Conservation Genetics
https://doi.org/10.1007/s10592-024-01601-1

RESEARCH

Using genomic data to estimate population structure of Gopher Tortoise (*Gopherus polyphemus*) populations in Southern Alabama

Alexander R. Krohn[1] · Brian Folt[2] · Joseph J. Apodaca[1] · Craig Guyer[3] · Jeffrey M. Goessling[4]

Received: 6 April 2023 / Accepted: 4 January 2024

Abstract

In the North American longleaf pine (*Pinus palustris*) ecosystem, the Gopher Tortoise (*Gopherus polyphemus*) is a keystone species that has declined significantly over the last century. Habitat degradation and fragmentation may have caused *G. polyphemus* to become separated into small, isolated local populations that suffer from decreased genetic diversity or inbreeding depression. Here we use genome-scale methods to sequence thousands of loci for 336 *G. polyphemus* individuals from 11 sites across southern Alabama to estimate population genetic structure and levels of genetic diversity. We found a pattern of isolation by distance among samples, where geographic distance predicted genetic difference. Principal components and structure analyses supported the existence of three weak genetic populations comprising individuals from (1) Fred T. Stimpson State Game Sanctuary and Perdido Wildlife Management Area, (2) Conecuh National Forest and Solon Dixon Forestry Education Center, and (3) Geneva State Forest Wildlife Management Area. We did not observe strong variation in genetic diversity or effective population size metrics among sampling locations or genetic populations identified by population structure analyses. Our results suggest that *G. polyphemus* historically operated on larger geographic scales than those considered by contemporary mark-recapture studies. Absence of variation in population genetic metrics suggests that either effects of fragmentation have not manifested themselves, or that the effects are similar across all locations. Given the common use of translocations in Gopher Tortoise management, we provide a framework for tortoise translocations based on our genomic data.

Keywords Genetic diversity · Isolation by distance · RADseq · Translocations

Introduction

Habitat loss and degradation are the primary causes of global biodiversity declines (Fischer and Lindenmayer 2007; Pereira et al. 2010; Wilson et al. 2016). Habitat loss typically leads to habitat fragmentation, where habitat is divided into smaller, more isolated patches surrounded by human-influenced landscapes that are unsuitable for many native species (Haddad et al. 2015). Fragmented landscapes can result in decreased habitat connectivity, wherein plant and animal populations experience decreased dispersal among patches, diminished reproduction within patches, and smaller population sizes through time (Fletcher et al. 2018). Decreased connectivity and population size can have myriad genetic consequences for populations (Templeton et al. 1990), including restrictions or cessation of gene flow (Delaney et al. 2010) and decreased genetic diversity (Schlaepfer et al. 2018; Lino et al. 2019; González et al. 2020), loss of diversity via genetic drift (Holderegger and

✉ Alexander R. Krohn
alex@tbconservation.org

Brian Folt
brian@csp-inc.org

Joseph J. Apodaca
jj@tbconservation.org

Craig Guyer
guyercr@auburn.edu

Jeffrey M. Goessling
goessljm@eckerd.edu

[1] Tangled Bank Conservation LLC, 196 E. Chestnut St., Suite A, Asheville, NC 28801, USA

[2] Conservation Science Partners, Inc., Truckee, CA 96161, USA

[3] Auburn University, Auburn, AL 36849, USA

[4] Eckerd College, St. Petersburg, FL 33711, USA

Published online: 20 January 2024

Di Giulio 2010), increased inbreeding depression, reduced adaptive potential, and increased frequency of deleterious alleles (Keyghobadi 2007). Through an interplay of these mechanisms, habitat fragmentation contributes to significant negative demographic effects on populations that can drive both population extirpation and ultimately even species extinctions.

In southeastern North America, habitat loss and degradation have caused the fire-maintained longleaf pine (*Pinus palustris*) ecosystem—once one of the most extensive ecosystems in the continent (Landers et al. 1995) and home to a global biodiversity hotspot (Noss et al. 2015)—to become a highly fragmented matrix that encompasses less than 3% of its historical range (Jose et al. 2006). The Gopher Tortoise (*Gopherus polyphemus*) is a keystone species in the longleaf pine ecosystem of the Southeastern United States (Guyer and Bailey 1993). Gopher Tortoises are relatively small (mean adult body mass is approximately 3 kg) compared to the giant tortoises in the family Testudinidae, but are extremely long lived, with a generation time around 60 years (Folt et al. 2021). Its status as a keystone species is determined primarily by the burrows that individuals create, maintain, and eventually abandon, the presence of which increases diversity within longleaf pine forests (Jackson and Milstrey 1989; Catano and Stout 2015). Secondarily, Gopher Tortoises achieve keystone status through their activities as grazers and frugivores, features that increase plant richness and diversity in understory plants (Richardson and Stiling 2019). Therefore, an important objective of managers seeking to conserve the longleaf pine ecosystem is to maximize the population persistence of *G. polyphemus*, a process that likely enhances habitat for other members of the rich longleaf pine biota. However, abundance of *G. polyphemus* has declined (Auffenberg and Franz 1982, McCoy et al. 2006) due largely to habitat degradation and fragmentation of the longleaf pine ecosystem, which has separated tortoises into small, isolated local populations that may suffer increased risk of extirpation through negative genetic demographic effects (e.g., drift, decreased genetic diversity, inbreeding depression) (Ennen et al. 2012; Hedrick and Garcia-Dorado 2016). However, such extirpation may be difficult to document because individual tortoises are long-lived, and habitat patches might remain occupied by the species despite a lack of recruitment required for long-term persistence in the patch.

Population genetic research on *G. polyphemus* to date has described phylogeographic and population genetic structure across the species' range (Osentoski and Lamb 1995; Schwartz and Karl 2006; Ennen et al. 2010, 2012; Richter et al. 2011; Gaillard et al. 2017), potential effects of insular processes on population genetics (Ennen et al. 2012; Winters et al. 2017), and effects of male body size on reproductive success (Moon et al. 2006; White et al. 2018). Yuan et al. (2019) linked relatedness-based inbreeding estimates to decreased hatching success of eggs, suggesting that small populations with increased inbreeding might experience decreased recruitment relative to larger, more genetically diverse populations (Yuan et al. 2019). If so, then the widespread fragmentation of tortoises into smaller, less connected local populations may increase inbreeding and decrease recruitment, contributing to population declines. While some recent studies have used mark-recapture methods to estimate population vital rates and population dynamics in *G. polyphemus* (Tuberville et al. 2008; Howell et al. 2020; Goessling et al. 2021; Folt et al. 2021; Hunter and Rostal 2021), they did not quantify genetic diversity of populations nor evaluate whether the populations are experiencing these predicted genetic effects of habitat fragmentation.

Negative effects of habitat fragmentation on populations operate at the scale of dispersal (Fletcher et al. 2018), but uncertainty in how to measure dispersal and what constitutes a population of *G. polyphemus* challenges our understanding of how the species might be influenced by fragmentation. One radiotelemetry study demonstrated that adult tortoises with established home ranges will occasionally emigrate (> 1 km linear distances) from a local population (Eubanks et al. 2003). If such populations (*sensu* Goessling et al. 2021) are effectively connected across the landscape by dispersal events, then genetic populations (*sensu* Gaillard et al. 2017) may occur over substantially larger areas within which deleterious effects of fragmentation may be offset by migration. However, to our knowledge no studies have evaluated how or to what degree dispersal effectively connects local populations within and among conservation lands that serve to prevent extirpation of tortoise populations. Further, given the long dispersal distances that may connect local populations of tortoises through effective migration and gene flow, widespread fragmentation of the upland longleaf pine ecosystem may be decreasing genetic connectivity and increasing the risk of negative genetic consequences for *G. polyphemus* populations. Therefore, in order to conserve this keystone species with such varied connections to diversity in the longleaf ecosystem, wildlife managers must better understand the genetic viability of local populations of *G. polyphemus*.

Genetic data allow us to measure population connectivity across wide spatial scales. While mark-recapture studies may take many years to quantify demographic rates for long-lived species (Gibbons and Semlitsch 1982), population genetic analysis can integrate the results of migration, reproduction, and fragmentation through either structured or unstructured populations. While mark-recapture studies measure contemporary movement (i.e., contemporary population structure), population genetic studies can reveal both contemporary and historical patterns (i.e., historical population structure), depending on the loci and the species' biology. Thus, genetic analysis provides a useful method to

determine whether contemporary effects of fragmentation are strong enough to disrupt historical connections among local populations (Epps and Keyghobadi 2015).

To this end, we used high-throughput methods to sequence thousands of genetic loci of 336 *G. polyphemus* individuals from 11 sites across southern Alabama, including seven sites within or immediately adjacent to Conecuh National Forest, a landscape managed to restore habitat quality and connectivity, and four sites that are smaller, more isolated by habitat fragmentation, and degraded in habitat quality. Our objectives were to delimit genetic populations of *G. polyphemus* in southern Alabama using population genetic methods and to quantify whether contemporary or historic processes better explained observed genetic structure within or among local populations. Combining our genetic data with previous mark-recapture data from the same sites (Goessling et al. 2021; Folt et al. 2021) allowed us to infer differences between observed genetic, and field-based estimates of population structure. We predicted that contemporary habitat fragmentation would best explain observed genetic structure, and that metrics of population genetic health (genetic diversity measured by heterozygosity and effective population size) would be greater for local populations of tortoises within or adjacent to a large and relatively unfragmented landscape, Conecuh National Forest, than for tortoises from sites in more fragmented landscapes.

Methods

Study sites – The study sites (Fig. 1) were chosen within key land tracts for *G. polyphemus* conservation across the non-federally protected range of the species in Alabama as animals were already being sampled for a study investigating the occurrence and prevalence of Upper Respiratory Tract Disease (URTD; Goessling et al. 2019). As part of that study, the original goal was to sample a minimum of 25 adult tortoises from each site. For some sites, intensive sampling of adjacent individuals was performed for animals that had been monitored long-term via capture-mark-recapture methods. For all other sites, samples were collected across large conservation properties lacking consistent long-term sampling efforts.

We outline the sites where we sampled tortoises in greater detail in Online Resource 1. Briefly, we sampled at six sites in Conecuh National Forest (hereafter Conecuh), a large federally managed forest that, since the 1970's, is being restored to have longleaf pine as the dominant overstory species (Aresco and Guyer 1999). The six sites that we sampled have been studied as part of 30-year mark-recapture studies and vary in their demographic trends (Goessling et al. 2021; Folt et al. 2021). We also sampled tortoises opportunistically that were encountered on roads outside of the six previously studied sites. Just north of Conecuh, we sampled tortoises at Solon Dixon Forestry Education Center (hereafter Solon Dixon). East of Conecuh, we sampled tortoises at Geneva State Forest (hereafter Geneva) and Rayonier Tract (hereafter Rayonier). At the time of fieldwork these sites were managed separately, but today they are managed together as part as the Geneva State Forest Wildlife Management Area. These two sites have small pockets of Gopher Tortoise habitat separated by clear cuts, fire-suppressed vegetation or wetlands. Much further west, we sampled tortoises at the Perdido River Wildlife Management Area (hereafter Perdido) and the Fred T. Stimpson State Game Sanctuary (hereafter Stimpson). Both these sites have smaller, isolated patches of habitat surrounded by wetlands or fire-suppressed areas.

Sample collection

We collected blood samples ($n = 336$) from *G. polyphemus* individuals as part of trapping efforts described in Goessling et al. (2019) and Folt et al. (2021). We collected blood from the subcarapacial sinus using a sterile 25 or 26 ga. needle affixed to a sterile pre-heparinized syringe; prior to venipuncture, skin was decontaminated with an alcohol wipe. We kept blood samples in microcentrifuge tubes on ice until they could be centrifuged to separate blood cells from plasma. Cells and plasma were then separated and frozen in liquid nitrogen before final storage in an ultracold freezer at − 80 °C. All blood samples were flash frozen in liquid nitrogen within 2 h of collection.

Genomic sequencing

For both objectives, we constructed a restriction-site-associated-DNA sequencing (RADseq) library to analyze genome-scale data. We extracted blood samples using DNAeasy kits (Qiagen, Hilden, Germany) following standard protocols. We verified extraction quality by gel electrophoresis and quantified each sample using a Qubit 3.0 Fluorometer (Life Technologies, Carlsbad, California). We prepared the RADseq library following a modified 3RAD protocol (Bayona-Vásquez et al. 2019). The library preparation steps were double enzyme digest, adapter ligation, limited cycle PCR, and a 1.2 concentration Serapure SpeedBead cleanup (Rohland and Reich 2012). In the double enzyme digest mixture, we increased the amount of genomic DNA to 10 µL and decreased dH_20 to 0.5 µL. We used the restriction enzymes ClaI, BamHi, and MspI (New England Biolabs, Ipswich, Massachusetts) for our digestion. To multiplex each sample, we used i5 and i7 iTru adapters and primers as dual internal indexes (Glenn et al. 2019). Following the SpeedBead clean-up, we visualized these libraries on a gel and quantified their concentrations using a Qubit. We pooled libraries to 100 ng/µL in pools of 48 individual libraries, cleaned

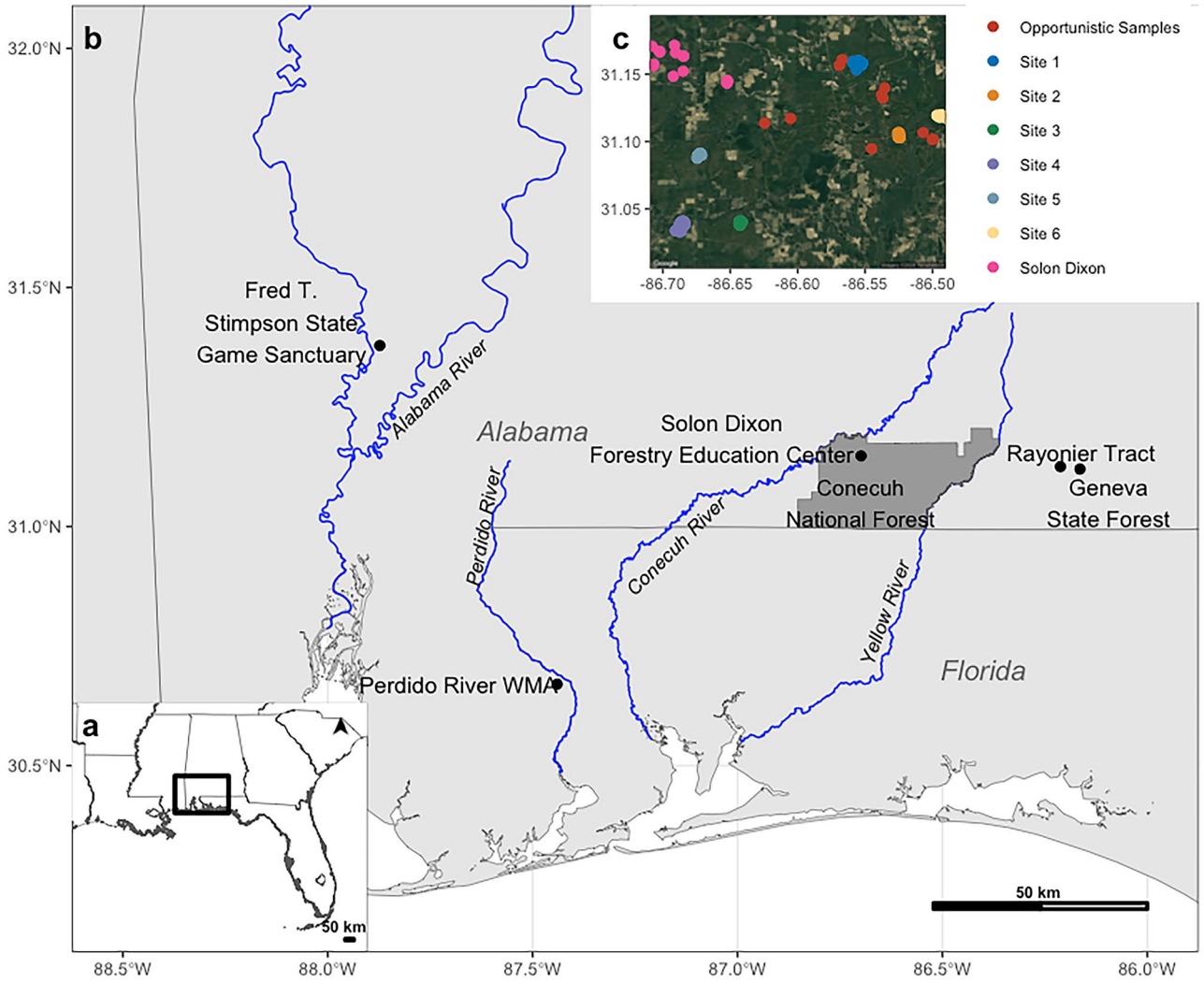


Fig. 1 Map of the study area and sampling locations. **a** The southeastern United States, with the study area indicated by the black rectangle. **b** Sampling locations (black dots) in southern Alabama, including Fred T. Stimpson State Game Sanctuary, Perdido Wildlife Management Area, Solon Dixon Forestry Education Center, Conecuh National Forest (administrative boundary in gray), Rayonier Tract, and Geneva State Forest. **c** Samples from six sites (colored dots) within Conecuh National Forest and Solon Dixon Forestry Education Center (Solon Dixon); opportunistic samples outside of these sites are indicated in red as 'opportunistic samples'. Satellite map is provided by Google through the R package ggmap (Kahle and Wickham 2013)

them again using SpeedBeads, and eluted the pools to 32 µL. We then size-selected the pools to the range of 400–600 bp using a Pippin Prep (Sage Science Inc., Beverly, Massachusetts). We quantified the final DNA concentration using a Qubit and then sent them to Genewiz (Azenta Life Sciences, South Plainfield, New Jersey) for sequencing on an Illumina NovaSeq lane with 150 bp paired-end reads.

Data filtering, clustering, and genotyping

From the raw sequenced reads, we removed adapter sequences, filtered reads, removed PCR duplicates, clustered reads into *de novo* RAD loci, aligned reads to the RAD loci, called SNPs, and generated genotype files using ipyrad version 0.9.81 (Eaton and Overcast 2020). We used the default settings for ipyrad except for the minimum depth required to call a base (10X, parameter file line numbers 11 and 12) and clustering threshold (parameter file line number 14). We tested a variety of clustering thresholds to determine which threshold best clustered likely homologous loci, while splitting likely paralogous loci (McCartney-Melstad et al. 2019). Following McCartney-Melstad et al. (2019) we calculated increases in divergence levels among SNPs per 100 km (changes in the slope of measured isolation by distance), variance explained by the first four Principal Component (PC) axes, and the Pearson's Correlation Coefficient

between missingness and genetic distance at four clustering thresholds: 0.80, 0.85, 0.90, 0.92. We chose the clustering threshold that maximized the number of SNPs, minimized change in the slope of isolation by distance, maximized variance explained by the first four PC axes, and minimized the correlation between genetic distance and missingness. Using these criteria, we created a variant call file (VCF) from ipyrad, which we used for subsequent analyses.

Each subsequent analysis is affected differently by missing data. We explain specific filtering thresholds below. In general, we endeavored to maximize the number of SNPs included while minimizing the effect of missing data on the analysis.

Isolation by distance

Before we began delimiting populations, we first tested for isolation by distance (Wright 1943). Isolation by distance is a pattern whereby geographically proximate individuals tend to have greater genetic similarity than geographically distant samples. Animals with limited dispersal and sampled across large distances, as in our study, may show strong patterns of isolation by distance. To delimit populations, it is important to disentangle the continuous process of isolation by distance from discrete patterns that result from barriers to gene flow (Novembre and Stephens 2008; Jombart et al. 2008; Frantz et al. 2009; Bradburd et al. 2018; Dutcher et al. 2020). To determine whether isolation by distance existed in our dataset, we plotted pairwise genetic differentiation (π) against geographic distance for all pairs of individuals. We calculated pairwise geographic distance from the most recent capture locations of each individual and estimated π using the 'snpgdsIBS()' function with package SNPRelate (Zheng et al. 2012) in Program R (R Core Team 2021). We used a Mantel test to determine whether pairwise genetic distance increased with pairwise geographic distance, using the 'mantel()' function with 999 permutations in the R package vegan version 2.6.4 (Oksanen et al. 2022). Although Mantel tests are likely more appropriate for pairwise distance matrices, we also ran a linear regression in R with genetic distance as the independent variable and geographic distance as the dependent variable for heuristic purposes.

Population structure

Clustering

We used two general methods to delimit population structure of *G. polyphemus*. First, we used spatial and non-spatial clustering methods to estimate the number of populations and pattern of admixture among individuals that best described the observed genetic variation. These clustering methods assign individuals to a given number of populations with no *a priori* location information. Second, we visualized overall population structure using spatial and non-spatial PCA. Because these methods have different requirements for missing data, we filtered the loci and SNPs accordingly. We explain the filtering processes below but report the results of these filters in the Results.

We used two clustering methods to determine the most likely number of populations and amount of admixture among individuals in our dataset. We used both a spatially explicit clustering method, conStruct (Bradburd et al. 2018), and a non-spatial clustering method, fastSTRUCTURE (Raj et al. 2014). conStruct uses a Markov Chain Monte Carlo simulation (MCMC) to model the admixture proportions of individuals that best explain the genetic data for a given number of populations (K; "layers" in conStruct), then tests how well the model explains the dataset. conStruct assumes that the relationship between allele frequencies and geographic distance is positive definite, and thus requires that there be few missing data. conStruct also requires a large training dataset for the MCMC, and thus requires many more SNPs than individuals. To this end, we filtered the original VCF stringently to accommodate conStruct analysis. First, using VCF tools, we only included biallelic sites that had less than 10% missing data using the program VCFTools (Danecek et al. 2011). We then selected a single SNP per RAD locus using the single_snp.py script (available at https://github.com/bmichanderson/RAD_scripts/) to increase the probability that SNPs in our dataset are independent. We then converted this VCF to the 'Structure' format using PGDSpider (Lischer and Excoffier 2012). Finally, we removed all individuals with more than 20% missing data. We attempted to include as many individuals as possible while maintaining positive definite covariances. However, where we still had too many individuals given our final number of SNPs, we chose the 10 individuals with the highest number of total loci per sampling location to ensure our final dataset contained few missing data, while still containing individuals from each sampling location. We ran the conStruct MCMC for 10,000 iterations, with 90% of the data used for model training, ran each value of K for 10 repetitions, and confirmed chain convergence and mixing before continuing with further analyses (Bradburd et al. 2018). fastSTRUCTURE also assigns individuals levels of admixture from a given number of populations, and tests the model's fit of the dataset, but it does not take geographic distance into account. For fastSTRUCTURE, we filtered the dataset using the same methods as in conStruct, but used slightly less stringent filters: we only included biallelic SNPs that had less than 20% missing data, included only a single SNP per RAD locus, and did not filter missingness on an individual level (Hodel et al. 2017). We ran fastSTRUCTURE using the logistical prior to better estimate fine-scale population structure.

In conStruct, we determined (1) whether a spatial or non-spatial model best described the data and (2) which value of K best described the data using both cross validation and layer contributions for conStruct following Bradburd et al. (2018). We used the included chooseK.py function in fastSTRUCTURE to determine which value of K best explained the structure in the dataset (Raj et al. 2014).

Because the strict data requirements for conStruct resulted in smaller sample sizes compared to the fastSTRUCTURE analyses (see Results), we visually compared results from a non-spatial conStruct model using fewer individuals to a fastSTRUCTURE with many more individuals to ensure that sub-setting individuals did not affect our conclusions.

To avoid over-interpreting population clustering analyses, we used a hierarchical approach and examined structure for all K values (Janes et al. 2017; Lawson et al. 2018). We first ran conStruct and fastSTRUCTURE on all individuals and evaluated support for K values from 1 to 7. We then removed the most divergent individuals, performed analyses again, and continued this process until it was clear the remaining individuals represented a single population (i.e., $K = 1$ best described the dataset).

Principal components analysis

Next, we visualized population structure using PCA. PCA derives composite orthogonal independent PC axes that best represent the greatest amount of variation in the SNP dataset, with individuals that are genetically most similar tending to cluster closer together. Unlike other methods, the PCA does not make assumptions of relatedness, Hardy-Weinberg equilibrium, etc., and can provide a relatively unbiased overview of genetic structure. We ran the PCA using ipyrad's built-in analysis tools (Eaton and Overcast 2020), which imputed the values of missing SNPs to minimize their effect on the analysis. For the PCA, we only included one SNP per RAD locus, and only included SNPs present in more than 50% of individuals. We used the "sampled" method of imputation for missing data, which randomly samples genotypes based on the frequency of alleles across all samples. We repeated the PCA for hierarchical structure levels where $K > 1$, as identified by the above fastSTRUCTURE and conStruct analyses.

Given evidence of isolation by distance in our data (see Results), we also used spatial PCA (sPCA) (Jombart et al. 2008) with package 'adegenet' (Jombart 2008) in R. Our sPCA used a spatially-explicit model to disentangle patterns of global allele frequency variation from local patterns and visualized underlying patterns of differentiation from a background of isolation by distance. Because sPCA is highly sensitive to missing data, we filtered our dataset stringently. First, we filtered SNPs to only include biallelic SNPs present in 90% of individuals using VCFTools. To increase the probability of each SNP being independent, we only included one SNP per RAD locus (as above). Next, we filtered individuals such that no individual had more than 40% missing data. To ensure that the amount of missing data per individual did not affect our conclusions, we ensured that there was no significant linear correlation between missingness per individual and PC1 or PC2. We assessed the correlation between missingness per individual, PC1 and PC2 by running a linear regression in R. We ran the sPCA with a geographic distance connection network consisting of nodes connected by a maximal linear geographic distance of 0.05, or approximately 5 km, a high upper limit of dispersal in *G. polyphemus* (Eubanks et al. 2003). We repeated the sPCA for hierarchical structure levels where $K > 1$, as identified by the above fastSTRUCTURE and conStruct analyses.

Population genetic metrics

For each of the study sites and delimited genetic populations from conStruct and sPCA analyses above, we calculated a variety of population genetic statistics to determine how populations varied in levels of genetic diversity, inbreeding, effective population size, and connectedness. Because population statistics, especially F_{ST}, are sensitive to missing data (Hodel et al. 2017), we filtered the original VCF to only include biallelic SNPs present in 85% of individuals, and only included one SNP per RAD locus (as above). To estimate genetic diversity, we calculated the observed heterozygosity (H_O; Nei 1987) and within population gene diversity (i.e., expected heterozygosity, H_S; Nei 1987; Goudet 2005) using the package 'hierfstat' (Goudet 2005) in R. To quantify levels of inbreeding, we calculated within-population subdivision (F_{IS}; Nei 1987) using the package 'hierfstat' (Goudet 2005). We calculated effective population size (N_e) in the program NeEstimator (Do et al. 2014) using the bias-corrected linkage disequilibrium method and reported values of N_e with a minor allele frequency cutoff of 0.05 (Waples and Do 2010). Finally, as a measure of population connectedness or differentiation, we calculated pairwise F_{ST} (Weir and Goudet 2017), again using the 'hierfstat' (Goudet 2005). For H_O, H_S, F_{IS}, and N_e, we grouped individuals by collection site and by the clusters identified by our population structure analyses (see Results). For F_{ST}, we only grouped individuals by the clusters identified by our structure analyses. We evaluated F_{ST} values as low (i.e., little genetic differentiation) when $F_{ST} < 0.05$ and moderate when $0.05 > F_{ST} > 0.15$ (Hartl and Clark 1997). For each pairwise cluster that we calculated F_{ST}, we also ran a permutation test to test whether the F_{ST} value was significantly higher than what one would expect by chance (i.e., if the individuals were randomly grouped). For each pairwise F_{ST}, we permutated the assignments of each individual 1,000 times and calculated F_{ST} each time. We compared this "null distribution" of F_{ST}

values to the actual F_{ST} value. If the actual F_{ST} value was higher than the 95% confidence interval of the null distribution, we considered the F_{ST} value significant at $\alpha = 0.05$. If the confidence interval contained 0, and the actual F_{ST} value was outside of the confidence interval, we could also say that the actual F_{ST} value was significantly different than 0.

Results

We sequenced 1.05 billion reads from 336 *G. polyphemus*. We determined that the clustering threshold of 0.88 maximized the number of SNPs, minimized change in the slope of isolation by distance, maximized variance explained by the first four PC axes, and minimized the correlation between genetic distance and missingness. After filtering, clustering, and genotyping with ipyrad, our final dataset contained a total of 510,648 SNPs on 140,609 RAD loci, although each analysis used a much smaller subset of these total SNPs due to differences in filtering parameters.

We found a significant relationship between pairwise genetic distance of individuals and geographic distance. Genetic distance increased linearly with geographic distance between individuals (Mantel $r = 0.55$, $P = 0.001$; Fig. 2), indicating an effect of isolation by distance. The linear regression between genetic and geographic distance is also significant ($R^2 = 0.36$, $F = 0.0003$, $P < 0.001$), with a slope indicating that genetic distance increases by 1% with every 100 km in geographic distance.

fastSTRUCTURE analyses of all individuals ($n = 336$; 3095 filtered SNPs) suggested that $K = 3$ best explained structure in the data (Fig. 3a). Examination of the admixture proportions at $K = 3$ (Fig. 3a) revealed that most individuals from Conecuh, Solon Dixon, Geneva and Rayonier comprised a single genetic population, while individuals from Perdido and Stimpson formed a second genetic population. A few individuals were diagnosed as coming from a third genetic population, mostly in Perdido, Conecuh - Site 2, and Rayonier. Overall, there was some, but limited, admixture between Perdido + Stimpson and Conecuh + Geneva + Rayonier + Solon Dixon. Examination of all the fastSTRUCTURE plots for these 336 individuals (Online Resource 2) revealed similar patterns: Perdido+Stimpson was generally distinct from the other locations, Conecuh+Solon Dixon formed a single genetic population; and Geneva + Rayonier was sometimes distinct from Conecuh + Solon Dixon.

As Perdido + Stimpson was the most consistently differentiated group, we removed individuals from those groups for our next hierarchical analysis. Removal of Perdido and Stimpson ($n = 280$ remaining, 2,717 filtered SNPs, run for $K = 1$ through $K = 7$) revealed that $K = 4$ best explained the structure in the data. At $K = 4$, Geneva + Rayonier was

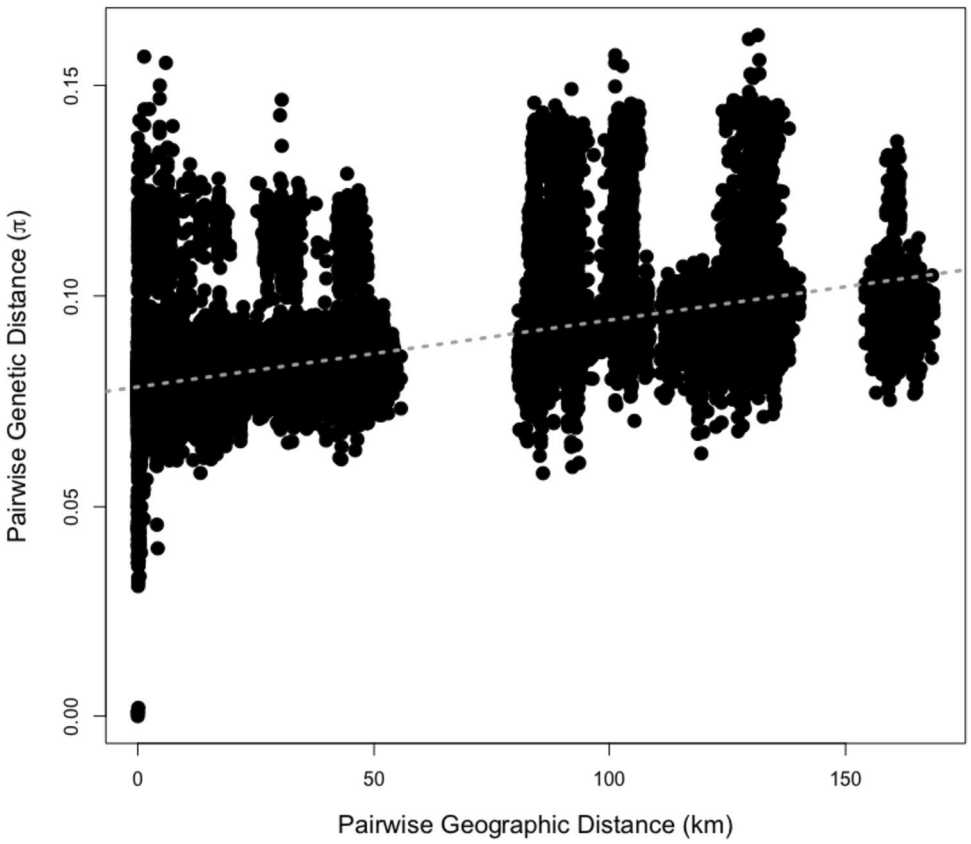
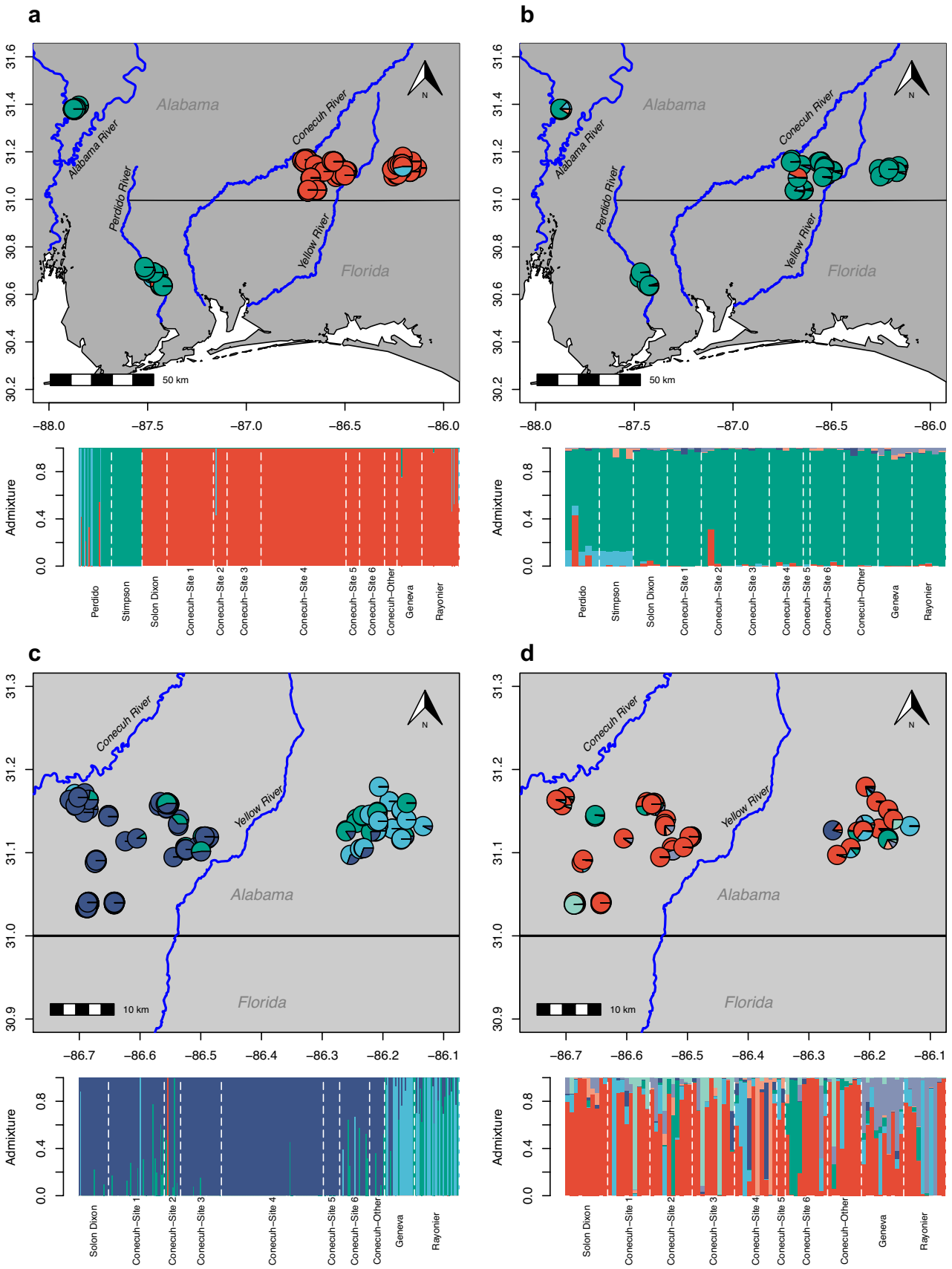


Fig. 2 Pairwise genetic distance by geographic distance for Gopher Tortoise (*Gopherus polyphemus*) samples ($n = 336$; 56,280 total comparisons). Pairwise genetic distance increases significantly with geographic distance (Mantel $r = 0.55$, $P = 0.001$). Regression line is shown for heuristic purposes


a
b
c
d
Alabama
Florida
Conecuh River
Alabama River
Perdido River
Yellow River
N
50 km
10 km
Admixture
Perdido
Stimpson
Solon Dixon
Conecuh–Site 1
Conecuh–Site 2
Conecuh–Site 3
Conecuh–Site 4
Conecuh–Site 5
Conecuh–Site 6
Conecuh–Other
Geneva
Rayonier

◂**Fig. 3** Cluster-based population assignment analyses for various subsets of Gopher Tortoise (*Gopherus polyphemus*) samples from sites in southern Alabama. Results from analyses are shown both as a map (top) and as a bar plot (bottom) for **a** all locations using non-spatial fastSTRUCTURE, **b** all locations using spatially explicit conStruct analyses. **c** all locations excluding Perdido and Stimpson using non-spatial fastSTRUCTURE and **d** all locations excluding Perdido and Stimpson using spatially explicit conStruct (**d**). Each pie chart or bar represents a single individual composed of admixture from differently colored inferred populations. Sites are: Conecuh National Forest (Conecuh), Fred T. Stimpson State Game Sanctuary (Stimpson), Geneva State Forest (Geneva), Perdido River Wildlife Management Area (Perdido), Rayonier Tract (Rayonier), and Solon Dixon Forestry Education Center (Solon Dixon). Samples from Conecuh National Forest were further divided into six sites ('Sites 1–6') and other opportunistic samples ('Other')

generally separate from Conecuh + Solon Dixon, but each site showed admixture with all four "populations" (Fig. 3c). This pattern generally held when examining all other fastSTRUCTURE plots for this subset (Online Resource 3). We also ran fastSTRUCTURE on Conecuh + Solon Dixon ($n = 225$, 4,043 filtered SNPs, K = 1 through K = 5 tested), Geneva + Rayonier ($n = 55$, 7,316 filtered SNPs, K = 1 through K = 4 tested), and Perdido + Stimpson ($n = 56$, 472 filtered SNPs, K = 1 through K = 4 tested). For each of these subsets, most individuals comprised a single genetic population, even when K > 1 (Online Resource 4).

conStruct analyses required us to subset the dataset of all individuals from 336 to 57 individuals, while retaining individuals from each sampling location (see Methods). Non-spatial conStruct runs (Online Resource 5a) were similar to fastSTRUCTURE results in that, for the most likely K (K = 6), most individuals fell into one of two genetic clusters: one comprising individuals from Conecuh + Solon Dixon + Geneva + Rayonier and a second comprising individuals from Perdido + Stimpson. Thus, we concluded that reducing the number of individuals from 336 in the fastSTRUCTURE analysis to 57 in the conStruct had little effect on inferences of genetic structure.

conStruct analyses of all individuals ($n = 57$, 8,422 filtered SNPs, tested at K = 1–7) revealed that the spatial model often, but not always, better fit the data than the non-spatial model and that the spatial model of K = 6 maximized the predictive accuracy of the model (Online Resource 6a). For all K values, a single population (K) contributed most to the dataset (Online Resource 6c). Accordingly, when plotted at K = 6, almost all individuals comprised a single genetic cluster (Fig. 3b). Of the remaining five populations, individuals from Perdido + Stimpson had approximately 10% of their admixture from one population, a smattering of individuals had small amounts of admixture from a second, and most individuals had small amount of admixture from the remaining three. Examination of all other values of K revealed a similar pattern of most individuals deriving from a single genetic population (Online Resource 7a).

As in the fastSTRUCTURE analyses, we next removed Perdido + Stimpson individuals. As above, conStruct analyses forced us to reduce the number of individuals in our analysis from 280 to 90 (see Methods). Non-spatial conStruct results (Online Resource 5b) were similar to those from fastSTRUCTURE. At the most likely K (K = 7), individuals from Geneva+Rayonier comprised a single genetic population, and individuals from Conecuh+Solon Dixon formed a second population, although they were heavily admixed with other locales. Thus, we concluded that reducing the number individuals from 280 in the fastSTRUCTURE analysis to 90 in the conStruct had little effect on conclusions regarding genetic structure.

Removing Perdido + Stimpson ($n = 90$, 3,691 filtered SNPs, tested at K = 1–7) revealed that the spatial model always outperformed the non-spatial model (Online Resource 6d), and that the predictive accuracy of the model increased with increasing values of K (Online Resource 6e). In general, for all models tested except K = 6, individuals formed one population (Online Resource 6f). When plotting K = 7, again most individuals formed a single genetic population, although there was considerable admixture with other populations (Fig. 3d). Individuals in Geneva + Rayonier derived from and showed admixture between two genetic populations (purple and blue in the figure), although some admixture was present for all clusters. Examination of all other values of K also revealed most individuals comprise a single genetic population even at K = 6, with admixture among the remaining populations (Online Resource 7b).

Despite finding that removal of Perdido + Stimpson yielded one likely genetic population, we continued subsetting as in the fastSTRUCTURE analyses for ease of comparison. conStruct analyses of samples from Stimpson + Perdido ($n = 20$: 6,921 filtered SNPs, tested at K = 1–4), Conecuh + Solon Dixon ($n = 70$: 2,776 filtered SNPs, tested at K = 1–5), and Geneva + Rayonier ($n = 20$, 11,096 filtered SNPs, tested at K = 1–4) each likely formed a single genetic population based on a combination of their predictive accuracy, layer contributions, and visualized admixture proportions (Online Resource 8).

The PCA of all individuals ($n = 336$, 18,991 SNPs) revealed two slightly separate groups (Fig. 4a). The first group, Stimpson + Perdido, separated from all other samples along PC1. The second group, Conecuh + Geneva + Rayonier + Solon Dixon, showed much more variation along PC2, rather than PC1. Still, the two groups are nearly contiguous in the PCA plot. There was considerable variation within Conecuh along PC2. Removing Stimpson and Perdido ($n = 280$: 21,327 filtered SNPs) revealed that most of the variation in Conecuh was due to

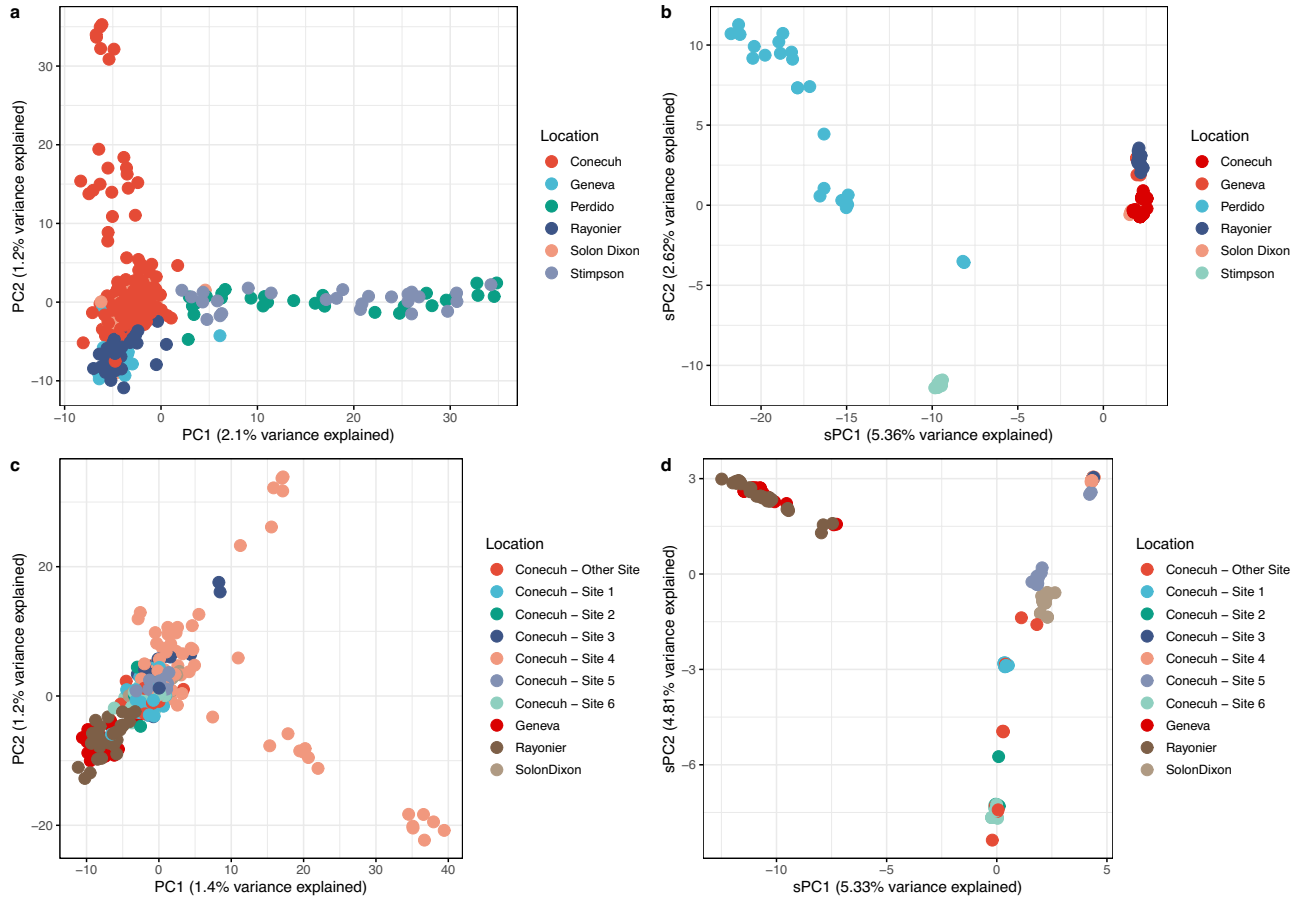


Fig. 4 Principal components analyses (PCA) and spatial principal component analyses (sPCA) for all Gopher Tortoises (**a** and **b**, respectively), and all Gopher Tortoises excluding individuals from Perdido and Stimpson (**c** and **d**, respectively). Dots represent one individual colored by their sampling location. Conecuh in a and b includes Sites 1 through 6, and other as shown in plots (**c** , **d**)

individuals from Site 4 (Fig. 4c). The other individuals generally formed a single group.

sPCA of all individuals ($n = 336$: 548 filtered SNPs) showed the same groups as the PCA, but with greater separation among them (Fig. 4b). Additionally, the sPCA of all individuals showed differentiation between Conecuh and Geneva + Rayonier, and between Stimpson and Perdido. Removing Stimpson and Perdido ($n = 278$: 1,241 filtered SNPs) revealed the same two groups (Fig. 4d), Conecuh + Solon Dixon and Geneva + Rayonier, but did not show the same differentiation for Site 4 as did the PCA. The sPCA analysis did show differentiation among sites within Conecuh (Fig. 4d) along sPC2. In general, individuals from each site clustered together, with maximal divergence along sPC2 occurring between overlapping individuals from Sites 3 and 4 and overlap of individuals from Sites 2 and 6. Individuals representing haphazard captures along roads connecting the six sites were distributed along sPC2. While most individuals from Site 5 clustered near all individuals from Solon Dixon, a few Site 5 individuals clustered with individuals from Sites 3 and 4.

We calculated a variety of population genetic statistics for 336 *G. polyphemus* individuals using 3095 filtered SNPs (Table 1). In general, most sites or genetically identified populations had similar levels of genetic diversity, little signs of inbreeding or outbreeding depression, and low N_e. F_{ST} values ranged from 0.01 to 0.10, indicating low to moderate genetic differentiation across the areas sampled (Hartl and Clark 1997) (Table 2). All F_{ST} values were significantly different than random, corroborating the population structure identified by conStruct, and were significantly different from 0.

Discussion

Understanding the spatial scale that connects local populations through demographic processes is critical to guide conservation. For genetic studies across relatively large areas,

Table 1 Summary population genetic statistics for Gopher Tortoise (*Gopherus polyphemus*) samples from Southern Alabama collected during 2013–2020. Samples were collected from state or federal lands, including Conecuh National Forest (Conecuh), Fred T. Stimpson State Game Sanctuary (Stimpson), Geneva State Forest (Geneva), Perdido River Wildlife Management Area (Perdido), Rayonier Tract (Rayonier), and Solon Dixon Forestry Education Center (Solon Dixon). The final three rows represent samples aggregated across study locations that were within population genetic groups identified by STRUCTURE analyses: all samples within Conecuh + Solon Dixon, Geneva + Rayonier, and Stimpson + Perdido. Columns are population genetic metrics: sample size (n), observed heterozygosity (H_O), expected heterozygosity (H_S), within population subdivision (F_{IS}), and effective population size (N_e). 95% confidence intervals for metrics are included within parentheses. Ne could not be calculated in some instances when there was insufficient variation among samples to distinguish signal from sampling error

Location	n	H_O	H_S	F_{IS}	N_e
Conecuh[a]	203	0.090	0.092	0.023	89.3 (87.3–91.4)
Site 1	41	0.089	0.091	0.018	28.3 (27.2–29.5)
Site 2	12	0.095	0.099	0.035	15.4 (14.3–16.7)
Site 3	30	0.091	0.089	− 0.026	17.5 (16.9–18.1)
Site 4	75	0.089	0.090	0.017	23.3 (22.9–23.8)
Site 5	12	0.088	0.089	0.006	27.7 (23.3–34.0)
Site 6	22	0.093	0.092	− 0.006	10.6 (10.2–11.0)
Stimpson	27	0.094	0.095	0.009	93.8 (76.6–120.0)
Geneva	22	0.095	0.098	0.029	172.7 (135.3–237.1)
Perdido	29	0.115	0.117	0.017	8.7 (8.4–8.9)
Rayonier	33	0.102	0.103	0.008	85.6 (78.4–94.3)
Solon Dixon	22	0.091	0.092	0.018	203.3 (152.8–301.1)
Conecuh + Solon Dixon[a]	225	0.090	0.092	0.023	112.0 (109.4–114.8)
Geneva + Rayonier	55	0.099	0.101	0.018	208.0 (186.4–234.8)
Stimpson + Perdido	56	0.105	0.109	0.034	46.7 (44.6–49.1)

[a]Includes samples opportunistically collected between Conecuh sites

Table 2 Pairwise F_{ST} values between genetic populations identified by analyses of population structure. Conecuh + Solon Dixon = samples from Conecuh National Forest and Solon Dixon Forestry Education Center; Geneva + Rayonier = samples from Geneva State Forest and Rayonier Tract; Perdido + Stimpson = samples from Perdido River Wildlife Management Area and Fred T. Stimpson State Game Sanctuary. Permutation tests show that all F_{ST} values are significantly higher than if the individuals were assembled randomly, and significantly different than 0

	Conecuh + Solon Dixon	Geneva + Rayonier	Perdido + Stimpson
Conecuh + Solon Dixon	–		
Geneva + Rayonier	0.0171	–	
Perdido + Stimpson	0.0961	0.1006	–

isolation by distance is often significant (Rousset 1997; Vekemans and Hardy 2004), including for reptiles in southeastern North America (e.g., Folt et al. 2019; Nikolakis et al. 2021). We observed a large-scale pattern of isolation by distance for *G. polyphemus* in south Alabama, both within and among groups of samples. However, our PCA, sPCA, and fastSTRUCTURE analyses of population structure consistently supported three genetic populations (Stimpson + Perdido, Conecuh + Solon Dixon, and Geneva + Rayonier) albeit with low levels of differentiation among them. Given that our samples were often separated by large rivers and distances (> 100 km) greater than those regularly traversed by a tortoise, we infer that the population structuring we observed primarily represents historical genetic populations that were structured by both landscape factors and distance.

Our results suggested an historic pattern of well-connected genetic populations over a large, once contiguous, longleaf pine ecosystem for *G. polyphemus* in southern Alabama. Given the low differentiation over large distances, we infer that these populations were once connected by effective dispersal. While previous demographic research on *G. polyphemus* largely focused on studying survival and reproduction within local populations (< 1 km^2) (Howell et al. 2020; Goessling et al. 2021; Folt et al. 2021; Hunter and Rostal 2021), the genetic populations and isolation by distance that we identified occurred at relatively large, regional scales (> 1,000 km^2). Even with small dispersal distances, this suggests that *G. polyphemus* individuals historically interacted on substantially larger scales than are documented by mark-recapture studies of local populations in contemporary landscapes. Our results are similar to those of Clostio et al. (2012), who also found isolation by distance to be an important factor across large spatial scales in Gopher Tortoises based on analysis of DNA microsatellites. F_{ST} values from both our study and previous studies (Gaillard et al.

2017) demonstrate low to moderate levels of genetic differentiation across > 100 km distances, a pattern that has also been documented for Desert Tortoises (*Gopherus agassizii*) (Shaffer et al. 2017; Dutcher et al. 2020).

Rivers often act as barriers to dispersal for terrestrial species and drive population genetic structuring for species in southeastern North America (Soltis et al. 2006). Two previous studies suggested that river barriers shaped historical population genetic structure for *G. polyphemus*, with the Alabama and Apalachicola drainages causing regional-scale genetic structuring of populations in the US Gulf Coastal Plain (Ennen et al. 2012; Gaillard et al. 2017). Our data refine assessment of river barriers between these two major drainages. Our strongest genetic differentiation occurs between the two sites west of the Perdido River (Stimpson + Perdido) and the other four sites, which are east of the Escambia/Conecuh River. This pattern corroborates a similar pattern across the Escambia/Conecuh River revealed by less-intensive genetic sampling reported in Gaillard et al. (2017). We also found genetic differentiation across the Yellow River, separating Conecuh + Solon Dixon from Geneva + Rayonier. While the magnitude of this genetic difference appears smaller than that across the Escambia/Conecuh River, these results were surprising. These results were surprising because the small drainage area of the Yellow River (1,300 km^2) appears to restrict gene flow in a way that the much larger Alabama River (114,000 km^2), which separates Stimpson from all other sites, does not.

If habitat fragmentation renders areas unpassable by individuals in contemporary local populations of *G. polyphemus*, this would interrupt historic metapopulation structure of local populations within the genetic populations created by river barriers. We expected to find strong evidence of such contemporary fragmentation within our samples. However, with a generation time of approximately 60 years (Folt et al. 2021), it may take many generations and thus hundreds of years for genetic differentiation to occur (Dutcher et al. 2023). We observed no strong variation in genetic diversity or effective population size among individuals from different sampling locations or among populations identified by population structure analyses. This was true in larger, well-managed properties, like Conecuh National Forest, and in isolated local populations with smaller census sizes embedded in less- managed landscapes, like Perdido. In fact, we observed an overall average H_O of ~ 0.10 across all sites. This study-wide heterozygosity estimate is lower than other studies using microsatellites or highly variable SNP datasets that estimated H_O of 0.2–0.74 (Sinclair et al. 2010, Richter et al. 2011, Elbers et al. 2017, Gaillard et al. 2017). However, those studies used loci purposely chosen to be hypervariable. Given that RADseq uses restriction enzymes to shear and sequence the genome on average once every 4096 base pairs (1/46 for a 6 bp restriction enzyme), it is unsurprising that our more random sample of the genome shows less variation than sites purposely chosen to be hypervariable. Given our larger (> 1000 sites) and more random sampling of the genome, we suspect that the lower genetic diversity we observed may more closely reflect genome-wide diversity. Thus, while it is unclear whether the observed genetic diversity is low or high for *G. polyphemus* given our unique data, the genetic diversity metrics were consistent among locations. Low variance in genetic diversity among sampling locations and populations may suggest that genetic diversity simply changes slowly in species with long generation times (Dutcher et al. 2023), or that *G. polyphemus* are experiencing similar genetic demographic conditions across the different study areas in southern Alabama.

After reducing the stronger effects of isolation by distance and population structure, we did find some evidence of genetic differentiation that could be explained by contemporary forces. Our study took advantage of dense sampling of individuals at six local populations within Conecuh National Forest, a site that has been managed to restore habitat quality and connectivity through longleaf restoration (Pudner et al. 2021). At the extremes of sPC2, sPCA within Conecuh showed some genetic differentiation between Sites 3 + 4 and Sites 2 + 6. Additionally, one individual caught at Site 5 clustered closest to Sites 3 + 4. Tortoises in Sites 3 + 4 and Sites 2 + 6 were separated by a minimum distance of 13 km and Hwy 137, a major artery for vehicular traffic that is a likely contemporary dispersal barrier based on observed tortoise fatalities (B.F., J.G., C.G., pers. observ.). Within the pairs, Sites 2 and 6 were separated by 2 km and a sparsely traveled paved road while Sites 3 and 4 were separated by 4 km and no obvious contemporary barrier. One possible explanation for this differentiation across sPC2 is contemporary road barriers and distance. Moreover, an individual captured at Site 5, but clustering genetically with Sites 3 + 4 could indicate migration from Sites 3 + 4 to Site 5, approximately 6 km distance. Given the distances between genetically distinct sites, locales in Conecuh may experience homogenizing gene flow across distances up to 6 km in managed landscapes lacking suspected contemporary barriers. Despite no documented migration among sites in over 30 years of mark recapture studies (Goessling et al. 2021; Folt et al. 2021), our incidental captures of tortoises while driving between sites may also be evidence of contemporary gene flow. Alternatively, among other explanations, this differentiation within Conecuh could be explained by historical connections among populations, or by other unsampled local populations that would make these breaks appear more like isolation by distance even above 6 km. For example, linear local populations along roadsides (Rautsaw et al. 2018) may enhance gene flow across managed landscapes.

We also found that genetic diversity did not vary consistently between sites differing in demographic trajectories.

Based on demographic models, Folt et al. (2021) found that tortoises from Sites 1, 3 and 4 have survival rates that imply stable populations, while Sites 2, 5, and 6 have demographic rates consistent with declining populations and considerable future extinction risk. Despite different apparent trajectories of population size in recent decades, estimates of genetic diversity did not vary strongly between these two groups of sites. Gene flow implied by the sPCA analyses may explain why no signature of population bottlenecks was observed within the six small local populations, especially those with high extinction risk.

We found consistently small estimates of N_e among sampling locations and genetic populations identified by our analyses. Although N_e might be best suited to genetic populations identified in our study (Waples 2022), 30 years of field work shows that tortoises interact more at the scale of the sampling locations (Goessling et al. 2021; Guyer et al. in press). These sites in Conecuh represent the smallest aggregations that might represent local populations (Goessling et al. 2021). Estimates of N_e for sites within Conecuh are reasonably close to the census population size of adults at the sites estimated in a recent study (Folt et al. 2021), suggesting that the census population size observed during surveys may be similar to the effective population size. However, at larger spatial scales (e.g., for all of Conecuh National Forest), estimated N_e is much less than population size estimates for the entire area (Goessling et al. 2021). Various factors can cause different effective and census population sizes, including differing sex ratios, fluctuating population sizes, or differing numbers of offspring per individual in the population (White et al. 2018; Yuan et al. 2019; Waples 2022). Regardless of the mechanism, we infer that this discrepancy indicates that only a small proportion of individuals in Conecuh contribute to the evolutionary trajectory of the Conecuh + Solon Dixon population. Despite population estimates that suggest between 3,500 and 7,700 individuals exist across Conecuh National Forest (Goessling et al. 2021), our results suggest the population is experiencing the same evolutionary pressures as a population with only 89 individuals.

Our results imply that *G. polyphemus* once interacted across vast areas of southern Alabama. We found that the predominant force acting on *G. polyphemus* genetic structure is isolation by distance but that three genetic populations separated by river barriers also exist. Replicate local populations for each of these genetic populations are present on public lands managed, in part, for wildlife conservation. Plans for tortoise conservation might prioritize maintenance of these local populations to preserve patterns of genetic diversity generated over the evolutionary history of the species in the state, as these genetic patterns are likely conserved across the species distribution. We found no evidence of strong inbreeding, even in local populations of small size. At first glance, this is good news. However, we note that the long generation times of *G. polyphemus* (Folt et al. 2021), sperm-storage (Moon et al. 2006), and avoidance of kin-mating (Yuan et al. 2019) may mask the appearance of inbreeding depression for centuries (Soltis et al. 2006). Nevertheless, our data provide additional evidence that local populations are large enough to maintain historical patterns of gene structure. Within Conecuh National Forest, migration of individuals across distances of at least 6 km appears to assist relatively isolated tortoise groups to avoid genetic deterioration expected of small populations. Management of existing public lands might attempt to match that at Conecuh National Forest to assure connectivity of local populations that will allow similar levels of apparent migration.

Given the importance of this keystone species in the longleaf pine ecosystem, managers often translocate *G. polyphemus* from degraded or developed areas to higher quality habitat. Our results inform how translocations might occur without disrupting the unique historical genetic structure. Isolation by distance is a strong statistical trend within Alabama and is likely important range wide for Gopher Tortoises. Thus, individuals closer to each other in space tend to be more closely related genetically, especially within the three identified genetic populations. Given the connectedness of these populations, and the genetic differences over large distances indicated by the isolation by distance, we recommend that managers (1) prioritize moving tortoises as short of a distance as possible from their original location, and (2) avoid mixing among populations. However, if it is necessary, tortoises can occasionally be moved among the three populations, as this likely occurred historically on rare occasions. We suggest that care should be taken to evaluate such translocations before they are made so that contemporary translocation mimics historical patterns of gene flow among genetic populations and does not genetically homogenize important patterns of population genetic structure.

The genetic structure that we observed across populations in southern Alabama likely reflects a history of high dispersal and population connectivity among a relatively contiguous longleaf pine landscape. As such, maintaining large swaths of suitable habitat will be useful to maintain the historic genetic population structure. In cases where that is not possible, management to increase migration and population connectivity among existing genetic populations, such as through the creation of habitat corridors, new management areas, and road passages, may be useful not only in Alabama, but also across the species range. While the days of contiguous longleaf pine savannas across the Southeastern Coastal Plains are gone for now, remnant patterns of that historic connectivity remain in the genomes of long-lived *G. polyphemus*.

Supplementary Information The online version contains supplementary material available at https://doi.org/10.1007/s10592-024-01601-1.

Acknowledgements Samples were collected under Alabama Scientific Collection Permits (number 2019128031068680) and Auburn University IACUC protocols (2013–2226; 2017-3102). We thank the Alabama Division of Wildlife and Freshwater Fisheries for funding this research. We thank Owen Edwards, Bill Hawthorne, Suzi Mersmann, Jorge Lopez-Perez, Jackson Roberts, James Stiles, Sierra Stiles, and Chava Weitzman for helping collect samples in the field and Louisa Collins for library preparations and sequencing. We thank Isaac Overcast for advice with data analysis.

Author contributions All authors contributed to the study design. BF, CG, and JMG collected samples in the field. JJA coordinated field samples and sequencing. ARK designed and ran the analysis. ARK and BF wrote the manuscript. All authors provided comments and feedback on all drafts of the manuscript.

Funding This research was supported by funds from the U.S. Fish and Wildlife Service that were administered by subcontract from the Alabama Division of Wildlife and Freshwater Fisheries (FAIN #F21AP03947).

Data availability Raw sequence data is archived at the Short Read Archive (SRA Project Number PRJNA947629), available 1 September 2023, or upon publication, whichever happens first. Scripts, and metadata, used to recreate the analysis are available on GitHub (https://github.com/alexkrohn/AL_GopherToroise_PopGen).

Declarations

Competing interest The authors declare no competing interests.

References

Aresco MJ, Guyer C (1999) Burrow abandonment by gopher tortoises inslash pine plantations of the Conecuh National Forest. J Wildlife Manag 63:26. https://doi.org/10.2307/3802484

Auffenberg W, Franz R (1982) The status and distribution of the gopher tortoise (*Gopherus polyphemus*). In: Bury RB (ed) North American tortoises: conservation and ecology. United States Fish and Wildlife Service Wildlife Research Report, Washington, DC, pp 95–126

Bayona-Vásquez NJ, Glenn TC, Kieran TJ et al (2019) Adapterama III: quadruple-indexed, double/triple-enzyme RADseq libraries (2RAD/3RAD). PeerJ 7:e7724. https://doi.org/10.7717/peerj.7724

Bradburd GS, Coop GM, Ralph PL (2018) Inferring continuous and discrete population genetic structure across space. Genetics 210:33–52. https://doi.org/10.1534/genetics.118.301333

Catano CP, Stout IJ (2015) Functional relationships reveal keystone effects of the gopher tortoise on vertebrate diversity in a longleaf pine savanna. Biodivers Conserv 24:1957–1974. https://doi.org/10.1007/s10531-015-0920-x

Clostio RW, Martinez AM, LeBlanc KE, Anthony NM (2012) Population genetic structure of a threatened tortoise across the southeastern United States: implications for conservation management: population genetic structure of threatened tortoise. Anim Conserv 15:613–625. https://doi.org/10.1111/j.1469-1795.2012.00557.x

Danecek P, Auton A, Abecasis G et al (2011) The variant call format and VCFtools. Bioinformatics 27:2156–2158. https://doi.org/10.1093/bioinformatics/btr330

Delaney KS, Riley SPD, Fisher RN (2010) A rapid, strong, and convergent genetic response to urban habitat fragmentation in four divergent and widespread vertebrates. PLoS One 5:e12767. https://doi.org/10.1371/journal.pone.0012767

Do C, Waples RS, Peel D et al (2014) NeEstimator v2: re-implementation of software for the estimation of contemporary effective population size (*ne*) from genetic data. Mol Ecol Resour 14:209–214. https://doi.org/10.1111/1755-0998.12157

Dutcher KE, Vandergast AG, Esque TC et al (2020) Genes in space: what Mojave desert tortoise genetics can tell us about landscape connectivity. Conserv Genet 21:289–303. https://doi.org/10.1007/s10592-020-01251-z

Dutcher KE, Nussear KE, Heaton JS et al (2023) Move it or lose it: predicted effects of culverts and population density on Mojave desert tortoise (*Gopherus agassizii*) connectivity. PLoS One 18:e0286820. https://doi.org/10.1371/journal.pone.0286820

Eaton DAR, Overcast I (2020) Ipyrad: interactive assembly and analysis of RADseq datasets. Bioinformatics 36:2592–2594. https://doi.org/10.1093/bioinformatics/btz966

Ennen JR, Kreiser BR, Qualls CP (2010) Low genetic diversity in several gopher tortoise (*Gopherus polyphemus*) populations in the Desoto National Forest, Mississippi. Herpetologica 66:31–38. https://doi.org/10.1655/08-083.1

Ennen JR, Kreiser BR, Qualls CP et al (2012) Mitochondrial DNA assessment of the phylogeography of the Gopher Tortoise. J Fish Wildlife Manag 3:110–122. https://doi.org/10.3996/102011-JFWM-063

Epps CW, Keyghobadi N (2015) Landscape genetics in a changing world: disentangling historical and contemporary influences and inferring change. Mol Ecol 24:6021–6040. https://doi.org/10.1111/mec.13454

Eubanks JO, Michener WK, Guyer C (2003) Patterns of movement and burrow use in a population of Gopher tortoises (*Gopherus polyphemus*). Herpetologica 59:311–321

Fischer J, Lindenmayer DB (2007) Landscape modification and habitat fragmentation: a synthesis. Glob Ecol Biogeogr 16:265–280. https://doi.org/10.1111/j.1466-8238.2007.00287.x

Fletcher RJ, Reichert BE, Holmes K (2018) The negative effects of habitat fragmentation operate at the scale of dispersal. Ecology 99:2176–2186. https://doi.org/10.1002/ecy.2467

Folt B, Bauder J, Spear S et al (2019) Taxonomic and conservation implications of population genetic admixture, mito-nuclear discordance, and male-biased dispersal of a large endangered snake, *Drymarchon couperi*. PLoS One 14:e0214439. https://doi.org/10.1371/journal.pone.0214439

Folt B, Goessling JM, Tucker A et al (2021) Contrasting patterns of demography and population viability among gopher tortoise populations in Alabama. J Wildlife Manag 85:617–630. https://doi.org/10.1002/jwmg.21996

Frantz AC, Cellina S, Krier A et al (2009) Using spatial bayesian methods to determine the genetic structure of a continuously distributed population: clusters or isolation by distance? J Appl Ecol 46:493–505. https://doi.org/10.1111/j.1365-2664.2008.01606.x

Gaillard D, Ennen JR, Kreiser BR et al (2017) Range-wide and regional patterns of population structure and genetic diversity in the Gopher Tortoise. J Fish Wildlife Manag 8:497–512. https://doi.org/10.3996/022017-JFWM-010

Gibbons JW, Semlitsch RD (1982) Survivorship and longevity of a long-lived vertebrate species: how long do turtles live? J Anim Ecol 51:523–527. https://doi.org/10.2307/3981

Glenn TC, Pierson TW, Bayona-Vásquez NJ et al (2019) Adapterama II: universal amplicon sequencing on Illumina platforms (TaggiMatrix). PeerJ 7:e7786. https://doi.org/10.7717/peerj.7786

Goessling JM, Guyer C, Godwin JC et al (2019) Upper respiratory tract disease and associated diagnostic tests of mycoplasmosis in Alabama populations of Gopher tortoises, *Gopherus polyphemus*.

PLoS One 14:e0214845. https://doi.org/10.1371/journal.pone.0214845
Goessling JM, Stober JM, Gyengo SG et al (2021) Implications from monitoring Gopher tortoises at two spatial scales. J Wildlife Manag 85:135–144. https://doi.org/10.1002/jwmg.21966
González AV, Gómez-Silva V, Ramírez MJ, Fontúrbel FE (2020) Meta-analysis of the differential effects of habitat fragmentation and degradation on plant genetic diversity. Conserv Biol 34:711–720. https://doi.org/10.1111/cobi.13422
Goudet J (2005) Hierfstat, a package for r to compute and test hierarchical F-statistics. Mol Ecol Notes 5:184–186. https://doi.org/10.1111/j.1471-8286.2004.00828.x
Guyer C, Bailey MA (1993) Amphibians and reptiles of longleaf pine communities. Proceedings of the Tall Timbers Fire Ecology Conference 18:139-158
Haddad NM, Brudvig LA, Clobert J et al (2015) Habitat fragmentation and its lasting impact on Earth's ecosystems. Sci Adv 1:e1500052. https://doi.org/10.1126/sciadv.1500052
Hartl DL, Clark AG (1997) Principles of population genetics. Sinauer Associates, Sunderland
Hedrick PW, Garcia-Dorado A (2016) Understanding inbreeding depression, purging, and genetic rescue. Trends Ecol Evol 31:940–952. https://doi.org/10.1016/j.tree.2016.09.005
Hodel RGJ, Chen S, Payton AC et al (2017) Adding loci improves phylogeographic resolution in red mangroves despite increased missing data: comparing microsatellites and RAD-Seq and investigating loci filtering. Sci Rep 7:17598. https://doi.org/10.1038/s41598-017-16810-7
Holderegger R, Di Giulio M (2010) The genetic effects of roads: a review of empirical evidence. Basic Appl Ecol 11:522–531. https://doi.org/10.1016/j.baae.2010.06.006
Howell HJ, Rothermel BB, White KN, Searcy CA (2020) Gopher Tortoise demographic responses to a novel disturbance regime. J Wildlife Man 84:56–65. https://doi.org/10.1002/jwmg.21774
Hunter EA, Rostal DC (2021) Fire management effects on long-term Gopher Tortoise population dynamics. J Wildlife Manag 85:654–664. https://doi.org/10.1002/jwmg.22033
Jackson, D.R., and E.G. Milstrey (1989) The fauna of Gopher Tortoise burrows. Pp. 86–98 In: Gopher tortoise relocation symposium proceedings. Diemer, J. (Ed.). Technical Report 5, Nongame wildlife program, Florida game and freshwater fish commission, Tallahassee.
Janes JK, Miller JM, Dupuis JR et al (2017) The $K=2$ conundrum. Mol Ecol 26:3594–3602. https://doi.org/10.1111/mec.14187
Jombart T (2008) Adegenet: a R package for the multivariate analysis of genetic markers. Bioinformatics 24:1403–1405. https://doi.org/10.1093/bioinformatics/btn129
Jombart T, Devillard S, Dufour A-B, Pontier D (2008) Revealing cryptic spatial patterns in genetic variability by a new multivariate method. Heredity 101:92–103. https://doi.org/10.1038/hdy.2008.34
Jose SE, Jokela J, Miller DF (2006) The Longleaf Pine Ecosystem: Ecology, Silviculture, and Restoration. Springer New York, NY. 438 pp.
Kahle D, Wichkam H (2013) ggmap: Spatial visualization with ggplot2. The R Journal 5(1):144-161
Keyghobadi N (2007) The genetic implications of habitat fragmentation for animals. Can J Zool 85:1049–1064. https://doi.org/10.1139/Z07-095
Landers JL, Van Lear DH, Boyer WD (1995) The longleaf pine forests of the Southeast: requiem or renaissance? J Forest 93:38–44. https://doi.org/10.1093/jof/93.11.38
Lawson DJ, van Dorp L, Falush D (2018) A tutorial on how not to over-interpret structure and admixture bar plots. Nat Commun 9:3258. https://doi.org/10.1038/s41467-018-05257-7
Lino A, Fonseca C, Rojas D et al (2019) A meta-analysis of the effects of habitat loss and fragmentation on genetic diversity in mammals. Mamm Biol 94:69–76. https://doi.org/10.1016/j.mambio.2018.09.006
Lischer HEL, Excoffier L (2012) PGDSpider: an automated data conversion tool for connecting population genetics and genomics programs. Bioinformatics 28:298–299. https://doi.org/10.1093/bioinformatics/btr642
McCartney-Melstad E, Gidiş M, Shaffer HB (2019) An empirical pipeline for choosing the optimal clustering threshold in RADseq studies. Mol Ecol Resour 19:1195–1204. https://doi.org/10.1111/1755-0998.13029
McCoy ED, Mushinsky HR, Lindzey J (2006) Declines of the gopher tortoise on protected lands. Biol Cons 128:120–127. https://doi.org/10.1016/j.biocon.2005.09.021
Moon JC, McCoy ED, Mushinsky HR, Karl SA (2006) Multiple paternity and breeding system in the Gopher Tortoise, *Gopherus polyphemus*. J Hered 97:150–157. https://doi.org/10.1093/jhered/esj017
Nei M (1987) Molecular evolutionary genetics. Columbia University Press
Nikolakis ZL, Orton RW, Crother BI (2021) Fine scale population structure within an eastern nearctic snake complex (*Pituophis melanoleucus*). Zool Scr 51(2):133–146
Noss RF, Platt WJ, Sorrie BA et al (2015) How global biodiversity hotspots may go unrecognized: lessons from the North American Coastal Plain. Divers Distrib 21:236–244. https://doi.org/10.1111/ddi.12278
Novembre J, Stephens M (2008) Interpreting principal component analyses of spatial population genetic variation. Nat Genet 40:646–649. https://doi.org/10.1038/ng.139
Oksanen JG, Simpson G, Blanchet FG et al (2022) Vegan: community ecology package. R package version 2.6-2
Osentoski MF, Lamb T (1995) Intraspecific phylogeography of the gopher tortoise, *Gopherus polyphemus*: RFLP analysis of amplified mtDNA segments. Mol Ecol 4:709–718. https://doi.org/10.1111/j.1365-294X.1995.tb00271.x
Pereira HM, Leadley PW, Proença V et al (2010) Scenarios for global biodiversity in the 21st century. Science 330:1496–1501. https://doi.org/10.1126/science.1196624
Pudner RC, Waddle H, Mersmann SP et al (2021) Changes in vegetation structure and Gopher Tortoise population structure after 17 years of restoration management. Nat Areas J 41:273–282. https://doi.org/10.3375/21-3
R Core Team (2021) R Foundation for statistical computing, Vienna.
Raj A, Stephens M, Pritchard JK (2014) Faststructure: variational inference of population structure in large SNP data sets. Genetics 197:573–589. https://doi.org/10.1534/genetics.114.164350
Rautsaw RM, Martin SA, Lanctot K, Vincent BA, Bolt MR, Seigel RA, Parkinson CL (2018) On the road again: assessing the use of roadsides as wildlife corridors for gopher tortoises (*Gopherus polyphemus*). J Herpetol 52:136–144
Richardson JC, Stiling P (2019) Gopher Tortoise herbivory increases plant species richness and diversity. Plant Ecol 220:383–391
Richter SC, Jackson JA, Hinderliter M et al (2011) Conservation genetics of the largest cluster of federally threatened Gopher Tortoise (*Gopherus polyphemus*) colonies with implications for species management. Herpetologica 67:406–419. https://doi.org/10.1655/HERPETOLOGICA-D-10-00044.1
Rohland N, Reich D (2012) Cost-effective, high-throughput DNA sequencing libraries for multiplexed target capture. Genome Res 22:939–946. https://doi.org/10.1101/gr.128124.111
Rousset F (1997) Genetic Differentiation and Estimation of Gene Flow from F-Statistics Under Isolation by Distance. Genetics 145:1219–1228. https://doi.org/10.1093/genetics/145.4.1219

Schlaepfer DR, Braschler B, Rusterholz H-P, Baur B (2018) Genetic effects of anthropogenic habitat fragmentation on remnant animal and plant populations: a meta-analysis. Ecosphere 9:e02488. https://doi.org/10.1002/ecs2.2488

Schwartz TS, Karl SA (2006) Population and conservation genetics of the gopher tortoise (*Gopherus polyphemus*). Conserv Genet 6:917–928. https://doi.org/10.1007/s10592-005-9078-5

Shaffer HB, McCartney-Melstad E, Ralph PL et al (2017) Desert tortoises in the genomic age: population genetics and the landscape. bioRxiv. https://doi.org/10.1101/195743

Sinclair CS, Dawes PJ, Seigel RA (2010) Genetic structuring of Gopher Tortoise (Gopherus polyphemus) populations on the Kennedy Space Center, Florida, USA. Herp Cons Bio 5:189–195

Soltis DE, Morris AB, McLachlan JS et al (2006) Comparative phylogeography of unglaciated eastern North America. Mol Ecol 15:4261–4293. https://doi.org/10.1111/j.1365-294X.2006.03061.x

Templeton AR, Shaw K, Routman E, Davis SK (1990) The genetic consequences of habitat fragmentation. Ann Missouri Botanical Gard 77:13. https://doi.org/10.2307/2399621

Tuberville TD, Norton TM, Todd BD, Spratt JS (2008) Long-term apparent survival of translocated gopher tortoises: a comparison of newly released and previously established animals. Biol Conserv 141:2690–2697. https://doi.org/10.1016/j.biocon.2008.08.004

Vekemans X, Hardy OJ (2004) New insights from fine-scale spatial genetic structure analyses in plant populations. Mol Ecol 13:921–935. https://doi.org/10.1046/j.1365-294X.2004.02076.x

Waples RS (2022) What is Ne, anyway? J Hered 113:371–379. https://doi.org/10.1093/jhered/esac023

Waples RS, Do C (2010) Linkage disequilibrium estimates of contemporary N_e using highly variable genetic markers: a largely untapped resource for applied conservation and evolution: precision and bias of LD estimates. Evol Appl 3:244–262. https://doi.org/10.1111/j.1752-4571.2009.00104.x

Weir BS, Goudet J (2017) A unified characterization of population structure and relatedness. Genetics 206:2085–2103. https://doi.org/10.1534/genetics.116.198424

White KN, Rothermel BB, Zamudio KR, Tuberville TD (2018) Male body size predicts reproductive success but not within-clutch paternity patterns in gopher tortoises (*Gopherus polyphemus*). J Hered 109(7):791–801. https://doi.org/10.1093/jhered/esy036

Wilson MC, Chen X-Y, Corlett RT et al (2016) Habitat fragmentation and biodiversity conservation: key findings and future challenges. Landsc Ecol 31:219–227. https://doi.org/10.1007/s10980-015-0312-3

Winters C, Ross J, Allman P (2017) Population genetics between an insular and coastal population of Gopher tortoises (*Gopherus polyphemus*) in Southwest Florida. Southeast Nat 16:369. https://doi.org/10.1656/058.016.0305

Wright S (1943) Isolation by distance. Genetics 28:114–138. https://doi.org/10.1093/genetics/28.2.114

Yuan ML, White KN, Rothermel BB et al (2019) Close-kin mating, but not inbred parents, reduces hatching rates and offspring quality in a threatened tortoise. J Evol Biol 32:1152–1162. https://doi.org/10.1111/jeb.13518

Zheng X, Levine D, Shen J et al (2012) A high-performance computing toolset for relatedness and principal component analysis of SNP data. Bioinformatics 28:3326–3328. https://doi.org/10.1093/bioinformatics/bts606