regions or on both sides of a break (NA). Orange are those species that were tested but were found to not be significant (NS). Species that were significant among breaks are in yellow. Blue gives species that were found to be significant within a break or region.

**Table 2.** Summary of AMOVA for all species. In each analysis, samples were partitioned according to regions (i.e., Ecoregion, Watershed) or by putative phylogeographic breaks (i.e., Coast, Appalachian Mountains, Apalachicola, and Mississippi River).

Species	Mississippi	Coast	Appalachian	Apalachicola	Ecoregion	Watershed
<i>Acipenser oxyrinchus</i>	NA	<0.001*	<0.001*	NA	<0.001*	<0.001*
<i>Apalone ferox</i>	NA	0.829	NA	0.618	0.246	0.618
<i>Apalone mutica</i>	0.033*	0.027*	NA	0.648	0.179	0.139
<i>Apalone spinifera</i>	0.313	<0.001*	NA	NA	0.172	<b>0.003*</b>
<i>Blarina brevicauda</i>	<0.001*	<0.001*	<0.001*	<0.001*	<0.001*	<0.001*
<i>Bugula neritina</i>	0.131	0.077	0.281	1	<b>0.031*</b>	0.316
<i>Busycon sinistrum</i>	<0.001*	<0.001*	NA	<0.001*	<0.001*	<0.001*
<i>Carcharhinus limbatus</i>	NA	0.094	NA	0.11	0.045*	<b>0.043*</b>
<i>Desmognathus wrightii</i>	NA	NA	NA	0.010*	NA	<b>0.004*</b>
<i>Eumeces fasciatus</i>	<0.001*	<0.001*	<b>0.001*</b>	<0.001*	<0.001*	<0.001*
<i>Eurycea bislineata</i>	NA	NA	0.071	0.224	0.109	0.216
<i>Eurycea cirrigera</i>	NA	<0.001*	<0.001*	<0.001*	<0.001*	<0.001*
<i>Eurycea junaluska</i>	NA	NA	0.059	NA	NA	NA

For each test, the *p*-value is reported, with significant results denoted by an '\*'. Those denoted by 'NA' are species that were not sampled on both sides of a split or in multiple ecoregions or watersheds. Blue denotes significant within, yellow among.

Table 2. Continued...

Species	Mississippi	Coast	Appalachian	Apalachicola	Ecoregion	Watershed
<i>Eurcyea multiplicata</i>	NA	NA	NA	NA	0.089	NA
<i>Eurcyea tymerensis</i>	NA	NA	NA	NA	0.286	NA
<i>Eurcyea wilderae</i>	NA	<b>0.001*</b>	<b>&lt;0.001*</b>	<b>&lt;0.001*</b>	<b>0.001*</b>	<b>0.001*</b>
<i>Farfantepenaeus aztecus</i>	0.639	0.594	NA	0.611	0.648	0.671
<i>Hypentelium nigricans</i>	<b>&lt;0.001*</b>	<b>0.025*</b>	0.189	0.435	<b>&lt;0.001*</b>	0.221
<i>Lampsilis altilis</i>	NA	0.824	0.342	NA	0.51	NA
<i>Lampsilis australis</i>	NA	NA	NA	NA	NA	NA
<i>Lampsilis ovata</i>	NA	NA	NA	NA	NA	NA
<i>Lampsilis perovalis</i>	NA	NA	NA	NA	0.764	NA
<i>Lampsilis teres</i>	NA	NA	NA	NA	NA	NA
<i>Litopenaeus setiferus</i>	<b>0.024*</b>	<b>0.003*</b>	NA	<b>0.001*</b>	<b>0.011*</b>	<b>0.009*</b>
<i>Pagarus longicarpus</i>	<b>&lt;0.001*</b>	<b>&lt;0.001*</b>	0.919	<b>&lt;0.001*</b>	<b>&lt;0.001*</b>	<b>&lt;0.001*</b>
<i>Pagarus pollicaris</i>	0.087	<b>&lt;0.001*</b>	0.146	<b>&lt;0.001*</b>	<b>&lt;0.001*</b>	<b>&lt;0.001*</b>
<i>Pantherophis obsoletus</i>	<b>0.002*</b>	0.065	0.364	0.068	<b>0.024*</b>	0.208
<i>Percina caprodes</i>	1	0.724	0.72	0.859	0.543	0.498
<i>Percina nasuta</i>	NA	NA	NA	NA	NA	NA
<i>Percina phoxocephala</i>	NA	NA	NA	NA	<b>&lt;0.001*</b>	NA
<i>Pseudacris brachyphona</i>	NA	<b>0.017*</b>	0.206	NA	<b>&lt;0.001*</b>	<b>&lt;0.001*</b>
<i>Pseudacris brimleyi</i>	NA	NA	NA	NA	1	NA
<i>Pseudacris clarkii</i>	NA	0.242	NA	NA	0.347	0.244
<i>Pseudacris crucifer</i>	<b>&lt;0.001*</b>	<b>&lt;0.001*</b>	<b>&lt;0.001*</b>	<b>&lt;0.001*</b>	<b>0.001*</b>	<b>&lt;0.001*</b>
<i>Pseudacris feriarium</i>	0.813	<b>&lt;0.001*</b>	<b>&lt;0.001*</b>	<b>&lt;0.001*</b>	0.089	<b>&lt;0.001*</b>
<i>Pseudacris nigrita</i>	NA	0.545	NA	0.552	0.664	0.664
<i>Pseudacris spnov</i>	<b>0.018*</b>	0.448	NA	NA	0.087	0.383
<i>Rana catesbeiana</i>	<b>&lt;0.001*</b>	<b>&lt;0.001*</b>	<b>&lt;0.001*</b>	<b>&lt;0.001*</b>	<b>&lt;0.001*</b>	<b>&lt;0.001*</b>
<i>Seriola dumerli</i>	NA	NA	NA	NA	<b>&lt;0.001*</b>	<b>&lt;0.001*</b>
<i>Spisula solidissima</i>	NA	<b>&lt;0.001*</b>	NA	0.001*	<b>&lt;0.001*</b>	<b>&lt;0.001*</b>
<i>Tryphlotriton spelaeus</i>	NA	NA	NA	NA	0.081	NA

For each test, the *p*-value is reported, with significant results denoted by an '\*'. Those denoted by 'NA' are species that were not sampled on both sides of a split or in multiple ecoregions or watersheds. Blue denotes significant within, yellow among.

procedure. The majority of species were assigned to two or three populations (Table 3). Results of PERMANOVA suggested population assignments were efficacious. For 42 of the 53 species that were split into populations ( $K > 1$ ), we found that population assignments explained significant portions of the genetic variation ( $p < 0.05$ , Table 3). Those with non-significant population assignments had either low sample sizes or low overall genetic variation. Fifteen species had significant PERMANOVA results but had beta-dispersion

values (Table 3) that suggested assigned groupings were not significantly differentiated. About half (27/53) of the species split into populations which had significant assignments that were not over-dispersed.

#### Species Distribution Modelling

We retained 31 species for species distribution modelling after removing those species with <25 unique occurrence records (Table S9). Output maps for each of these species display habitat suitability indices for



**Table 3.** Summary of population assignment results and assessment of population assignment utility. Shown for each species are the number of individuals, the model of sequence evolution, the final 'K' chosen, the *P*-value from the PERMANOVA, and the mean population membership probability. For each species, populations were assigned by calculating a distance matrix based on an appropriate model of sequence evolution, analyzing this matrix with PCoA, using the principal components to calculate successive K-means, and choosing an optimum K based on BIC vs. K plots (Fig. S1). Efficacy of population assignments were assessed via PERMANOVA and DAPC. Complete records found in Table S11.

Species	Number of individuals	Model of sequence evolution	K selected by comparing BIC vs. K plots	<i>P</i> from PERMANOVA	population membership probability
<i>Acer rubrum</i>	37	F81	2	<0.001	0.972
<i>Acipenser oxyrinchus</i>	614	K80	1	NA	1.000
<i>Ambystoma tigrinum</i>	54	K81	3	<0.001	0.840
<i>Apalone ferox</i>	8	K81	2	0.250	NA
<i>Apalone mutica</i>	7	K81	2	0.029	1.000
<i>Apalone spinifera</i>	21	K81	2	<0.001	0.971
<i>Apios americana</i>	63	K81	2	<0.001	1.000
<i>Blarina brevicauda</i>	74	K81	3	<0.001	1.000
<i>Bryopsis sp.</i>	65	F81	5	<0.001	1.000
<i>Bugula neritina</i>	20	K80	2	0.006	1.000
<i>Busycon sinistrum</i>	31	K81	2	<0.001	0.971
<i>Carcharhinus limbatus</i>	10	K81	2	0.008	1.000
<i>Daphnia obtusa</i>	36	K81	2	<0.001	1.000
<i>Desmognathus wrightii</i>	28	TN93	2	<0.001	1.000
<i>Dicerandra spp.</i>	28	F81	3	1.000	NA
<i>Emerita talpoida</i>	3	F81	1	NA	NA
<i>Eumeces fasciatus</i>	82	TN93	5	<0.001	1.000
<i>Eurycea bislineata</i>	52	TN93	2	<0.001	1.000
<i>Eurycea cirrigera</i>	247	TN93	12	<0.001	1.000
<i>Eurycea junaluska</i>	6	TN93	2	0.167	1.000
<i>Eurycea multiplicata</i>	44	TN93	2	<0.001	0.982
<i>Eurycea taylori</i>	14	TN93	2	0.070	1.000
<i>Eurycea wilderae</i>	125	TN93	3	<0.001	1.000
<i>Fagus grandifolia</i>	22	F81	2	<0.001	1.000
<i>Farfantepenaeus aztecus</i>	76	K80	2	<0.001	1.000
<i>Gracilaria tikvahiae</i>	21	TN93	2	0.047	1.000
<i>Hypentelium nigricans</i>	72	TN93	2	<0.001	1.000
<i>Lampsilis altilis</i>	5	F81	2	0.200	NA
<i>Lampsilis australis</i>	5	F81	2	0.200	1.000
<i>Lampsilis ovata</i>	2	F81	1	NA	NA
<i>Lampsilis perovalis</i>	4	F81	2	0.250	NA
<i>Lampsilis teres</i>	2	F81	1	NA	NA
<i>Liquidambar styraciflua</i>	111	F81	2	<0.001	0.984
<i>Litopenaeus setiferus</i>	91	K80	2	<0.001	1.000
<i>Pagrus longicarpus</i>	67	K81	3	<0.001	1.000
<i>Pagrus pollicaris</i>	21	K81	2	<0.001	1.000

Table 3. Continued...

Species	Number of individuals	Model of sequence evolution	K selected by comparing BIC vs. K plots	P from PERMANOVA	population membership probability
<i>Pantherophis obsoletus</i>	37	K81	2	<0.001	1.000
<i>Percina caprodes</i>	13	TN93	2	0.007	1.000
<i>Percina nasuta</i>	6	TN93	2	0.133	1.000
<i>Percina phoxocephala</i>	34	TN93	3	<0.001	0.830
<i>Prunus spp.</i>	223	K81	2	<0.001	1.000
<i>Pseudacris brachyphona</i>	25	K81	2	<0.001	1.000
<i>Pseudacris brimleyi</i>	8	K81	2	0.036	1.000
<i>Pseudacris clarkii</i>	9	TN93	2	0.012	1.000
<i>Pseudacris crucifer</i>	61	K80	5	<0.001	1.000
<i>Pseudacris feriarium</i>	60	TN93	2	<0.001	1.000
<i>Pseudacris nigrita</i>	19	TN93	2	0.006	1.000
<i>Pseudacris spnov</i>	27	TN93	2	<0.001	1.000
<i>Rana catesbeiana</i>	211	TN93	2	<0.001	1.000
<i>Seriola dumerli</i>	16	K81	2	<0.001	1.000
<i>Sphagnum bartlettianum</i>	11	F81	2	0.090	1.000
<i>Spisula solidissima</i>	52	K81	3	<0.001	1.000
<i>Tilia americana</i>	298	K81	3	<0.001	0.992
<i>Trillium cuneatum</i>	281	F81	7	1.000	1.000
<i>Tryphlotriton spelaeus</i>	12	K81	2	0.084	1.000
<i>Uniola paniculata</i>	130	TN93	4	<0.001	1.000
<i>Xerula furfuracea</i>	25	K80	2	0.041	1.000

the models based on these data (Individual maps in Fig. S4). The output maps displaying the cumulative stable habitat (including combinations of major functional groups of Amphibians, Reptiles, Terrestrial, Terrestrial & Freshwater), that existed in both the current day and LGM predictions are in (Fig. 4). Amphibians and Reptiles had the greatest amount of stable habitat extending in a band from Texas across the Florida panhandle to North Carolina, while the Freshwater and Plants categories were mainly localized within the Gulf region and in the Carolinas. The complete Terrestrial & Freshwater group had the greatest amount of stable habitat along the Gulf Coast and near North Carolina. In marine species there was stable habitat along the Gulf of California, along the tip of the Florida Peninsula, and localized areas throughout the Caribbean and up into the east coast; these models were based on three species, so there was a maximum overlap of three.

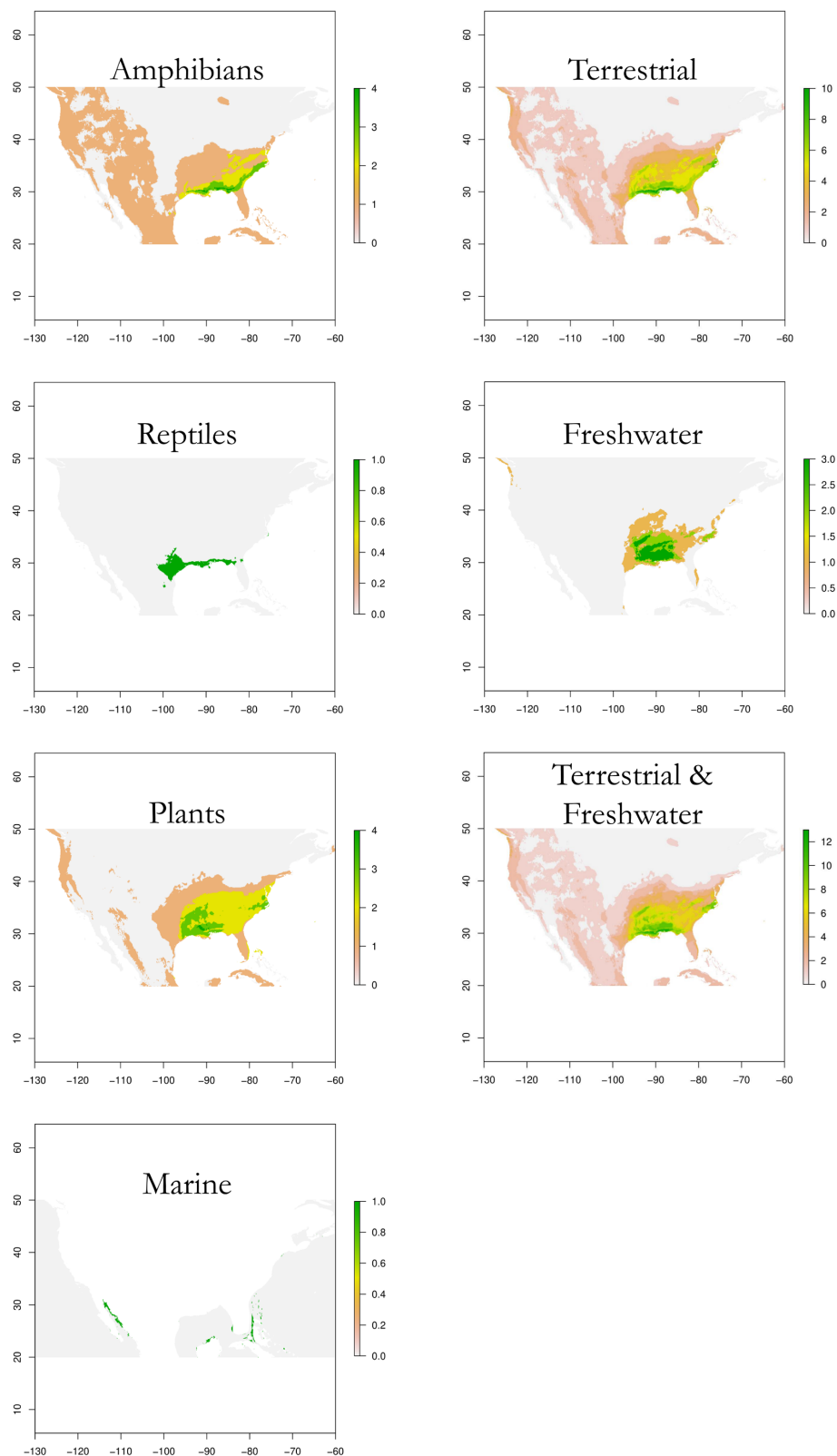
#### Multiple Matrix Regression with Randomization

The MMRR required both a genetic distance matrix (requiring more than 3 individuals that exhibit genetic differences) and a species distribution model, generated previously. Using these criteria, we were able to include 28 species in this analysis. The multiple matrix regression highlighted the differences

between species, with varying combinations of our independent factors (IBD, IBE, and IBI) being found significant (Fig. 5). Additionally, those variables that were correlated (>0.5) and were both significant, were checked using partial mantel tests. This analysis was able to explain between 3% and 86% of the variation in genetic distances using varying combinations of IBD, IBE and IBI, averaging 27% and a median of 14% variation explained. In short, IBD was found to be a significant factor in 71% of species, IBE in 29%, and IBI in 32% of species. Interestingly, 14% of species' genetic diversity was not significantly influenced by any of these factors. There were no apparent trends based on taxonomy or other characteristics of species in terms of which factors explained genetic variation in a given species.

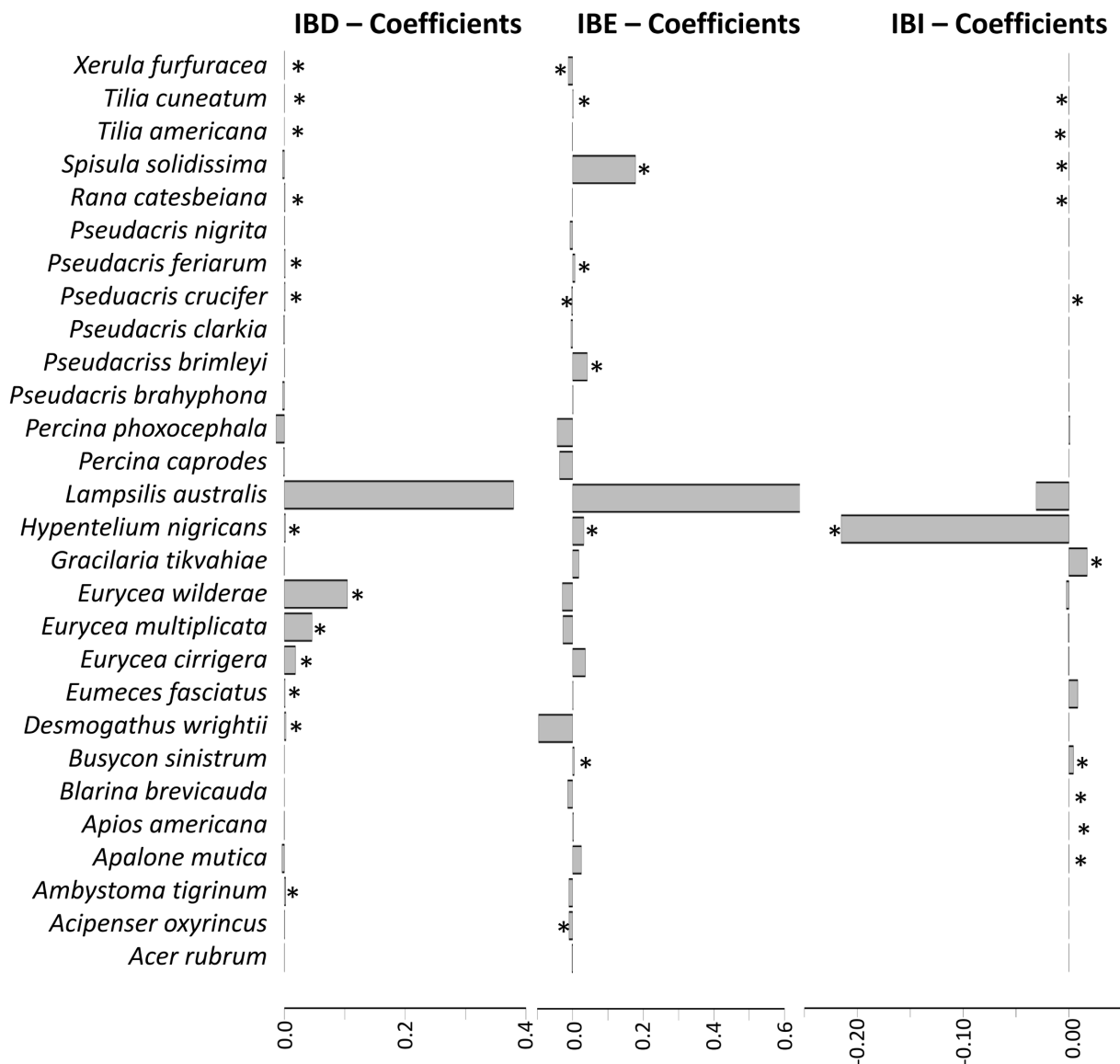
#### Random Forest

Across all tests the random forest analysis was generally not effective at building a predictive model that could correctly assign species to the predetermined categories (Table S11), a result that may result from the small number of species present in our data matrix (Table S12). The use of the machine learning framework provides a check against overinterpretation of similar results in particular species.



**Figure 4.** Regions of habitat stability mapped for terrestrial (Amphibians/Plants/Reptiles) and aquatic (Freshwater/Marine) organisms. Each figure depicts the stability of the habitat in each grid cell since the last glacial maximum. Scales bars (to right of each figure) show the color scale representing the number of species with overlapping stable habitat within each taxon type.





**Figure 5.** Results of multiple matrix regression across species. Shown for each species are the  $p$ -values and coefficients for three tests (IBD, IBE, IBI) with significant results with an \*. Full detailed numbers can be found in Table S11.

## Discussion

Data repurposing (Sidlauskas et al. 2010) provides a more comprehensive approach than meta-analysis for comparative phylogeography because data collected from species that share a geographic distribution can be analyzed and interpreted in a common framework. While the species examined here have all been investigated previously (e.g., Soltis 2000, Avise 2006), the specific contexts of these studies differ, which makes it difficult to evaluate common factors. Our results support the viewpoint that many forces combine to influence genetic diversity within species (e.g., Zhang et al. 2018). While this result is not unexpected, it is striking to observe in a comparative context across disparate taxa. Three factors (geographic

separation, habitat stability, and barriers to dispersal) largely explain gross patterns of intraspecific genetic variation in the species analyzed here.

The first factor that contributes to intraspecific genetic structure is the physical separation of individuals. Genetic isolation by geographic distance (e.g., Wright, 1943) is statistically significant in all but five cases (82%; Fig. 5), and these species generally contain fewer samples, which may lead to false negatives and is consistent with previously reported biases (Meirmans 2012, Pelletier and Carstens 2018). We identified contemporary climatic conditions as a significant factor in 8/28 (29%) species exhibited significant IBE (Fig. 5), a result that is broadly consistent with the AMOVAs that were conducted with species partitioned by 'ecoregions' (19/35 species, a higher proportion

than were significantly partitioned by 'watershed'; Table 2). Within many species, genetic diversity is correlated with differences in the current environment across the range of the species, as assumed by many landscape genetic investigations (Sexton et al. 2013).

Following previous work, we also explored the impact of the phylogeographic breaks that are expected to structure genetic diversity over deeper time scales. In a slight majority of species (18/35; 51%), genetic variance was significantly partitioned across putative biogeographic barriers in our AMOVA results (Table 2). Since these barriers were derived from the literature on species included in this comparison set, this result is circular for some species (see Soltis et al. 2006 for these species). However, many of these same species also exhibited isolation by habitat instability, which also speaks to the relative importance of historical processes in structuring genetic diversity. Unlike Vasconcellos et al. (2019), we found that habitat instability over time does not appear to explain how intraspecific genetic diversity is distributed across the landscape. While 9/28 species exhibited significant isolation by instability, in nearly all of these cases, other factors (i.e., IBD or IBE) were also found to be significant. The nucleotide diversity map (Fig.2) supports this idea. If habitat stability were an important predictor, we would see an "out of refugia" pattern due to those areas being the most stable over long periods of time. Although the combined regression models explained a large (>50%) proportion of the variation in several species, only one of these (*Xerula furfuracea*) included IBI. Overall, this suggests that IBI is not an important contributor to population genetic structure, at least in this region. However, it may play a role in other regions that were more directly impacted by glaciation.

In most cases, either geographic distance or physical barriers were important in determining genetic differentiation; this is supported by both broad-scale phylogeographic and fine-scale landscape approaches. Despite the near-ubiquitous importance of physical barriers and geographic distance, the influence of environmental factors in the process of genetic differentiation should not be ignored. On the contrary, our results implicate a complex mixture of processes related to geographic distance, environment, and instability, which are responsible for affecting within-species genetic differentiation. A more concise view of these complexities may be attained by observing species that were in both the AMOVA and the MMRR analyses (see Table S13). Each species with a significant inter-population AMOVA result (7 species; Table S13) also had significant MMRR results. In only one of these cases was IBD the only significant MMRR result (*Apalone mutica*), supporting the role of processes other than geographic distance or geographic barriers in structuring genetic diversity. In several cases (11; Table S14), the MMRR analyses found support for an environmental (IBE), distance-related (IBD), and/or instability-related (IBI) factor without support for a phylogeographic barrier hypothesis. Clearly, geographic distance/phylogeographic

breaks are important facets to intraspecific genetic differentiation, but an overreliance on these types of hypotheses would ignore other variables explaining this differentiation, and in turn could prevent the recognition of patterns that may exist between taxonomic groups of interest. These results highlight the need for the integration of both phylogeographic and landscape genetic approaches in investigations that seek to discover processes affecting genetic differentiation within species. They also suggest that sampling bias is a potential limitation. For example, data from some of the species used here were able to be included in only some analyses due to low sampling numbers in some regions.

The mixed results presented here offer the possibility for hypotheses that could be tested in future studies in this region. In closely related species, we could observe differences between both time scales influencing genetic structuring but also the factors present in the same time scale. To us, this indicates that we cannot simply apply understanding from one species and expect it to hold for another species, even a closely related one. Evolution is a complex process and it can be difficult to measure every feature that may be under selection, leading to genetic structuring, especially for non-model systems. For example, two species of frogs may be evolutionarily similar, exist in the same time and space, but have experienced vastly different evolutionary pressures leading to variations in the factors influencing their genetic structure. One of the two species may be slightly more impacted by environmental pressures that are hard to detect, so it may be affected by IBE where the other species may not be, but these types of relationships are hard to detect when only using one type of analysis.

Our results suggest that both broad historical and fine scale contemporary processes are likely to influence genetic diversity within a given species. Rissler (2016) suggested that landscape genetics and phylogeography were headed towards a conceptual unification due to advances in sequencing technology and analytical approaches. As observed here, a more complete understanding of the forces that influence genetic diversity might be gained through a conceptual unification of these two methods, particularly if data from a given species is analyzed using approaches drawn from each. Our study highlights the multiple processes that likely influence genetic structure in a given species and suggests that any investigation conducted from a single perspective (i.e., historical or contemporary; phylogeographic or landscape genetic) is likely to provide an incomplete understanding of the focal system.

## Data Accessibility

All code used is freely available for download at [https://github.com/jgwieringa/IBI\\_and\\_others](https://github.com/jgwieringa/IBI_and_others) and all sequencing data can be collected from the source manuscripts listed in Table 1.

## Acknowledgements

The authors thank Soltis et al. (2006) for compiling the review that served as a starting point for our own analyses. This work is supported by the following means: EMF (process #88881.170016/2018-1) is supported through the Coordenação de Apoio à Formação de Pessoal de Nivel Superior (CAPES) for his doctoral fellowship, MVD (process #88881.190071/2018-01) is supported by CAPES Programa de Doutorado-sandúfche no Exterior (PDSE), MLS is support by the National Science Foundation Graduate Research Fellowship Program under Grant #: DG-1343012, HG is supported by the Ohio State University Helen M and Milton O Lee Fellowship, DJP is supported by the Ohio State University Graduate Enrichment Fellowship, and DD is supported by the OSU University Fellowship. MRB is supported by National Science Foundation Graduate Research Fellowships Program #DGE-1343012. Some analyses were conducted using computational resources from the Ohio Supercomputer Center (PAA0202).

## Supplementary Materials

The following materials are available as part of the online article from <https://escholarship.org/uc/fb>

**Table S1.** Population assignment analyses using four methods.

**Table S2.** Data collection for all species and estimated parameter values.

**Table S3.** AMOVA results for ecoregion analyses.

**Table S4.** AMOVA results for watershed analyses.

**Table S5.** AMOVA results for coastal analyses.

**Table S6.** AMOVA results for Appalachian analyses.

**Table S7.** AMOVA results for Apalachicola analyses.

**Table S8.** AMOVA results for Mississippi analyses.

**Table S9.** Georeferenced localities for all individuals used in manuscript.

**Table S10.** AUC evaluation for each species from the SDM analyses.

**Table S11.** Detailed results from the MMR analysis.

**Table S12.** Detailed results from the Random Forest analysis.

**Table S13.** Data matrix used in the Random Forest analysis.

**Table S14.** Summary of results from MMR and AMOVA.

**Figure S1.** Graphical summary of data analysis strategy.

**Figure S2.** BIC vs K plots for all species giving the recommended number of K for each species.

**Figure S3.** PCOA visualization for population groupings.

**Figure S4.** Species distribution models for all species in present and last glacial maximum climates.

**Appendix S1.** Supplementary bibliography.

**Appendix S2.** Analysis pipeline scripts and associated readme files.

## References

- Agriculture and Agri-Food Canada [AAFC] (2017) Terrestrial ecoprovinces of Canada. National Ecological Framework for Canada. Available at [http://sis.agr.gc.ca/cansis/nsdb/ecostrat/gis\\_data.html](http://sis.agr.gc.ca/cansis/nsdb/ecostrat/gis_data.html)
- Anderson, M. J. (2006) Distance-based tests for homogeneity of multivariate dispersions. *Biometrics*, 62, 245-253.
- Austin, J. D., Loughheed, S. C. & Boag, P. T. (2004) Discordant temporal and geographic patterns in maternal lineages of eastern north American frogs, *Rana catesbeiana* (Ranidae) and *Pseudacris crucifer* (Hylidae). *Molecular Phylogenetics and Evolution*, 32, 799-816.
- Avice, J. C., Arnold, J., Ball, R. M., Bermingham, E., Lamb, T., Neigel, J. E., Reeb, C. A. & Saunders, N. C. (1987) Intraspecific phylogeography: the mitochondrial DNA bridge between population genetics and systematics. *Annual Review of Ecology and Systematics*, 18, 489-522.
- Avice, J.C. (2000) *Phylogeography: the history and formation of species*. Harvard University Press.
- Barrow, L. N., Soto-Centeno, J. A., Warwick, A. R., Lemmon, A. R., Lemmon, E. M. (2017) Evaluating hypotheses of expansion from refugia through comparative phylogeography of south-eastern Coastal Plain amphibians. *Journal of Biogeography*, 44, 2692-2705.
- Berendzen, P. B., Simons, A. M. & Wood, R. M. (2003) Phylogeography of the northern hogsucker, *Hypentelium nigricans* (Teleostei: Cypriniformes): Genetic evidence for the existence of the ancient Teays River. *Journal of Biogeography*, 30, 1139-1152.
- Biau, G. (2012) Analysis of a random forests model. *Journal of Machine Learning Research*, 13, 1063-1095.
- Bivand, R., Archer, E., Baddeley, A., Bearman, N., Callahan, J., Forrest, D., Friendly, M., Giraudoux, P., Golicher, D., Gomez, V. R., Hausmann, P., Hufthammer, K. O., Jagger, T., Johnson, K., Luque, S., MacQueen, D., Niccolai, A., Pebesma, E., Lamigueiro, O. P., Plunkett, E., Short, T., Snow, G., Stabler, B., Stokely, M. & Turner, R. (2019) R Package 'maptools.' <https://cran.r-project.org/web/packages/maptools/index.html>
- Bivand, R., Keitt, T., Rowlingson, B., Pebesma, E., Summer, M., Higmans, R., Rouault, E., Warmerdam, F., Ooms, J. & Rundel, C. (2019)



- R Package 'rgdal.' <https://CRAN.R-project.org/package=rgdal>
- Bivand, R., Stuetz, R., Ove, K., Giraudoux, P., Davis, M. & Santilli, S. (2019) R Package 'rgeos'. <https://cran.r-project.org/web/packages/rgeos/index.html>
- Bonett, R. M. & Chippindale, P. T. (2004) Speciation, phylogeography and evolution of life history and morphology in plethodontid salamanders of the *Eurycea multiplicata* complex. *Molecular Ecology*, 13, 1189-1203.
- Brant, S. V. & Ortí, G. (2003) Phylogeography of the Northern short-tailed shrew, *Blarina brevicauda* (Insectivora: Soricidae): past fragmentation and postglacial recolonization. *Molecular Ecology*, 12, 1435-1449.
- Bremer, J. R. A., Ditty, J. G., Turner, J. S. & Saxton, B. L. (2010) Molecular species identification of commercially important penaeid shrimp from the Gulf of Mexico using a multiplex haplotype-specific PCR assay. *Biochemical Systematics and Ecology*, 38, 715-721.
- 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.
- Campbell, D. C., Serb, J. M., Buhay, J. E., Roe, K. J., Minton, R. L. & Lydeard, C. (2005) Phylogeny of North American amblesines (Bivalvia, Unionoida): prodigious polyphyly proves pervasive across genera. *Invertebrate Biology*, 124, 131-164.
- Carnaval, A. C., Hickerson, M. J. & Haddad, C. F. B. (2009) Stability predicts genetic diversity in the Brazilian Atlantic Forest hotspot. *Science*, 323, 785-790.
- Carstens, B. C. & Richards, C. L. (2007) Integrating coalescent and ecological niche modeling in comparative phylogeography. *Evolution*, 61, 1439-1454.
- Cola, V. Di, Broennimann, O., Petitpierre, B., Breiner, F. T., Amen, M. D., Randin, C., Engler, R., Pottier, J., Pio, D., Dubuis, A., Mateo, R. G., Hordijk, W., Salamin, N. & Guisan, A. (2017) ecospat: an R package to support spatial analyses and modeling of species niches and distributions. *Ecography*, 40, 774-787.
- Commission for Environmental Cooperation Working Group [CECWG] (1997) Ecological Regions of North America: Toward a Common Perspective.
- Crespi, E. J., Rissler, L. J. & Browne, R. A. (2003) Testing Pleistocene refugia theory: phylogeographical analysis of *Desmognathus wrighti*, a high-elevation salamander in the southern Appalachians. *Molecular Ecology*, 12, 969-984.
- Darriba, D., Taboada, G. L., Doallo, R. & Posada, D. (2015) jModelTest 2: more models, new heuristics and high-performance computing. *Nature Methods*, 9, 6-9.
- Emmanuel, P., Jombart, T., Brian, K., Schliep, K. & Potts, A. (2018) R Package 'pegas.' Environmental Systems Research Institute [ESRI] (2014) ArcGIS Desktop: Release 10. Redlands, CA.
- Excoffier, L., Smouse, P. E. & Quattro, J. M. (1992) Analysis of molecular variance inferred from metric distances among DNA haplotypes: application to human mitochondrial DNA restriction data. *Genetics*, 131, 479-491.
- Excoffier, L., Laval, G. & Schneider, S. (2005) Arlequin (version 3.0): an integrated software package for population genetics data analysis. *Evolutionary bioinformatics*, 1, 117693430500100003.
- Felsenstein, J. (1981) Evolutionary Trees from DNA Sequences: A Maximum Likelihood Approach, *Journal of Molecular Evolution*, 368-376.
- Fick, S. E. & Hijmans, R. J. (2017) WorldClim 2: new 1-km spatial resolution climate surfaces for global land areas. *International Journal of Climatology*, 37, 4302-4315.
- Geber, M. A. (2011) Ecological and evolutionary limits to species geographic ranges. *The American Naturalist*, 178, S1-S5.
- Gold, J. R. & Richardson, L. R. (1998) Population structure in greater amberjack, *Seriola dumerili*, from the Gulf of Mexico and the western Atlantic Ocean. *Fishery Bulletin*, 96, 767-778.
- Gonzales, E., Hamrick, J. L. & Chang, S. M. (2008) Identification of glacial refugia in south-eastern North America by phylogeographical analyses of a forest understorey plant, *Trillium cuneatum*. *Journal of Biogeography*, 35, 844-852.
- Goslee, S. C. & Urban, D. L. (2007) The ecodist Package for Dissimilarity-based Analysis of Ecological Data. *Journal of Statistical Software*, 22, 1-19.
- Graham, C. H., VanDerWal, J., Phillips, S. J., Moritz, C. & Williams, S. E. (2010) Dynamic refugia and species persistence: tracking spatial shifts in habitat through time. *Ecography*, 33(6), 1062-1069.
- Grunwald, C., Maceda, L., Waldman, J., Stabile, J. & Wirgin, I. (2008) Conservation of Atlantic

- sturgeon *Acipenser oxyrinchus oxyrinchus*: delineation of stock structure and distinct population segments. *Conservation Genetics*, 9, 1111-1124.
- Guindon, S. & Gascuel, O. (2003) A Simple, Fast, and Accurate Algorithm to Estimate Large Phylogenies. *Systematic Biology*, 52, 696-704.
- Gurgel, C. F. D., Fredericq, S. & Norris, J. N. (2004) Phylogeography of *Gracilaria tikvahiae* (Gracilariaceae, Rhodophyta): a study of genetic discontinuity in a continuously distributed species based on molecular evidence. *Journal of Phycology*, 40, 748-758.
- Hao, Y. J., Zhao, Q., Wang, S. X. & Yang, Z. L. (2016) What is the radicate *Oudemansiella* cultivated in China? *Phytotaxa*, 286, 1-12.
- Hare, M. P. & Weinberg, J. R. (2005) Phylogeography of surfclams, *Spisula solidissima*, in the western North Atlantic based on mitochondrial and nuclear DNA sequences. *Marine Biology*, 146, 707-716.
- Hijmans, R. J., van Etten, J., Sumner, M., Cheng, J., Bevan, A., Bivand, R., ... Hijmans, M. R. J. (2019) R Package 'raster'.
- Hijmans, R. J., Cameron, S. E., Parra, J. L., Jones, G. & Jarvis, A. (2005) Very high-resolution interpolated climate surfaces for global land areas. *International Journal of Climatology*, 25, 1965-1978.
- Hijmans, R. J. & Elith, J. (2017) Species distribution modeling with R. CRAN Project, pgs. 1-79. <https://cran.r-project.org/web/packages/dismo/vignettes/sdm.pdf>
- Hijmans, R. J., Phillips, S., Leathwick, J. & Elith, J. (2017) R Package 'dismo'. <https://cran.r-project.org/web/packages/dismo/index.html>
- Hijmans, R., Sumner, M., Macqueen, D., Lemon, J., Brien, J. O. & Rourke, J. O. (2018) R Package 'sp'.
- Hodel, R. G. & Gonzales, E. (2013) Phylogeography of sea oats (*Uniola paniculata*), a dune-building coastal grass in southeastern North America. *Journal of Heredity*, 104, 656-665.
- Howes, B. J., Lindsay, B. & Lougheed, S. C. (2006) Range-wide phylogeography of a temperate lizard, the five-lined skink (*Eumeces fasciatus*). *Molecular Phylogenetics and Evolution*, 40, 183-194.
- Hugall, A., Moritz, C., Moussalli, A. & Stanisci, J. (2002) Reconciling paleodistribution models and comparative phylogeography in the Wet Tropics rainforest land snail *Gnarosiphia bellendenkerensis* (Brazier 1875). *Proceedings of the National Academy of Sciences USA*, 99, 6112-6117.
- Joly, S. & Bruneau, A. (2004) Evolution of triploidy in *Apios americana* (Leguminosae) revealed by genealogical analysis of the histone H3-D gene. *Evolution*, 58, 284-295.
- Jombart, T. & Ahmed, I. (2011) adegenet 1.3-1: new tools for the analysis of genome-wide SNP data. *Bioinformatics*, 27, 3070-3071.
- Jombart, T. & Collins, C. (2015) A tutorial for Discriminant Analysis of Principal Components (DAPC) using adegenet 2.0.0.
- Jombart, T., Lyon, D. & Biome, L. De. (2008) adegenet: A R package for the multivariate analysis of genetic markers. *Bioinformatics*, 24, 1403-1405.
- Keeney, D. B., Heupel, M. R., Hueter, R. E. & Heist, E. J. (2005) Microsatellite and mitochondrial DNA analyses of the genetic structure of blacktip shark (*Carcharhinus limbatus*) nurseries in the northwestern Atlantic, Gulf of Mexico, and Caribbean Sea. *Molecular Ecology*, 14, 1911-1923.
- Kimura, M. (1981) Estimation of evolutionary distances between homologous nucleotide sequences b. *Proceedings of the National Academy of Sciences USA*, 78, 454-458.
- Kimura, M. (1980) A Simple Method for Estimating Evolutionary Rates of Base Substitutions Through Comparative Studies of Nucleotide Sequences. *Journal of Molecular Evolution*, 16, 111-120.
- Kozak, K. H., Blaine, R. A. & Larson, A. (2006) Gene lineages and eastern North American palaeodrainage basins: phylogeography and speciation in salamanders of the *Eurycea bislineata* species complex. *Molecular Ecology*, 15, 191-207.
- Krellwitz, E. C., Kowallik, K. V. & Manos, P. S. (2001) Molecular and morphological analyses of *Bryopsis* (Bryopsidales, Chlorophyta) from the western North Atlantic and Caribbean. *Phycologia*, 40, 330-339.
- Legendre, P., Lapointe, F. I. & Casgrain, P. (1994) Modeling Brain Evolution from Behavior: A Permutational Regression Approach. *Evolution*, 48, 1487-1499.
- Lemmon, E. M., Lemmon, A. R., Collins, J. T., Lee-Yaw, J. A. & Cannatella, D. C. (2007) Phylogeny-based delimitation of species boundaries and contact zones in the trilling chorus frogs

- (*Pseudacris*). *Molecular Phylogenetics and Evolution*, 44, 1068-1082.
- Liaw, A. & Wiener, M. (2002) Classification and regression by random Forest. *R news*, 2, 18-22.
- Lydeard, C., Minton, R. L. & Williams, J. D. (2000) Prodigious polyphyly in imperiled freshwater pearly-mussels (Bivalvia: Unionidae): a phylogenetic test of species and generic designations. *Geological Society, London, Special Publications*, 177, 145-158.
- Maggioni, R., Rogers, A. D., Maclean, N. & D’Incao, F. (2001) Molecular phylogeny of western Atlantic *Farfantepenaeus* and *Litopenaeus* shrimp based on mitochondrial 16S partial sequences. *Molecular Phylogenetics and Evolution*, 18, 66-73.
- Mateo, G., Vanderpoorten, A. & Laenen, B. (2013) Modeling species distributions from heterogeneous data for the biogeographic regionalization of the European bryophyte flora. *PLoS ONE*, 8, e55648.
- McCarthy, D. M. & Mason-Gamer, R. J. (2016) Chloroplast DNA-based phylogeography of *Tilia americana* (Malvaceae). *Systematic Botany*, 41, 865-880.
- McGovern, T. M. & Hellberg, M. E. (2003) Cryptic species, cryptic endosymbionts, and geographical variation in chemical defences in the bryozoan *Bugula neritina*. *Molecular Ecology*, 12, 1207-1215.
- McLachlan, J. S., Clark, J. S. & Manos, P. S. (2005) Molecular indicators of tree migration capacity under rapid climate change. *Ecology*, 86, 2088-2098.
- McMillen-Jackson, A. L. & Bert, T. M. (2003) Disparate patterns of population genetic structure and population history in two sympatric penaeid shrimp species (*Farfantepenaeus aztecus* and *Litopenaeus setiferus*) in the eastern United States. *Molecular Ecology*, 12, 2895-2905.
- McRae, B. H. (2006) Isolation by resistance. *Evolution*, 60, 1551-1561.
- Meirmans, P. G. (2012) The trouble with isolation by distance. *Molecular Ecology*, 21, 2839-2846.
- Morris, A. B., Ickert-Bond, S. M., Brunson, D. B., Soltis, D. E. & Soltis, P. S. (2008) Phylogeographical structure and temporal complexity in American sweetgum (*Liquidambar styraciflua*; Altingiaceae). *Molecular Ecology*, 17, 3889-3900.
- Natural Resources Canada, United States Geological Survey, and Instituto Nacional de Estadística y Geografía [NRC, USGS, and INEG] (2010a) North American Atlas – Basin Watersheds. Available at: <https://www.sciencebase.gov/catalog/item/4fb697b2e4b03ad19d64b47f>
- Natural Resources Canada, United States Geological Survey, and Instituto Nacional de Estadística y Geografía [NRC, USGS, and INEG] (2010b) North American Atlas – Hydrography. Available at: <https://www.sciencebase.gov/catalog/item/4fb55df0e4b04cb937751e02>
- Near, T. J. (2008) Rescued from synonymy: a redescription of *Percina bimaculata* Haldeman and a molecular phylogenetic analysis of logperch darters (Percidae: Etheostomatinae). *Bulletin of the Peabody Museum of Natural History*, 49, 3-19.
- Nosil, P. (2004) Reproductive isolation caused by visual predation. *Proceedings of the Royal Society B*, 271, 1521-1528.
- Nylander, J. A. A. (2004) MrModeltest v2. Program distributed by the author. Evolutionary Biology Centre, Uppsala University.
- Oliveira, L. O., Huck, R. B., Gitzendanner, M. A., Judd, W. S., Soltis, D. E. & Soltis, P. S. (2007) Molecular phylogeny, biogeography, and systematics of *Dicerandra* (Lamiaceae), a genus endemic to the southeastern United States. *American Journal of Botany*, 94, 1017-1027.
- Oksanen, J., Blanchet, F. G., Friendly, M., Kindt, R., Legendre, P., Mcglinn, D., ... Szoecs, E. (2018) R Package ‘vegan’.
- Orsini, L., Vanoverbeke, J., Swillen, I., Mergeay, J. & De Meester, L. (2013) Drivers of population genetic differentiation in the wild: isolation by dispersal limitation, isolation by adaptation and isolation by colonization. *Molecular Ecology*, 22, 5983-5999.
- Papadopoulou, A. & Knowles, L.L. (2016) Toward a paradigm shift in comparative phylogeography driven by trait-based hypotheses. *Proceedings of the National Academy of Sciences USA* 113, 29, 8018-8024.
- Paradis, E., J. Claude & K. Strimmer. (2004) APE: Analyses of Phylogenetics and Evolution in R language. *Bioinformatics*, 20, 289-290.
- Pebesma, E. & Bivand, R. S. (2005) *Classes and Methods for Spatial Data: the sp Package*. 21pp.
- Penton, E. H., Hebert, P. D. N. & Crease, T. J. (2004) Mitochondrial DNA variation in North American populations of *Daphnia obtusa*: continentalism



- or cryptic endemism? *Molecular Ecology*, 13, 97-107.
- Pelletier, T. A. & Carstens, B. C. (2018) Geographical range size and latitude predict population genetic structure in a global survey. *Biology Letters*, 14, 20170566.
- Petersen, R. H. & Hughes, K. W. (2010) The *Xerula/Oudemansiella* complex (Agaricales). *Nova Hedwigia, Beihefte Beih*, 137.
- Phillips, S. J., Dudík, M. & Schapire, R. E. (2017) Maxent software for modeling species niches and distributions (Version 3.4.1). Available at [http://biodiversityinformatics.amnh.org/open\\_source/maxent/](http://biodiversityinformatics.amnh.org/open_source/maxent/)
- Rios, N. E. & Bart, H. L. (2010) GEOLocate (Version 3.22) Belle Chasse, LA: Tulane University Museum of Natural History.
- Rissler, L. J. (2016) Union of phylogeography and landscape genetics. *Proceedings of the National Academy of Sciences USA*, 113, 8079–8086.
- Robison, H. W., Cashner, R. C., Raley, M. E. & Near, T. J. (2014) A new species of darter from the Ouachita Highlands in Arkansas related to *Percina nasuta* (Percidae: Etheostomatinae). *Bulletin of the Peabody Museum of Natural History*, 55, 237-253.
- Roe, K. J., Hartfield, P. D. & Lydeard, C. (2001) Phylogeographic analysis of the threatened and endangered superconglutinate-producing mussels of the genus *Lampsilis* (Bivalvia: Unionidae). *Molecular Ecology*, 10, 2225-2234.
- Rozas, J., Ferrer-mata, A., S, J. C., Guirao-rico, S., Librado, P., Ramos-onsins, E., ... & Gene, D. De. (2017) DnaSP 6: DNA Sequence Polymorphism Analysis of Large Data Sets. *Molecular Biology and Evolution*, 34, 3299–3302.
- Satler, J. D., Carstens, B. C. (2016) Phylogeographic concordance factors quantify phylogeographic congruence among co-distributed species in the *Sarracenia alata* pitcher plant system. *Evolution*, 70, 1105-1119.
- Sbrocco, E. J. & Barber, P. H. (2013) MARSPEC: ocean climate layers for marine spatial ecology. *Ecological Archives*, 94, 27705.
- Sexton, J. P., Hangartner, S. B. & Hoffmann, A. A. (2014) Genetic isolation by environment or distance: which pattern of gene flow is most common? *Evolution*, 68, 1-15.
- Shaw, A. J., Cox, C. J. & Boles, S. B. (2005) Phylogeny, species delimitation, and recombination in *Sphagnum* section *Acutifolia*. *Systematic Botany*, 30, 16-33.
- Shaw, J. & Small, R. L. (2005) Chloroplast DNA phylogeny and phylogeography of the North American plums (*Prunus* subgenus *Prunus* section *Prunocerasus*, Rosaceae). *American Journal of Botany*, 92, 2011-2030.
- Sidlauskas, B., Ganapathy, G., Hazkani-Covo, E., Jenkins, K. P., Lapp, H., McCall, L. W., ... & Kidd, D. M. (2010) Linking big: the continuing promise of evolutionary synthesis. *Evolution*, 64, 871-880.
- Soltis, P. S. (2006) Comparative phylogeography of unglaciated eastern North America. *Molecular Ecology*, 15, 4261–4293.
- Spalding, M. D., Fox, H. E., Allen, G. R., Davidson, N., Ferdaña, Z. A., Finlayson, M. A. X., ... Robertson, J. (2007) Marine Ecoregions of the World: A bioregionalization of coastal and shelf areas. *BioScience*, 57, 573–583.
- Sullivan, J., Arellano, E. & Rogers, D. S. (2000) Comparative phylogeography of Mesoamerican highland rodents: concerted versus independent response to past climatic fluctuations. *The American Naturalist*, 155, 755-768.
- Tam, Y. K., Kornfield, I. & Ojeda, F. P. (1996) Divergence and zoogeography of mole crabs, *Emerita* spp. (Decapoda: Hippidae), in the Americas. *Marine Biology*, 125, 489-497.
- Tamura, K. & Nei, M. (1993) Estimation of the number of nucleotide substitutions in the control region of mitochondrial DNA in humans and chimpanzees. *Molecular biology and evolution*, 10, 512-526.
- Thompson, J. D., Higgins, D. G. & Gibson, T. J. (1994) CLUSTAL W: improving the sensitivity of progressive multiple sequence alignment through sequence weighting, position-specific gap penalties and weight matrix choice. *Nucleic Acids Research*, 22, 4673–4680.
- United Nations Environment Programme-World Conservation Monitoring Centre [UNEP-WCMC] (2002) Mountains and tree cover in mountain regions. Available at <https://www.unep-wcmc.org/resources-and-data/mountains-and-tree-cover-in-mountain-regions>
- United States Environmental Protection Agency [USEPA] (2012) U.S. Level III and IV Ecoregions. Research Triangle Park, NC. Available at <https://catalog.data.gov/dataset/u-s-level-iii-and-iv-ecoregions-u-s-epa>



- Urbanek, S. (2019) R Package 'rJava'. van Proosdij, A. S. J., Sosef, M. S. M., Wieringa, J. J. & Raes, N. (2016) Minimum required number of specimen records to develop accurate species distribution models. *Ecography*, 39, 542–552.
- Vasconcellos, M. M., Ortiz, E. M., Weber, J. N., Cannatella, D. C. & Rodrigues, M. T. (2019) Isolation by instability: Historical climate change shapes population structure and genomic divergence of treefrogs in the Neotropical Cerrado savanna. *Molecular Ecology*, 28, 1748–1764.
- Vazquez-Bader, A. R., Carrero, J. C., Garcia-Varela, M., Gracia, A. & Laclette, J. P. (2004) Molecular phylogeny of superfamily Penaeoidea Rafinesque-Schmaltz, 1815, based on mitochondrial 16S partial sequence analysis. *Journal of Shellfish Research*, 23, 911–918.
- Wang, I. J. (2013) Examining the Full Effects of Landscape heterogeneity on spatial genetic variation: A multiple matrix regression approach for quantifying geographic and ecological isolation. *Evolution*, 67, 3403–3411.
- Weisrock, D. W. & Janzen, F. J. (2000) Comparative molecular phylogeography of North American softshell turtles (Apalone): implications for regional and wide-scale historical evolutionary forces. *Molecular Phylogenetics and Evolution*, 14, 152–164.
- Wise, J., Harasewych, M. G. & Dillon, R. T. (2004) Population divergence in the sinistral whelks of North America, with special reference to the east Florida ecotone. *Marine Biology*, 145, 1167–1179.
- Wright, S. (1943) Isolation by distance. *Genetics*, 28, 114–138.
- Yang, Z. L., Zhang, L. F., Mueller, G. M., Kost, G. W. & Rexer, K. H. (2009) A new systematic arrangement of the genus *Oudemansiella* s. str. (Physalacriaceae, Agaricales). *Mycosystema*, 28, 1–13.
- Young, A., Torres, C., Mack, J. & Cunningham, C. (2002) Morphological and genetic evidence for vicariance and refugium in Atlantic and Gulf of Mexico populations of the hermit crab *Pagurus longicarpus*. *Marine Biology*, 140, 1059–1066.
- Zhang, Y. H., Wang, I. J., Comes, H. P., Peng, H. & Qiu, Y. X. (2016) Contributions of historical and contemporary geographic and environmental factors to phylogeographic structure in a Tertiary relict species, *Emmenopterys henryi* (Rubiaceae). *Scientific reports*, 6, 24041.

Submitted: 01 October 2019

First decision: 20 November 2019

Accepted: 07 January 2020

Edited by Janet Franklin