Title

Population genomic analysis of the Speckled Dace species complex identifies three distinct lineages in California

Journal

Transactions of the American Fisheries Society, 151(6)

ISSN

0002-8487

Authors

Su, Yingxin
Moyle, Peter B
Campbell, Matthew A
et al.

Publication Date

2022-11-01

DOI

10.1002/tafs.10388

Peer reviewed

Received: 22 April 2022 | Revised: 31 August 2022 | Accepted: 2 September 2022

DOI: 10.1002/tafs.10388

FEATURE ARTICLE

Population genomic analysis of the Speckled Dace species complex identifies three distinct lineages in California

Yingxin Su[1] | **Peter B. Moyle**[2,3] | **Matthew A. Campbell**[1] | **Amanda J. Finger**[1] | **Sean M. O'Rourke**[1] | **Jason Baumsteiger**[3] | **Michael R. Miller**[1,3]

[1]Department of Animal Science, University of California–Davis, Davis, California 95616, USA

[2]Department of Wildlife, Fish, and Conservation Biology, University of California–Davis, Davis, California 95616, USA

[3]Center for Watershed Sciences, University of California–Davis, Davis, California 95616, USA

Correspondence

Yingxin Su

Email: syxsu@ucdavis.edu

Abstract

Objective: Speckled Dace *Rhinichthys osculus* is small cyprinoid fish that is widespread in western North America. In California and elsewhere it is currently treated as a single species with multiple subspecies, many undescribed. However, these subspecies may represent evolutionary lineages that are cryptic species because they cannot be distinguished using standard morphometric techniques. In this study, we attempt to determine evolutionary lineages within California populations of Speckled Dace using the population genetic and genomic information.

Methods: We used restriction site-associated DNA sequencing to extract thousands of single-nucleotide polymorphisms across the genome to identify genetic differences among all the samples from 38 locations in the western USA, with a focus on California. We performed principal component analysis, admixture analysis, estimated pairwise values of the genetic differentiation index F_{ST}, and constructed molecular phylogenies to characterize population genetic and phylogenetic relationships among sampled Speckled Dace populations.

Result: Our analyses detected three major lineages of Speckled Dace in California that align with geography: (1) Sacramento River, central California coast, Klamath River, and Warner Basin; (2) Death Valley and Lahontan Basin; and (3) Santa Ana River basin, in southern California. These lineages fit well with the geologic history of California, which has promoted long isolation of populations of Speckled Dace and other fishes.

Conclusion: The presence of distinct evolutionary lineages indicates that Speckled Dace in California should be managed with distinct population segments to preserve within-species diversity. This study highlights the importance of genetic analyses for conservation and management of freshwater fishes.

KEYWORDS

conservation genetics, conservation genomics, conservation management, genetics and genomics, RAD sequencing, taxonomy and systematics, population genetics, population genomics

INTRODUCTION

The Speckled Dace *Rhinichthys osculus* is a small (usually <10 cm TL) cyprinoid (Cypriniformes, Leuciscidae) fish that is widely distributed across western North America. It is found in northern Mexico; southern, central, and northern California; the Great Basin; and the Pacific Northwest to southwestern Canada (Moyle 2002; Smith et al. 2017). Despite its wide distribution, the Speckled Dace is considered to be one highly variable species, albeit with numerous subspecies, many of which are undescribed and of uncertain taxonomic status (Moyle 2002; Smith et al. 2017). We refer to the species, therefore, as the "Speckled Dace complex" (SDC). The SDC diverged from the Longnose Dace *Rhinichthys cataractae* of eastern and northwestern North America over 6 million years ago (mya; Spencer et al. 2008). The common ancestor of the SDC was presumably initially isolated from Longnose Dace in the ancestral Columbia River and then spread throughout the western USA, parts of Mexico, and British Columbia as the result of geologic events that connected and disconnected watersheds (Smith et al. 2017). Speckled Dace complex populations are found in a wide array of habitats from desert springs to large rivers and lakes to (most typically) small- to medium-sized streams. Their morphology is variable but generally reflects the habitat in which a particular population lives. For example, narrow caudal peduncles and large pectoral fins characterize swift-water populations, while more robust bodies, thicker caudal peduncles, and smaller pectoral fins characterize slow-water populations (Sada et al. 1995; Page and Burr 2011; Smith et al. 2017). Consequently, morphological and meristic differences do not reflect phylogenetic relationships among populations of the SDC, resulting in a long and complex taxonomic history (Smith et al. 2017).

The most comprehensive study of the systematics of the SDC is that by Smith et al. (2017), who compared dace populations from throughout western North America using mitochondrial DNA (mtDNA), morphology, fossils, and the geologic record of the entire region. Although their analyses indicated multiple lineages, they concluded that there was considerable, if sporadic, gene flow among populations, reflecting complex geologic events that promoted both connectivity and isolation. Smith et al. (2017) suggested that gene flow has prevented the formation of morphologically distinct populations that might be defined as species through the process of reticulate evolution.

Overall, genetic studies have produced mixed results as to whether or not any lineages in the SDC are distinct enough to be designated as species or subspecies. The default position is to follow Spencer et al. (2008) and Smith et al. (2017) that the SDC is a single species throughout its range because the various populations lack diagnostic characteristics that would allow them to be described as distinct phylogenetic entities. This default position is particularly problematic for California, a region that is rich in endemic fish species, many of which are threatened with extinction (Moyle 2002; Leidy and Moyle 2021). Notably, California SDC populations are among those most distant from the region of origin in the Columbia River and are also among the most southern populations of the taxon. Thus, California SDC populations reflect their remarkable record of colonizing new regions during the wetter periods of the Pleistocene and then adapting to new conditions as the waters they colonized became smaller and more isolated (Smith et al. 2017).

Impact Statement

Genomic analysis shows three distinct lineages within Speckled Dace in California, revealing genetic divergence of populations and showing the need for population-specific conservation measures.

In this paper, we analyze Speckled Dace relationships using restriction site-associated DNA (RAD) sequencing data. This approach is well suited for analyzing the SDC because it uses thousands of loci distributed across the genome from each individual rather than only a single locus or handful of loci as were used with earlier methods. For further discussion of this approach to resolving issues with identifying cryptic fish species, see Baumsteiger et al. (2017).

We investigated the following question using RAD sequencing to examine relationships among populations of Speckled Dace collected over much of the SDC's wide range: based on genetic distinctness, should members of the SDC in California be treated as a single lineage or as multiple lineages for conservation and management?

Previous genetic studies of Speckled Dace

Historically, many populations of Speckled Dace were described as separate species. For example, Jordan and Evermann (1896) listed nine species, which had mostly been described based partially on their isolation from other populations and partially on morphological and meristic characters, even though these characters overlapped among populations. However, the presence of many isolated populations of Speckled Dace with similar adaptations to local environments and, hence, convergent morphologies suggests that cryptic species exist within the SDC and that some of the recognized subspecies

(listed in Smith et al. 2017) could be recognized as species (Evermann and Meek 1896).

The advent of molecular genetic techniques resulted in renewed efforts to examine diversity within the SDC. Resulting genetic information was used to test hypotheses of evolutionary relationships among populations and to generate biogeographic scenarios relating Speckled Dace evolution to the development of the western aquatic landscape (Smith et al. 2017). Thus far, mtDNA analysis has been the primary genetic approach used to investigate the systematics of Speckled Dace (Oakey et al. 2004; Smith et al. 2017). Smith et al. (2017) compared dace populations from throughout western North America and concluded that while geographically based lineages existed, as shown by Oakey et al. (2004), there was no basis for declaring them separate species. Oakey et al. (2004) used restriction sites in the mitochondrial genome of dace distributed across the western USA to construct a molecular phylogeny. They found a close match between mtDNA patterns and the geologic history and isolation of drainage basins, concluding that the SDC consisted of three main evolutionary lineages: (1) Colorado River basin and Los Angeles Basin; (2) Great Basin (Snake River, Bonneville, Death Valley, and Lahontan basins); and (3) Columbia and Klamath–Pit rivers (Oakey et al. 2004). Pfrender et al. (2004) showed that mtDNA patterns reflected long isolation of populations in five river basins in Oregon and suggested that some of the lineages were distinct enough to be considered species. In contrast, Billman et al. (2010) did not find species-level differences in mtDNA among SDC members found in Great Basin waterways (Snake River, Bonneville, and Lahontan basins).

More narrowly, Ardren et al. (2010) applied mtDNA analysis to dace from throughout the Warner Basin and concluded that the lineages could collectively qualify as a species. Hoekzema and Sidlauskas (2014) also examined SDC fish from the Warner Basin along with dace from five other isolated Great Basin populations in Oregon. They used mtDNA and nuclear DNA (nuclear s7 intron) and found that dace in the Warner Basin were different—potentially at the species level—from dace in the other five basins.

Recognizing the limitations of mtDNA for determining evolutionary lineages, Mussmann et al. (2020) compared isolated populations of Speckled Dace from throughout the Death Valley region, in the Owens and Amargosa River basins, using double-digest RAD. They found that the region has four distinct evolutionary lineages of Speckled Dace, with each of the lineages being recognizable as a distinct population segment for management purposes.

METHODS

Sampling and DNA sequencing

We obtained samples from 38 locations across several major zoogeographic regions throughout the range of Speckled Dace (Figure 1; Table S.1 available in the Supplement in the online version of this article). Samples from Butte Lake in Lassen Volcanic National Park were included to determine whether the population is native or introduced. Fin clips were taken from live adults or from the whole fish stored in ethanol, and the fin clips were dried on Whatman qualitative filter paper and stored at room temperature. The DNA was extracted from fin clips with a magnetic bead-based protocol (Ali et al. 2016) and quantified using Quant-iT PicoGreen Double-Stranded DNA Reagent (Thermo Fisher Scientific) with an FLx800 Fluorescence Reader (BioTek Instruments). Genomic DNA was used to generate *Sbf*I RAD libraries (Ali et al. 2016) and was sequenced with paired-end, 100-bp reads on an Illumina HiSeq 2500. Demultiplexing was performed, requiring an exact match with well and plate barcodes (Ali et al. 2016). Sequencing coverage was assessed at the 50-bp position of each de novo RAD contig (see below) across all individuals using the depth function in SAMtools (Li et al. 2009).

Restriction site-associated DNA de novo assembly and alignments

To generate a reference sequence for Speckled Dace, we performed RAD de novo assembly on eight individuals from the Walker River (Data S.1–S.3 available in the Supplement in the online version of this article). Specific details of the de novo assembly methods are provided by Baumsteiger et al. (2017). Briefly, a bioinformatic pipeline including a genome assembler was used to construct a partial reference for Speckled Dace. After de novo assembly, the mem algorithm in the Burrows–Wheeler Aligner (BWA; Li and Durbin 2009) was used to align each sample to the reference under the default parameters. SAMtools was used to convert sequence alignment map (SAM) files to binary alignment map (BAM) files, calculate the percentage of aligned reads, remove PCR duplicates, filter for the proper pairs, and merge the alignments if needed (Li et al. 2009). After this process, we removed low-coverage individuals with less than 70,000 mapped reads.

Genetic population structure

Principal components analysis

To begin investigating population structure, we used Analysis of Next Generation Sequencing Data (ANGSD

FIGURE 1 Map of sampling sites in which Speckled Dace were collected for this study. The location represented by each number and the number of individuals sampled are detailed in Table S.1.

version 1.9) to identify single-nucleotide polymorphisms (SNPs; -SNP_eval 1e-12), calculate genotype likelihoods using a SAMtools model (-GL 1), infer major and minor alleles directly from the genotype likelihoods (-doMajorMinor 1), and estimate allele frequencies assuming fixed major and minor alleles (-doMaf 1; Korneliussen et al. 2014). Only reads with a mapping quality score above 20 (-minMapQ 20) and only bases with a quality score above 20 (-minQ 20) were used in this process. Furthermore, we only included only SNPs with a minor allele frequency of at least 0.01 (-minMaf 0.01) and that were represented in at least 50% of the included samples (-minInd 88). These SNPs were then used to calculate a covariance matrix (-doCov 1), which was used to generate eigenvalues and eigenvectors for principal components analysis (PCA). The percentage of total genetic variation explained by each principal component (PC) was calculated, and PCs explaining a relatively large proportion of genetic variation were plotted with ggplot2 (Wickham 2009). To view the substructure within groups from the initial PCA, subsequent PCAs were performed on samples from each group using the same methods described above. The rangewide PCA was also colored by the RAD sequencing libraries to check whether the library effect influenced the final results.

Admixture

To further assess population structure in Speckled Dace, we used the same parameters as above to generate a Beagle file with ANGSD 1.9 for admixture analysis. The Beagle

output file was then used as the input file for NGSadmix (Skotte et al. 2013). The parameter *K* (the number of clusters into which samples are partitioned by NGSadmix) was tested from 2 to 9, and each run had a minor allele frequency filter of 0.01. After population structure was initially characterized, we repeated the procedure as described above on subsets of samples in order to determine substructure within each group.

Pairwise F_{ST}

To quantify the genetic divergence among populations, we calculated genomewide pairwise values of the genetic differentiation index F_{ST} for population units identified by the analysis above and/or the sample collection locations. The folded site allele frequencies (SAFs) were estimated for each group with RealSFS in ANGSD 1.9. The SAF files for the pairwise locations were the input to estimate two-dimensional site frequency spectrum (SFS). The SFS files were then indexed by the SAF files to generate F_{ST} files. We then estimated the weighted genomewide F_{ST} values from the F_{ST} files with the stats function set in RealSFS.

Molecular phylogeny

To further investigate the relationships among different genetic lineages, a rangewide phylogenetic tree was generated using SVDQuartets (Chifman and Kubatko 2014, 2015) and IQ-TREE version 1.6.12 (Schrempf et al. 2019). Relict Dace *Relictus solitarius* and Tui Chub *Siphateles bicolor* were used as outgroups to root molecular phylogenies (Data S.1–S.3). For SVDQuartets, samples were pooled by the locations where they were collected (subregion 1 of Table S.1). If a significant genetic difference was shown between locations within a geographic region in PCA or admixture analysis, the region was labeled by the subregions if they divided into two tips in the phylogeny (subregion 2 of Table S.1).

We used ANGSD 1.9 to perform genotype calling, and we used the same parameters as mentioned above except that we generated a variant call format (VCF) file (-dovcf 1). We used BCFtools version 1.13 to prune the SNPs with r^2 values greater than 0.90 within each RAD contig (https://samtools.github.io/bcftools/bcftools.html). The pruned VCF file was transformed into NEXUS format by vcf2phylip (https://github.com/edgardomortiz/vcf2phylip). The pruned NEXUS file was analyzed by SVDQuartets within PAUP* version 4.0 (Pfenninger and Posada 2007). We used a multispecies coalescent model to construct the phylogeny with 1,000,000 random quartets and 100 bootstraps.

We used the same SNP alignment with IQ-TREE as we did with SVDQuartets. The optimal substitution model was determined with ModelFinder under the Bayesian information criterion (-m MFP selected TVM + F + R4; Kalyaanamoorthy et al. 2017), and 1,000 bootstraps were performed with ultrafast bootstrap approximation (-bb 1000; Hoang et al. 2017). The resulting consensus tree was visualized with ggtree (Yu et al. 2017). Individuals with a substantial degree of missing data (<10,000 contigs with mapped reads) were pruned from the consensus tree for presentation.

RESULTS

Sequencing, de novo restriction site-associated DNA assembly, and alignment

Demultiplexed sequence data are available from the Sequence Read Archive (National Center for Biotechnology Information) under BioProject PRJNA851170, and code for analyses is available at https://github.com/yingxins/speckled-dace/tree/Script. The final assembly contained 17,639 contigs, with a mean contig length of 456.20, a maximum length of 788, and a minimum length of 89 (Data S.1–S.3). After filtering individuals with sequencing and mapping quality, there were 175 individuals and 421,929 SNPs for rangewide analysis. The mean individual coverage (i.e., the average coverage across all of the contigs in one individual) was 7.69, with a maximum of 24.88, a minimum of 2.50, and an SD of 3.92 (Figure S.1 available in the Supplement in the online version of this article). For further analyses, there were 76 individuals and 108,746 SNPs for group 1; 67 individuals and 196,412 SNPs for group 2; and 32 individuals and 142,578 SNPs for group 3.

Rangewide genetic structure

Across all Speckled Dace samples rangewide, the first two PCs explained 16.7% of the total variance and divided our samples into three clusters, and the result was not influenced by the library effect (Figure 2A; Figure S.2). Because the percentage of genetic variation explained by PCA is highly influenced by the minor allele frequency cutoff (we selected 0.01 as a cutoff) and given that our samples covered several distinct lineages, 16.7% is a considerable amount of genetic variation to be explained. Group 1 (Figure 2A, upper right) consists of Speckled Dace populations from the Sacramento River, central California coast, Klamath River, Warner Basin, and Butte Lake. Group 2 (Figure 2A, upper left) is made up of Speckled Dace from the Amargosa River, Long Valley, Owens River basin, and

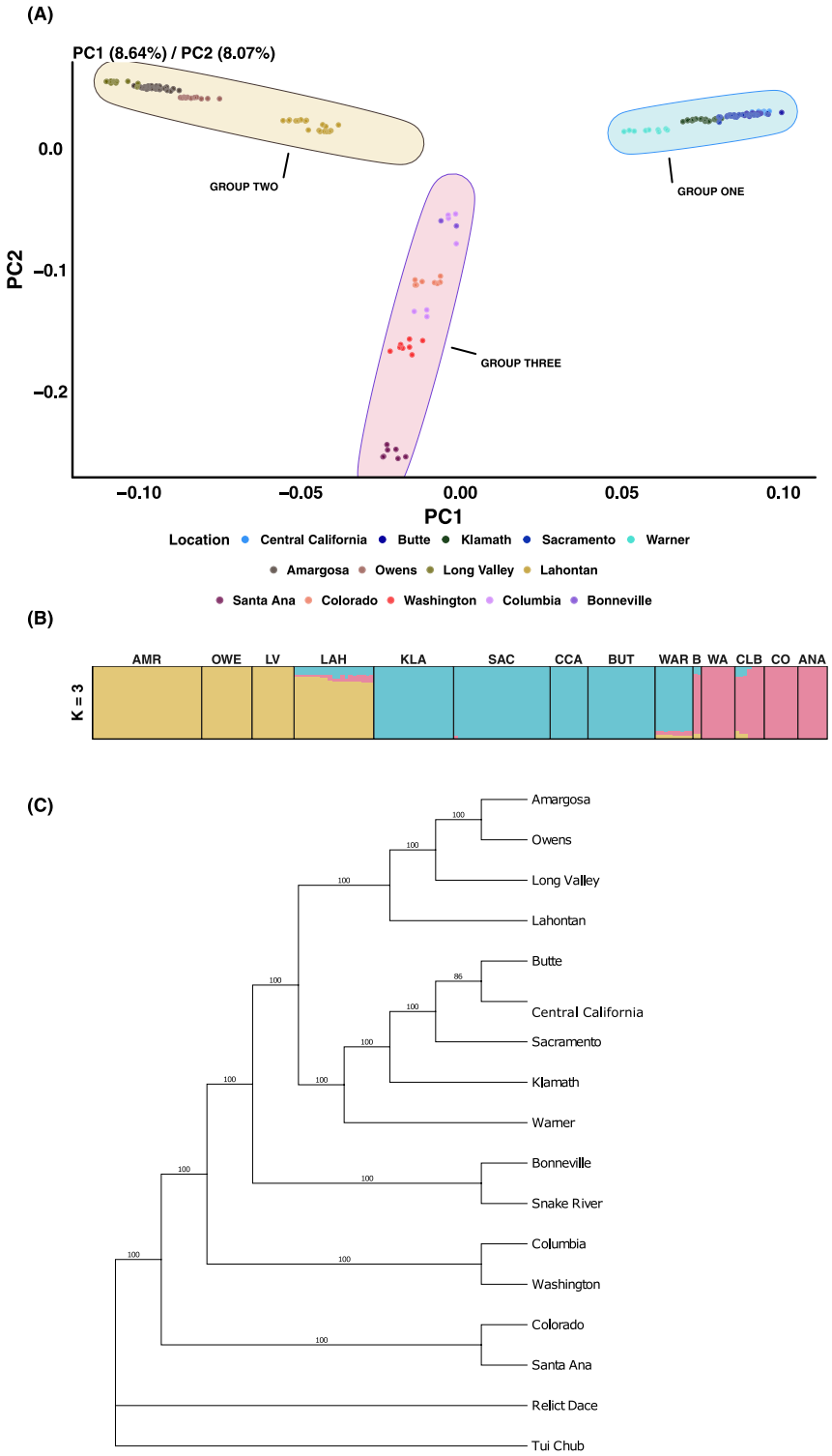

FIGURE 2 Rangewide Speckled Dace population structure. (A) Principal components analysis of all samples is depicted. Color represents the location from which the Speckled Dace were collected; 16.7% of the genetic variation is explained in total (principal component 1 [PC1] explains 8.6% of the variation, while PC2 explains 8.1% of the variation). Three groups are distinguishable: group 1 includes Speckled Dace from the Sacramento River (SAC), central California coast (CCA), Klamath River (KLA), Butte Lake (BUT), and Warner Basin (WAR); group 2 includes Speckled Dace from the Amargosa River (AMR), Owens River (OWE), Long Valley (LV), and Lahontan Basin (LAH); and group 3 includes Speckled Dace from the Santa Ana River (ANA), Washington coast (WA), Columbia River (CLB), Bonneville Basin (B), and Colorado River basin (CO). (B) Admixture analysis of all samples at a *K*-value of 3 (i.e., we assumed that the current populations were admixed by three populations in the past) is presented. (C) SVDQuartets results for all samples are shown. Groups 1 and 2 are monophyletic and are the sister groups of each other, while Santa Ana River Speckled Dace were clustered with Speckled Dace from the lower Colorado River basin and are the sister group of all other Speckled Dace included in this study.

Lahontan Basin. Group 3 (Figure 2A, lower middle) is composed of populations from the Santa Ana River as well as locations outside of California, such as the Bonneville Basin, Washington coast, Columbia River, and lower Colorado River. Speckled Dace from the four regions outside of California showed that the California populations we sampled are distinct from populations in the remainder of the SDC range.

We next used an admixture analysis of rangewide samples to complement our PCA. The admixture analysis was run with K ranging from 2 to 9 (Figure S.4). At a K-value of 3, members of each group in the admixture analysis comprised groups 1, 2, and 3 as indicated by PCA (Figure 2B). Furthermore, pairwise F_{ST} calculated between different populations varied from 0.16 (Speckled Dace from the Owens and Amargosa rivers) to 0.68 (Speckled Dace from the Amargosa and Santa Ana rivers; Table S.2). Taken together, these results revealed that the Speckled Dace groups in California have highly variable levels of genetic divergence and that taxonomic revision is warranted.

Our SVDQuartets rangewide phylogenetic analyses indicated that the Speckled Dace in California are mainly distributed into two monophyletic groups, with the exception of Santa Ana River Speckled Dace, which represent a distinct evolutionary lineage (Figure 2C). Similar to the results of PCA and admixture analysis, Speckled Dace from the Sacramento River, central California coast, Klamath River, Warner Basin, and Butte Lake belong to the same monophyletic group (group 1; bootstrap support = 100%). Lahontan Basin, Long Valley, Amargosa River, and Owens River Speckled Dace constitute another monophyletic group (group 2; bootstrap support = 100%), which is the sister group of group 1. Speckled Dace collected from the Santa Ana River are the sister lineage of Speckled Dace from the lower Colorado River drainages (bootstrap support = 100%). Speckled Dace from the Santa Ana River and lower Colorado River are the earliest-branching lineage in the species tree, followed by subsequent branching of (1) Speckled Dace from the Washington coast and the Columbia River, (2) Speckled Dace from the Snake River and Bonneville Basin, and (3) all other California Speckled Dace.

The phylogeny generated by IQ-TREE indicates three main lineages of California Speckled Dace (Figure S.5): (1) group 1 as previously defined; (2) group 2 as previously defined, with groups 1 and 2 being distinct lineages that are sister to each other; and (3) Speckled Dace from the Santa Ana River, which are sister to Speckled Dace from the lower Colorado River drainage (placement and monophyly bootstrap support = 100%), with this combined lineage being sister to groups 1 and 2 (bootstrap support = 100%). Speckled Dace from the Columbia River and Washington coast combined are the earliest-branching lineage in this phylogeny (placement and monophyly bootstrap support = 100%) and are sister to all other Speckled Dace. Individuals sequenced from the Snake River ($n = 4$) and Bonneville Basin ($n = 2$) were filtered out due to a low number of contigs with aligned reads. Overall, the genetic structure/divergence of Speckled Dace in California is hierarchical, with multiple levels of genetically distinct lineages (Figure 2).

Group 1: Klamath River, Central California Coast, Sacramento River, and Warner Basin

Group 1 includes Speckled Dace collected from the Klamath and Sacramento rivers and central California coast plus Speckled Dace collected from Butte Lake and Warner Basin. After our rangewide data showed that group 1 was distinct, we performed additional PCA and admixture analysis using only group 1 samples. For this PCA, the first three PCs explain the largest proportion of the genetic variation (Figure S.3). Principal component 1 splits the Warner Basin population from populations in the other regions; PC2 separates the Klamath River population from the Sacramento River populations (Figure 3A); and PC3 separates the central California coast populations from the Sacramento River basin populations (Pit River, Goose Lake, and Sacramento River; Figure 3B). The Butte Lake population clusters with Speckled Dace from the Sacramento River in all of the PCs, indicating genetic similarity. Admixture analysis supports the PCA (Figure 3C): all of the locations are separated in different K-values. More specifically, admixture analysis splits out the Warner Basin and central California coast populations when K is equal to 2; the Klamath River and Sacramento River populations are distinct when K is equal to 3. At a K of 4, the Butte Lake population is separated from the Sacramento River population. Pairwise F_{ST} analysis also supports the results from PCA and admixture analysis; the highest F_{ST} values are found between Warner Basin and the other locations (mean = 0.32) and between the central California coast and the other locations (mean = 0.28; Table S.2), whereas the F_{ST} value between the Klamath and Sacramento River populations is 0.18. The F_{ST} value between the Sacramento River and Butte Lake populations is only 0.084. The central California coast population has relatively low pairwise F_{ST} values with the Sacramento River and Butte Lake (0.19 and 0.23, respectively).

The rangewide SVDQuartets analysis (Figure 2C) and the IQ-TREE phylogeny (Figure S.5) are concordant with the above, placing Speckled Dace from the Klamath River, central California coast, Sacramento River, and Warner Basin into one clade (bootstrap support = 100%). The Warner Basin population diverges first within the group 1 clade in both the SVDQuartets species tree and the IQ-TREE phylogeny. Subsequent to the further phylogenetic

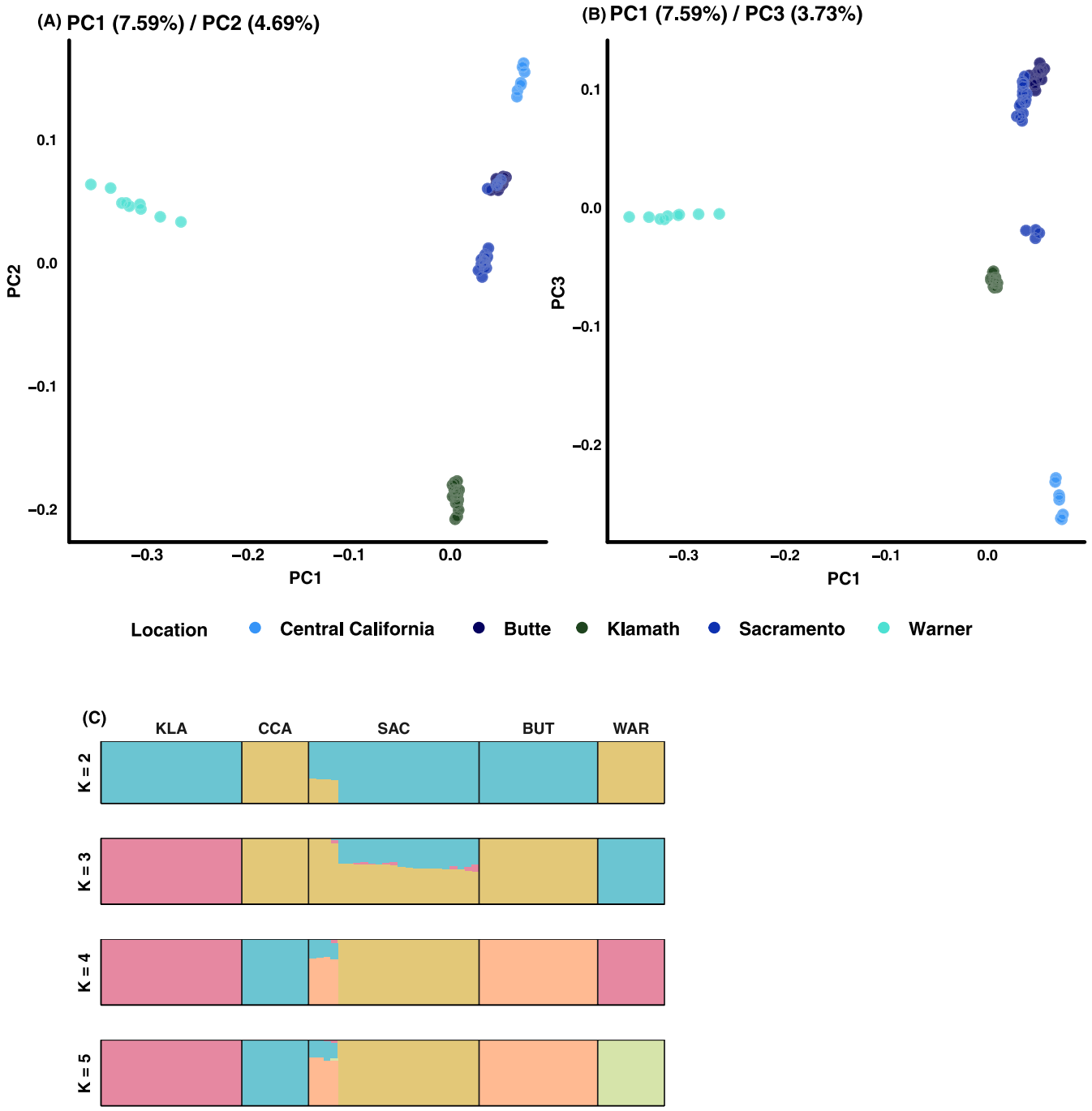

FIGURE 3 Population structure of Speckled Dace from the Sacramento River, central California coast, Klamath River, and Warner Basin: (A) principal components analysis of samples in group 1, with color representing the location of the Speckled Dace (12.3% of the genetic variation is explained by principal component 1 [PC1; 7.6%] and PC2 [4.7%]); (B) principal components analysis of samples in group 1 when genetic variation is explained by PC1 (7.6%) and PC3 (3.7%); and (C) admixture analysis of samples in group 1 for *K*-values of 2–5. The abbreviation labels represent locations as defined in Figure 2.

branching of the lineages, in both the SVDQuartets species tree and the IQ-TREE phylogeny, Speckled Dace from the Klamath River population separate out as distinct. A well-supported clade of Speckled Dace representing the Klamath River population is present in the IQ-TREE phylogeny. The remaining sampling locations—Butte Lake, central California coast, and Sacramento River—are not monophyletic with regard to sampling location when analyzed in a concatenated phylogenetic framework. Instead, these three geographic regions are mixed together into a single, well-supported group 1 subclade, which may be caused by lower genetic differentiation (Figure S.5).

Group 2: Death Valley and Lahontan Basin

Speckled Dace in group 2 include samples from three locations in the Death Valley region (Amargosa River, Owens River, and Long Valley) and four locations in the Lahontan Basin. To investigate the genetic structure within group 2, we performed PCA and admixture analysis on these samples. The first two PCs explain the largest proportion of the genetic variation (Figure S.3): Amargosa and Owens River populations are very close to each other in both PCs; both PC1 and PC2 split Lahontan Basin and Long Valley populations from Owens and Amargosa River populations (Figure 4A). Admixture analysis supports the PCA results. The Lahontan Basin population is split from all other Speckled Dace when K is equal to 2, and the Long Valley population is split from the Owens and Amargosa River populations when K is equal to 3. At K-values of 4 and 5, we observed the local substructure in the Amargosa River population, which is not discussed in this paper (Figure 4B; but see Mussmann et al. 2020). Although not as obvious as in group 1, the F_{ST} results support the PCA and admixture analysis. The F_{ST} value between Speckled Dace collected from the Owens and Amargosa rivers is 0.16, which is the lowest of all group 2 pairwise F_{ST} values. This is consistent with their close proximity in the PCA and differentiation at higher K-values using admixture analysis. The F_{ST} values between Long Valley and the Owens River and between Long Valley and the Amargosa River are 0.38 and 0.30, respectively, which agrees with their separation in the PCA and their early split in the admixture analysis (Table S.2). However, although the Lahontan Basin population is the first lineage to separate in the admixture analysis, it does not have the highest pairwise F_{ST} values; the F_{ST} value between the Amargosa River and Lahontan Basin is 0.33, which is higher than the F_{ST} values for the Lahontan Basin and the Owens River (0.25) and for the Lahontan Basin and Long Valley (0.26).

The rangewide SVDQuartets analysis is concordant with PCA and admixture analysis for group 2. The Lahontan Basin population, which split at a K-value of 2, is the sister group of all of the Death Valley populations. The result of IQ-TREE is also similar to the result of SVDQuartets but exhibits structuring between the Martis Creek, Humboldt River, and Walker River sampling locations from the Lahontan Basin (Figure S.5). The Owens and Amargosa rivers, which show little genetic divergence in the PCA and admixture analysis, are sister lineages in the SVDQuartets species tree and form a monophyletic clade in the IQ-TREE phylogeny. The position of the Speckled Dace from Long Valley in the SVDQuartets species tree and in the IQ-TREE phylogeny, as a sister lineage to Owens and Amargosa River Speckled Dace, is also supported by the admixture analysis, where Long Valley splits after the Lahontan Basin but before the Amargosa and Owens rivers (bootstrap support = 100%). Although PCA and pairwise F_{ST} indicate that the Long Valley population is a separate lineage from the Amargosa and Owens River populations, this incongruence could be due to overestimation of genetic divergence caused by genetic drift in a small population under long isolation.

Group 3: Santa Ana River

The only California Speckled Dace lineage in group 3 is from the Santa Ana River, which clusters with non-California Speckled Dace (Table S.1). To investigate the distinctiveness of the Santa Ana River population, we performed PCA and admixture analysis and estimated pairwise F_{ST} for sample collections in group 3. The PCA and admixture analysis show that the Santa Ana River population exhibits striking genetic differences from the non-California Speckled Dace in group 3. In the PCA for group 3, the largest proportion of the genetic variation is explained by PC1 and PC2 (Figure S.3). Both PC1 and PC2 split the Santa Ana River population from all other Speckled Dace lineages. Admixture analysis for group 3 was run for K-values from 2 to 5, and the Santa Ana River population split from the other locations (lower Colorado River basin, Bonneville Basin, Columbia River basin, and Washington coast) at K-values of 3–5 (Figure 5B). In addition, the Santa Ana River population has high pairwise F_{ST} values with all other California and non-California Speckled Dace (Table S.2).

The rangewide SVDQuartets analysis places the Santa Ana River population as sister to the population from the lower Colorado River basin (bootstrap support = 100%). IQ-TREE also places the Santa Ana River population and lower Colorado River basin population as sister lineages (bootstrap support = 100%) and indicates that each of these two populations are monophyletic as well (bootstrap support = 100%). The results of admixture analysis and PCA support the genetic affinity between Speckled Dace from the Santa Ana River population and those from the Colorado River basin. The lower Colorado River basin population clusters with the Santa Ana River population when K is equal to 2, and the lower Colorado River basin population is the closest lineage to the Santa Ana River population in the PCA. However, due to the limited number of samples of non-California Speckled Dace, we did not further explore the relationship between non-California Speckled Dace and Santa Ana River Speckled Dace.

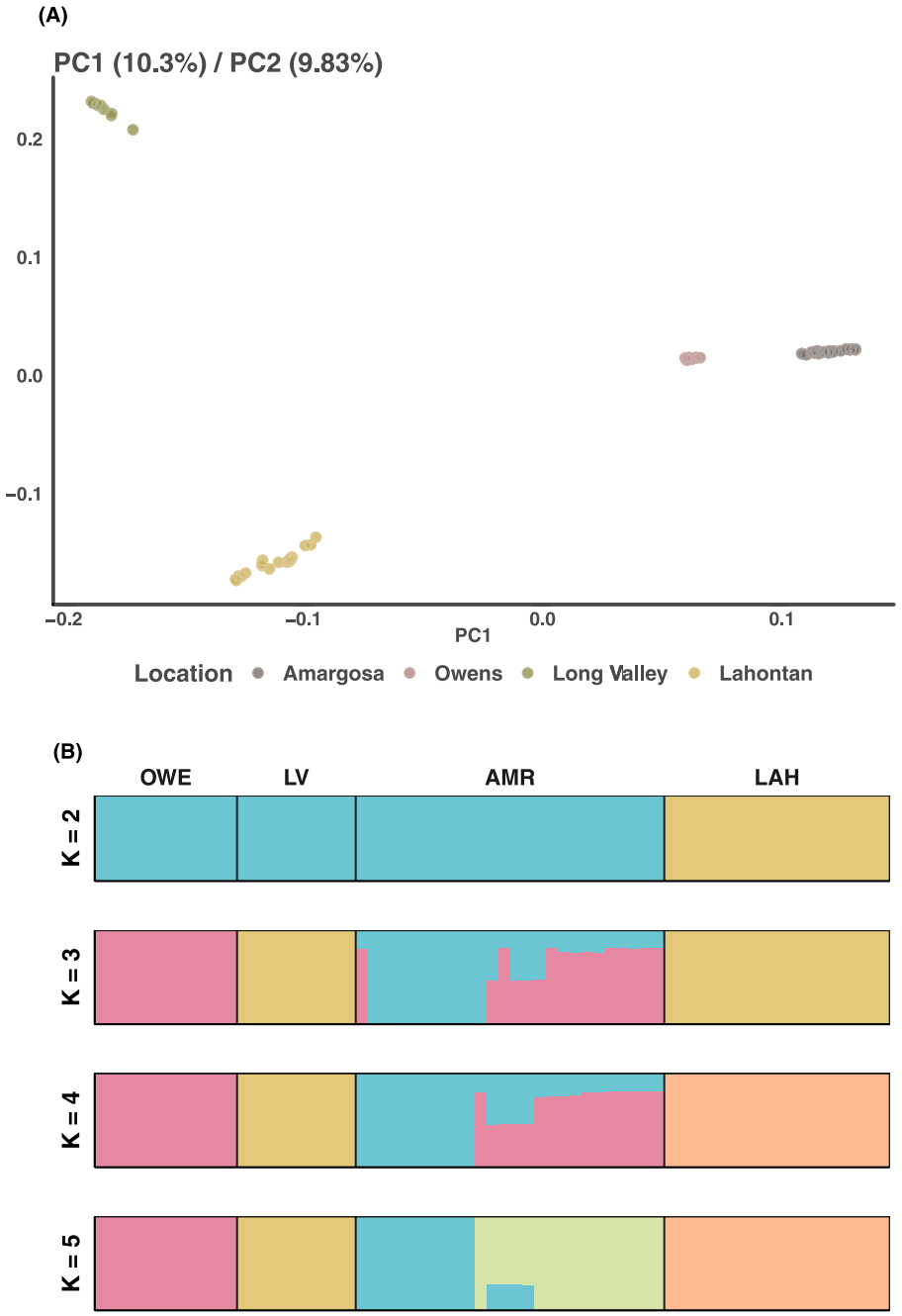

FIGURE 4 Population structure of Speckled Dace from Death Valley and the Lahontan Basin: (A) principal components analysis of samples in group 2, with color representing the location where Speckled Dace were collected (20.1% of the genetic variation is explained by principal component 1 [PC1; 10.3%] and PC2 [9.8%]); and (B) admixture analysis of samples in group 2 for *K*-values of 2–5. The abbreviation labels represent the locations as defined in Figure 2.

DISCUSSION

The Speckled Dace has multiple lineages

Our genomic data analyses show that the Speckled Dace has hierarchical, genetically distinct lineages. In other words, they are genetically divergent at different levels as opposed to having relatively uniform relatedness as might be expected for a single widespread population. Our genetic analysis of California populations divides them into three lineages with sublineages. These lineages and sublineages coincide with zoogeographic regions that are largely isolated from one another and that contain other endemic fishes, suggesting long isolation (Moyle 2002). If allopatry sustains the genetic divergence for a sufficient duration, phenotypic and genotypic differences will

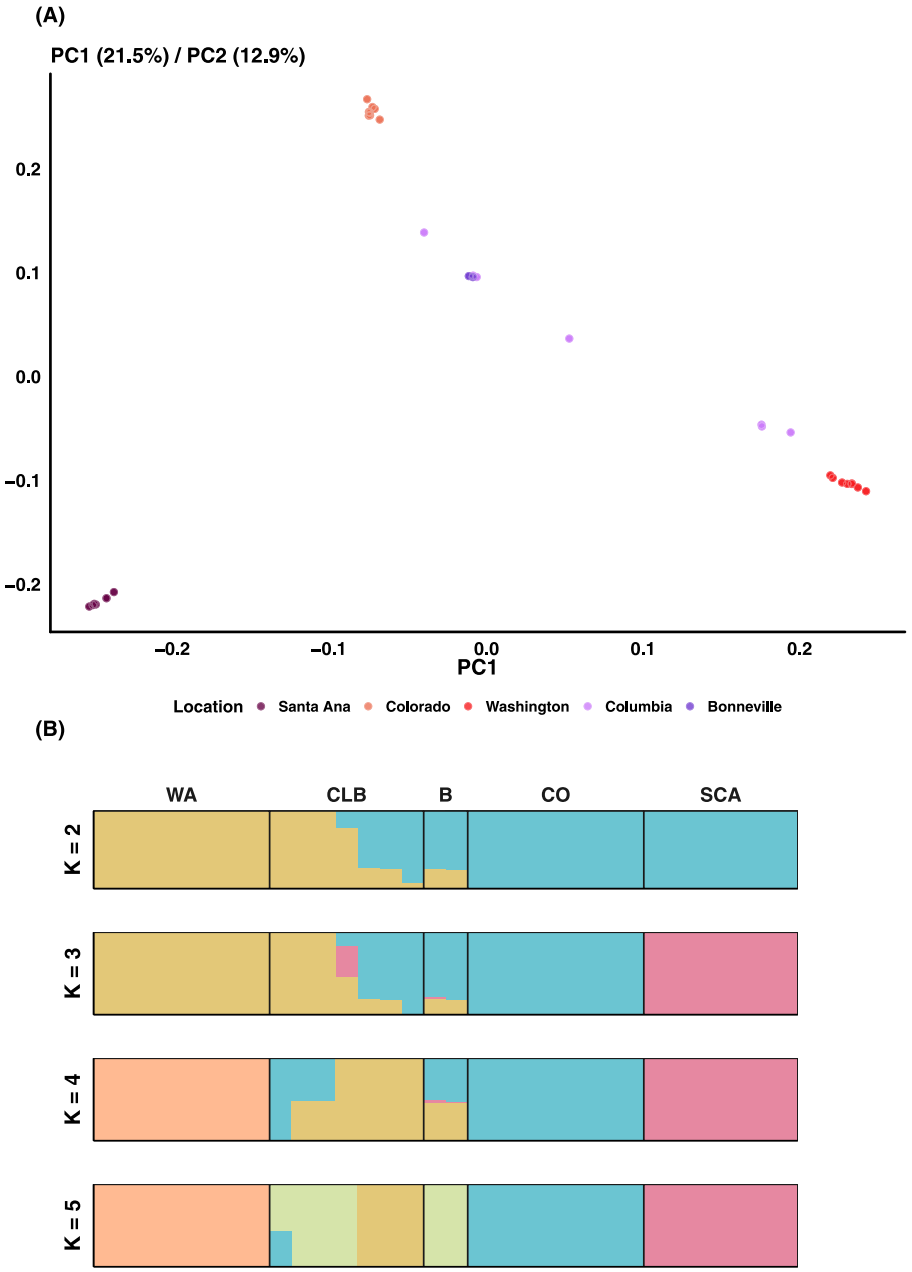

FIGURE 5 Population structure of Speckled Dace from the Santa Ana River basin and from locations outside of California: (A) principal components analysis of samples in group 3, with color representing the location where Speckled Dace were collected (34.4% of the genetic variation is explained in total; principal component 1 [PC1] explains 21.5% of the variation, while PC2 explains 12.9% of the variation); and (B) admixture analysis of samples in group 3 for *K*-values of 2–5. The abbreviations correspond to locations defined in Figure 2. Both principal components analysis and admixture analysis support Speckled Dace from the Santa Ana River as distinct from non-California Speckled Dace, although they do have a distant relationship to Speckled Dace from the lower Colorado River basin.

presumably accumulate. Hence, it is unlikely that (1) the split lineages can merge into a single lineage again and (2) genetic differences will be lost to hybridization upon secondary contact. Thus, we assume that hybrid individuals from two distinct lineages would be poorly adapted to the ecological system in which they occur and that they would have reduced fitness (Coyne and Orr 2004). We therefore find it appropriate to label geographically isolated lineages with large genomic differences as distinct lineages and to label geographically isolated lineages with less genomic differentiation as sublineages (Freudenstein et al. 2016). These lineages and sublineages will be a precondition of the formal species delimitation of taxa within the SDC. A more comprehensive definition of species in the SDC will be presented in a separate paper that formally describes the species and subspecies in California.

Speckled Dace from the Klamath River, Sacramento River, Central California Coast, and Warner Basin are a genetically distinct lineage

In all of the analyses, Speckled Dace from the Klamath River, Sacramento River, central California coast, and Warner Basin are a monophyletic lineage (group 1). These dace have relatively low F_{ST} values within the lineages compared to F_{ST} values between them and other populations. For example, F_{ST} values between the Sacramento River population and the Klamath River and Warner Basin populations are 0.17 and 0.29, respectively, but the F_{ST} values between the Sacramento River population and Speckled Dace from the Amargosa River, Owens River, and Long Valley (Death Valley populations) are 0.56, 0.51, and 0.44, respectively. The Klamath and Sacramento River populations have less genetic divergence from each other than either does from the Warner Basin population, but the geographical basins in which each occurs are all well-defined and contain other endemic fishes. This means that the isolation of these three basins has been long enough for diversification, although the Sacramento and Klamath River lineages have closer ties to each other than either has to the Warner Basin lineage. The Klamath River has not drained south into the ancestral Sacramento River system since the end of the Pliocene (~3 mya). Major geologic activity, especially vulcanism, has occurred almost continuously in the northern Sacramento River and southern Klamath River basins (Colman et al. 2004). It is probable that the vulcanism led to repeated drainage captures of headwater streams and their fishes, which allowed for intermittent gene flow. Interbasin connectivity was presumably less frequent between the Klamath/Sacramento River basin streams and those in the Warner Basin. The uplift of the Warner Range, starting around 3 mya, resulted in the permanent separation of the basins by the late Pleistocene (1.0–0.1 mya; Egger and Miller 2011).

The large geographic extent of the Klamath and Sacramento River drainages has resulted in some geographic population structure in each basin, creating genetically distinct population segments that need further investigation for determination of taxonomic status (Oakey et al. 2004). For example, Speckled Dace from the central California coast (San Luis Obispo Creek, Santa Maria River, and Monterey Bay drainages) show enough genetic differentiation that this group of populations could be recognized as a genetically distinct sublineage within dace populations from the Sacramento River drainage.

Speckled Dace from Butte Lake are an introduced population

Butte Lake is located in Lassen Volcanic National Park and historically drained into the Lahontan Basin, so Speckled Dace from Butte Lake were assumed to be genetically related to the Speckled Dace in the Lahontan Basin. However, Speckled Dace from Butte Lake have much greater similarity to Speckled Dace from the central California coast and the Sacramento River than to Lahontan Basin Speckled Dace in all analyses. Therefore, we classify Speckled Dace from Butte Lake as part of the central California coast or Sacramento River population and hypothesize that the population most likely represents a bait bucket introduction. Butte Lake drains northward from Mount Lassen through Butte Creek (which also has Speckled Dace) and may have been connected at one time to the Eagle Lake watershed in the Lahontan Basin, although frequent lava flows have obscured drainage patterns. The three other fishes present in Butte Lake—Tahoe Sucker *Catostomus tahoensis*, Lahontan Redside *Richardsonius egregius*, and Tui Chub—are Lahontan Basin fishes, lending credence to the bait bucket hypothesis.

Speckled Dace from Death Valley and the Lahontan Basin are a single lineage

In the rangewide PCA, admixture analysis, and phylogenetic analyses, Speckled Dace from Death Valley (Owens River, Amargosa River, and Long Valley) and the Lahontan Basin are most closely related to each other (group 2). The two geographic regions—Death Valley and the Lahontan Basin—are isolated geological basins, and Speckled Dace reflect this division in the analyses of group 2 (Figure 4B). Similar to Speckled Dace from the Sacramento and Klamath rivers, we also consider Speckled Dace from Death Valley and the Lahontan Basin to be a single lineage composed of two distinct sublineages with a common ancestor.

Within the Death Valley lineages, the Amargosa and Owens River populations only show small genetic differences, a finding that is consistent with the work of Mussmann et al. (2020). Smith et al. (2017) found that Speckled Dace from the Amargosa River shared haplotypes with dace from the Owens Valley. Dace from Oasis Valley, Nevada (the headwaters of the Amargosa River), and from Ash Meadows (Bradford Spring) are sister lineages of Speckled Dace from the Owens River. Unlike the situation for Speckled Dace from the Klamath and Sacramento rivers, the Owens and Amargosa River watersheds are internal drainages that were connected via a

chain of large lakes during extended wet periods in the late Pleistocene. Given the results of our analyses and their relatively recent geographic separation and isolation, we place Speckled Dace from the Amargosa and Owens rivers as one sublineage. Although Speckled Dace from Long Valley are in the same watershed as the Owens Valley, dace from Long Valley are genetically distinct from dace in the Owens and Amargosa rivers. Mussmann et al. (2020) observed a similar pattern in the phylogeny and F_{ST} estimates. This is probably the result of genetic drift due to isolation of small dace populations in small streams flowing into the Long Valley Caldera. Climatic shifts and vulcanism in the southern Mono Lake basin subsequently isolated the Owens River basin from the Lahontan Basin. Here, we treat the Long Valley population as a sublineage under Lahontan Basin Speckled Dace.

Speckled Dace from the Walker River, Humboldt River, eastern Sierra Nevada streams, and Death Valley system streams are one lineage: the Lahontan Speckled Dace *Rhinichthys osculus robustus*, which is a widely recognized taxon (Rutter 1904; Deacon and Williams 1984; Moyle 2002). Although Lahontan Basin Speckled Dace split at a *K*-value of 2 in the admixture analysis for group 2, the F_{ST} values between Lahontan Basin and Owens River Speckled Dace and between Lahontan Basin and Long Valley Speckled Dace are somewhat small (0.25 and 0.26, respectively) and are even lower than the F_{ST} value between Long Valley and Owens River Speckled Dace.

The historic habitat of Long Valley Speckled Dace was a series of hot spring outflows and connected marshes in the remnants of the Long Valley Caldera. The caldera was created by vulcanism approximately 767,000 years ago and then filled with water and sediment (Hildreth and Fierstein 2016). The dace (and other fishes) presumably colonized these habitats in the late Pleistocene, when the Lahontan Basin was connected to the Owens River basin via outlets of large pluvial lakes. When the climate changed, the lakes dried up. Complex events then created the Owens River gorge, with high waterfalls, draining the caldera lake and isolating the fish that became Long Valley Speckled Dace about 100,000 years ago (Hildreth and Fierstein 2016).

Low F_{ST} values suggest that either (1) a hybridization event took place between Speckled Dace from the Lahontan Basin and Owens River, creating Long Valley Speckled Dace; or (2) Lahontan Basin and Long Valley Speckled Dace share their early evolutionary history and were later separated by geologic change. The presence of the Owens Tui Chub *S. bicolor snyderi* and Owens Sucker *Catostomus fumeiventris* in Long Valley supports the second hypothesis because the closest relatives of both taxa are in the Lahontan Basin (Moyle 2002). Therefore, we consider Lahontan Speckled Dace to have given rise to two genetically distinct lineages in the Death Valley region: Speckled Dace in Death Valley (Amargosa and Owens River systems) and Speckled Dace in Long Valley.

Speckled Dace from the Santa Ana River are a distinct lineage

Our rangewide analyses revealed that Speckled Dace from the Santa Ana River basin are strikingly different from all other populations of Speckled Dace in California (group 3). Speckled Dace in the Santa Ana River share more genetic similarities with Speckled Dace from the lower Colorado River basin, Bonneville Basin, Washington coast, and the Columbia River than with other dace lineages in California. Due to the small number of samples, the genetic diversity within non-California basins is not discussed in this paper. The evolutionary history of Speckled Dace from the Santa Ana River can be linked most closely with Speckled Dace from the lower Colorado River basin because they did not split from each other in the admixture analysis with all the samples at *K*-values of 3–8 (Figure S.4). According to Smith et al. (2017), Speckled Dace collected from the Colorado River basin and those collected from the Santa Ana River basin are sister lineages in the Colorado group, with relatively weak bootstrapping support in the mtDNA phylogeny. Regarding pairwise F_{ST} values, we find that Santa Ana River basin Speckled Dace have high genetic divergence from both California and non-California Speckled Dace. The lower bootstrapping reported by Smith et al. (2017) was likely caused by high genetic divergence and relatively low diversity in mtDNA.

In our study, we clarify the genetic distinctness of Speckled Dace from the Santa Ana River basin. All analyses show that these dace have remarkably large genetic differences from other populations (Table S.2). Due to their unique genetic structure, Santa Ana River Speckled Dace are clearly a distinct lineage. This same basic conclusion was reached by Cornelius (1969), who conducted a detailed study of the morphometrics and meristics of Santa Ana River basin Speckled Dace as well as dace from neighboring streams (Sacramento River basin), the Virgin River (lower Colorado River basin), and Lake Tahoe (Lahontan Basin). His study was the first to link the origins of Speckled Dace from the Santa Ana River basin to the lower Colorado River basin. Details of how this connection occurred can be found in McClay and Bonora (2001), Axen and Fletcher (2010), and Dorsey and Langenheim (2015).

CONCLUSIONS AND CONSERVATION IMPLICATIONS

If we view the SDC as a single lineage, it is a species that does not merit special consideration for conservation because of its wide distribution and large population size. However, our genetic analyses show that Speckled Dace in California have a hierarchical order of divergence and that the different levels of genetic distinctness divide California Speckled Dace into multiple lineages. More specifically, our genetic analyses place all California populations into three major genetic lineages: (1) Speckled Dace from the Sacramento River, central California coast, Klamath River, and Warner Basin; (2) Speckled Dace from Death Valley, Long Valley, and the Lahontan Basin; and (3) Speckled Dace from the Santa Ana River and nearby rivers. Each distinct population within the three lineages can represent a sublineage under the main lineage. Each of these lineages needs further study to locate populations within them that need special protection; indeed, the naming of a lineage as a species or subspecies can improve its protection.

In this study, the populations across geographical regions are genetically divergent at different levels, depending on the time and degree of isolation from other populations of Speckled Dace. However, populations in different geographical regions face different environmental threats. For example, the Amargosa River region in Death Valley is one of the hottest and driest places in North America, where fish depend on springs that draw ancient water from underground aquifers; pumping water from these aquifers threatens to deplete this small flow (Robbins 2017; Belcher et al. 2019). In the Death Valley region, the Ash Meadows Speckled Dace was listed as endangered in 1984, which is probably the main reason that it still exists. The Center for Biological Diversity has filed a petition (2020) to have endangered status applied to all dace populations in the Death Valley region based on Mussmann et al. (2020).

In the Los Angeles region, Santa Ana River Speckled Dace persist despite urbanization, which has eliminated much of their habitat throughout the Santa Ana, San Gabriel, and Los Angeles River systems and has altered much of what is left. The dace presently inhabit small, isolated streams and creeks in the Santa Ana and San Gabriel River watersheds, where they are vulnerable to the effects of fire, droughts, and floods (SAWPA 2004; Nunziata et al. 2013). A petition to list the Santa Ana River Speckled Dace as threatened under the federal Endangered Species Act in 1996 was rejected, largely due to the lack of a formal species description (USFWS 1996). The genetic and evolutionary distinctiveness of Santa Ana River Speckled Dace should no longer be an issue; therefore, listing is justified.

These taxonomic issues will be further explored in a paper devoted solely to the taxonomy of the SDC in California. We can now combine our knowledge of genetic divergence with that of ecosystem status and characteristics to design distinct conservation management and policy strategies for different populations of Speckled Dace.

The focus of this paper is Speckled Dace in California, so how our findings relate to Speckled Dace outside of California is not discussed. However, it seems likely that there are non-California lineages that can also be designated (or redesignated) as species or subspecies when genomic methods are applied with careful sampling. Further genomic research in other zoogeographic areas will undoubtedly reveal heretofore unrecognized genetic structure.

ACKNOWLEDGMENTS

We thank Amber Manfree for making the maps (Figure 1). The following people provided fish samples for this study: Scot Lucas (Klamath River), Mollie Ogaz (Sacramento River system), Nick Buckmaster and Steve Parmenter (Owens Valley and Lahontan Basin), Kevin Guadalupe and Gregory Munson (Virgin River, Oasis Valley, and Amargosa River), Hal Fairfield (Death Valley region), Jennifer Pareti (Santa Ana River), and Claire Ingel (southern California).

CONFLICT OF INTEREST

There is no conflict of interest declared in this article.

ETHICS STATEMENT

None of the fin clips from the wild fish were collected by the authors. Fin clips were donated or collected by individuals listed in the Acknowledgments.

DATA AVAILABILITY STATEMENT

All data are fully available without restriction. Demultiplexed sequence data is available on from the NCBI Sequence Read Archive under BioProject PRJNA851170 and code for analyses is available at https://github.com/yingxins/speckled-dace/tree/Script.

ORCID

Yingxin Su https://orcid.org/0000-0002-5536-6996
Matthew A. Campbell https://orcid.org/0000-0002-5826-0329
Jason Baumsteiger https://orcid.org/0000-0001-7921-5722

REFERENCES

Ali, O. A., O'Rourke, S. M., Amish, S. J., Meek, M. H., Luikart, G., Jeffres, C., & Miller, M. R. (2016). RAD capture (Rapture): Flexible and efficient sequence-based genotyping. *Genetics*, *202*(2), 389–400.

Ardren, W., Baumsteiger, J., & Allen, C. (2010). Genetic analysis and uncertain taxonomic status of threatened Foskett Spring Speckled Dace. *Conservation Genetics*, *11*, 1299–1315.

Axen, G., & Fletcher, J. (2010). Late Miocene–Pleistocene extensional faulting, Northern Gulf of California, Mexico and Salton Trough, California. *International Geology Review*, *40*(3), 217–244.

Baumsteiger, J., Moyle, P. B., Aguilar, A., O'Rourke, S. M., & Miller, M. R. (2017). Genomics clarifies taxonomic boundaries in a difficult species complex. *PLOS ONE*, *12*(12), Article e0189417.

Belcher, W. R., Sweetkind, D. S., Hopkins, C. B., & Poff, M. E. (2019). *Hydrogeology of Lower Amargosa Valley and Groundwater discharge to the Amargosa Wild and Scenic River, Inyo and San Bernardino counties, California, and adjacent areas in Nye and Clark counties, Nevada* (Report 2018-5151). U.S. Geological Survey.

Billman, E. J., Lee, J. B., Young, D. O., McKell, M. D., Evans, R. P., & Shiozowa, D. K. (2010). Phylogenetic divergence in a desert fish: Differentiation of Speckled Dace within the Bonneville, Lahontan, and upper Snake River basins. *Western North American Naturalist*, *70*(1), 39–47.

Chifman, J., & Kubatko, L. (2014). Quartet inference from SNP data under the coalescent model. *Bioinformatics*, *30*(23), 3317–3324.

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

Colman, S. M., Platt Bradbury, J., & Rosenbaum, J. G. (2004). Paleolimnology and paleoclimate studies in Upper Klamath Lake, Oregon. *Journal of Paleolimnology*, *31*(2), 129–138.

Cornelius, R. H. (1969). The systematics and zoogeography of *Rhinichthys osculus* (Girard) in southern California [Master's thesis, California State University Fullerton].

Coyne, J. A., & Orr, H. A. (2004). *Speciation*. Sinauer Associates.

Deacon, J. E., & Williams, J. E. (1984). Annotated list of the fishes of Nevada. *Proceedings of the Biological Society of Washington*, *97*(1), 113–118.

Dorsey, R. J., & Langenheim, V. E. (2015). Crustal-scale tilting of the central Salton block, southern California. *Geosphere*, *11*(5), 1365–1383.

Egger, A. E., & Miller, E. L. (2011). Evolution of the northwestern margin of the basin and range: The geology and extensional history of the Warner Range and environs, northeastern California. *Geosphere*, *7*(3), 756–773.

Evermann, B. W., & Meek, S. E. (1896). A report upon salmon investigations in the Columbia River basin and elsewhere on the Pacific coast in 1896. *U.S. Fish Commission Bulletin*, *17*, 15–84.

Freudenstein, J. V., Broe, M. B., Folk, R. A., & Sinn, B. T. (2016). Biodiversity and the species concept—lineages are not enough. *Systematic Biology*, *66*(4), 644–656.

Hildreth, W., & Fierstein, J. (2016). *Long Valley Caldera Lake and reincision of Owens River Gorge* (Scientific Investigations Report 2016-5120). U.S. Geological Survey.

Hoang, D. T., Chernomor, O., von Haeseler, A., Minh, B. Q., & Vinh, L. S. (2017). UFBoot2: Improving the ultrafast bootstrap approximation. *Molecular Biology and Evolution*, *35*(2), 518–522.

Hoekzema, K., & Sidlauskas, B. L. (2014). Molecular phylogenetics and microsatellite analysis reveal cryptic species of Speckled Dace (Cyprinidae: *Rhinichthys osculus*) in Oregon's Great Basin. *Molecular Phylogenetics and Evolution*, *77*, 238–250.

Jordan, D. S., & Evermann, B. W. (1896). *The fishes of North and Middle America* (Vol. 47). U.S. National Museum Bulletin.

Kalyaanamoorthy, S., Minh, B. Q., Wong, T. K. F., von Haeseler, A., & Jermiin, L. S. (2017). ModelFinder: Fast model selection for accurate phylogenetic estimates. *Nature Methods*, *14*(6), 587–589.

Korneliussen, T. S., Albrechtsen, A., & Nielsen, R. (2014). ANGSD: Analysis of next generation sequencing data. *BMC Bioinformatics*, *15*, Article 356.

Leidy, R., & Moyle, P. (2021). Keeping up with the status of freshwater fishes: A California (USA) perspective. *Conservation Science and Practice*, *3*(8), Article e474.

Li, H., & Durbin, R. (2009). Fast and accurate short read alignment with Burrows–Wheeler transform. *Bioinformatics*, *25*(14), 1754–1760.

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

McClay, K., & Bonora, M. (2001). Analog models of restraining stepovers in strike-slip fault systems. *AAPG Bulletin*, *85*(2), 233–260.

Moyle, P. B. (2002). *Inland fishes of California*. University of California Press.

Mussmann, S. M., Douglas, M. R., Oakey, D. D., & Douglas, M. E. (2020). Defining relictual biodiversity: Conservation units in Speckled Dace (Leuciscidae: *Rhinichthys osculus*) of the greater Death Valley ecosystem. *Ecology and Evolution*, *10*(19), 10798–10817.

Nunziata, S. O., Lance, S. L., Jones, K. L., Nerkowski, S. A., & Metcalf, A. E. (2013). Development and characterization of twenty-three microsatellite markers for the freshwater minnow Santa Ana Speckled Dace (*Rhinichthys osculus*, Cyprinidae) using paired-end Illumina shotgun sequencing. *Conservation Genetics Resources*, *5*(1), 145–148.

Oakey, D. D., Douglas, M. E., & Douglas, M. R. (2004). Small fish in a large landscape: Diversification of *Rhinichthys osculus* (Cyprinidae) in western North America. *Copeia*, *2004*(2), 207–221.

Page, L. M., & Burr, B. M. (2011). *Peterson field guide to freshwater fishes of North America north of Mexico* (2nd ed.). Houghton Mifflin Harcourt.

Pfenninger, M., & Posada, D. (2007). Phylogeographic history of the land snail *Candidula unifasciata* (Helicellinae, Stylommatophora): Fragmentation, corridor migration, and secondary contact. *Evolution: International Journal of Organic Evolution*, *56*(9), 1776–1788.

Pfrender, M. E., Hicks, J., & Lynch, M. (2004). Biogeographic patterns and current distribution of molecular-genetic variation among populations of Speckled Dace, *Rhinichthys osculus* (Girard). *Molecular Phylogenetics and Evolution*, *30*(3), 490–502.

Robbins, J. (2017, December 18). The Amargosa River defies the desert. *The New York Times*. https://www.nytimes.com/2017/12/18/science/amargosa-death-valley-mojave-pupfish.html

Rutter, C. (1904). Notes on fishes from streams and lakes of northeastern California not tributary to the Sacramento Basin. *U.S. Fish Commission Bulletin*, *22*, 145–148.

Sada, D. W., Britten, H. B., & Brussard, P. F. (1995). Desert aquatic ecosystems and the genetic and morphological diversity of Death

Valley system Speckled Dace. In J. L. Nielsen (Ed.), *Evolution and the aquatic ecosystem: Defining unique units in population conservation* (Symposium 17, pp. 350–359). American Fisheries Society.

SAWPA (Santa Ana Watershed Project Authority). (2004). *Old, Grand Prix, and Padua Fires (October, 2003): Burn impacts to water systems and resources*. SAWPA.

Schrempf, D., Minh, B. Q., von Haeseler, A., & Kosiol, C. (2019). Polymorphism-aware species trees with advanced mutation models, bootstrap, and rate heterogeneity. *Molecular Biology and Evolution*, *36*(6), 1294–1301.

Skotte, L., Korneliussen, T. S., & Albrechtsen, A. (2013). Estimating individual admixture proportions from next generation sequencing data. *Genetics*, *195*(3), 693–702.

Smith, G., Chow, J., Unmack, P., Markle, D., & Dowling, T. (2017). *Evolution of the Rhinichthys osculus complex (Teleostei: Cyprinidae) in western North America* (Miscellaneous Publication 204, Vol. 2). University of Michigan, Museum of Zoology.

Spencer, J. E., Smith, G. R., & Dowling, T. E. (2008). Middle to late Cenozoic geology, hydrography, and fish evolution in the American Southwest. In M. C. Reheis, R. Hershler, & D. M. Miller (Eds.), *Late Cenozoic drainage history of the southwestern Great Basin and lower Colorado River region: Geologic and biotic perspectives* (Special Paper 439, pp. 279–299). Geological Society of America.

USFWS (U.S. Fish and Wildlife Service). (1996). Endangered and threatened wildlife and plants: 90-day finding on a petition to list the Santa Ana Speckled Dace, Santa Ana Sucker, and the Shay Creek Threespine Stickleback as endangered. *Federal Register*, *61*(132), 36021–36023.

Wickham, H. (2009). *Ggplot2: Elegant graphics for data analysis*. Springer.

Yu, G., Smith, D. K., Zhu, H., Guan, Y., & Lam, T. T.-Y. (2017). ggtree: An R package for visualization and annotation of phylogenetic trees with their covariates and other associated data. *Methods in Ecology and Evolution*, *8*(1), 28–36.

SUPPORTING INFORMATION

Additional supporting information can be found online in the Supporting Information section at the end of this article.