Conservation Genetics
https://doi.org/10.1007/s10592-022-01457-3

RESEARCH ARTICLE

Islands in the desert: assessing fine scale population genomic variation of a group of imperiled desert fishes

D. Cooper Campbell[1,3] · **D. T. Camak**[2] · **K. R. Piller**[1]

Received: 18 August 2021 / Accepted: 29 May 2022

Abstract
The genus *Crenichthys* (Teleostei: Goodeidae) is an imperiled group of desert spring specialist fishes currently containing two species and five subspecies, found within only a few of the relictual springs distributed throughout the Great Basin of North America. Threatened by multiple forms of human disturbance, including habitat destruction, invasive species, and pollution, the need to better understand their population structure is immediate. This is further emphasized by previous research that demonstrated that the current taxonomy of *Crenichthys* needs re-evaluation and that genetic substructure may be present. The genus also represents a perfect opportunity to better understand desert spring habitats. These unique ecosystems often contain a suite of endemics, trapped within individual isolated springs distributed throughout a desert. The assumption is often that each spring will contain genetically distinct populations, however, this is not always true. We used single nucleotide polymorphisms (SNPs) to describe the genetic diversity and structure among populations of the genus *Crenichthys* with the intent to better understand the patterns of diversity within desert endemic fishes. Our results corroborated previous research suggesting genetic divergence between two groups within both *C. baileyi* and *C. nevadae*. It further demonstrated that many of the populations are genetically similar, likely due to a combination of short divergence time and possible past admixture.

Keywords Population genetics · Goodeidae · SNPs · Great Basin · Endangered species

Introduction

Desert spring habitats represent some of the most imperiled aquatic ecosystems in the world (Meffe and Vrijenhoek 1988). These aquatic "islands" often support a unique suite of rare and endemic species, which has resulted in desert spring habitats being referred to as evolutionary cradles (Murphy et al. 2015). The unique biodiversity that they harbor is often relictual in nature, representing a species formerly widespread during wetter periods. Geologic phenomena, such as stream capture, channel erosion, and volcanic uplift (Smith et al. 2002), as well as climate change and human modifications, have fragmented once interconnected, perennial freshwater systems into modern-day arid and isolated, island-like habitats (Axelrod 1979; Smith 1981). Conserving and managing the biodiversity within these desert springs has long been a challenge, as information on basic ecology, life history, and genetic diversity for many species is often lacking or is limited (Minckley and Deacon 1991; Minckley and Marsh 2016).

The Great Basin of North America is an example of an ecologically and evolutionarily unique region. It is a large, arid endorheic basin that has been subjected to substantial tectonic and climatic events that have shaped and modified its inclusive aquatic systems since the beginning of the subduction of the Farallon plate during the Tertiary period (McKee 1971). Because of this and the age of the Great Basin, there has been ample opportunity for allopatric

✉ D. Cooper Campbell
dcampbell@tulane.edu

D. T. Camak
dcamak@unm.edu

K. R. Piller
kyle.piller@selu.edu

[1] Dept. of Biological Sciences, Southeastern Louisiana University, 70402 Hammond, LA, United States

[2] Dept. of Biology, University of New Mexico, 87131 Albuquerque, NM, United States

[3] Dept. of Ecology and Evolutionary Biology, Tulane University, 70118 New Orleans, LA, United States

Published online: 25 July 2022

speciation and diversification to occur in spring habitats and restructured rivers throughout the system. In this sense, the Great Basin has served as both a driver of diversification, as well as a refugial habitat that has preserved a unique array of aquatic biodiversity. Unfortunately, little is known about the patterns of genetic variation and connectivity across this heterogeneous landscape (Riddle et al. 2014), particularly in regard to the aquatic species that occupy desert spring habitats in the basin.

Aquatic endemicity within the Great Basin is high, with unique representatives of several fish families including Leuciscidae, Cyprinodontidae, and Goodeidae (Williams and Wilde 1981; Miller et al. 1989; Sada and Vinyard 2002). The genus *Crenichthys* (Cyprinodontiformes: Goodeidae: Empetrichthyinae) (Gilbert 1893) is solely distributed within freshwater spring and pool habitats of the Great Basin of Nevada (United States of America). The majority of its closest relatives (Goodeinae) occupy the Mesa Central of Central Mexico, where species diversity is high (Lyons et al. 2019). Formation of the Sonoran Desert during the Tertiary period is believed to have played a substantial role in the isolation of the two subfamilies of goodeid fishes, Empetrichthyinae and Goodeinae, respectively inhabiting the Great Basin and the Mesa Central (Doadrio and Dominguez 2004).

Crenichthys is composed of two nominal species with five subspecies designated for one: *Crenichthys baileyi* (five subspecies: *C. b. albivalis*, *C. b. baileyi*, *C. b. grandis*, *C. b. moapae*, and *C. b. thermophilus*) and *Crenichthys nevadae* (Williams and Wilde 1981; Parenti 1981). Campbell and Piller (2017) suggested that a taxonomic revision for the genus is needed as the current taxonomic arrangement does not agree with the evolutionary history. In addition, species tree analyses suggest that there may also be two undescribed species within the genus, one within *C. baileyi* and one within *C. nevadae* (Campbell and Piller 2017). *Crenichthys baileyi* is distributed across disconnected endorheic springs that make up the pluvial White River in southeastern Nevada, which ultimately runs into the Pahranagat Wash and then to Lake Mead. Similarly, *Crenichthys nevadae* is distributed across disjunct endorheic springs west of the White River between Railroad Valley and Pancake Range. Because of their geographic isolation and threats from invasive species, subspecies of *C. baileyi* are listed under the US Endangered Species Act, while *C. nevadae* is listed as threatened throughout its range (Fig. 1) (La Rivers 1994; Scoppettone et al. 2004; Jelks et al. 2008; Guadalupe 2012).

As a result of its restricted distribution and unique evolutionary history within the Great Basin, *Crenichthys* represents an excellent aquatic model to study gene flow and genetic diversity within a desert system. From a historical perspective, *Crenichthys* has been isolated in multiple disjunct systems within the Great Basin for a minimum of 10,000 to 30,000 years (Hubbs et al. 1974). Despite this short evolutionary time-period, known morphological differences among the subspecies exist, but fail to distinguish between subspecies due to character overlap. Furthermore, the amount of genetic variation within populations has never been quantified for this group. Aquatic desert species have high conservation value and are particularly susceptible to imperilment (Minckley and Deacon 1968; Minckley and Marsh 2016), and understanding population genetic structure is an important step to adequately manage and conserve biological resources (Moran 2002; Schwartz et al. 2006).

The overall objective of this study is to describe genetic structure and diversity among populations of *C. baileyi* and *C. nevadae*, using single nucleotide polymorphisms (SNPs). SNPs are useful molecular markers because numerous loci can be genotyped, providing the statistical power to detect shallow population structure and large number of individuals may not be necessary (Helyar et al. 2011; Larson et al. 2013). Imperiled species are often difficult to manage because of the lack of information on population genetic variation, effective population sizes, genetic structure, and general demographic processes (Westemeier et al. 1998; DeSalle and Amato 2004; Russello et al. 2015). Therefore, it is expected that the results of this study will fill in the genetic data gap for *Crenichthys* and will provide a comparative system for other aquatic desert species. By understanding the underlying patterns of genetic variation and genetic structure among populations, scientifically informed management practices can be employed, and a stronger species recovery plan can be implemented in the future.

Methods

Sample locality and extraction

Fin clips of individuals from nearly every known population of *C. baileyi* and *C. nevadae* (17 different localities) were collected, preserved in 95% ethanol, and provided to us by the Nevada Department of Wildlife (Table 1). Samples of *C. baileyi* were collected from ten localities spanning the entire range of this species. Samples of *C. nevadae* were obtained from seven populations, including a potentially undescribed species from the Duckwater River (Fig. 1) (Campbell and Piller 2017). Genomic DNA was extracted from 201 individuals of *C. baileyi* and 119 individuals of *C. nevadae* using the Wizard® SV 96 Genomic DNA Purification System (Promega Inc.) following the manufacturer's recommendations. All extracted DNA samples were stored at -80°C before being sent for data collection. DNA from all samples was quantified using a Qubit 2.0 Fluorometer

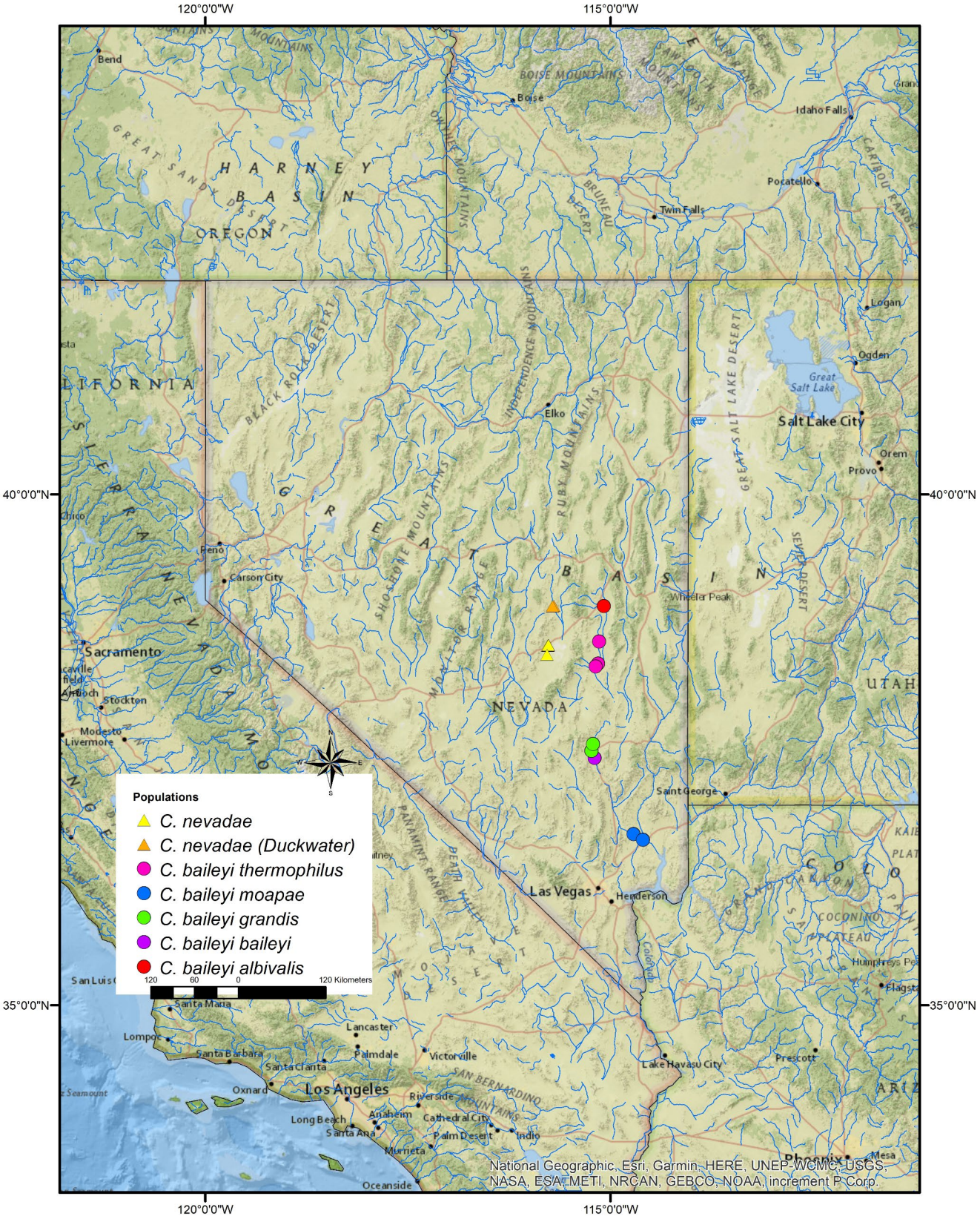

Fig. 1 Topographic map of the Great Basin in Nevada. Points represent the GPS locations where the samples for each population were collected. Labeled features represent significant physiological features separating populations. Colors and shapes correspond to the DAPC

Table 1 Taxon, population ID label, collection site, latitude, longitude, and number of individuals sequenced using RADseq and included in the final analysis

Taxon	Taxon label	Locality Label	Collection Site	Latitude	Longitude	N
C. baileyi thermophilus	Cbt	MRHS	Moon River Hot Springs	38.379382	-115.150381	15
		MH	Moorman Hot Springs	38.351817	-115.181545	8
		HC	Hot Creek	38.593969	-115.139386	14
C. b. moapae	Cbm	MR	Muddy River	36.712041	-114.714350	20
		HVP	Hidden Valley Pond	36.654803	-114.597732	21
C. b. grandis	Cbg	CSS	Crystal Springs South	37.531767	-115.233648	17
		CSN	Crystal Springs North	37.532211	-115.233349	10
		HS	Hiko Spring	37.598306	-115.215299	21
C. b. baileyi	Cbb	AS	Ash Spring	37.461052	-115.193526	17
C. b. albivalis	Cba	PS	Preston Spring	38.933514	-115.081779	24
Crenichthys nevadae	Cn	BSLR	Big Springs Loches Ranch	38.554172	-115.774762	17
		NSLR	North Springs Loches Ranch	38.560210	-115.765420	16
		LWD	Little Warm Duckwater	38.935874	-115.699145	16
		SSD	School Spring Duckwater	38.932801	-115.715342	7
		RS	Reynolds Spring	38.554964	-115.766944	16
		HR	Haycorral Loches Ranch	38.557341	-115.764230	12
		THS	Terrace Hot Spring	38.464644	-115.782834	15

(Invitrogen, Life Technologies). Samples with concentrations lower than required for sequencing were concentrated using a Savant™ DNA 120 SpeedVac™ Concentrator (Thermo Fisher Inc.) to increase concentration before 70 µl of each sample was submitted for restriction site-associated DNA sequencing (RADseq).

SNP genotyping and filtering

Library preparation and *de novo* sequencing of RAD makers was performed by Floragenex (Eugene, OR; http://www.floragenex.com/) using a protocol based on Baird et al. (2008), Emerson et al. (2010), and Hohenlohe et al. (2010). Briefly, *de novo* reference sequences were constructed for both *C. baileyi* and *C. nevadae* by Floragenex using BWA v0.6.1 (Li and Durbin 2009). Reads were aligned to the *de novo* reference using BOWTIE v1.1.1 (Langmead et al. 2009). After alignment, BCFtools, part of SAMtools v.0.1.16 (Li et al. 2009), was used to call genotype variants and produce variant call formatted (VCF) files.

Resulting VCF files were filtered using a highly conservative approach. All quality control filters were applied using VCFtools v 0.1.13 (Danecek et al. 2011). Variants were restricted to biallelic SNP markers with minor allele counts ≥ 3, a minor allele frequency ≥ 0.05, a Phred-based quality score ≥ 20, and successfully genotyped in ≥ 75% of the individuals. A site was removed if genotypes had read depths < 5. An individual was removed if it was missing ≥ 50% of data across loci. A genotype call rate was also assessed per population. A locus was removed if data was missing for ≥ 20% of individuals per population. Finally, we filtered loci that deviated from Hardy-Weinberg equilibrium (HWE). We used a minimum threshold of $p < 0.05$ in at least 25% of the total number of populations sampled.

Candidate loci under natural selection (i.e. outlier loci) were detected using a Bayesian F_{ST}-based method in BayeScan v. 2.1 (Foll and Gaggiotti 2008; Foll et al. 2010; Fischer et al. 2011). Chain parameters were left at default (5000 iterations, 50,000 burn-in) as well as model parameters except the prior odds for the neutral model was set to 100 to decrease false positives. Two separate runs were performed. Chain convergence was examined graphically and chain variances were compared using Gelman and Rubin's diagnostic (Gelman and Rubin 1992) in R. Outlier loci were subsequently removed from the data.

We filtered loci with a strong signal of linkage disequilibrium. The square of the correlation coefficient between genotypes (r^2) was calculated between each locus pair. Locus pairs with $r^2 \geq 0.6$ were considered strongly linked loci. A single variant was randomly pruned from each pairwise locus comparison with $r^2 > 0.6$.

Genetic diversity and structure

Arlequin v3.5.2.2 (Excoffier and Lischer 2010) was used to calculate the number of polymorphic sites (per population), observed heterozygosity (H_O; per locus and population), expected heterozygosity (H_E; per locus and population), nucleotide diversity (π), and Watterson's theta based on segregating sites (θ S; Watterson 1975) using the default parameters for both *C. baileyi* and *C. nevadae*. Standard deviations for both observed and expected heterozygosity (H_O SD, H_E SD) were also obtained for both species.

Pairwise fixation indices (F_{ST}) were calculated with the R package *StAMPP* (Pembleton et al. 2013) following Weir

Table 2 Observed measures of genetic diversity for all ten populations of C. baileyi over 85 SNP loci. HE, expected heterozygosity; HO, observed heterozygosity; HE SD, expected heterozygosity standard deviation; HO SD, observed heterozygosity standard deviation;, nucleotide diversity; Θ S, Watterson's estimator

π								
Population		**Polymorphic sites**	**H_E**	**H_O**	**H_ESD**	**H_OSD**	**π**	**Θ S**
MRHS	61	0.30		0.38	0.13	0.22	0.22	15.40
MH	48	0.37		0.48	0.12	0.20	0.21	13.96
HC	63	0.30		0.38	0.13	0.21	0.22	15.64
MR	64	0.34		0.46	0.12	0.21	0.26	14.87
HVP	64	0.36		0.50	0.11	0.21	0.28	14.87
CSS	62	0.29		0.36	0.11	0.17	0.21	15.16
CSN	58	0.35		0.46	0.11	0.19	0.25	16.35
HS	62	0.28		0.34	0.11	0.16	0.19	13.84
AS	64	0.24		0.26	0.11	0.14	0.17	14.71
PS	61	0.32		0.44	0.13	0.24	0.24	13.75

and Cockerham's method (θ; Weir and Cockerham 1984). Associated p-values and 95% confidence intervals were estimated between populations using 1,000 bootstraps. A Benjamini-Hochberg procedure was used to control the false discovery rate at 10% (Benjamini and Hochberg 1995).

We used discriminant analysis of principal components (DAPC) to analyze the genetic structure of *C. baileyi* and *C. nevadae* without any assumptions of population structure *a priori* in the R package *adegenet* (Jombart 2008; Jombart et al. 2010; Jombart and Ahmed 2011). We used a cross-validation procedure to determine the number of PCA axes to retain. A total of 999 replicates were carried out at each level of principal component (hereafter PCs) retention. The number of PCs associated with the lowest mean square error were retained in the DAPC.

To further infer genetic structure, ancestry probabilities (Q-values) were estimated per individual using a maximum likelihood model-based method implemented in ADMIXTURE v1.22 (Alexander et al. 2009). Fifteen genetic clusters (*K*) were tested and a fifteen-fold cross validation procedure was run per cluster. The number of clusters with the lowest standard error was taken as the optimal value of *K*. The analysis was then run using the optimal *K* value with the default parameters. The resulting Q-values were plotted as a bar plot using the R package *ggplot2* (Wickham 2016) and associated packages *grid*, *gridExtra*, and *reshape2* (Wickham 2007).

Results

SNP Discovery

A total of 2.1×10^8 reads was sequenced for *C. baileyi* with an average of 1.0×10^6 reads per sample (std. dev. = 7.9×10^5). A total of 90,793 variants were detected across all populations before filtering. As for *C. nevadae*, a total of 1.1×10^8 reads was sequenced with an average of 8.9×10^5 reads per sample (std. dev. = 6.3×10^5). A total of 51,009 variants were detected before filtering. Floragenex called genotypes for 729 SNP loci for *C. baileyi* and 631 loci for *C. nevadae*. Of the 320 individuals sequenced, 37 individuals had ≥ 50% missing data and were removed from subsequent analyses. After applying quality control filters, removing outlier loci, and removing strongly linked loci, 85 SNPs and 180 individuals were retained from the populations of *C. baileyi*, while 77 SNPs and 103 individuals were retained from *C. nevadae*'s populations. These variants were used for downstream analysis.

Genetic diversity

Values for observed measures of genetic diversity can be found in Table 2 for *C. baileyi* and Table 3 for *C. nevadae*. All population abbreviations can be referred to in Table 1. The number of polymorphic sites varied between 48 (MH) and 64 loci (MR) per population in *C. baileyi* and varied between 51 (SSD) and 72 loci (NSLR) per population in *C. nevadae*. Measures of observed heterozygosity (H_O) did not vary significantly among populations of *C. baileyi* (0.263–0.499) or *C. nevadae* (0.313–0.553). The same was true for measures of expected heterozygosity (H_E) among *C. baileyi* (0.242–0.365) and *C. nevadae* (0.288–0.402). Although measures of observed heterozygosity were greater than expected heterozygosity for all populations in both species, they were all within one standard deviation of the expected heterozygosity. Measures of nucleotide diversity (π) and Watterson's Estimator (Θ S) did not vary significantly in either species as well.

Discriminate analysis principal components (DAPC)

For *C. baileyi*, the DAPC assigned individuals from ten localities into three genetic clusters ($K=3$). One cluster

Table 3 Observed measures of genetic diversity for all seven populations of C. nevadae over 77 SNP loci. HE, expected heterozygosity; HO, observed heterozygosity; HE SD, expected heterozygosity standard deviation; HO SD, observed heterozygosity standard deviation;, nucleotide diversity; Θ S, Watterson's estimator

π							
Population	**Polymorphic sites**	**H_E**	**H_O**	**H_ESD**	**H_OSD**	**π**	**Θ S**
BSLR	70	0.32	0.41	0.13	0.20	0.27	17.12
NSLR	72	0.32	0.40	0.13	0.20	0.29	17.88
LWD	56	0.35	0.48	0.14	0.24	0.25	13.70
SSD	51	0.40	0.55	0.14	0.27	0.26	16.04
RS	71	0.31	0.39	0.13	0.21	0.26	17.63
HR	71	0.34	0.45	0.14	0.24	0.31	19.02
THS	68	0.29	0.31	0.13	0.18	0.23	16.40

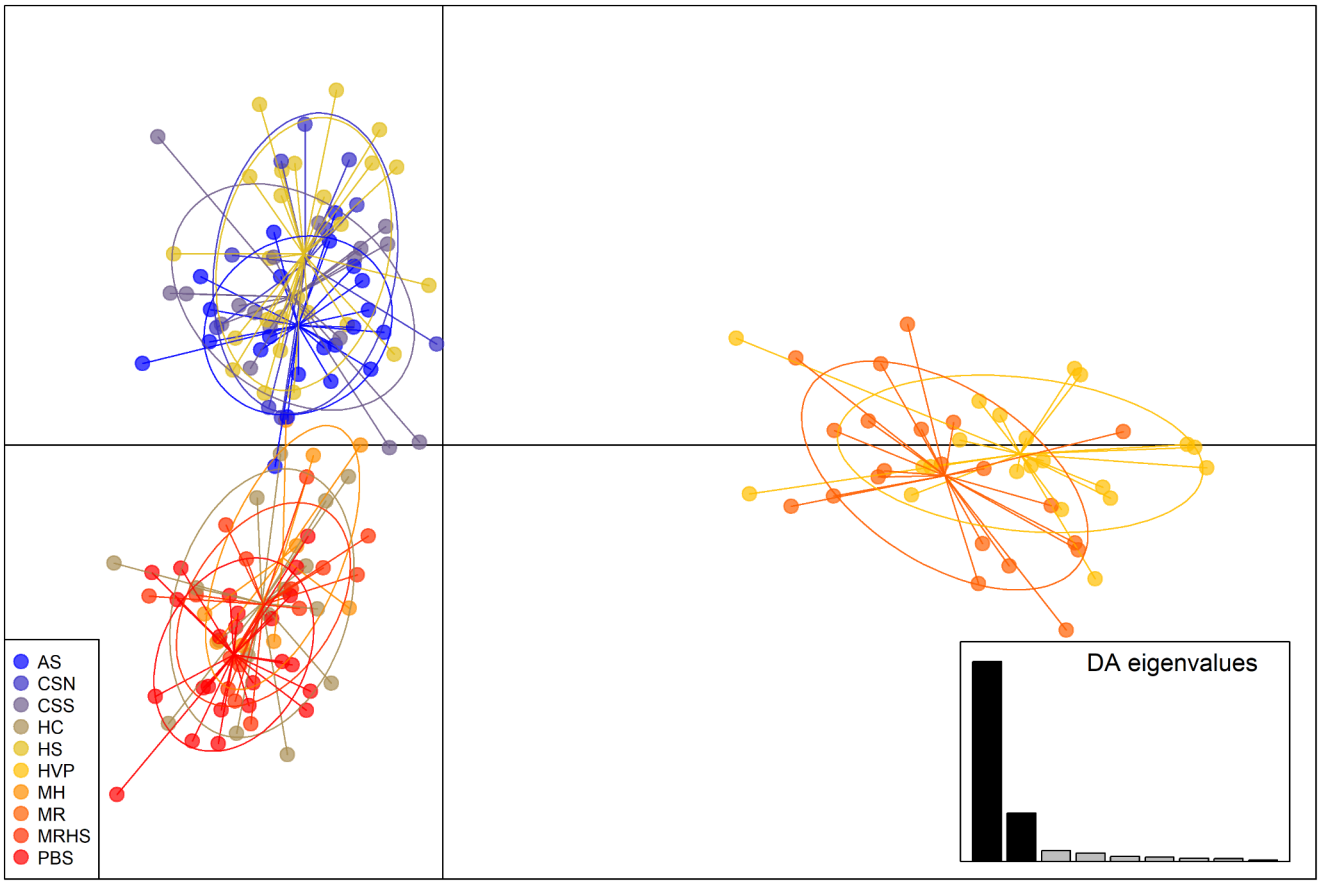

Fig. 2 Discriminant analysis of principal components for populations of *Crenichthys baileyi* based on 85 SNPs. Nine linear discriminant functions and 38 principal components (PC) were retained conserving 81.9% of the total variance. The first PC (PC1) accounted for 11.62% of the total variance, PC2 accounted for 6.33% of the total variance. Populations clustered into three groups when comparing functions 1 and 2: one containing MR and HVP, the two populations of *C. b. moapae*, a second containing MRHS, MH, HC and PBS, the three populations of *C. b. thermophilus* and one of *C. b. albivalis*, and a last one containing CSS, CSN, HS, and AS, the three populations of *C. b. grandis* and one of *C. b. baileyi*

contains individuals from MR and HVP (*C. b. moapae*). A second cluster contains individuals from MRHS, MH, HC, and PBS (*C. b. thermophilus* and *C. b. albivalis*). The final cluster contains all individuals from CSS, CSN, HS, and AS (*C. b. grandis* and *C. b. baileyi*) (Fig. 2). For the analysis, nine linear discriminant functions and 38 principal components (PC) were retained conserving 81.9% of the total variance. The first PC (PC1) accounted for 11.62% of the total variance, PC2 accounted for 6.33% of the total variance. PC1 separated the cluster containing the two populations of *C. b. moapae* from the clusters containing all the other members of *C. baileyi*. PC2 separated the cluster containing *C. b. thermophilus* and *C. b. albivalis* and the cluster containing *C. b. grandis* and *C. b. baileyi*.

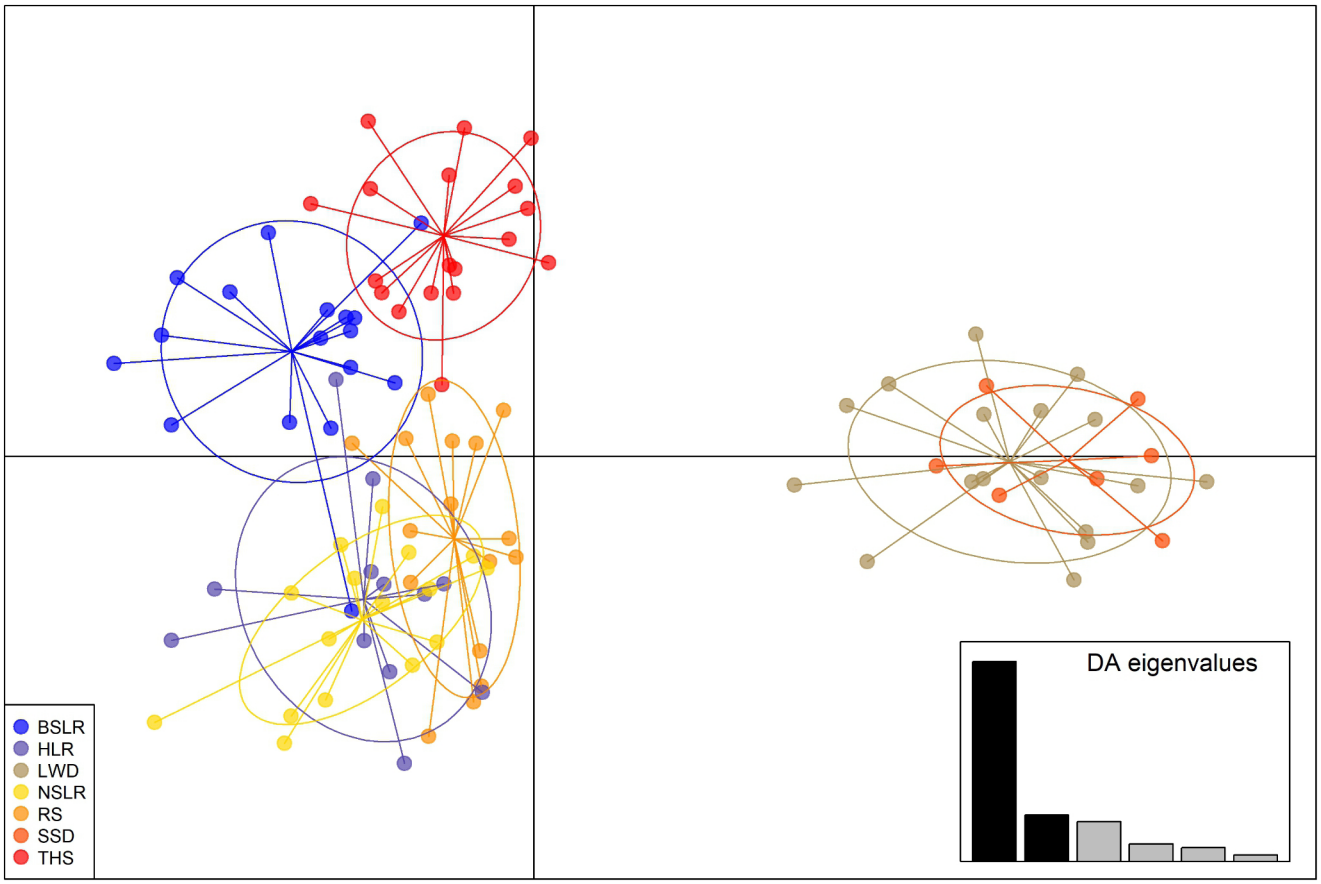

Fig. 3 Discriminant analysis of principal components for *Crenichthys nevadae* based on 77 SNPs. Six linear discriminant functions and 46 principal components (PC) were retained conserving 92.1% of the total variance. PC1 accounted for 7.87% of the total variance, PC2 accounted for 5.59% of the total variance. Populations clustered into two main groups: one containing LWD and SSD, the two Duckwater populations of *C. nevadae*, and the other containing BSLR, NSLR, HLR, RS, and THS, the remaining populations of *C. nevadae*. We can also see separation between the populations of BSLR and NSLR

When considering *C. nevadae*, DAPC assigned individuals from the seven localities into two main genetic clusters ($K=2$). The first cluster contains individuals of *C. nevadae* from the two Duckwater populations (SSD; LWD). This cluster corresponds to the potential undescribed species identified by Campbell and Piller (2017). The second cluster contains the remaining localities of *C. nevadae* (Fig. 3). For the analysis, six linear discriminant functions and 46 principal components (PC) were retained conserving 92.1% of the total variance. PC1 accounted for 7.87% of the total variance and separated the two main clusters, one containing the two Duckwater populations of *C. nevadae* (SSD, LWD) and the other containing the remaining populations.

ADMIXTURE Analysis

Following cross-validation tests, the optimal number of clusters retained for *C. baileyi* was $K=3$. Clusters identified by ADMIXTURE were comparable to those identified by DAPC. The first cluster included all the individuals from AS, CSN, CSS, HS (*C. b. grandis* and *C. b. baileyi*), the second contained all the individuals from MH, HC, MRHS, PBS (*C. b. thermophilus* and *C. b. albivalis*), and the third contained all the individuals from MR and HVP (*C. b. moapae*) (Fig. 4). There was significant admixture between individuals in clusters one and two, while almost none was present with cluster 3. Within clusters 1 and 2, the populations PBS and MH also had reduced admixture in comparison to other populations.

As for *C. nevadae*, the optimal number of clusters retained was $K=2$. Similar to *C. baileyi*, clusters identified by ADMIXTURE matched the clusters suggested by the DAPC, with the members of the Duckwater populations (SSD and LWD) forming one cluster (K2) and the other five populations (BSLR, NSLR, RS, HLR, THS) forming another cluster (K1). There was very little admixture between the two clusters (Fig. 5).

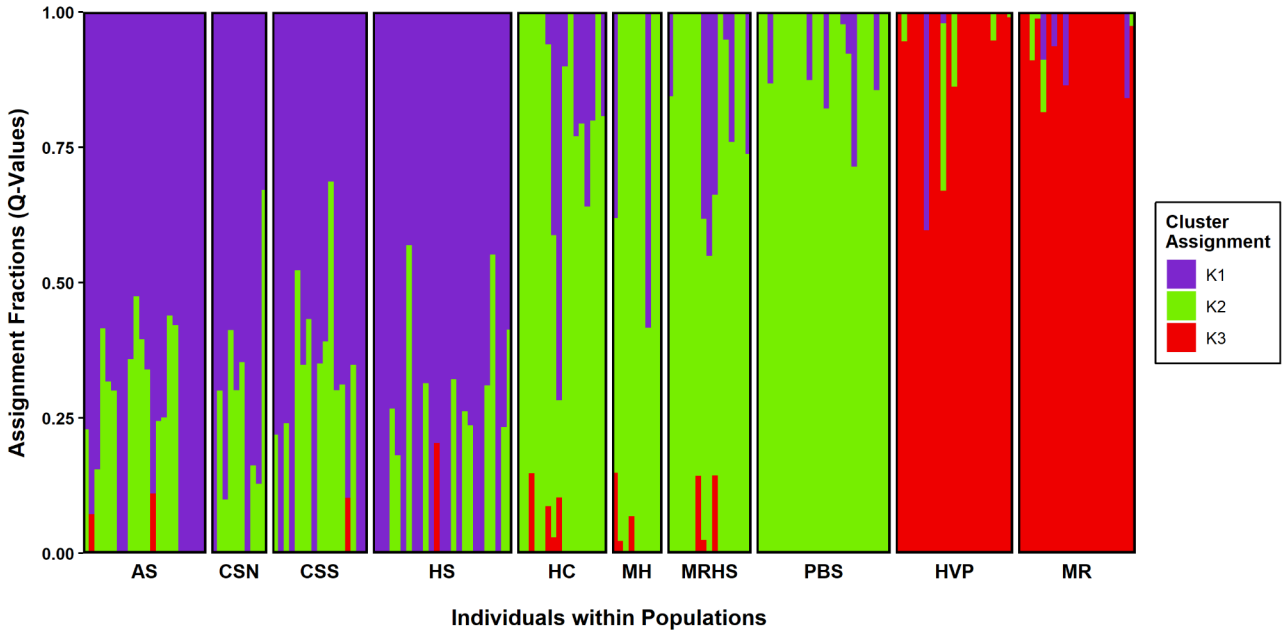

Fig. 4 Individual clustering analysis obtained using ADMIXTURE of 182 *Crenichthys baileyi* for $K=3$. Colors correspond to separate genetic clusters. Each vertical bar corresponds to an individual and their respective probability of assignment to each cluster

Estimation of F_{ST}

Pairwise F_{ST} values for *C. baileyi* ranged from zero between CSS and CSN to 0.123 between the *C. b. moapae* population at HVP and the *C. b. baileyi* population at AS (Table 4). Significant F_{ST} values corrected using the Benjamini-Hochberg procedure ($p<0.05$) supported the results obtained with DAPC and ADMIXTURE. Values for the populations HVP and MR were an order of magnitude greater than the other results, suggesting these populations have differentiated greatly from the other populations of *C. baileyi*.

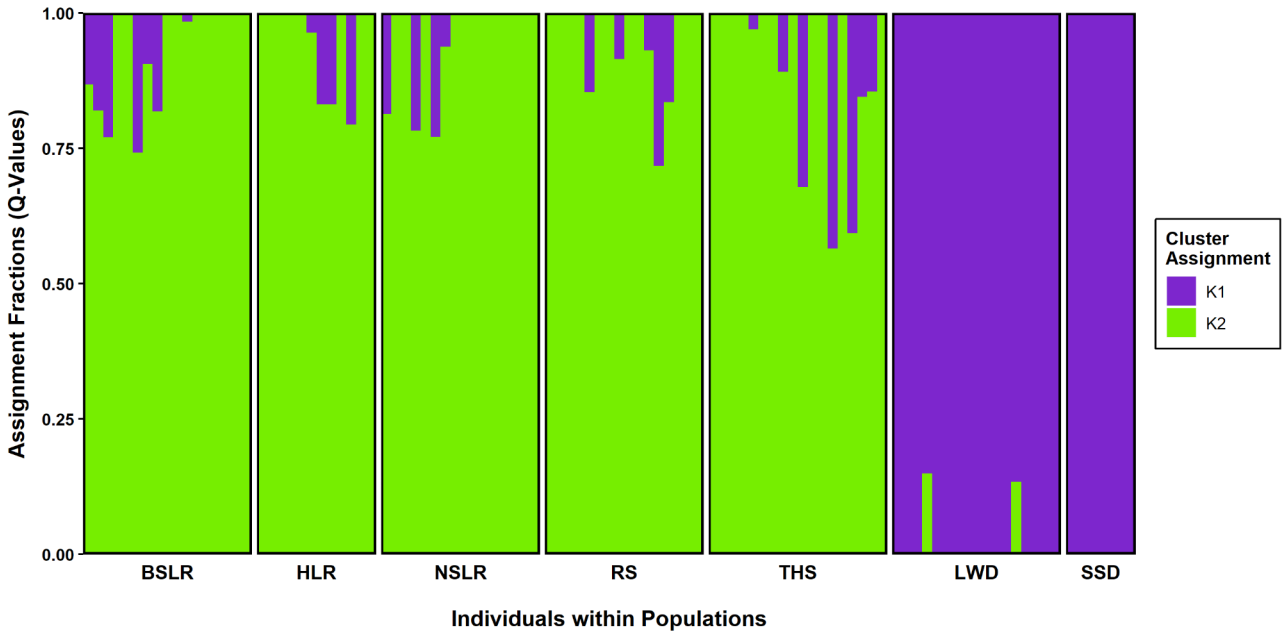

Fig. 5 Individual clustering analysis obtained using ADMIXTURE of 104 *Crenichthys nevadae* for $K=2$. Colors correspond to separate genetic clusters. Each vertical bar corresponds to an individual and their respective probability of assignment to each cluster

Table 4 Pairwise FST values based on Weir and Cockerham (1984) for Crenichthys baileyi below the diagonal. Values with a * indicate significance after a Benjamini-Hochberg procedure ($p < 0.05$). 95% confidence intervals are in the supplementary data Table S1

	AS	CSN	CSS	HS	HC	MH	MRHS	PBS	HVP	MR
AS	---									
CSN	0.024*	---								
CSS	0.016*	0.000	---							
HS	0.011*	0.000	0.000	---						
HC	0.046*	0.038*	0.031*	0.047*	---					
MH	0.057*	0.056*	0.046*	0.063*	0.021*	---				
MRHS	0.042*	0.042*	0.035*	0.045*	0.000	0.024*	---			
PBS	0.059*	0.040*	0.045*	0.055*	0.001	0.025*	0.004	---		
HVP	0.123*	0.104*	0.114*	0.120*	0.104*	0.110*	0.107*	0.110*	---	
MR	0.107*	0.102*	0.110*	0.112*	0.095*	0.105*	0.098*	0.104*	0.004	---

For *C. nevadae*, pairwise F_{ST} values ranged from zero between RS and both HLR and NSLR, to 0.091, between SSD and THS (Table 5). Similar to *C. baileyi*, significant F_{ST} values for *C. nevadae* corrected using the Benjamini-Hochberg procedure ($p < 0.05$) supported the results obtained with the DAPC and ADMIXTURE. Values were greatest between the Duckwater populations (LWD, SSD) and the other five populations of *C. nevadae*.

as samples from nearly every known geographic locality were examined.

Population Genetic structure and Gene Flow

In our study, the number of SNPs recovered after filtering was low in comparison to other studies (Corander et al. 2013; Reitzel et al. 2013; Sovic et al. 2018). However, despite the relatively low numbers of loci, a sufficient amount of variation was recovered to successfully describe

Table 5 Pairwise FST as described by Weir and Cockerham (1984) for Crenichthys nevadae below the diagonal. Values with a * indicate significance after a Benjamini-Hochberg procedure ($p < 0.05$). Confidence intervals are in the supplementary data Table S2

	RS	BSLR	HLR	NSLR	THS	LWD	SSD
RS	---						
BSLR	0.006	---					
HLR	0.000	0.002	---				
NSLR	0.000	0.012*	0.005	---			
THS	0.011*	0.010*	0.023*	0.019*	---		
LWD	0.049*	0.072*	0.066*	0.067*	0.070*	---	
SSD	0.077*	0.089*	0.076*	0.083*	0.091*	0.014*	---

Discussion

Historically, the management of *Crenichthys*, an imperiled group of desert fishes, has utilized information on age, growth, reproductive biology, and census size to develop management and conservation strategies (Williams and Wilde 1981; Williams and Williams 1981). Although these approaches provide invaluable data, they only tell a portion of the story. Modern approaches that incorporate genetic data have become extremely relevant and informative for conservation (Meffe 1986; Ryman and Utter 1987; Schwartz et al. 2006). Prior to Campbell and Piller (2017) and this present study, there were uncertainties regarding basic taxonomy, genetic structure, and genetic variation for *Crenichthys*, all of which have hampered conservation efforts. This present study represents the most comprehensive review of genetic structure and genetic variation within *Crenichthys*, the genetic structure within *Crenichthys*. Other studies, using small numbers of SNP loci, have also been successful in recovering informative levels of genetic variation, which is testament to the power and usefulness of SNPs (Hannelius et al. 2008; Waples and Do 2010; Rašić et al. 2014; Torres-Martinez and Emery 2016; Caldu-Primo et al. 2017).

The DAPC and ADMIXTURE results are in agreement with the phylogenetic results of Campbell and Piller (2017) and supports the existence of three groups within *C. baileyi*. The results of the analysis of pairwise F_{ST} matched these results and some of the results of the DAPC. The results of this analysis are highly dependent on recent demography and sampling, however, so they should not be taken as the final word on these fishes population structure. Pairwise F_{ST} values for the population of *C. b. baileyi* (AS) were significant when compared to all other populations of *C. baileyi*, but did not appear in the DAPC, while significant values for

the Moorman Hot Spring (MH) population were corroborated by separation in the DAPC.

Geographic isolation and reduced gene flow have previously been shown to play major roles in the differentiation of other organisms in aquatic desert systems (i.e. Death Valley Model of Meffe and Vrijenhoeck, 1988; Houston et al. 2015) leading to differentiation across short evolutionary time scales. This is the pattern recovered for *C. baileyi* in our study, matching previous work by Witt et al. (2006) in native amphipods within the fragmented springs of the White River. Hubbs et al. (1974) suggested that populations of *Crenichthys* have been isolated in the Great Basin for the last 10,000–30,000 years, with individual springs having potentially been isolated for 10,000 years or more, which has resulted in genetic divergence among several of the subspecies (Williams and Wilde 1981). The geographic separation of *C. baileyi* populations has previously been discussed in Campbell and Piller (2017), which showed that the three groups obtained here are geographically separated by > 40 km of desert with no connection between the three groups. The springs' temperature and oxygen levels also vary among one another, but are relatively constant within each individual spring. This may explain the divergence between individual populations as a result of adaptation, in addition to genetic drift from isolation. Both AS and MH maintain high temperatures (35ºC and 37ºC on average respectively), while MH also has very low dissolved oxygen (0.7 ppm) (Sumner and Sargent 1940; Williams and Wilde 1981). Because of this, these populations may represent genetically distinct groups that should be managed independently, especially MH, which is more geographically isolated. Since AS was only significant in our F_{ST} analysis, and had lower values, it may represent an overestimation as a result of reduced sample size. Pairwise F_{ST} values can be overestimated when sample sizes are small (<1% of the total population) and can affect results obtained with molecular data, as stated previously (DeSalle and Amato 2004; Morin et al. 2009; Willing et al. 2012).

As for *C. nevadae*, the results from the DAPC and ADMIXTURE analyses suggest there are two groups within this species, corroborating the previous phylogenetic results of Campbell and Piller (2017). The results of the F_{ST} analysis showed significant differences between THS and all other populations, and between both NSLR and BSLR, which were both seen in the DAPC. However, significant differences between both Duckwater populations (LWD, SSD) were shown in the F_{ST} analysis as well, but were not seen in the DAPC.

Crenichthys nevadae occupies an area formerly known as Lake Railroad, which desiccated into two areas within the Railroad Valley since the Pleistocene (Williams and Williams 1981). The northern region, near the Duckwater Shoshone Reservation, is currently occupied by the Duckwater populations of *C. nevadae*, whereas the southern area, approximately 43 km south, is inhabited by the remaining five populations examined in this study. The desiccation and fragmentation of ancient Lake Railroad likely was the driving force leading to the differentiation of the two groups of isolated *C. nevadae*. The genetic differentiation between groups in *C. nevadae* is lower than within *C. baileyi*, however, so the period of isolation may not have been short. This idea is supported by the work of Hubbs and Miller (1948), which described how Lake Railroad likely desiccated during the Holocene, while the White River System began fragmenting in the early Pleistocene. These factors, and genetic drift via isolation, likely led to the differentiation of these two groups following the Pleistocene. The separation of THS recovered in the F_{ST} and DAPC analyses may be a result of the population being introduced and be the result of a founder effect or genetic drift over time since its introduction (*pers. comm*. Nevada Department of Wildlife; Mayr 1942; Planes and Lecaillon 1998; Janac et al. 2017). Based on our diversity results, the population does not seem appear to be inbred ($H_O > H_E$). The separation recovered between BSLR and NSLR seems to be the result of genetic divergence rather than a reduction in genetic diversity, as neither have a particularly low sample size or values for genetic diversity, and neither appear to be greatly separated geographically.

Although we expected reduced genetic variation to be detected within populations of *C. baileyi* and *C. nevadae*, the genetic diversity results suggest otherwise. All obtained values for observed heterozygosity were greater than expected heterozygosity for all populations of both species, within one standard deviation of one another. Nucleotide diversity and Watterson's estimator also did not vary greatly among populations of *C. baileyi* or *C. nevadae*, suggesting roughly equal levels of genetic variation among populations. These results support the Death Valley Model (Meffe and Vrijenhoeck, 1988) described above, and further support that both species may have rapidly evolved following the Pleistocene glaciation (Gillooly et al. 2001; Lynch 2010; Martin et al. 2016; Martin and Höhna 2017).

Conclusions

At the present time, two species (*C. baileyi* and *C. nevadae*) and five subspecies of *C. baileyi* (*albivalis, baileyi, grandis, thermophilus*, and *moapae*) are managed as unique evolutionary entities, with this scenario being based on the morphological diagnoses of the species and subspecies (Williams and Wilde 1981). Both Campbell and Piller (2017), as well as the data presented herein, support this

current taxonomic-based management scenario, but also argue that additional units should be considered separately, and that further taxonomic work is necessary. Based on our data, we recommend a management scheme corresponding to the major genetic clusters and divergent populations identified in this study, in addition to the taxonomic groups. We described three genetic clusters within *C. baileyi* and two within *C. nevadae*. These five groups are supported by the results of the DAPC, ADMIXTURE, and F_{ST}. Within these clusters, at a minimum, the divergent populations within each species should be further managed independently (MH and AS for *C. baileyi*; BSLR, NSLR, SSD and THS for *C. nevadae*) based on the results of the DAPC and F_{ST}. We recommend future work targets these individual populations to help improve conservation management of these species.

Supplementary Information The online version contains supplementary material available at https://doi.org/10.1007/s10592-022-01457-3.

Acknowledgements We would like to thank the Nevada Department of Wildlife for providing the samples for this study. We would also like to thank the biologists Brandon Senger, Kevin Guadalupe, James Harter, and Mark Beckstrand for funding, support, and discussions throughout this study. Also, we would like to thank the members past and present of the Piller Lab at Southeastern Louisiana University for their continued support, advice, and assistance with this study. We would like to extend a special thanks to Jerry Kattawar III for assisting with the DNA extractions and sample preparation. The funders had no role in study design, molecular data collection and analysis, decision to publish, or preparation of the manuscript.

Authors' contributions DCC and KRP conceived the project idea. KRP secured the project funding. DCC performed DNA extraction and laboratory preparation of samples and was responsible for data analysis. DTC assisted with data analyses and draft editing. DCC spearheaded the writing with substantial contributions DTC and KRP.

Funding Funding was provided by the Nevada Department of Wildlife and the National Science Foundation (DEB 1,354,930) to KRP.

Data Availability, material, and code The data and datasets generated/analyzed during the current study are available on DataDryad (https://doi.org/10.5061/dryad.prr4xgxpc).

Declarations

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

Ethics approval Southeastern Louisiana University, IACUC 0002.

Consent for publication All authors have consented to this papers publication.

References

Alexander DH, Novembre J, Lange K (2009) Fast model-based estimation of ancestry in unrelated individuals. Genome Res 19:1655–1664

Axelrod DI (1979) Age and origin of Sonoran Desert vegetation. Occ Pap California Acad Sci 132:1–74

Baird N, Etter PD, Atwood TS et al (2008) Rapid SNP discovery and genetic mapping using sequenced RAD markers. PLoS ONE 3:1–7

Benjamini Y, Hochberg Y (1995) Controlling the false discovery rate: a practical and powerful approach to multiple testing. J Royal Stat Soc Ser B Methodological 57:289–300

Caldu-Primo J, Mastretta-Yanes A, Wegier A, Pinero D (2017) Finding a needle in a haystack: Distinguishing Mexican maze landraces using a small number of SNPs. Front Genet 8:1–12

Campbell DC, Piller KR (2017) Let's jump in: A phylogeographic study of the Great Basin springfish and poolfishes, *Crenichthys* and *Empetrichthys* (Cyprinodontiformes: Goodeinae). PLoS ONE 12:1–21

Corander J, Majander KK, Cheng L, Merila J (2013) High degree of cryptic population differentiation in the Baltic Sea herring *Clupea harengus*. Mol Ecol 22:2931–2940

Danecek P, Auton A, Abecasis G et al (2011) The variant call format and VCFtools. Bioinformatics 27:2156–2158

DeSalle R, Amato G (2004) The expansion of conservation genetics. Nat Rev Genet 5:702–712

Doadrio I, Dominguez O (2004) Phylogenetic relationships within the fish family Goodeidae based on cytochrome b sequence data. Mol Phylogenet Evol 31:416–430

Emerson KJ, Merz CR, Catchen JM et al (2010) Resolving postglacial phylogeography using high-throughput sequencing. Proceedings of the National Academy of Sciences 107:16196–16200

Excoffier L, Lischer HEL (2010) Arlequin suite ver 3.5: A new series of programs to perform population genetics analyses under Linux and Windows. Mol Ecol Resour 10:564–567

Fischer MC, Foll M, Excoffier L, Heckel G (2011) Enhanced AFLP genome scans detect local adaptation in high-altitude populations of a small rodent (*Microtus arvalis*). Mol Ecol 20:1450–1462

Foll M, Gaggiotti OE (2008) A genome scan method to identify selected loci appropriate for both dominant and codominant markers: A Bayesian perspective. Genetics 180:977–993

Foll M, Fischer MC, Heckel G, Excoffier L (2010) Estimating population structure from AFLP amplification intensity. Mol Ecol 19:4638–4647

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

Gillooly J, Brown J, West G, Savage V (2001) Effects of size and temperature on metabolic rate. Science 293.:2248–2251

Gilbert CH (1893) Report on the fishes of the Death Valley expedition collected in southern California and Nevada in 1891, with descriptions of new species. North Am Fauna 7:233

Guadalupe K (2012) Nevada Department of Wildlife native fish and amphibian field trip report XXXIII. 2:81–87

Hannelius U, Salmela E, Lappalainen T et al (2008) Population substructure in Finland and Sweden revealed by the use of special coordinates and a small number of unlinked autosomal SNPs. BMC Genet 9:1–12

Helyar SJ, Hemmer-Hansen J, Bekkevold D et al (2011) Application of SNPs for population genetics of non-model organisms:new opportunities and challenges. Mol Ecol Resour 11:123–136

Hohenlohe PA, Bassham S, Etter PD et al (2010) Population genomics of parallel adaptation in threespine stickleback using sequenced RAD tags. PLoS ONE Genetics 6:1–23

Houston D, Evans R, Shiozawa D (2015) Pluvial drainage patterns and Holocene desiccation influenced the genetic architecture of Relict Dace, Relictus solitarius (Teleostei: Cyprinidae). PLoS ONE 10:1–12

Hubbs CL, Miller RR (1948) The zoological evidence; correlation between fish distribution and hydrographic history in the desert basins of western United States. Bull Univ Utah 38:17–166

Hubbs CL, Miller RR, Hubbs LC (1974) Hydrographic history and relict fishes of the north-central Great Basin. Calif Acad Sci 7:1–259

Janac M, Bryja J, Ondrackova M et al (2017) Genetic structure of three invasive gobiid species along the Danube-Rhine invasion corridor: Similar distributions, different histories. Aquat Invasions 12:551–564

Jelks HL, Walsh SJ, Burkhead NM et al (2008) Conservation status of imperiled North American freshwater and diadromous Fishes. Fisheries 33:372–407

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

Jombart T, Devillard S, Balloux F (2010) Discriminant analysis of principal components: a new method for the analysis of genetically structured populations. BMC Genet 11:94

Jombart T, Ahmed I (2011) adegenet 1.3-1: new tools for the analysis of genome-wide SNP data. Bioinformatics 27:3070–3071

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:R25

La Rivers I (1994) Fishes and Fisheries of Nevada. Reprint. University of Nevada Press

Larson WA, Seeb LW, Everett MV et al (2013) Genotyping by sequencing resolves shallow population structure to inform conservation of Chinook Salmon (*Oncorhynchus tshawytscha*). Evol Appl 7:355–369

Li H, Durbin R (2009) Fast and accurate short read alignment with Burrows-Wheeler Transform. Bioinformatics 25:1754–1760

Li H, Handsaker B, Wysoker A et al (2009) The sequence alignment/map format and SAMtools. Bioinformatics 25:2078–2079

Lynch M (2010) Evolution of the mutation rate. Trends Genet 26:345–352

Lyons J, Piller KR, Artigas-Azas JM et al (2019) Distribution and current conservation status of the Mexican Goodeidae (Actinopterygii, Cyprinodontiformes). ZooKeys 885:115–158

Martin CH, Crawford JE, Turner BJ, Simons LH (2016) Diabolical survival in Death Valley. Recent pupfish colonization, gene flow, and genetic assimilation in the smallest species range on earth. Proceedings of the Royal Society Biological Sciences 283:23–34

Martin CH, Höhna S (2017) New evidence for the recent divergence of Devil. 's Hole pupfish and the plausibility of elevated mutation rates in endangered taxa. Mol Ecol 27:831–838

Mayr E (1942) Systematics and the Origin of Species. Columbia University Press, New York

McKee EH (1971) Tertiary Igneous chronology of the Great Basin of Western United States-Implications for tectonic models. Geol Soc Am Bull 82:3497–3502

Meffe GK (1986) Conservation genetics and the management of endangered fishes. Fisheries 11:14–23

Meffe GK, Vrijenhoek RC (1988) Conservation genetics in the management of desert fishes. Conserv Biol 2:157–169

Miller RR, Williams JD, Williams JE (1989) Extinctions of North American fishes during the past century. Fisheries 14:22–28

Minckley WL, Deacon JE (1968) Southwestern fishes and the enigma of "endangered species". Science 159:1424–1431

Minckley WL, Deacon JE (1991) Battle against Extinction. Native Fish Management in the American West University of Arizona Press, Tuscon, AZ,537pp

Minckley WL, Marsh PC (2016) Inland fishes of the Greater Southwest: Chronicle of a vanishing biota. University of Arizona Press, Tuscon, AZ, p 478

Moran P (2002) Current conservation genetics: Building an ecological approach to the synthesis of molecular and quantitative genetic methods. Ecol Freshw Fish 11:30–55

Morin PA, Martien KK, Taylor BL (2009) Assessing the statistical power of SNPs for population structure and conservation studies. Mol Ecol Resour 9:66–73

Murphy NP, Guzik MT, Cooper SJB, Austin AD (2015) Desert spring refugia: Museums of diversity or evolutionary cradles? Zoolog Scr 44:693–701

Parenti LR (1981) A phylogenetic and biogeographic analysis of Cyprinodontiform fishes (Teleostei, Atherinomorpha). Bull Am Museum Nat History 168:335–357

Pembleton LW, Cogan NOI, Forster JW (2013) StAMPP: an R package for calculation of genetic differentiation and structure of mixed-ploidy level populations. Mol Ecol Resour 13:946–952

Planes S, Lecaillon G (1998) Consequences of the founder effect in the genetic structure of introduced island coral reef fish populations. Biol J Linn Soc 63:537–552

Rašić G, Filipović I, Weeks AR, Hoffmann AA (2014) Genome-wide SNPs lead to strong signals of geographic structure and relatedness patterns in the major arbovirus vector. Aedes aegypti BMC Genomics 15:1–12

Reitzel AM, Herrera S, Layden MJ et al (2013) Going where traditional markers have not gone before: Utility of and promise for RAD sequencing in marine invertebrate phylogeography and population genomics. Mol Ecol 22:2953–2970

Riddle BR, Jezkova T, Hornsby AD, Matocq MD (2014) Assembling the modern Great Basin mammal biota: Insights from molecular biogeography and the fossil record. J Mammal 95:1107–1127

Russello MA, Waterhouse MD, Etter PD, Johnson EA (2015) From promise to practice: Pairing non-invasive sampling with genomics in conservation. PeerJ 3. DOI https://doi.org/10.7717/peerj.1106

Ryman N, Utter F (1987) Population Genetics and Fisheries Management. University of Washington Press, Seattle, WA

Sada DW, Vinyard GL (2002) Anthropogenic changes in biogeography of Great Basin aquatic biota. Smithson Contrib Earth Sci 33:277–293

Schwartz MK, Luikart G, Waples RS (2006) Genetic monitoring as a promising tool for conservation and management. Trends in Ecology and Evolution 22:25–32

Scoppettone GG, Rissler PH, Shea S (2004) A fish survey of the White River, Nevada. Western North American Naturalist 64:45–52

Smith GR, Dowling TE, Gobelet KW et al (2002) Biogeography and timing of evolutionary events among Great Basin fishes. Smithson Contrib Earth Sci 33:405

Smith ML (1981) Late Cenozoic fishes in the warm deserts of North America. A reinterpretation of

desert adaptations. Pp. 11–38 in Fishes in North American Deserts. R. J. Naiman and D. L. Soltz

(editors). John Wiley, New York

Sovic M, Fries A, Martin SA, Gibbs HL (2018) Genetic signatures of small effective population size and demographic declines in an endangered rattlesnake, *Sistrurus catenatus*. Evol Appl 12:664–678

Sumner FB, Sargent MC (1940) Some observations on the physiology of warm spring fishes. Ecology 21:45–54

Torres-Martinez L, Emery NC (2016) Genome-wide SNP discovery in the annual herb, *Lasthenia fremontii* (Asteraceae): Genetic resources for the conservation and restoration of a California vernal pool endemic. Conserv Genetic Resour 8:145–158

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. Evol Appl 3:244–262
Watterson GA (1975) On the number of segregating sites in genetical models without recombination. Theor Popul Biol 7:256–276
Weir BS, Cockerham CC (1984) Estimating F-statistics for the analysis of population structure. Evolution 38:1358–1370
Westemeier RL, Brawn JD, Simpson SA et al (1998) Tracking the long-term decline and recovery of an isolated population. Science 282:1695–1698
Wickham H (2007) Reshaping data with the reshape package. J Stat Softw 21:1–20
Wickham H (2016) ggplot2: Elegant graphics for data analysis. Springer-Verlag, New York, NY
Williams JE, Wilde GR (1981) Taxonomic status and morphology of isolated populations of the White River Springfish, *Crenichthys baileyi* (Cyprinodontidae). Southwest Nat 25:485–503
Williams CE, Williams JE (1981) Distribution and status of native fishes of the Railroad Valley system, Nevada. Transactions of the California-Nevada Section,Wildlife Society48–51
Willing E, Dreyer C, Oosterhout C (2012) Estimates of genetic differentiation measured by F_{ST} do not necessarily require large sample sizes when using many SNP markers. PLoS ONE 7:1–7
Witt J, Threloff D, Hebert P (2006) DNA barcoding reveals extraordinary cryptic diversity in an amphipod genus: Implications for desert spring conservation. Mol Ecol 15:3073–3082