Conservation Genetics (2021) 22:811–821
https://doi.org/10.1007/s10592-021-01364-z

RESEARCH ARTICLE

Geographic patterns of genomic variation in the threatened Salado salamander, *Eurycea chisholmensis*

Population genetics of the Salado salamander

Chris C. Nice[1] · **James A. Fordyce**[2] · **V. Alex Sotola**[3] · **Justin Crow**[4] · **Peter H. Diaz**[4]

Received: 31 August 2020 / Accepted: 13 April 2021 / Published online: 24 April 2021
© The Author(s), under exclusive licence to Springer Nature B.V. 2021

Abstract
Aquatic, karst and spring endemic organisms have become a focus of conservation efforts as human population densities and demand for groundwater increase. This is especially true of Texas salamanders in the genus *Eurycea* that have been the subject of investigations of patterns of genetic differentiation in order to understand their systematics and to inform conservation planning. Here we generated data from several thousand single nucleotide polymorphisms (SNPs) to quantify within- and among-population genetic variation in the northernmost species, *Eurycea chisholmensis*, the Salado salamander, which is listed as a federally threatened species. We used approximate Bayesian computation and a method based on linkage disequilibrium to estimate effective population size, N_e. Levels of differentiation were low, but revealed a primary division between northern and southern populations with no evidence of gene exchange between them. Genetic diversity was similar across all sampling locations and estimates of N_e were largely congruent across the two methods and indicate population sizes large enough to maintain genetic variation, at least over the short term. These results suggest that two management units comprise the range of *E. chisholmensis* but that further sampling in intervening areas is required to precisely delineate, and determine the nature of, the boundary of these units.

Keywords Genomic differentiation · Gene flow · Population genomics · Approximate Bayesian computation · *Eurycea* · Plethodontidae

Introduction

Quantifying geographic patterns of genetic variation is a necessary and critical first step in investigations of evolutionary processes. Understanding these patterns also constitutes a foundation for management decisions and conservation planning (Mills 2012). Local, state, and federal governing authorities rely on the best available science to determine recovery criteria and regulatory actions needed to recover or protect threatened or endangered species. Patterns of standing genetic variation and differentiation can be used to delineate units for conservation (e.g. populations, population segments, or groups of populations) (Moritz 1994; Vogler and Desalle 1994; Fraser and Bernatchez 2001) and inform captive assurance and captive breeding programs, a key recovery effort implemented for certain species. For example, the International Union for Conservation of Nature and Natural Resources listed captive programs among the top response priorities relevant to amphibian conservation globally (Gascon et al. 2007). These colonies do not replace other recovery efforts but are set up to prevent catastrophic loss of species while conservation actions take place. Maintaining the genetic variation reflective of the wild population

✉ Chris C. Nice
ccnice@txstate.edu

[1] Department of Biology, Population and Conservation Biology Program, Texas State University, San Marcos, TX 78666, USA

[2] Department of Ecology and Evolutionary Biology, University of Tennessee, Knoxville, TN 37996, USA

[3] Department of Genetics, University of Georgia, Athens, GA 30602, USA

[4] United States Fish and Wildlife Service, Texas Fish and Wildlife Conservation Office, San Marcos, TX 78666, USA

is fundamental to the success of these captive programs over time so that a genetically representative and viable population might be released (Frankham 2010). Thus, an understanding of the distribution and organization of genetic variation that can be provided by molecular genetics data is critical for conservation planning.

Molecular genetics methods also provide estimates of effective population size (N_e) that are used to parameterize evolutionary models, or, in combination with age or stage class data, can be used for indirect estimates of population size. Estimates of N_e can be especially informative for cryptic organisms or species for which direct counts are difficult. Here we investigate patterns of population genomic variation in samples of the federally threatened Salado salamander, *Eurycea chisholmensis* (Chippindale et al. 2000) to provide critical information for management decisions.

Salamanders of the subgenus *Paedomolge* (genus *Eurycea* (Plethodontidae) in Texas are aquatic, neotenic (paedomorphic) salamanders that commonly have restricted ranges, particularly in the Edwards Plateau region of central Texas (Hillis et al. 2001). This region contains one of the most diverse groundwater systems in the world that is home to a large number of locally endemic plant and animal species, a great many of which are a focus of conservation efforts (Longley 1981; Hutchins 2018) due to climate variation and increasing water withdrawls for municipal and commercial use (Loáiciga et al. 2000; National 2015). Several of the approximately 14 species of *Paedomolge* are listed as threatened or endangered by the State of Texas and the US Fish and Wildlife Service (Devitt et al. 2019). Chippindale et al. (2000) described *E. chisholmensis* based on morphology and patterns of allozymic and mitochondrial genetic variation. This species was identified as part of a northern group of *Paedomolge* that occupy spring and cave sites in the northern segment of the Edwards Plateau. In the same publication, Chippindale et al. (2000) described the closely related *E. naufragia*, called the Georgetown salamander. The ranges of these two species were, at the time of their description, not well understood (Chippindale et al. 2000), with *E. chisholmensis* known only from near the city of Salado in Bell County, Texas, and *E. naufragia* described from Williamson County, Texas to the south of Salado Springs. Recent investigations using genotyping-by-sequencing (GBS) methods to generate population genomics data for *Eurycea* across the Edwards Plateau have identified the North Fork of the San Gabriel (including Lake Georgetown) as the boundary between *E. naufragia* (to the south) and *E. chisholmensis* (to the north) (Devitt et al. 2019), although the two species are closely related (see Figs. 2, 5, S2, S11 and S12A in Devitt et al. (2019)).

Similar to the other species of *Paedomolge*, *E. chisholmensis* is associated with karst formations, including spring openings and caves. As a consequence of their relatively low dispersal capacity and habitat limitations (Pierce et al. 2014; Bendik 2017; Gutierrez et al. 2018) they are expected to exhibit differentiation over short distances or among sites associated with distinct underground flow patterns (spring sources) (Lucas et al. 2009). Further, *E. chisholmensis* is known to be "elusive" (Chippindale et al. 2000) and difficult to enumerate. In the few places where salamanders can readily be observed at the surface, density estimates are lower than 200 individuals (Pierce et al. 2014; Diaz et al. 2015; Diaz and Warren 2019). Given their elusive nature, threatened status, concerns about potential impacts of changes in water availability in the Edwards and Trinity Aquifers, and the potential for local impacts on habitat quality, we undertook a population genomics investigation of *E. chisholmensis* to explore fine scale population structure and patterns of connectivity among localities. This investigation of patterns of population genetic variation was designed to provide a foundation for future management and conservation plans for *E. chisholmensis*. We combined relatively large sample sizes per locality and fine-scale sampling with a GBS approach to quantify geographic patterns of genetic variation and estimate effective population sizes for these salamanders.

Materials and methods

DNA sequencing and data collection

Tail clips were collected during the course of repeated surveys from each of seven spring sites (Fig. 1) as a part of ecological monitoring activities in 2017 and 2018 (Diaz and Warren 2019). Samples from 180 salamanders were collected. After capture, a small portion from the distal region of the tail was removed using sterile equipment. Information on the sex of individuals was not obtainable because sexual characteristics were not visible (i.e. individuals were captured out of breeding season) or individuals were too small for sex to be determined. Tissue was preserved in 95% EtOH and DNA was extracted using the DNeasy Blood and Tissue Kit (Qiagen Inc., Alameda, CA, USA). Reduced representation genomic libraries were prepared following the methods of Parchman et al. (2012) and Gompert et al. (2014). In brief, genomic DNA was digested with the restriction enzymes EcoR1 and Mse1. Illumina adapters with unique 8-10bp individual multiplex identifier (MID) sequences were ligated to the resulting fragments. Fragments were amplified with two rounds of PCR using iProof high fidelity polymerase (BioRad, Inc.). PCR products were then pooled. Size selection of fragments between 300 and 450 bp was performed using a BluePippin (Sage Science Inc., Beverly, MA, USA) and the resulting fragments were sequenced on one lane of Illumina HiSeq 2500 (SR 1×100) at the University

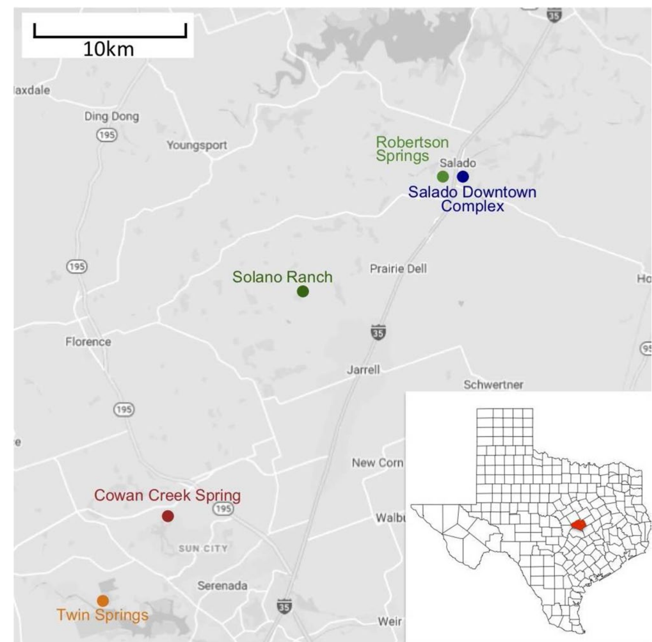

Fig. 1 Map of sampling localities. Larger circles indicate sampling localities, colors of circles match those in Fig. 2. The locality "Salado Downtown Complex" represents three spring sites in close proximity: Anderson Spring, Big Boiling Spring, and Side Spring

of Texas at Austin Genomic Sequencing and Analysis Facility (GSAF; Austin, Texas, USA).

We used BOWTIE version 1.1.2 (Langmead et al. 2009) to remove phiX reads. Custom scripts were employed to remove identifier sequences (MIDs) from each read and to filter short reads as well as reads that contained Mse1 adapter sequence. The resulting 223,337,965 sequence reads were written to individual files in FASTQ format. Five individuals produced less than 300,000 reads and were excluded from further analyses, including one individual each from Twin Springs, Solana Ranch Spring and Robertson Springs, and two individuals from Cowan Creek Spring. The reads from the remaining 175 individuals (Table 1) were filtered and assembled using the de novo strategy described in the dDocent variant calling pipeline (Puritz et al. 2014) and CD-hit (Fu et al. 2012). To streamline the assembly of sequence reads, we discarded sequence reads with less than four copies per individual and reads that were shared among fewer than four individuals. The resulting filtered reads were assembled with a homology threshold of 90% (other thresholds from 80 to 95% were investigated but produced very similar assemblies; data not presented). The consensus reads from this de novo assembly were used as the basis for a reference-based assembly of all reads using the Burrows-Wheeler Aligner (BWA version: 0.7.12) with up to four mismatches allowed (preliminary assemblies with from 2 to 6 mismatches allowed produced similar numbers of assembled reads; data not presented). Variable sites, also known as single nucleotide polymorphisms (SNPs), were identified using BCFTOOLS version 1.9 (Li et al. 2009) using the *mpileup* and *call* commands, ignoring indels and retaining variable sites if the posterior probability that the nucleotide was invariant was < 0.05. Custom scripts were used for further filtering to improve data quality that excluded variable sites with sequence depth less than 350 reads (an average of two reads per site per individual, 2X) and greater than 38,277 reads (equal to the mean sequence depth across sites plus two standard deviations; this reduces reads from potentially paralogous loci), less than at least 20 reads of the alternative allele, mapping quality less than 30, an absolute value of the mapping quality rank sum test greater than 2.5, an absolute value of the read position rank sum test greater than 2, absolute value of the base quality rank sum test greater than 3, minor allele frequency less than 0.05, or missing data for more than 35 individuals (20% of individuals). To minimize linkage disequilibrium (LD) among loci, one variable site per contig was chosen randomly and retained.

Clustering analysis and genotype estimation

The Bayesian clustering algorithm ENTROPY (Gompert et al. 2014) was used to estimate admixture proportions for each individual, allele frequencies for each cluster (population) and posterior genotype probabilities. ENTROPY is similar to the STRUCTURE algorithm (Pritchard et al. 2000) but uses pre-calculated maximum likelihood estimates of genotypes from BCFTOOLS as input. A series of ENTROPY models were fit to the data varying the number of *a priori* clusters or source populations (k) from two to seven. For each model run, two MCMC simulations of 330,000 steps were performed with a burnin of 30,000 steps while retaining values from every 10th step (total of 30,000 steps for each chain). To assess model performance and check that the models reached a stable sampling distribution,

we calculated Gelman and Rubin's convergence diagnostic and effective sample sizes (Gelman and Rubin 1992; Brooks and Gelman 1998) for each chain with the package CODA version 0.19-1 in R (Plummer et al. 2006; R Core Team 2020). We rejected model results when the mean Gelman and Rubin's convergence diagnostic across individual values of the admixture proportion, q, was greater than 1.12 (Gelman and Rubin 1992; Brooks and Gelman 1998), mean effective sample size was less than 500, or when no individuals were assigned to a cluster. Models for k = 2 and 3 satisfied these criteria, while all other models failed. Attempts to employ more MCMC steps, longer burnin, and different thinning strategies failed to rescue models for k = 4-7 (data not presented). Models for k = 2 and 3 exhibited the lowest consistent Deviance Information Criterion scores across chains (see Supplemental Figure S8). Patterns of genomic variation among individuals were illustrated by ordination using Principal Component Analysis (PCA) on the posterior genotype probabilities. Patterns of admixture were illustrated with barplots reflecting ancestry proportions. All analyses were performed in R (R Core Team 2020).

Relatedness

Because it is possible that individual salamanders might have been recaptured after tail re-growth over the course of the sampling and such repeat sampling of individuals would compromise estimation of within- and between-population genetic parameters, we employed two approaches to identify repeat samples. First, we compared all 175 individuals by calculating a pairwise genetic distance between individuals. Specifically, we multiplied the probability of each genotype being homozygous for the reference allele, heterozygous, or homozygous for the alternative allele, by 0, 1 or 2, respectively, and then calculated the difference between each pair of individuals at each locus and averaged this genetic distance over all loci (Alberici da Barbiano et al. 2013). These genotypic distances are expected to be zero for identical individuals (i.e. a recaptured individual). Second, we measured relatedness (Wang 2017) with two estimators using the RELATED package (Pew et al. 2015) in R (R Core Team 2020). We used the moment estimator of Lynch and Ritland (1999) and the likelihood estimator of Wang (2007) to calculate relatedness among individuals within the populations (clusters) identified using ENTROPY (see Results). Because the infile format for RELATED requires called genotypes, the posterior genotype probabilities estimated with ENTROPY for the full data set (8731 loci) were binned to convert them to three discrete genotypes for each locus. Posterior genotype probabilities of ≤ 0.66 were coded as 0101, ≥ 0.67 and ≤ 1.33 were coded as 0102, and ≥ 1.34 were coded as 0202.

Genetic diversity

To quantify differentiation among sites, we calculated genome-average Nei's G_{ST} (an analog of the standard measure of differentiation, F_{ST}, (Nei 1973)) for all pairwise combinations of sites. G_{ST} was calculated as the average across all loci and 9999 permutations were used to calculate 95% confidence intervals. All of the above analyses were performed with custom scripts in R and Perl. We then used BCFTOOLS version 0.1.9 to calculate two measures of genetic diversity for samples from each of the seven localities. We estimated Watterston's θ (based on the number of segregating sites) and Tajima's π (nucleotide diversity or heterozygosity) using the expectation-maximization algorithm with 20 iterations (Li 2011) which was sufficient for values to converge for all localities. These diversity estimators were calculated for each of the seven sampling localities.

Effective population size

We employed two strategies to estimate contemporary N_e for the populations (clusters) identified using ENTROPY (see Results) including: (1) a model-based approximate Bayesian computation (ABC) approach that relies on simulations (Beaumont et al. 2002; Beaumont 2008; Csilléry et al. 2010), and (2) the effective population size estimation procedure based on linkage disequilibrium (Weir 1979; Hill 1981; Waples 1991; Wang 2005; Waples 2006; Hare et al. 2011; Waples et al. 2016). For the ABC approach, we used the DIYABC software (Cornuet et al. 2014) to calculate N_e for each major genetic group (i.e. northern group, and the two southern localities, see Results). DIYABC uses approximate Bayesian computation methods to simulate data under a set of evolutionary scenarios. Here we used a simple scenario of simultaneous divergence for the three groups. To minimize computational time requirements, we reduced the data set in terms of both number of loci and number of individuals using the following procedure: First, the posterior genotype probabilities estimated with ENTROPY for the full data set (8731 loci) were again binned to convert them to three discrete genotypes coded as 0, 1 and 2 for the number of alternative alleles in each genotype. Posterior genotype probabilities of ≤ 0.66 were converted to 0, ≥ 0.67 and ≤ 1.33 were rounded to 1, and ≥ 1.34 were rounded to 2. After calling genotypes in this fashion, a symmetric Procrustes analysis was performed to compare the matrix of PCA scores for called genotypes with the matrix of PCA scores for the posterior genotype probabilities using the *protest* function of the VEGAN v2.5-6 package in R (Oksanen et al. 2019). The two data sets were highly similar (Procrustes $r = 0.9982$) indicating that "called" genotypes effectively represent the observed genetic variation. Second, we determined the "minimum"

number of loci required to capture the patterns of geographic genetic variation. This is not common practice in GBS studies (but see Jahner et al. (2016)), however, reducing the data set to a minimum number of loci can be effective for reducing the computational burden in ABC analyses. To do this, we used the *sample* function in R to draw random replicate datasets consisting of 50, 100, 250, 500, 1000, 1250, and 1500 SNPs with 10 replicates each (i.e. 10 data sets of 50 loci, etc.). PCAs were conducted on each randomly drawn data set separately and the results were compared using symmetric Procrustes analyses (again with the *protest* function in VEGAN) to the full data set and the Procrustes statistic, *r*, was plotted versus the number of loci in each replicate data set (Fig. S11). From these analyses, we chose 1000 SNP loci as our minimum data set size which balanced information retention (mean of Procrustes $r = 0.9919$, SE = 0.00035) with data reduction. Next, to further reduce computational time, 20 individuals were randomly selected from each of the three genetic groups revealed by clustering analysis (the northern group of localities, Cowan Creek Spring and Twin Springs; see Results). An additional Procrustes analysis was conducted on the PCA of the final data set of 60 total individuals for 1000 randomly selected SNPs versus the total 8731 SNPs. The reduced data set was found to be highly similar to the full data (Procrustes $r = 0.9919$). Using the DIYABC software (Cornuet et al. 2014), uniform prior distributions for N_e from 5 to 20,000 were set for each of the three groups. Ten million simulations were run to generate data sets to calculate N_e for each genetic group, and the default minimum allele frequency criterion (Hudson's algorithm; Hudson (2002)) was used. We calculated the mean and variance of the gene diversity across all populations and loci (Nei 1987) to compute posterior estimates and distributions of parameters (N_e) based on the nearest 10,000 simulated data sets. Finally, after conducting the simulations, we performed a PCA comparing the prior and posterior distributions of the summary statistics to the observed data to ensure proper model fit.

In the second approach to estimating N_e, we employed the LD method. This estimator is based on the relationship between disequilibrium and N_e and includes corrections for sample size and bias associated with loci linked on chromosomes when a reference genome or other genomic tools are not available (Waples et al. 2016). We used VCFTOOLS ver. 0.1.13 (Danecek et al. 2011) to calculate the squared correlation of allele frequencies between loci, r^2, using 9999 permutations to assess confidence. We then subtracted the amount of disequilibrium expected due to sampling error to obtain an adjusted disequilibrium, $r^{2'}$, which was then used in the calculation of N_e, including corrections for sample size and for the number of chromosomes, which we assumed to be 14 for *E. chisholmensis* following Bogart (1967). Specifically, we employed the equations of Waples (2006) and Waples et al. (2016). (Specific details of these calculations can be found in the Supplementary Material.)

Results

16,854,095 reads were retained after filtering variants that comprised a SNP data set of 8,955 loci for 175 individuals (mean of 96,309 reads per individual). Mean sequence depth was 10.7 reads per locus per individual with an average of 7.9% missing data per locus. Preliminary analyses revealed 224 loci that exhibited LD that was approximately five times higher than all other loci (see Supplementary Material). This group of loci might be contained within a chromosomal inversion or be located on a sex chromosome, however, we lack a reference genome or information on the sex of individuals to test such hypotheses. These 224 loci were excluded and the analyses aimed at answering our main questions were performed on a final data set of 8731 loci.

Ordination of individual multilocus genotypes using PCA revealed two major groupings of individuals: a northern group including individuals from the Salado Springs complex (Anderson Spring, Big Boiling Spring, Side Spring) and Robertson Spring and Solana Ranch Springs, and a southern group consisting of Cowan Creek Spring and Twin Springs (Figs. 1 and 2a). The first PC axis, representing the north-south split explained 7.53% of the genotypic variance. The second PC axis captured differentiation within each of these groups and explained 1.02% of the variation, while the third PC axis showed some differentiation between the individuals from the southern group (Fig. 2b) and explained 0.97% of the variation. The fourth PC axis explained 0.90% of the variance and revealed differentiation within the northern group with the Solana Ranch Springs individuals forming a separate group, though the level of differentiation is quite low (Fig. 2c). A three-dimensional plot of the first three PC axes (Fig. 2d) provides visual evidence that the northern and southern groups represent the most substantial partitioning of variance compared to minor differentiation within each of these two groups. We found no evidence of repeated sampling of individuals or close relatives (Figs. S1–S7) and all individuals were retained for analyses.

The division between northern and southern sites was also observed in estimates of admixture proportions. DIC scores (Fig. S8) and MCMC diagnostics indicated that models for k = 2 and 3 outperformed other clustering solutions. Models with higher clustering (i.e. greater than k = 3) fit the data poorly and diagnoses of these models indicated that chain convergence was not achieved. For the model of k = 2, the northern and southern groups are distinct with no evidence of gene exchange between them (Fig. 3). The absence of individuals with intermediate admixture proportions

indicates little evidence of contemporary gene flow between the groups. The model for k = 3 showed the separation of Cowan Creek Springs from Twin Springs within the southern group (Fig. 3). One individual collected at Twin Springs was assigned to the Cowan Creek Spring cluster with 100% ancestry.

Patterns of differentiation estimated as Nei's G_{ST} conformed to the clustering analyses. G_{ST} values were generally low among all sites with a mean of 0.0121 (Table 2). The largest values were observed for comparisons between northern and southern localities. Mean pairwise G_{ST} within the northern group was very low, at 0.0078, and the pairwise G_{ST} between the two southern localities was lower still at 0.0042, despite the structuring recovered in the clustering analysis at k = 3 (Fig. 3). The average between-group (northern vs. southern) G_{ST} was moderate, at 0.0215. Genetic diversity within sites, measured as nucleotide diversity (expected heterozygosity), π, and Watterson's θ, were similar for all localities (Fig. 4, Table S1). No locality had particularly high or low diversity.

The two approaches to estimating N_e produced roughly similar results. ABC produced estimates of between approximately 3000 to 5000 for N_e for each of the three groups identified using ENTROPY at k = 3 (Table 3). Estimates of N_e based on linkage disequilibrium for each group were smaller than the ABC estimates (between

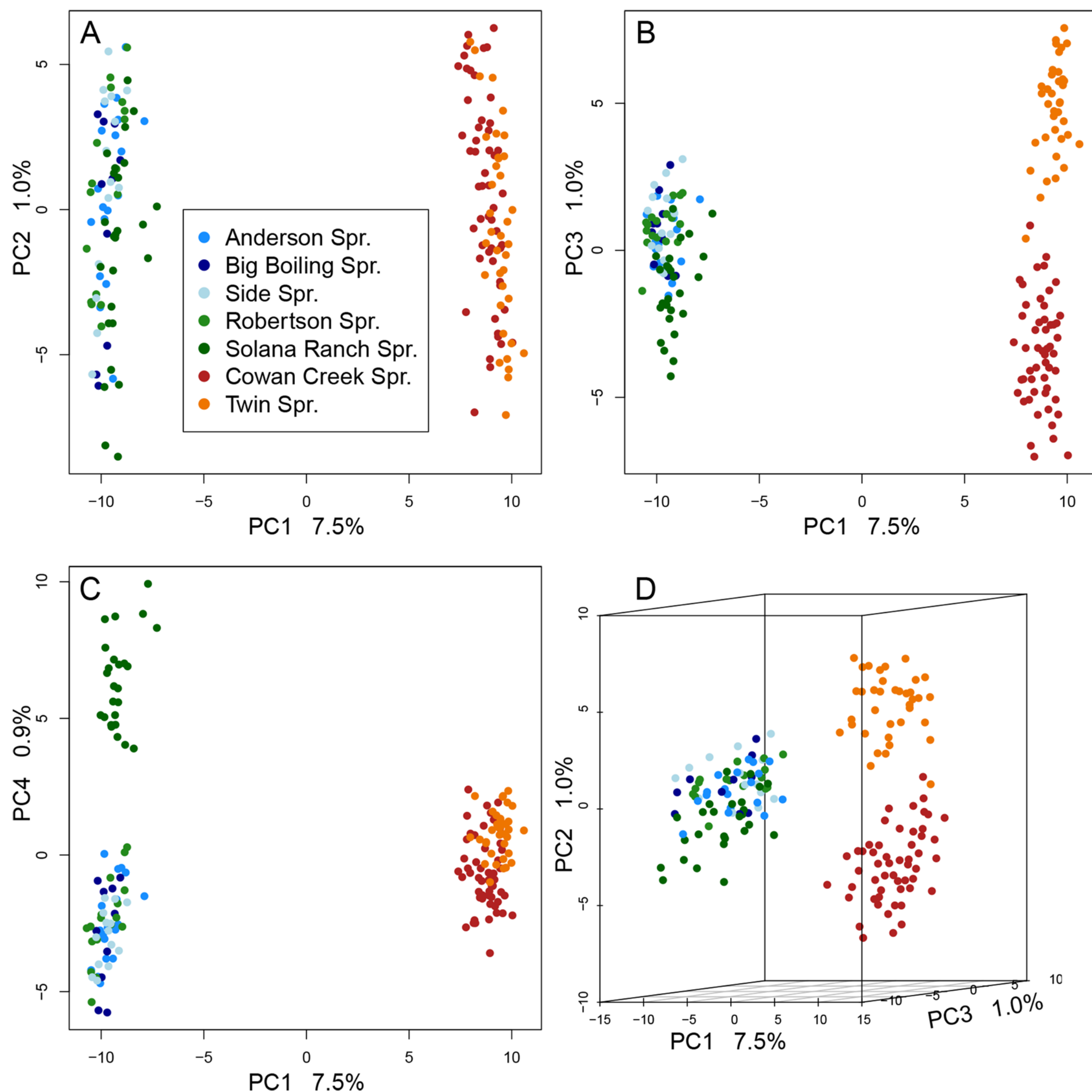

Fig. 2 Principal Components Analysis of posterior genotype estimates summarizing variation among 175 *Eurycea chisholmensis* individuals at 8731 SNP loci. **a** Ordination of PC 1 versus PC 2, **b** Ordination of PC 1 versus PC 3, **c** Ordination of PC 1 versus PC 4, **d** Three dimensional ordination of PCs 1,2 and 3. Each individual is represented by a point colored by locality

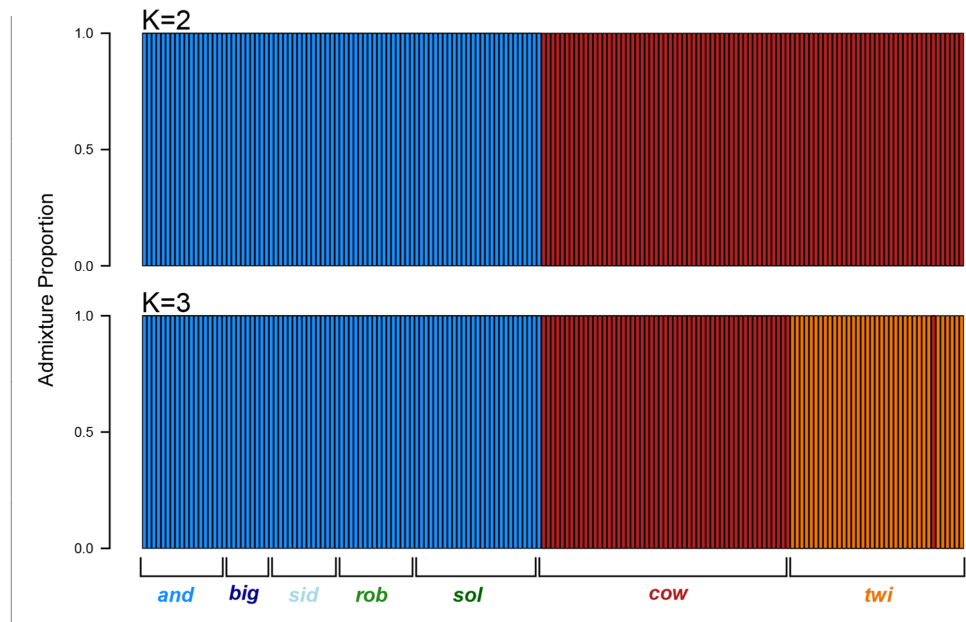

Fig. 3 Barplots of admixture proportions estimated with ENTROPY showing estimates for k = 2 and k = 3. Each bar represents an individual and the colors within bars show the proportion of ancestry from each cluster for each individual. Locality labels are: and = Anderson Spring, big = Big Boiling Spring, sid = Side Spring, rob = Robertson Springs, sol = Solano Ranch Spring, cow = Cowan Creek Spring, twi = Twin Springs

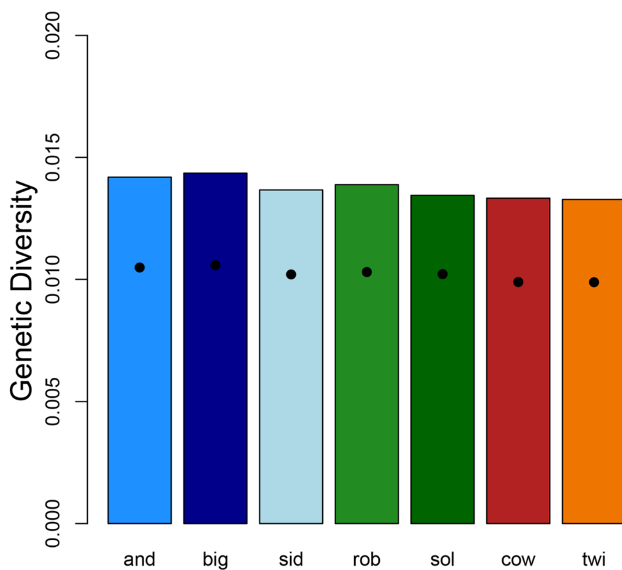

Fig. 4 Genomic diversity estimated from 8731 loci for 175 *Eurycea chisholmensis* from seven localities. Nucleotide diversity as Tajima's π (expected heterozygosity) is plotted as bars. Watterson's θ is indicated as points for each site. Locality labels are: and = Anderson Spring, big = Big Boiling Spring, sid = Side Spring, rob = Robertson Springs, sol = Solano Ranch Spring, cow = Cowan Creek Spring, twi = Twin Springs (Table 1)

approximately 1700 and 2500), but fell within the 95% credible intervals (equal-tailed probability interval) from the ABC analyses. It should be noted that the bounds of confidence intervals estimated from permutations of r^2 were indistinguishable from the observed values, which is an indication of the ability of very large numbers of pairwise comparisons (i.e. pairwise comparisons of 8731 loci) to generate precise estimates.

Discussion

Currently no federal recovery plan or species status assessment exists for *E. chisholmensis*. Since the listing of *E. chisholmensis* as threatened under the Endangered Species Act in 2014 (US 2014), the only accompanying federal regulatory actions have been the designation of critical habitat in 2020 (US 2020). The Endangered Species Act specifically requires the identification, protection, management, and recovery of species of plants and animals in danger of extinction. Fulfilling this responsibility requires the protection and conservation of not only individual organisms and populations, but also the genetic and ecological resources that listed species represent (US 2000). This study was undertaken with the main objective of providing detailed information on the distribution of genetic variation and estimates of N_e for *E. chisholmensis* to inform management plans.

A major feature of genetic variation in these salamanders is the division between the northern populations in the Salado Creek drainage (Salado Springs complex [Anderson Spring, Big Boiling Spring, Side Spring] and Robertson Spring and Solana Ranch Springs), and the southern populations (Cowan Creek Spring and Twin Springs) in the Berry Creek area (Fig. 2). This geographic division within *E. chisholmensis* was also observed in Devitt et al. (2019)'s larger survey of genomic variation in *Eurycea*. The exact location of this primary division will remain unclear until

Table 1 Sampling details

Abbreviation	Locality name	n
And	Anderson Spring (Salado Downtown Complex)	18
Big	Big Boiling Spring (Salado Downtown Complex)	10
Sid	Side Spring (Salado Downtown Complex)	14
Rob	Robertson Springs	17
Sol	Solana Ranch Spring	26
Cow	Cowan Creek Spring	53
Twi	Twin Springs	37
	Total	175

Locality names and samples sizes (n) for individuals genotyped for each locality

intervening localities for *E. chisholmensis* can be sampled. Within the northern group, there is little differentiation. Pairwise G_{ST} values are quite low among all pairs of northern populations (Table 2). In the individual analyses, there is evidence that the Solana Ranch Spring population is slightly differentiated from the other northern localities. In the PCA analysis of individual genotypes, PC axis 4 explains less than 1% of the genetic variance, but does differentiate Solana Ranch Spring from the other northern localities (Fig. 2c). Nevertheless, this subtle differentiation was not detected by the clustering algorithm and admixture proportions are homogenous across all northern individuals sampled (Fig. 3). These patterns (low pairwise G_{ST} values, low variance explained on PC axis 4, homogeneous admixture proportions across all northern sampling localities) suggest considerable contemporary, or relatively recent, gene exchange among all of these northern localities. This is somewhat surprising given the expectation of generally higher differentiation over shorter distances for these salamanders due to the nature of their habitats (springs and karst features) and limited dispersal (Lucas et al. 2009; Bendik 2017; Pierce et al. 2014; Gutierrez et al. 2018).

The southern localities form a separate group of populations that are moderately differentiated from the northern group. Average G_{ST} values between northern and southern localities were roughly two times higher than G_{ST} values within these two groups (Table 3). However, this between-group differentiation is still modest and approximately half the level of differentiation observed among populations of *Eurycea* salamanders from the Comal Springs complex in central Texas (Lucas et al. 2016). The two southern localities separate into distinct clusters at k = 3 (Fig. 3). A single individual collected at Twin Springs shows complete ancestry with Cowan Creek Spring individuals (Fig. 3). Dispersal over this distance (approximately 6 km) is well outside distances reported for *Eurycea* (Pierce et al. 2014; Gutierrez et al. 2018) and seems incredibly unlikely. It is, of course, possible that sample labels were mistakenly switched during processing, though we have no evidence for this. However, as with the northern group, there is no evidence of admixture with individuals showing 100% assignments to each cluster. The lack of admixture indicates that the putative "migrant" individual is a recent migrant and not the offspring of migrants. Despite this apparent individual migrant, there is no evidence of gene flow in the form of admixed individuals among the southern populations, suggesting that migration is highly unlikely between these two southern sites.

Genetic diversity is homogenously distributed across all localities (Fig. 4). Diversity in *E. chisholmensis* populations, measured as nucleotide diversity (expected heterozygosity), π, and Watterson's θ (number of segregating sites), are comparable to (Mandeville et al. 2015) or higher than (Sotola et al. 2019) levels of diversity found in some freshwater fishes, and higher than diversity observed in terrestrial insects (Gompert et al. 2014; Bell et al. 2017; Driscoe et al. 2019). Unfortunately, comparable measures of π and θ are not available for other *Eurycea* populations. Estimates of N_e based on both ABC and linkage disequilibrium (Table 3) suggest that populations include a minimum of a few thousand individuals. These numbers are higher than estimated for some other species of salamanders (e.g. *Ambystoma leorae*; (Sunny et al. 2014) and suggest population sizes are large enough to maintain their genetic variation. While these estimates of N_e provide some information

Table 2 Pairwise genome-average G_{ST} values

	And	Big	Sid	Rob	Sol	Cow	Twi
And	–	(0.0081–0.0087)	(0.0068–0.0073)	(0.0059–0.0063)	(0.0065–0.0074)	(0.0185–0.0215)	(0.0204–0.0234)
Big	0.0084	–	(0.0090–0.0097)	(0.0082–0.0088)	(0.0088–0.0099)	(0.0213–0.0242)	(0.0230–0.0260)
Sid	0.0071	0.0094	–	(0.0066–0.0070)	(0.0075–0.0085)	(0.0198–0.0226)	(0.0216–0.0245)
Rob	0.0061	0.0085	0.0068	–	(0.0066–0.0075)	(0.0189–0.0217)	(0.0206–0.0235)
Sol	0.0069	0.0093	0.0080	0.0070	–	(0.0177–0.0205)	(0.0195–0.0224)
Cow	0.0199	0.0227	0.0212	0.0204	0.0191	–	(0.0039–0.0046)
Twi	0.0218	0.0244	0.0229	0.0221	0.0209	0.0042	–

Averages are below the diagonal and 95% confidence intervals are below the diagonol. Population abbreviations follow those in Table 1

Table 3 Estimates of effective population size ($\widehat{N_e}$) using approximate Bayesian computation (ABC) implemented in DIYABC (Cornuet et al. 2014) and the LD method (Waples 2006; Waples et al. 2016)

Locality	ABC $\widehat{N_e}$ (95% CI)	LD $\widehat{N_e}$
Northern Group	3250 (518–10,000)	1785
Southern Group		
Cowan Creek Spring	4960 (770–12,400)	2442
Twin Springs	3130 (489–8950)	1877

Estimates are provided for the three populations (clusters) identified by ENTROPY (see Fig. 4). 95% credible intervals are included for the ABC estimates. Permutational 95% confidence intervals for the LD estimates were indistinguishable from the estimates and are not presented

about contemporary population sizes, some caution is warranted. Scaling N_e to the actual, or census, population size is difficult without more information on the age-structure in these populations (Frankham 1995; Kalinowski and Waples 2002; Mills 2012). As such, our estimates constitute what can be considered a minimum estimate of population size since the ratio of N_e to census population size is ≤ 1.

This study provides valuable information about the genetic structure and diversity of the threatened *E. chisholmensis*, but the picture for *E. chisholmensis* is not complete. Of particular interest is the location and nature of the boundary between the northern group of localities and southern populations. Potential populations in the intervening habitats were not sampled in this study, nor have they been extensively sampled in previous investigations. This phylogeographical boundary could be sharp and reflect watershed or drainage boundaries, or could exhibit a more gradual pattern of isolation by distance. While more sampling is needed here, the location of this boundary is close to the boundary of Williamson and Bell counties in Texas, meaning that the extent of coordination between the counties for effective management might be dependent on the nature of this phylogeographical boundary. Bell County currently has an active groundwater conservation district responsible for the quality and quantity of the aquifer water. In contrast, Williamson County has not developed a groundwater conservation district but has instead established a self sustaining karst foundation that is tasked with acquiring land containing state or federally listed species (either aquatic or terrestrial) and maintaining the land as a preserve. At this time Twin Springs is the one preserve in Williamson County that contains *E. chisholmensis*. Evidence of connectivity or patterns of isolation-by-distance between the northern and southern populations might require joint conservation planning between Williamson and Bell counties. Future resampling would also be especially useful for monitoring N_e using estimators that take advantage of the changes in allele frequencies between temporal samples (Jorde and Ryman 1995, 2007; Wang and Whitlock 2003). Such monitoring is likely to be increasingly important as climate variation, including droughts, and human population density, and attendant water withdrawls, continue to increase in this region of Texas. Furthermore, similar estimates of N_e and genetic diversity within- and among-populations from other *Eurycea* and other karst-endemic organisms would provide a comparative framework in which to design, implement, and evaluate conservation efforts for the biodiversity of the Edwards Plateau region.

Supplementary Information The online version contains supplementary material available at https://doi.org/10.1007/s10592-021-01364-z.

Acknowledgements We thank C. Mier Y Teran for assistance in the lab, and Ashley Wall and Ryan Jones for assistance in the field. Jim Ott and the Martin/Nice/Ott lab group provided valuable comments and discussion. Three anonymous reviewers and the Associate Editor provided valuable criticism. We also thank Shane Flaherty and the Texas State University LEAP High Performance Computing Cluster. This research was funded by the Clearwater Underground Water District. The views expressed in this paper are the authors' and do not necessarily reflect the view of the U.S. Fish and Wildlife Service.

Author contributions All authors contributed to the study conception and design, material preparation, data collection and analysis. The first draft of the manuscript was written by Chris C. Nice and all authors commented on previous versions of the manuscript. All authors read and approved the final manuscript.

Data availability Genotypic data are available on the Dryad Digital Repository at: https://datadryad.org/stash/share/lw_ECi_Rpm0hbwTPX91iUlKCFEgeTsQHbW1Zd7dyn54.

Code availability Custom scripts available from the corresponding author.

Declarations

Conflict of interest The authors declare that they have no conflict of interest.

Funding Clearwater Underground Water District

Conflicts of interest None

Ethical approval Include appropriate approvals or waivers.

Consent to participate: Not applicable.

Consent for publication Not applicable.

References

Alberici da Barbiano L, Gompert Z, Aspbury AS, Gabor CR, Nice CC (2013) Population genomics reveals a possible history of backcrossing and recombination in the gynogenetic fish *Poecilia formosa*. Proc Nat Acad Sci 110(34):13797–13802

Beaumont MA (2008) Joint determination of topology, divergence time and immigration in population trees. In: Matsumura S, Forster P, Renfrew C (eds) Simulations, genetics and human prehistory. McDonald Institute for Archaeological Research, Cambridge, pp 135–154

Beaumont MA, Zhang W, Balding DJ (2002) Approximate Bayesian computation in population genetics. Genetics 162(4):2025–2035

Bendik NF (2017) Demographics, reproduction, growth, and abundance of jollyville plateau salamanders (*Eurycea tonkawae*). Ecol Evol 7(13):5002–5015

Bell KL, Hamm CA, Shapiro AM, Nice CC (2017) Sympatric, temporally isolated populations of the pine white butterfly, *Neophasia menapia*, are morphologically and genetically differentiated. PLoS ONE 12(5):e0176989

Bogart JP (1967) Life history and chromosomes of some of the neotenic salamanders of the Edward's plateau. PhD thesis, University of Texas at Austin

Brooks SB, Gelman A (1998) General methods for monitoring convergence of iterative simulations. J Comput Graph Stat 7:434–455

Chippindale PT, Price AH, Wiens JJ, Hillis DM (2000) Phylogenetic relationships and systematic revision of central Texas Hemidactyliine Plethodontid salamanders. Herpetological Monographs, pp 1–80

Cornuet JM, Pudlo P, Veyssier J, Dehne-Garcia A, Gautier M, Leblois R, Marin JM, Estoup A (2014) Diyabc v2.0: a software to make approximate Bayesian computation inferences about population history using single nucleotide polymorphism, dna sequence and microsatellite data. Bioinformatics 30(8):1187–1189

Csilléry K, Blum MG, Gaggiotti OE, François O (2010) Approximate Bayesian computation (ABC) in practice. Trends Ecol Evol 25(7):410–418

Danecek P, Auton A, Abecasis G, Albers CA, Banks E, DePristo MA, Handsaker RE, Lunter G, Marth GT, Sherry ST et al (2011) The variant call format and VCFtools. Bioinformatics 27(15):2156–2158

Devitt TJ, Wright AM, Cannatella DC, Hillis DM (2019) Species delimitation in endangered groundwater salamanders: implications for aquifer management and biodiversity conservation. Proc Natl Acad Sci 116(7):2624–2633

Diaz P, Warren JB (2019) Salado salamander monitoring final report 2019. Texas Fish and Wildlife Conservation Office, US Fish and Wildlife Service

Diaz P, Montagne MA, Gibson RJ (2015) Salado salamander monitoring final report 2015. Texas Fish and Wildlife Conservation Office, US Fish and Wildlife Service

Driscoe AL, Nice CC, Busbee RW, Hood GR, Egan SP, Ott JR (2019) Host plant associations and geography interact to shape diversification in a specialist insect herbivore. Mol Ecol 28(18):4197–4211

Frankham R (1995) Conservation genetics. Ann Rev Genet 29(1):305–327

Frankham R (2010) Challenges and opportunities of genetic approaches to biological conservation. Biol Conserv 143(9):1919–1927

Fraser DJ, Bernatchez L (2001) Adaptive evolutionary conservation: towards a unified concept for defining conservation units. Mol Ecol 10(12):2741–2752

Fu L, Niu B, Zhu Z, Wu S, Li W (2012) Cd-hit: accelerated for clustering the next-generation sequencing data. Bioinformatics 28(23):3150–3152

Gascon C, Collins JP, Moore RD, Church DR, McKay J, Mendelson III JR (2007) Amphibian conservation action plan: proceedings IUCN/SSC Amphibian Conservation Summit 2005. IUCN

Gelman A, Rubin DB (1992) Inference from iterative simulation using multiple sequences. Stat Sci 7:457–511

Gompert Z, Lucas LK, Buerkle CA, Forister ML, Fordyce JA, Nice CC (2014) Admixture and the organization of genetic diversity in a butterfly species complex revealed through common and rare genetic variants. Mol Ecol 23(18):4555–4573. https://doi.org/10.1111/mec.12811

Gutierrez AM, Guess ST, Pierce BA (2018) Within-spring movement of the Georgetown salamander (*Eurycea naufragia*). Herpetol Conserv Biol 13(2):383–390

Hare MP, Nunney L, Schwartz MK, Ruzzante DE, Burford M, Waples RS, Ruegg K, Palstra F (2011) Understanding and estimating effective population size for practical application in marine species management. Conserv Biol 25(3):438–449

Hill WG (1981) Estimation of effective population size from data on linkage disequilibrium. Genet Res 38(3):209–216

Hillis DM, Chamberlain DA, Wilcox TP, Chippindale PT (2001) A new species of subterranean blind salamander (Plethodontidae: Hemidactyliini: *Eurycea*: *Typhlomolge*) from Austin, Texas, and a systematic revision of central Texas paedomorphic salamanders. Herpetologica pp 266–280

Hudson RR (2002) Generating samples under a Wright-Fisher neutral model of genetic variation. Bioinformatics 18(2):337–338

Hutchins BT (2018) The conservation status of Texas groundwater invertebrates. Biodivers Conserv 27(2):475–501

Jahner JP, Gibson D, Weitzman CL, Blomberg EJ, Sedinger JS, Parchman TL (2016) Fine-scale genetic structure among greater sage-grouse leks in central Nevada. BMC Evol Biol 16(1):1–13

Jorde PE, Ryman N (1995) Temporal allele frequency change and estimation of effective size in populations with overlapping generations. Genetics 139(2):1077–1090

Jorde PE, Ryman N (2007) Unbiased estimator for genetic drift and effective population size. Genetics 177(2):927–935

Kalinowski ST, Waples RS (2002) Relationship of effective to census size in fluctuating populations. Conserv Biol 16(1):129–136

Langmead B, Trapnell C, Pop M, Salzberg SL (2009) Ultrafast and memory-efficient alignment of short DNA sequences to the human genome. Genome Biol 10(3):R25

Li H (2011) A statistical framework for SNP calling, mutation discovery, association mapping and population genetical parameter estimation from sequencing data. Bioinformatics 27(21):2987–2993

Li H, Handsaker B, Wysoker A, Fennell T, Ruan J, Homer N, Marth G, Abecasis G, Durbin R et al (2009) The sequence alignment/map format and samtools. Bioinformatics 25(16):2078–2079

Loáiciga H, Maidment D, Valdes J (2000) Climate-change impacts in a regional karst aquifer, Texas, USA. J Hydrol 227(1–4):173–194

Longley G (1981) The Edwards Aquifer: Earth's most diverse groundwater ecosystem? Int J Speleol 11(1):12

Lucas L, Gompert Z, Ott J, Nice CC (2009) Geographic and genetic isolation in spring-associated *Eurycea* salamanders endemic to the Edward's Plateau region of Texas. Conservation Genetics pp 1309–1319

Lucas LK, Gompert Z, Gibson JR, Bell KL, Buerkle CA, Nice CC (2016) Pervasive gene flow across critical habitat for four narrowly endemic, sympatric taxa. Freshw Biol 61(6):933–946

Lynch M, Ritland K (1999) Estimation of pairwise relatedness with molecular markers. Genetics 152(4):1753–1766

Mandeville EG, Parchman TL, McDonald DB, Buerkle CA (2015) Highly variable reproductive isolation among pairs of Catostomus species. Mol Ecol 24(8):1856–1872

Mills LS (2012) Conservation of wildlife populations: demography, genetics, and management. Wiley, Chichester

Moritz C (1994) Defining ‘evolutionarily significant units’ for conservation. Trends Ecol Evol 9(10):373–375

National Research Council (2015) Review of the Edwards Aquifer habitat conservation plan: report 1. National Academies Press

Nei M (1973) Analysis of gene diversity in subdivided populations. Proc Natl Acad Sci 70(12):3321–3323

Nei M (1987) Molecular evolutionary genetics. Columbia University Press, New York

Oksanen J, Blanchet F, Freindly M, Kindt R, Legendre P, McGlinn D, Minchin P, O’Hara R, Simpson G, Solymos P, Stevens M, Szoecs E, Wagner H (2019) vegan: community ecology package version 2.5-6. https://cran.r-project.org/web/packages/vegan/index.html

Parchman TL, Gompert Z, Mudge J, Schilkey FD, Benkman CW, Buerkle CA (2012) Genome-wide association genetics of an adaptive trait in lodgepole pine. Mol Ecol 21(12):2991–3005

Pew J, Muir PH, Wang J, Frasier TR (2015) related: an R package for analysing pairwise relatedness from codominant molecular markers. Mol Ecol Resour 15(3):557–561

Pierce BA, McEntire KD, Wall AE (2014) Population size, movement, and reproduction of the Georgetown salamander. *Eurycea naufragia*. Herpetol Conserv Biol 9(1):137–145

Plummer M, Best N, Cowles K, Vines K (2006) CODA: convergence diagnosis and output analysis for MCMC. R News 5(1):7–11

Pritchard JK, Stephens M, Donnelly P (2000) Inference of population structure using multilocus genotype data. Genetics 155:945–959

Puritz JB, Hollenbeck CM, Gold JR (2014) dDocent: a RADseq, variant-calling pipeline designed for population genomics of non-model organisms. PeerJ 2:e431

R Development Core Team (2020) R: a language and environment for statistical computing. R Foundation for Statistical Computing, Vienna

Sotola VA, Ruppel DS, Bonner TH, Nice CC, Martin NH (2019) Asymmetric introgression between fishes in the Red River basin of Texas is associated with variation in water quality. Ecol Evol 9(4):2083–2095

Sunny A, Monroy-Vilchis O, Fajardo V, Aguilera-Reyes U (2014) Genetic diversity and structure of an endemic and critically endangered stream river salamander (Caudata: Ambystoma leorae) in Mexico. Conserv Genet 15(1):49–59

US Fish and Wildlife Service (2000) Policy regarding controlled propagation of species listed under the endangered species act. Federal Register 65(183):56916–56922

US Fish and Wildlife Service (2014) Endangered and threatened wildlife and plants; determination of threatened species status for the Georgetown salamander and Salado salamander throughout their ranges; final rule. Federal Register 79:10236–10293

US Fish and Wildlife Service (2020) Endangered and threatened wildlife and plants; designation of critical habitat for the Georgetown and Salado salamanders. Federal Register 85:57578–57613

Vogler AP, Desalle R (1994) Diagnosing units of conservation management. Conserv Biol 8(2):354–363

Wang J (2005) Estimation of effective population sizes from data on genetic markers. Phil Trans R Soc B 360(1459):1395–1409

Wang J (2007) Triadic IBD coefficients and applications to estimating pairwise relatedness. Genet Res 89(3):135–153

Wang J (2017) Estimating pairwise relatedness in a small sample of individuals. Heredity 119(5):302–313

Wang J, Whitlock MC (2003) Estimating effective population size and migration rates from genetic samples over space and time. Genetics 163(1):429–446

Waples RS (1991) Genetic methods for estimating the effective size of Cetacean populations. Report of the International Whaling Commission (Special Issue) 13:279–300

Waples RS (2006) A bias correction for estimates of effective population size based on linkage disequilibrium at unlinked gene loci. Conserv Genet 7(2):167

Waples RK, Larson WA, Waples RS (2016) Estimating contemporary effective population size in non-model species using linkage disequilibrium across thousands of loci. Heredity 117(4):233

Weir BS (1979) Inferences about linkage disequilibrium. Biometrics pp 235–254