12-2018

Diversification Across a Dynamic Landscape: Phylogeography and Riverscape Genetics of Speckled Dace (Rhinichthys osculus) in Western North America

Steven Michael Mussmann
University of Arkansas, Fayetteville

Diversification Across a Dynamic Landscape: Phylogeography and Riverscape Genetics of Speckled Dace (*Rhinichthys osculus*) in Western North America

A dissertation submitted in partial fulfillment
of the requirements for the degree of
Doctor of Philosophy in Biological Sciences

by

Steven Mussmann
Iowa State University
Bachelor of Science in Anthropology, 2005
Colorado State University
Bachelor of Science in Biology, 2008
University of Illinois at Urbana-Champaign
Master of Science in Biology, 2011

December 2018
University of Arkansas

The dissertation is approved for recommendation to the Graduate Council.

Michael E. Douglas, Ph.D.
Dissertation Director

Marlis R. Douglas, Ph.D.
Dissertation Co-Director

David P. Philipp, Ph.D.
Committee Member

Andrew J. Alverson, Ph.D.
Committee Member

Abstract

Evolution occurs at various spatial and temporal scales. For example, speciation may occur in historic time, whereas localized adaptation is more contemporary. Each is required to identify and manage biodiversity. However, the relative abundance of Speckled Dace (*Rhinichthys osculus*), a small cyprinid fish in western North America (WNA) and the study species for this dissertation, establishes it an atypical conservation target, particularly when contrasted with the profusion of narrowly endemic forms it displays. Yet, the juxtaposition of ubiquity versus endemism provides an ideal model against which to test hypotheses regarding the geomorphic evolution of WNA. More specifically, it also allows the evolutionary history of Speckled Dace to be contrasted at multiple spatial and temporal scales, and interpreted in the context of contemporary anthropogenic pressures and climatic uncertainty.

Chapter II dissects the broad distribution of Speckled Dace and quantifies how its evolution has been driven by hybridization/ introgression. Chapter III narrows the geographic focus by interpreting Speckled Dace distribution within two markedly different watersheds: The Colorado River and the Great Basin. The former is a broad riverine habitat whereas the latter is an endorheic basin. Two biogeographic models compare and contrast the tempo and mode of evolution within these geologically disparate habitats.

Chapter IV employs a molecular clock to determine origin of Speckled Dace lineages in Death Valley (CA/NV), and to contrast these against estimates for a second endemic species, Devil's Hole Pupfish (*Cyprinodon diabolis*). While palaeohydrology served to diversify *Rhinichthys*, its among-population connectivity occurred contemporaneously. These data also provide guidance for assessing the origin of the Devil's Hole Pupfish, a topic of considerable contention.

The final two chapters present bioinformatic software that facilitates the analysis of single-nucleotide-polymorphism (SNP) DNA data (used herein). Chapter V describes COMP-D, a program designed to assess introgression among lineages, whereas Chapter VI presents programmatic modifications to BAYESASS that allow migration to be quantified from SNP datasets.

These five studies provide an in-depth understanding of contemporary and historical processes that shape aquatic biodiversity in environments prone to anthropogenic disturbance. They also highlight the complexities of evolutionary mechanisms and their implications for conservation in a changing world.

Acknowledgements

This dissertation would not have been possible without the dedicated work of numerous desert fish researchers over the past 30 years, including university representatives and employees of federal, state, and tribal agencies. These people are too numerous to list *in toto*, but a few outstanding individuals graciously provided samples during my tenure at the University of Arkansas: Kevin Guadalupe and Jeff Petersen (Nevada Department of Wildlife); Dr. Bret Harvey (Humboldt State University); Claire Ingel, and Steve Parmenter (California Department of Fish & Wildlife); Dr. Ernest Keely and Dr. Janet Loxterman (Idaho State University); Eric Miskow (Nevada Natural Heritage Program); Rodney Nakamoto (U.S. Forest Service); Dr. Brian Sidlauskas (Oregon State University).

Special recognition is afforded to Dr. David Oakey, a former Douglas Lab member who wrote his dissertation on genetic variation in *Rhinichthys*. His work provided the foundation for my dissertation. Much of what I have accomplished would never have been possible without his dedicated research efforts, and the sampling trips he conducted with Dr. Matt Andersen (now USGS). Many of the samples analyzed herein were obtained through their efforts, and as such, provide insights into genetic diversity within now-extinct populations of *Rhinichthys*.

Many thanks are also due to the numerous people I met at various universities and institutions throughout my graduate career. My fellow graduate students in the Department of Biological Sciences at the University of Arkansas provided endless support through the duration of my PhD. I would like to acknowledge my fellow lab members, especially Dr. Max R. Bangs and Tyler K. Chafin for their assistance and feedback throughout the design and analysis of the project. Special thanks are due to Dr. Mark Davis at the Illinois Natural History Survey for the help and support he provided to me as a Master's student. My gratitude is also extended to my

coworkers at the Southwestern Native Aquatic Resources & Recovery Center in Dexter, New Mexico.

This research would not have been possible without the computational and financial support provided by the University of Arkansas. The Arkansas High Performance Computing Center (AHPCC) and the Nebula Cloud Platform provided hardware support and computational assistance. Additional resources were secured through the NSF-funded Jetstream cloud-computing resource (XSEDE Allocation TG-BIO160065). Funding was provided through generous fellowships and endowments: The Bruker Professorship in Life Sciences (Marlis Douglas), 21st Century Chair in Global Change Biology (Michael Douglas), a Doctoral Academy Fellowship, and the Professor Delbert Schwartz Endowed Graduate Fellowship. Teaching assistantships were provided by the Department of Biological Sciences at the University of Arkansas.

Last but not least, I acknowledge the emotional support of my parents, and their understanding of my educational choices. Without them, none of this would have been possible.

Table of Contents

List of Published Papers

Chapter V:

Mussmann SM, Douglas MR, Douglas ME (2018) COMP-D: A program for comprehensive computation of D-statistics and population summaries of reticulated evolution. Conservation Genetics Resources *In Review.*

I. Introduction

Evolutionary processes important to conservation occur on multiple temporal and spatial scales (Conover et al. 2006). Disciplines such as systematics and taxonomy describe biodiversity units resulting from processes occurring on geologic timescales, whereas population genetics focuses on mechanisms that influence biodiversity on contemporary, localized spatial scales. These disciplines are united through the field of conservation genetics, which utilizes molecular techniques to quantify biodiversity at fine spatiotemporal scales (Hopken et al. 2013; Mussmann et al. 2017) and provides deep historical perspectives by differentiating species (Douglas et al. 2006, 2009). The study of biodiversity at different scales helps identify and prioritize units of conservation concern (Volkmann et al. 2014; Welch and Beaulieu 2018).

The accelerated loss of biodiversity is a critical issue in the Anthropocene (Lewis and Maslin 2015), with direct and potentially irreversible consequences for the well-being of humankind. This topic is germane to the American west, in which many aquatic habitats have been engineered to suit humanity's collective needs, often at great expense to native fauna (Minckley and Deacon 1968). The intermountain west is a rugged landscape dissected by canyons, sparse natural lakes, few large rivers, and many intermittent streams in endorheic basins (Smith 1978; Minckley et al. 1986; Minckley and Unmack 2000). Vast expanses receive <50cm of rainfall annually (Vörösmarty et al. 2010), concomitant with an evaporation rate that often exceeds mean annual rainfall (Meyers and Nordenson 1962; Vörösmarty et al. 2010).

The geologic history of western North America has profoundly shaped the distribution of its fishes. Tectonism and volcanic activity destroyed connections among basins and created others anew (Minckley et al. 1986). Modern fish assemblages emerged in late Miocene/ early Pliocene (Miller 1958), with similarities to modern fauna established by the latter (Miller 1965;

Smith 1981). The intervening period saw the extinction of many species now found east of the continental divide (Miller 1958), with relictual Cypriniform, Cyprinodontiform, and Salmoniform fishes comprising the residual western fauna (Minckley et al. 1986; Smith et al. 2010). The Cascade, Sierra Nevada, and Rocky Mountains effectively prevented recolonization following extirpation events (Smith et al. 2002). This is particularly evident within the Great Basin, which contains over 150 drainage basins and 160 mountain ranges (Smith 1978).

Native western fauna adapted to survive a dry climate and harsh landscape (Minckley et al. 1986), however these fragile ecosystems have been disrupted by anthropogenic activity (Cayan et al. 2010). Water is a precious and increasingly rare commodity (MacDonald 2007; Woodhouse et al. 2010). Despite environmental realities, it is often diverted for agriculture and urbanization to the detriment of native aquatic biodiversity (Christensen et al. 2004; Jelks et al. 2008; Gleick 2010). Minckley (1991) stated that "many western fishes appear to require little more than water to survive, but they do need water." This perhaps over simplifies the plight of western fishes, but profoundly describes one of several anthropogenic threats (Miller et al. 1989; Unmack and Minckley 2008). Another hazard concomitant with hydrological changes is the frequent, intentional introductions of exotic species. This exacerbates the ongoing homogenization of a unique, depauperate, and often narrowly endemic Western fauna (Rahel 2000), while underscoring the necessity of its conservation (Schade and Bonar 2005).

The natural and anthropogenic factors that impact evolutionary processes provide an interesting system, however selection of a study organism to address both contemporary and historical evolutionary processes at multiple spatial scales is difficult. Few species provide adequate resolution for this purpose, being narrowly endemic or otherwise restricted in range. Furthermore, many western fishes are imperiled, with 60 species or subspecies (~25%) currently

listed under the Endangered Species Act (ESA) (USFWS 2016). Several of these also sustain unique evolutionarily significant units (ESUs) listed under the ESA, thus many more are eliminated for consideration due to the inherent difficulties with studying endangered species.

Speckled Dace

Speckled Dace (SPD), *Rhinichthys osculus*, (Cypriniformes: Cyprinidae) has a broad distribution in western North America, being found in all major drainage basins west of the continental divide (Miller and Miller 1948). It has adapted to a variety of habitats, ranging from the extreme heat and aridity of the Mojave and Sonoran deserts, to high elevations and colder climates of the Pacific Northwest. SPD was long suspected to be the sister species of *R. falcatus* based upon presence of a frenum (Hubbs et al. 1974), however this character is not unique to these species (Woodman 1992) and genetic evidence suggests *R. cataractae* as a closer relative (Schönhuth et al. 2012).

Organisms with such wide geographic distributions and life history characteristics are atypical targets of conservation (Soulé 1985; Metrick and Weitzman 1996), and *R. osculus* as a whole is not threatened with extinction (NatureServe 2013), however history has demonstrated that even widely distributed, abundant species may contain unique extinction-prone lineages in highly restricted habitats. Two such lineages were eliminated in the 20th century due to overuse of water (Las Vegas Dace: *R. deaconi*) and introduction of non-native species (Grass Valley SPD: *R. o. reliquus*) (Miller et al. 1989). Four described subspecies of *R. osculus* are listed under the ESA: Ash Meadows SPD (*R. o. nevadensis*), Clover Valley SPD (*R. o. oligoporus*), Independence Valley SPD (*R. o. lethoporus*), and Kendall Warm Springs SPD (*R. o. thermalis*) (USFWS 2016). Other subspecies are in imminent danger of extinction (Moyle et al. 2015) but

have not been afforded federal protection. Genetic investigations have identified cryptic variation in SPD, indicating the full extent of intraspecific diversity has not yet been documented (Hoekzema and Sidlauskas 2014; Wiesenfeld et al. 2018).

Recent advances in molecular techniques have brought genomic approaches to the forefront of conservation genetics, providing researchers with unprecedented access to cutting-edge molecular tools (Allendorf 2016). Methods once restricted to use in organisms with *a priori* genomic knowledge are now affordable for non-model organisms (Andrews and Luikart 2014; Puritz et al. 2014). Restriction-site associated DNA sequencing (RAD-seq) methods are most often applied for these purposes (Baird et al. 2008; Peterson et al. 2012; Ali et al. 2016). These methods provide a means of reducing genome complexity while simultaneously identifying molecular markers useful at multiple time-scales, including deep-scale phylogenetics (Eaton et al. 2017) and population-level studies (Davey and Blaxter 2010).

This dissertation provides an extensive study of fine-scale population dynamics, anthropogenic impacts on genetic diversity, and evolutionary processes affecting SPD in geological time. Modern genomic techniques and high-performance computing resources are used to investigate diversification of this fish across a large landscape. Results impact conservation and management of multiple fish species throughout the West, and extend our knowledge on the depth of variation found within geographically widespread species. This is accomplished in Chapter II by providing a comprehensive phylogeographic analysis of SPD through its native range. A near-comprehensive sampling of SPD and several outgroup species (N=307) was assessed using 15,762 double-digest RAD (ddRAD) loci. Phylogenetic analysis and evaluation of historical introgression revealed that SPD evolution has conformed to known palaeohydrology of western North America. ESUs are delineated for SPD, and the taxonomic

status of several species are evaluated. Results provide a comparative framework for other aquatic species of the west.

Chapter III provides a fine-scale population genetic study of SPD in the Colorado River and Great Basin. Population-level sampling of 116 sites yielded 1,130 SPD for analysis using ddRAD. Hypotheses concerning contemporary mechanisms of evolution were tested in a widely distributed riverine species, with inter- and intra-basin population dynamics being elucidated. SPD of the Great Basin conformed to a "Death Valley" model of intra-species diversity driven by basin isolation. Those in the Colorado River Basin conformed to a "stream hierarchy" model that is described by stream network branching patterns. Previously unidentified diversity was recorded for SPD, and anthropogenic influence was evaluated in light of evolutionary patterns.

Chapter IV investigates the timescale for divergence within *Rhinichthys* with special emphasis on SPD subspecies of the Death Valley ecosystem. Samples from eight sites in the Owens and Amargosa basins yielded 112 samples that were evaluated using 12,556 ddRAD loci. Multispecies coalescent analysis identified six ESUs within the Death Valley region. Molecular clock analysis revealed that geological events occurring over the past million years provide the best explanation for the modern distribution of SPD diversity across the region. This contradicts younger ages (i.e., hundreds to a few thousand years) recovered for sympatric *Cyprinodon* species.

Chapters V and VI describe new and updated computational programs for analyzing genome-scale SNP data. Chapter V describes COMP-D: a program written in C++ that takes advantage of parallel computing resources to calculate Patterson's D-statistic and its derivatives (Partitioned-D and D-foil). Chapter VI describes BA3-SNPs: an update to BAYESASS 3 (Wilson and Rannala 2003) that assesses contemporary migration using SNP data.

These five studies combine to assess native aquatic biodiversity of the western US at varying spatial and temporal scales, and describe new tools that aid in accomplishing those goals. Reduced representation genomic techniques (i.e., ddRAD sequencing) allow this to be achieved on an unprecedented scale. The results demonstrate historical and contemporary evolutionary processes acting on natural populations that continue to shape biodiversity in the Anthropocene. Implications for the conservation of narrowly endemic SPD subspecies are discussed. The patterns elucidated through these studies provide guidance for the conservation of sympatric species throughout western North America.

References

Ali OA, O'Rourke SM, Amish SJ, et al (2016) RAD capture (rapture): flexible and efficient sequence-based genotyping. Genetics 202:389. doi: 10.1534/genetics.115.183665

Allendorf FW (2016) Genetics and the conservation of natural populations: allozymes to genomes. Mol Ecol 26:420–430. doi: 10.1111/mec.13948

Andrews KR, Luikart G (2014) Recent novel approaches for population genomics data analysis. Mol Ecol 23:1661–1667

Baird NA, Etter PD, Atwood TS, et al (2008) Rapid SNP discovery and genetic mapping using sequenced RAD markers. PLoS ONE 3:1–7. doi: 10.1371/journal.pone.0003376

Cayan DR, Das T, Pierce DW, et al (2010) Future dryness in the southwest US and the hydrology of the early 21st century drought. Proc Natl Acad Sci U S A 107:21271–21276

Christensen NS, Wood AW, Voisin N, et al (2004) The effects of climate change on the hydrology and water resources of the Colorado River basin. Clim Change 62:337–363

Conover DO, Clarke LM, Munch SB, Wagner GN (2006) Spatial and temporal scales of adaptive divergence in marine fishes and the implications for conservation. J Fish Biol 69:21–47. doi: 10.1111/j.1095-8649.2006.01274.x

Davey JW, Blaxter ML (2010) RADSeq: next-generation population genetics. Brief Funct Genomics 9:416–423. doi: 10.1093/bfgp/elq031

Douglas ME, Douglas MR, Schuett GW, Porras LW (2006) Evolution of rattlesnakes (Viperidae; *Crotalus*) in the warm deserts of western North America shaped by Neogene vicariance and Quaternary climate change. Mol Ecol 15:3353–3374. doi: 10.1111/j.1365-294X.2006.03007.x

Douglas ME, Douglas MR, Schuett GW, Porras LW (2009) Climate change and evolution of the New World pitviper genus *Agkistrodon* (Viperidae). J Biogeogr 36:1164–1180. doi: 10.1111/j.1365-2699.2008.02075.x

Eaton DA, Spriggs EL, Park B, Donoghue MJ (2017) Misconceptions on missing data in RAD-seq phylogenetics with a deep-scale example from flowering plants. Syst Biol 66:399–412

Gleick PH (2010) Roadmap for sustainable water resources in southwestern North America. Proc Natl Acad Sci U S A 107:21300–21305. doi: 10.1073/pnas.1005473107

Hoekzema K, Sidlauskas BL (2014) Molecular phylogenetics and microsatellite analysis reveal cryptic species of speckled dace (Cyprinidae: *Rhinichthys osculus*) in Oregon's Great Basin. Mol Phylogenet Evol 77:238–250. doi: 10.1016/j.ympev.2014.04.027

Hopken MW, Douglas MR, Douglas ME (2013) Stream hierarchy defines riverscape genetics of a North American desert fish. Mol Ecol 22:956–971. doi: 10.1111/mec.12156

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

Jelks HL, Walsh SJ, Burkhead NM, et al (2008) Conservation status of imperiled North American freshwater and diadromous fishes. Fisheries 33:372–407. doi: 10.1577/1548-8446-33.8.372

Lewis SL, Maslin MA (2015) Defining the Anthropocene. Nature 519:171

MacDonald GM (2007) Severe and sustained drought in southern California and the West: present conditions and insights from the past on causes and impacts. Quat Int 173–174:87–100. doi: 10.1016/j.quaint.2007.03.012

Metrick A, Weitzman ML (1996) Patterns of behavior in endangered species preservation. Land Econ 72:1–16

Meyers JS, Nordenson TJ (1962) Evaporation from the 17 western states. US Government Printing Office

Miller RR (1958) Origin and affinities of the freshwater fish fauna of western North America. In: Zoogeography. American Society for the Advancement of Science, Washington, DC, pp 187–222

Miller RR (1965) Quaternary freshwater fishes of North America. In: The Quaternary of the United States. Princeton University Press, Princeton, NJ, pp 569–581

Miller RR, Miller RG (1948) The contribution of the Columbia River system to the fish fauna of Nevada: five species unrecorded from the state. Copeia 1948:174–187. doi: 10.2307/1438452

Miller RR, Williams JD, Williams JE (1989) Extinctions of North American fishes during the past century. Fisheries 14:22–38

Minckley W, Deacon JE (1968) Southwestern fishes and the enigma of "endangered species." Science 159:1424–1432

Minckley W, Hendrickson D, Bond C (1986) Geography of western North American freshwater fishes: description and relationships to intracontinental tectonism. In: Hocutt CH, Wiley EO (eds) The Zoogeography of North American Freshwater Fishes. John Wiley & Sons, New York, pp 519–613

Minckley W, Unmack PJ (2000) Western springs: their faunas, and threats to their existence. In: Freshwater Ecoregions of North America: A Conservation Assessment. Island Press, Washington, DC, pp 52–53

Minckley WL (1991) Native fishes of the Grand Canyon region: an obituary. In: Colorado River Ecology and Dam Management. National Academy Press, Washington, DC, pp 124–177

Moyle P, Quiñones RM, Katz J, Weaver J (2015) Fish species of special concern of California. California Department of Fish and Wildlife

Mussmann SM, Douglas MR, Anthonysamy WJB, et al (2017) Genetic rescue, the greater prairie chicken and the problem of conservation reliance in the Anthropocene. R Soc Open Sci 4:. doi: 10.1098/rsos.160736

NatureServe (2013) Rhinichthys osculus (Speckled Dace). In: IUCN Red List Threat. Species 2013. http://www.iucnredlist.org/details/62205/0. Accessed 4 Aug 2016

Peterson BK, Weber JN, Kay EH, et al (2012) Double digest RADseq: an inexpensive method for *de novo* SNP discovery and genotyping in model and non-model species. PLoS ONE 7:1–11. doi: 10.1371/journal.pone.0037135

Puritz JB, Matz MV, Toonen RJ, et al (2014) Demystifying the RAD fad. Mol Ecol 23:5937–5942

Rahel FJ (2000) Homogenization of fish faunas across the United States. Science 288:854–856. doi: 10.1126/science.288.5467.854

Schade CB, Bonar SA (2005) Distribution and abundance of nonnative fishes in streams of the western United States. North Am J Fish Manag 25:1386–1394. doi: 10.1577/M05-037.1

Schönhuth S, Shiozawa DK, Dowling TE, Mayden RL (2012) Molecular systematics of western North American cyprinids (Cypriniformes: Cyprinidae). Zootaxa 3586:1–303

Smith G, Dowling T, Gobalet K, et al (2002) Biogeography and timing of evolutionary events among Great Basin fishes. Gt Basin Aquat Syst Hist 33:175–234

Smith GR (1978) Biogeography of intermountain fishes. Gt Basin Nat Mem 2:17–42

Smith GR (1981) Late Cenozoic freshwater fishes of North America. Annu Rev Ecol Syst 12:163–193

Smith GR, Badgley C, Eiting TP, Larson PS (2010) Species diversity gradients in relation to geological history in North American freshwater fishes. Evol Ecol Res 12:693–726

Soulé ME (1985) What is conservation biology? A new synthetic discipline addresses the dynamics and problems of perturbed species, communities, and ecosystems. BioScience 35:727–734

Unmack PJ, Minckley W (2008) The demise of desert springs. In: Aridland Springs in North America: Ecology and Conservation. University of Arizona Press, Tucson, AZ, pp 11–34

USFWS (2016) Endangered Species. In: US Fish Wildl. Serv. Online Endanger. Species Database. https://www.fws.gov/endangered/. Accessed 4 Aug 2016

Volkmann L, Martyn I, Moulton V, et al (2014) Prioritizing populations for conservation using phylogenetic networks. PLoS ONE 9:e88945. doi: 10.1371/journal.pone.0088945

Vörösmarty CJ, McIntyre PB, Gessner MO, et al (2010) Global threats to human water security and river biodiversity. Nature 467:555–561. doi: 10.1038/nature09440

Welch . Jessica, Beaulieu . Jeremy (2018) Predicting extinction risk for data deficient bats. Diversity 10:. doi: 10.3390/d10030063

Wiesenfeld JC, Goodman DH, Kinziger AP (2018) Riverscape genetics identifies speckled dace (*Rhinichthys osculus*) cryptic diversity in the Klamath–Trinity Basin. Conserv Genet 19:111–127. doi: 10.1007/s10592-017-1027-6

Wilson GA, Rannala B (2003) Bayesian inference of recent migration rates using multilocus genotypes. Genetics 163:1177–1191

Woodhouse CA, Meko DM, MacDonald GM, et al (2010) A 1,200-year perspective of 21st century drought in southwestern North America. Proc Natl Acad Sci U S A 107:21283–21288. doi: 10.1073/pnas.0911197107

Woodman DA (1992) Systematic relationships within the cyprinid genus *Rhinichthys*. In: Mayden RL (ed) Systematics, Historical Ecology, and North American Freshwater Fishes. Stanford University Press, Stanford, CA, pp 374–391

II. Small Fish in a Large Landscape Revisited: The Role of Introgression and Hybridization on the Evolution of *Rhinichthys osculus* in Western North America

Introduction

An understanding of riverine phylogeography is fundamental for the conservation and management of aquatic organisms. It interprets ecological processes (Räsänen and Hendry 2008; Kisel and Barraclough 2010), defines biogeographic patterns (Lessios et al. 2001; Hubert and Renno 2006), and recognizes barriers to gene flow (Case and Taper 2000; Nosil et al. 2005). Connectivity among populations (or lack thereof) also demarcates historical dispersal, isolation, and secondary contact within basins (Green et al. 1996; Stevens and Hogg 2003; Abbott et al. 2013). The latter aspect can also distinguish hybridization and introgression, particularly in cases where species are recently evolved (Coyne and Orr 1998; Edmands 2002).

Biologists once considered hybridization a rarity among animals, even delineating species according to the reproductive viability of their offspring (Mayr 1942). However, evidence of hybridization (Arnold 1992; Dowling and Secor 1997) and its importance in evolution (Grant and Grant 1992) became widely acknowledged as genetic methods evolved (Dasmahapatra et al. 2012). Hybridization and introgression are now rightfully acknowledged as important in promoting adaptive divergence more quickly than mutation alone (Barton 2001; Gompert et al. 2012; Abbott et al. 2013). Species boundaries are considered semi-permeable for many organisms (Harrison and Larson 2014), with hybridization occurring in approximately 10% of animals, with greater prevalence among certain taxa (Mallet 2005). Fishes in particular maintain species boundaries despite horizontal transfer of genes among lineages (Wagner et al. 2013; Harrison et al. 2017).

The evolution of many western North American fishes has been driven by dispersal, isolation, and secondary contact with subsequent hybridization and introgression (Unmack et al. 2014; Broughton and Smith 2016; Dowling et al. 2016). The most famous examples reside within the Colorado River Basin *Gila,* and include species of hybrid origin (*G. seminuda*: Demarais et al. 1992), and introgression of mitochondrial haplotypes from one species to another (*G. cypha* into *G. robusta*: Gerber et al. 2001). Increasing aridity during the Pleistocene and Holocene was likely a driving force that coalesced species into refugia where interbreeding occurred (Dowling and DeMarais 1993). Introgressive hybridization has also played an important role in the evolution of *Catostomus* sp. (Cypriniformes: Catostomidae) that are found within these same arid regions (Bangs et al. 2018).

Speckled Dace (*Rhinichthys osculus*) stands in contrast to other western fishes because it is not considered an exemplar of hybridization. Its notoriety instead stems from dispersal capabilities that have produced numerous morphologically variable subspecies within all major drainages of western North American (Miller and Miller 1948). It exists in allopatry throughout much of its range, save for the Snake and Columbia river drainages, where it is sympatric with *R. cataractae*, *R. falcatus*, and *R. umatilla*. The latter is believed the result of an historic hybridization event between *R. osculus* and *R. falcatus* (Haas 2001). This fosters an opportunity to study ecological and evolutionary processes across a large landscape (Oakey et al. 2004).

Complex geological (Minckley et al. 1986) and climatic events (Cayan et al. 2010; Woodhouse et al. 2010) since the late Miocene have impacted all fishes of the region (Smith and Dowling 2008). Dace, in particular, specialize in headwater environments (i.e., second- and third-order streams: Moyle 2002), but also occupy desert springs, large natural lakes, and big rivers (Sigler and Sigler 1987). This ecological diversity has not only produced isolated endemic

forms but also the potential for hybridization upon secondary contact (Malde 1968). Few attempts have been made to quantify range-wide genetic differentiation or construct a phylogeny for Speckled Dace (but see Oakey et al. 2004; Smith et al. 2017). Studies have instead focused on variability within or among a few modern basins (Pfrender et al. 2004; Smith and Dowling 2008; Ardren et al. 2010; Wiesenfeld et al. 2018), with special emphasis on unique forms inhabiting endorheic basins (Sada et al. 1995; Billman et al. 2010; Hoekzema and Sidlauskas 2014).

This study explores phylogenetic relationships and patterns of introgression across the western landscape using the Speckled Dace (SPD) as a study species. Data were sampled from throughout the species' genome, and subsequently evaluated using classical phylogenetic and contemporary species tree methods as a means to interpret historical introgression and dissect conflicting gene tree topologies (Maddison 1997; Maddison and Knowles 2006; Durand et al. 2011). The monophyly of *R. osculus* and its relationship to other *Rhinichthys* species is explored in the context of hybridization, introgression, and the changing western landscape. The taxonomic status of a related species (*Tiaroga cobitis*) is also discussed. In addition, the conservation implications with regard to subspecific lineages and evolutionary significant units (ESU) are delineated.

Methods

Sampling

The entire geographic range of SPD was evaluated, and a comprehensive sampling accomplished (N=282; Figure 1), with but a few subspecies absent [i.e., Big Smoky Valley (*R. o. lariversi*), Diamond Valley (*R. o.* ssp.), Independence Valley (*R. o. lethoporus*), Kendall Warm

Springs (*R. o. thermalis*), and Foskett Springs SPD (*R. o.* ssp.)]. Extant species of *Rhinichthys* (N=13) were also comprehensively covered (save Western Blacknose Dace, *R. obtusus*). Several western North American cyprinid species (N=12) including *Agosia chrysogaster* (Longfin Dace), *Iotichthys phlegethontis* (Least Chub), *Meda fulgida* (Spikedace), *Mylocheilus caurinus* (Peamouth Chub), *Richardsonius balteatus* (Redside Shiner), and *Tiaroga cobitis* (Loach Minnow) were utilized as outgroup taxa.

Data Collection

Whole genomic DNA extractions were performed using one of four methods: Gentra Puregene DNA Purification Tissue kit, QIAGEN DNeasy Blood and Tissue Kit, QIAamp Fast DNA Tissue Kit, or CsCl-gradient. Quality of extracted DNA was assessed visually on a 2.0% agarose gel and quantified with a Qubit 2.0 fluorometer (Thermo-Fisher Scientific, Inc.). Preparation of double digest Restriction-Site Associated DNA (ddRAD) libraries followed the methods of Peterson et al. (2012). Barcoded samples (100 ng DNA each) were pooled in sets of 48 following Illumina adapter ligation, then size selected (375 to 425 bp in length; Chafin et al. 2017) using the Pippin Prep System (Sage Science). Size-selected DNA was subjected to 12 cycles of PCR amplification using Phusion high-fidelity DNA polymerase (New England Bioscience) according to manufacturer protocols. Subsequent quality checks to confirm successful amplification of DNA were performed with the Agilent 2200 TapeStation and qPCR. Three indexed libraries (N=144 samples) were pooled per lane for 100bp single-end sequencing using the Illumina HiSeq 2000 (University of Wisconsin Biotechnology Center) or HiSeq 4000 (University of Oregon Genomics & Cell Characterization Core Facility).

Samples from the Bear, Columbia, and Snake Rivers (N=41) were selected for DNA barcoding so as to clarify relationships within this basin. Additional *Rhinichthys* sequences (N=45) were downloaded from Genbank (Table S1: Hubert et al. 2008; April et al. 2011) for comparison. The FishF1 (5' TCAACCAACCACAAAGACATTGGCAC 3') and FishR1 (5' TAGACTTCTGGGTGGCCAAAGAATCA 3') primers and PCR protocols of Hubert et al. (2008) were used to amplify a 648bp region of the *cytochrome c oxidase I* (*COI*) mitochondrial gene. PCR cleanup was accomplished via ExoSAP-IT (Thermo Fisher Scientific, Inc., Waltham, MA). PCR products were fluorescently labeled using BigDye v. 3.1 (Applied Biosystems Inc., Forest City, CA) according to manufacturer specifications, and sequenced on an ABI 3730XL DNA Analyzer (W.M. Keck Center for Comparative and Functional Genomics, University of Illinois at Urbana-Champaign). Sequences were edited and aligned manually in Sequencher v. 5.4 (Gene Codes Corporation, Ann Arbor, MI) with haplotype network reconstruction via statistical parsimony (Clement et al. 2000) in POPART (Leigh and Bryant 2015).

ddRAD Alignment

Illumina reads were de-multiplexed and filtered for quality in STACKS (Catchen et al. 2013). All reads with uncalled bases or Phred quality scores <10 were discarded. Reads that passed quality filtering but had ambiguous barcodes were recovered when possible (≤ 1 mismatched nucleotide). A clustering threshold of 0.85 was utilized in PYRAD (Eaton 2014) to perform *de novo* assembly of ddRAD loci for all samples. Reads with >4 low quality bases (Phred quality score < 20) were removed from analysis. A minimum of 5 reads was required for a locus to be called for an individual. A filter to remove putative paralogs was applied by

discarding loci with heterozygosity >0.6. Only loci present in at least 50% of ingroup samples (N=141) were retained.

Phylogenetic Analysis

The unique analytical challenges presented by RADseq data have promoted disagreement on proper phylogenetic approaches (Edwards et al. 2016; Springer and Gatesy 2016). This study used a compendium of traditional methods and a contemporary species tree approach that models incomplete lineage sorting by allowing polymorphic states within populations, rather than following assumptions of the traditional DNA substitution model in which a taxon is fixed for a specific nucleotide at a locus (Schrempf et al. 2016).

The traditional approach used Maximum Likelihood (ML) and Bayesian inference, with complete sequences of all recovered loci concatenated for each analysis. A ML tree was calculated in ExaML (Kozlov et al. 2015) using the per-site rate category model (i.e., the GTRCAT model equivalent from RAxML: Stamatakis 2014). One thousand bootstrap replicates were performed to assess statistical support for all nodes in the resulting tree, using a custom pipeline available at <https://github.com/stevemussmann/ExaML_pipeline>. ExaBayes (Aberer et al. 2014) was employed to conduct Bayesian inference. Due to computational constraints, we reduced the number of samples (N=155) by randomly subsampling one individual per locality or outgroup species. The GTR model with rate heterogeneity was applied, with two simultaneous runs (5 million generations) employing 12 MCMC chains each to promote convergence. Summary statistics calculated by ExaBayes and visual analysis of trace plots in Tracer (Rambaut et al. 2018) resulted in the first 25% of sampled trees being discarded as burn-in. Finally, a species tree was calculated using the reversible polymorphism-aware phylogenetic

model (POMO: Schrempf et al. 2016) in IQ-TREE (Nguyen et al. 2014). A virtual population size of 19 was assumed to model genetic drift. Mutations were assumed to follow a GTR substitution model, and rate heterogeneity was modeled with four rate categories. An ultrafast algorithm (Hoang et al. 2018) was used to perform 1000 bootstrap replicates.

Analyses of Introgression and Hybridization

Three methods were utilized to assess the impact of horizontal gene transfer on the evolution of SPD. The first tests the hybrid origin hypothesis for *R. umatilla* (*R. falcatus* x *R. osculus*: Haas 2001; McPhail 2007). This was accomplished using HYDE (Blischak et al. 2018), which employs a coalescent-based approach that compares a putative hybrid taxon against two candidate parental species and an outgroup (*R. atratulus*). All possible hybrid combinations (N=142) were evaluated, with significance assessed at a Bonferroni adjusted threshold ($\alpha = 2.9 \times 10^{-8}$).

To evaluate the extent and relative age of hybridization between *R. c. dulcis* and *R. osculus*, a hybrid index (Buerkle 2005) was calculated using the est.h function in the INTROGRESS R package (Gompert and Buerkle 2010). This was performed following the recovery of a non-monophyletic *R. cataractae* during phylogenetic analysis. *Rhinichthys* samples were selected from putative parental populations where the species exist in allopatry. Samples from Fossil Creek (Verde River drainage, AZ) were chosen to represent *R. osculus*. *Rhinichthys cataractae dulcis* could not be obtained from a putatively "pure" population, so individuals from the South Platte River, Colorado (*R. c. cataractae*) were utilized instead. Data were filtered to retrieve unlinked SNPs among putative parental populations. Interspecific heterozygosity was calculated

using the calc.intersp.het function in INTROGRESS, and results were visualized using the triangle.plot function.

Finally, the three phylogenies were contrasted to identify topological discordance that may indicate mixing among SPD lineages. Patterson's D-statistic (Durand et al. 2011) as implemented in COMP-D was utilized to test if introgression has impacted phylogenetic analysis. These tests focused on three geographic regions known to have Pliocene or Pleistocene connections with surrounding basins: The Klamath, Snake, and Virgin river drainages. Tests were performed for each using *R. atratulus* as an outgroup. Additional tests were performed to detect introgression of *Agosia chrysogaster* or *Tiaroga cobitis* into *Rhinichthys*, with *Richardsonius balteatus* as outgroup. A Z-score was calculated from 1,000 bootstrap replicates for each test and statistical significance was evaluated at a Holm-Bonferroni adjusted threshold ($\alpha = 0.0001$).

Results

DNA Barcoding

The mtDNA analysis recovered 96 segregating sites in the 648bp region of *COI*, with 87 (90.6%) parsimony-informative. Figure 2 shows the haplotype network recovered from these sequences. Clear separation (4.3% sequence divergence) occurred between haplotypes of the most closely related Middle Snake River *R. osculus* and *R. cataractae.* Scant divergence was found in *R. falcatus* (max = 0.6%) or *R. umatilla* (max = 0.3%) when compared with Middle Snake River *R. osculus*. *Rhinichthys falcatus* and *R. umatilla* showed a maximum of 0.9% sequence divergence.

Haplotypes of *R. cataractae* were observed in 5 individuals identified as *R. osculus* from the Deschutes River (Columbia Drainage, OR), and likely represent misidentifications in the field. Past connections between the Bonneville and Snake River were also evident in mitochondrial haplotypes. Two samples from Mink Creek (Bear River tributary, ID) in the Bonneville Basin were identical to Dempsey Creek (Portneuf River tributary, ID) in the Upper Snake River drainage. Other samples from the Bonneville basin (Bear Lake, UT) were separated from these samples by 7.7% sequence divergence.

pyRAD

Two datasets were generated for phylogenetic analysis. All individuals had a minimum average sequencing depth of 9.8x per locus, and a maximum of 73.3x. Mean depth per locus for EXAML and IQ-TREE inputs was 41.8x. Mean depth for the EXABAYES input was 44.8x. Variable numbers of loci were recovered due to filtering methods applied during the genotyping step of PYRAD. Input files for EXAML and IQ-TREE contained 15,762 loci each, with 311 to 14,596 loci per individual ($\bar{x}$ = 11,374.5, σ = 2,363.2). Individuals selected for EXABAYES had 2,573 to 17,690 loci (total = 19,755, $\bar{x}$ = 13,584.4, σ = 2,579.5). This dataset was subjected to the same filtering parameters as EXAML and IQ-TREE, but with fewer individuals (N=155), and greater sequencing depth ($\bar{x}$ = 44.8x), resulting in greater numbers of filtered loci. Samples genotyped at <10,000 loci were often other (non-*Rhinichthys*) species. On average, these samples ranged from 4,670.7 (EXAML; IQ-TREE) to 5890.2 loci (EXABAYES). Missing data were randomly distributed among loci for all data files.

Phylogenetic Analysis

Meda fulgida was used to root the phylogeny based upon previous studies (Schönhuth et al. 2012, 2018). Relationships among outgroup taxa were similar regardless of the method used for phylogenetic analysis (Figure 3). The only exception involves *Agosia chrysogaster* as sister to all taxa except *M. fulgida* in the IQ-TREE analysis. *Rhinichthys atratulus* was recovered in all analyses as sister to all other *Rhinichthys.* Generally, clades were recovered that linked *R. evermanni*, *R. cataractae*, and at least one Millicoma River *R. osculus* sample. However, *R. c. dulcis* was an exception in that it was occasionally recovered as sister to *R. osculus* (EXABAYES and IQ-TREE). Relationships among *R. c. cataractae*, *R. evermanni*, and *R. osculus* "Millicoma River" also changed according to method. In EXAML and EXABAYES, *R. c. cataractae* was sister to the *R. evermanni* and the *R. osculus* "Millicoma River" clade. For IQ-TREE, *R. osculus* "Millicoma River" was sister to *R. c. cataractae* and *R. evermanni*.

Relationships at the deepest nodes within *R. osculus* were mostly concordant among the different methods. Figure 4 represents the EXABAYES result, with Figure 5 showing deviations from this topology for trees produced by other methods. Four major clades were recovered and are referred to as the Columbia River, Lahontan, Snake River, and Colorado River clades. The Columbia River clade consists of the mainstem Columbia River, lower Snake River, and coastal drainages of Oregon. The Lahontan clade was composed of many endorheic basins in California, Nevada, and southeastern Oregon. This includes the coastal-draining Klamath, Sacramento, and San Benito drainages of California, as well as tributaries of the Middle Snake that once connected to the Lahontan system (e.g., the Owyhee River and Salmon Falls Creek).

The Snake River clade also reflects past hydrological connections. This clade includes the Bear and Weber rivers of the Bonneville Basin, in addition to the Middle and Upper Snake

River. *Rhinichthys falcatus* and *R. umatilla* were recovered as members of this clade by all phylogenetic methods save IQ-TREE, where *R. falcatus* was sister to the Colorado River clade. The Deschutes and Crooked rivers (Columbia River drainage) were also recovered in this clade.

The fourth clade represents the modern Colorado River Basin. It also includes the coastal-draining Santa Ana and San Gabriel rivers of Southern California, the Sevier River, and the Snake Valley region of the former Bonneville Basin. In all analyses, the Colorado River clade was sister to the Snake River clade. The Lahontan clade was sister to this relationship, and the Columbia River was sister to the clade formed by these three groups.

Relationships within each clade varied according to phylogenetic reconstruction (Figure 5). Results from IQ-TREE mostly agreed with those from ExaBayes and ExaML throughout the Columbia and Lahontan Clades. However, there were minor divergences within the Upper Colorado River Basin, and greater discordance in the Snake River. Most discordant relationships centered around two areas of shifting hydrological connections: The Snake and Virgin river drainages.

Snake River

Relationships were most variable within the Snake River clade, reflecting the complex geological history of this region (Figure 6). Few contemporary hydrologic patterns are apparent, and no single river basin involved (i.e., the Bear, Columbia, or Snake) was recovered as monophyletic by any method. However, the Middle Snake River seemingly formed one clade, while the Upper Snake River (above Shoshone Falls) and the northeastern Bonneville Basin formed another.

Rhinichthys falcatus and *R. umatilla* were also recovered in the Snake River clade, except for the IQ-TREE analysis that recovered *R. falcatus* as sister to the Colorado River clade. Our *R. falcatus* samples represented three rivers (Columbia, Fraser, and Wilamette), and were monophyletic under all analyses. The single *R. umatilla* sample was collected from the Similkameen River (Columbia River Basin, BC, Canada) and was sister to the Deschutes River (Columbia River tributary, OR) in all cases.

Virgin River

The relationship of the Virgin River clade varies from one analytical method to the next. Three analyses (EXABAYES, EXAML, and IQ-TREE) recovered the Virgin River and its tributaries as monophyletic. Two of these (EXABAYES and EXAML) designated it as sister to the Upper Colorado River/ Bill Williams River clade. IQ-TREE recovered it as sister to the Sevier River and Snake Valley clade, which was sister to the remainder of the Colorado River Basin.

Relationships within the Virgin River system again varied by analytical method (Figure 7). The pluvial White River was found as monophyletic. Meadow Valley Wash was monophyletic and seemingly reflected a close relationship with Moody Wash. The Moapa River was sister to the White River in EXABAYES and EXAML analyses but more closely related to the Virgin River mainstem in IQ-TREE. The mainstem Virgin River was never recovered as monophyletic in any analysis.

Introgression

Evidence for introgression was investigated in two outgroup taxa: *Agosia chrysogaster* and *Tiaroga cobitis*. Each was compared to *R. osculus* samples from the Lower Colorado River

Basin, where they are now sympatric or may have historically coexisted. Little evidence was found for the mixing of these lineages in the Gila, Verde, or Agua Fria rivers. However, 92.9% of tests involving the Bill Williams River indicated possible introgression between *R. osculus* and *A. chrysogaster* (mean ABBA=17.5, mean BABA=10.3). Only 55% of tests involving the Bill Williams and *T. cobitis* were significant (mean ABBA=23.1, mean BABA=22.4). The appearance of introgression may be a statistical artifact derived from the low numbers of ABBA and BABA sites detected.

Introgression was also investigated as a potential cause of discordance for the Klamath, Snake, and Virgin River clades. Less than 1% of tests comparing the Klamath River to surrounding basins indicated significant gene transfer. However, statistically significant D-statistic values were observed when coastal Oregon drainages were compared with the Pit (36.6% significant) and San Benito rivers (42.2% significant), indicating past hydrologic connections (Figure 8A). The Sacramento River revealed similar connections to the Pit (38.9%) and San Benito (40.2%) rivers. The Willamette also showed evidence of mixing with Pit (42.8%) and San Benito (43.7%). All tests involving Rattlesnake Creek of the Mahleur (Oregon Lakes) basin also indicated connections to the Willamette, Sacramento, and coastal drainages.

Signals of introgression in the Snake River clade were the most complex, but not surprisingly so, given its complicated hydrologic history (Figure 8B). Strong evidence was present for mixing between the Bonneville (Bear and Weber rivers) and the Upper Green River (Colorado River Basin). Sixty-two percent of tests comparing the Upper Green and Bear rivers were significant, whereas 79.7% of tests were significant for the Weber River. Nearly all tests investigating introgression among the Bonneville Basin and Middle Snake were significant (93.9%), whereas all tests were between the Bonneville and upper Snake River.

Elevated levels of lateral gene transfer were noted for both *R. falcatus* and *R. umatilla* (Figure 8C). Most tests comparing *R. falcatus* with the Columbia Basin (56.9%), Bonneville Basin (62.8%), Middle Snake (67%), and Upper Snake (51.6%) were significant. Nearly all tests comparing *R. umatilla* to the same drainages were also significant (>=99%), with the exception of the Columbia where 70% were significant.

The Virgin River and its tributaries also showed a complex pattern of mixing (Figure 8D). Beaver Dam Wash showed evidence of mixing with the White River (73.2%), Moapa (82.8%), and Meadow Valley Wash (40.2%). The latter was also linked to the Moapa in 74.5% of tests. The coastal rivers of Southern California were found to have mixed with Meadow Valley Wash in 72.4% of tests. Moapa also showed very low levels of mixing with two other rivers, the pluvial White River (4.6%), and the mainstem Virgin River (4.7%).

Hybridization

Three taxa from locations with sympatric *R. cataractae* and *R. osculus* were evaluated using a hybrid index: *R. c. dulcis* from the Columbia River, *R. osculus* from the Deschutes River, and *R. osculus* from the Millicoma River (Figure 9). The *R. c. dulcis* samples were found to be *R. osculus* x *R. cataractae* hybrids. The Deschutes River samples were mostly *R. osculus* in origin but with a low level of *R. cataractae* genes that may have historically entered this population. Lastly, the two *R. osculus* samples from the Millicoma River differed: One showed strong evidence for a mostly *R. cataractae* genome, while the other was mostly *R. osculus*, clustering close to the Deschutes River samples.

HyDe was used to evaluate the putative hybrid origin of *R. umatilla* (i.e., whether it had a *R. falcatus* x *R. osculus* origin). Of the 142 tests, only 11 (7.7%) were significant (Table 1).

Parental *R. osculus* populations were represented by the Bear (N=1), Colorado (N=6), and Snake (N=4) rivers. No populations from the Columbia Basin, where the *R. umatilla* sample was collected, produced any significant results.

Discussion

The phylogenetic analyses indicate that the isolation and reconnection of western North American drainage basins has driven the evolution of SPD. This is reflected in the phylogenies where deep nodes agreed but more contemporary relationships varied. These trees also demonstrated numerous hybridization and introgression events that have similarly promoted the evolution of other western North American fishes. Below, we discuss the manner by which geologic and hydrologic events have impacted our four clades. We also define the taxonomic status of *R. umatilla*, *R. falcatus*, and *T. cobitis,* and benchmark the manner by which interspecific mixing has promoted the evolutionary trajectory of our study species.

Columbia River Clade

All phylogenetic methods recovered the mainstem Columbia, lower Snake, Willamette, and coastal Oregon drainages as a monophyletic group, thus supporting the subspecific status of *R. osculus nubilus* (Blackside Speckled Dace: McPhail and Lindsey 1986). Exceptions were the Deschutes and Crooked rivers that instead aligned more closely with the Middle Snake drainage. This suggests that historical connections across the Oregon Lakes region (Behnke 1979; Taylor 1985) may have promoted the sympatry of Middle Snake and Columbia forms in the Deschutes system.

The Columbia clade is further divided into Columbia/ lower Snake and Willamette/ Coastal Oregon clades. The Willamette and Coastal Oregon clades lack modern connections yet still appear admixed, supporting a past connection among the Willamette River and modern Oregon coastal drainages. *Rhinichthys cataractae* displays a similar pattern in this region (McPhail and Taylor 2009).

Lahontan Clade

Numerous complex relationships are manifest in this clade, as derived from drainage isolation and reconnection over millions of years (Minckley et al. 1986). *Rhinichthys osculus robustus* (Lahontan Speckled Dace) is widespread in the Humboldt River and its tributaries, as well as several isolated systems such as the Warner Subbasin and the Smoke Creek system. However, it is paraphyletic with the undescribed *R. o.* ssp. "Monitor Valley" (proximate to the Humboldt River) and with Coils Creek, despite a lack of a modern connection among these systems.

The ESA-listed Clover Valley Speckled Dace (*R. o. oligoporus*) is monophyletic and separated from other dace of the area. All the above-mentioned Lahontan subspecies form a clade that is sister to the Amargosa and Owens Valley systems that may also contain at least five subspecies. Pliocene connections between the Humboldt River and Mono Lake that connect to the Death Valley region support this argument (Hubbs and Miller 1948; Smith 1978; Miller and Smith 1981; Reheis and Morrison 1997).

The Lahontan clade also includes the coastal-draining basins of northern California. The Klamath Basin, containing Klamath River Speckled Dace (*R. o. klamathensis*), is monophyletic in most analyses when the neighboring endorheic Silver Lake Basin is included. These samples

were analyzed by Hoekzema and Sidlauskas (2014) who also hypothesized an historical connection to the Klamath Basin (Hubbs and Miller 1948). Relationships within the Klamath Basin are also congruent with those of Wiesenfeld et al. (2018).

Historical connections also foster the relationship among the Klamath, Sacramento, San Benito, and Pit rivers (Repenning et al. 1995; Link et al. 2002, 2005; Hershler and Liu 2004). All of these, as well as the Oregon Lakes region, seemingly connected to the pre-modern Snake River (Smith 1975; Taylor 1985). The "Sacramento" fauna of the San Benito River results from a prior connection to San Francisco Bay (Taylor 1985), and a possible headwater transfer with the San Joaquin River (Snyder 1905). The Pit River, a modern tributary of the Sacramento, was subject to orogeny and volcanism during the Pliocene/ Pleistocene that connected it with other rivers of the Klamath-Cascade region (Magill and Cox 1981; Minckley et al. 1986). These connections help explain the association between the Pit River, San Benito River, and Yreka Creek of the Klamath Basin, again without contemporary connectivity.

Paleoriver drainages in the Klamath region are now obscured, but evidence for an early Pleistocene connection to the Snake River is underscored by the presence of deposits with high potassium feldspar content that originated from batholithic sources in Idaho (Aalto 2006; Anderson 2008). However, the timing of this connection is tenuous (Anderson 2008) and our data provide scant evidence of introgression involving the Klamath and surrounding basins.

The Sacramento, Humboldt, and Snake rivers were connected as recently as the Pliocene when the Snake began draining Lake Idaho into the Columbia, prompting a reduction in flow that severed the Sacramento-Humboldt-Snake linkage (Houston 2009). These connections are well supported in our data, particularly given the high degree of introgression detected among them. The precise location of the Humboldt-Snake River connection is unknown but

hypothesized to have existed across the northwestern Bonneville Basin (Taylor 1985; Link et al. 2002; Smith et al. 2002), and is supported by the close affinity of the Thousand Springs Creek drainage with Salmon Falls Creek (a Snake River tributary), and subsequently with the remainder of the Lahontan clade. The sister relationship of the Owyhee, a modern tributary of Snake River, also supports this relationship with the Lahontan clade.

Snake River Clade

Rhinichthys o. carringtoni is a monophyletic group in the Upper Snake, Bear, and Weber rivers, and its taxonomic history typifies the complexities of *R. osculus* nomenclature (Gilbert and Evermann 1894; Snyder 1905; Jordan et al. 1930). The name was formerly applied to forms found within additional geographic regions, to include the Lahontan Basin and south-central California (Jordan and Evermann 1896). It has also been applied to forms in the Middle Snake River (Deacon and Williams 1984), however the region seems much more complex taxonomy, given that *R. falcatus* and *R. umatilla* are included.

The late Pleistocene connection between the Snake River and the Bonneville Basin is apparent in our analyses. The Bear River was captured by Bonneville Basin from the Upper Snake approximately 35,000 years ago (Minckley et al. 1986; Johnson 2002; Smith et al. 2002), with Lake Bonneville subsequently filling, then eventually spilling over at Red Rocks Pass, Idaho (Oviatt 1997; Link et al. 1999). The Bear was also connected with the Upper Colorado River (Smith et al. 2002), as supported by the distribution of Cutthroat Trout (Loxterman and Keeley 2012) and evidence herein of introgression involving the Bonneville Basin and Upper Green River.

Colorado River Clade

Southern portions of the Bonneville Basin are included in this clade, as well as coastal drainages of southern California, again underscoring the deep history of hydrological connections among disjunct basins. The Gila River and its tributaries, to include the type locality for *R. o. osculus* (i.e., San Pedro River: Girard 1856), form a monophyletic group, as do the southern California coastal drainages that have been previously hypothesized as a unique subspecies (Cornelius 1969; Hubbs 1979).

Speckled dace in the Colorado River Basin, north of the Gila River drainage, has been broadly viewed as *R. o. yarrowi* (Minckley 1973), but with some isolated subspecies in the western portions of the Virgin River system being excluded (i.e., *R. o. velifer*, *R. o. moapae*, and *R. o.* ssp. "Meadow Valley Wash"). It is clearly a polyphyletic taxon and thus in need of revision. The Little Colorado River *R. o. yarrowi* also displays a close relationship with the Gila/ Southern California clade. Minckley (1973) recognized that *R. o. osculus* and *R. o. yarrowi* "intergrade chaotically" along the Mogollon Rim, suggesting a history of repeated stream capture events across this geographic feature (described in Douglas et al. 2016).

The Bill Williams River, Grand Canyon, and Upper Colorado River form a clade, with the Bill Williams distinct from the others, reflecting its isolated nature. The Grand Canyon and Upper Colorado form a large, homogenous group within which it is difficult to discern clear relationships. Rivers are separated by very short branches, potentially indicating a recent genetic bottleneck as suggested for other native fishes in the region (Douglas et al. 2003; Borley and White 2006; Hopken et al. 2013).

The Virgin River system is a composite that reflects pluvial connections among disjunct stream segments, and includes subspecies of the pluvial White River (*R. o. velifer*: Gilbert 1893),

Moapa River (*R. o. moapae*: Williams 1978), and Meadow Valley Wash (*R. o.* ssp.: Deacon and Williams 1984). The White River was once connected to the Virgin River (Minckley et al. 1986) as well as to lakes in the north and west during the Pleistocene (Hubbs and Miller 1948). The Moapa River had a Pleistocene connection to the White River system, as well as an intermittent connection to Meadow Valley Wash. It was a tributary of the Virgin until the construction of Hoover Dam (Hubbs and Miller 1948). Patterns of introgression, with mixing among a modern Virgin River tributary (Beaver Dam Wash), the White River, Moapa, and Meadow Valley Wash, demonstrate the biological results that stem of these connections.

The potential for a Mid- to Late-Pleistocene connection between the Lower Colorado and southern California coast is provided by a series of pluvial lakes in the Mojave Desert (Smith 1966; Enzel et al. 2003; Roskowski et al. 2010). Morphological similarities are also apparent between *R. o. yarrowi* and an undescribed southern California subspecies (Cornelius 1969). Meadow Valley Wash was the only Virgin River stream that displayed significant levels of introgression with Southern California.

The southern portion of the Bonneville Basin is nested within the Colorado River clade, a situation also reflected within other western fishes (*Gila* sp.: Miller 1958). The Sevier River (*R. o. adobe*; Jordan 1889) is closely aligned with other isolated drainages in the southern Bonneville Basin, and forms a clade sister to the entire Virgin River system. Samples from Snake Valley have been attributed to a potential subspecies (Smith 1978; Miller 1984) but the close affinity of this taxon to the Sevier drainage necessitates additional analyses.

Middle Snake River Species

A paraphyletic *R. osculus* is underscored by the placement of *R. falcatus* and *R. umatilla* in all of our phylogenetic trees. However, the placement of *R. falcatus* was inconsistent, a situation interpreted as the result of violating PoMo assumptions. This species represents an amalgamation of Columbia, Fraser, and Wilamette river samples, and as such violates the assumption that taxa represent breeding populations of individuals (Schrempf et al. 2016).

Both *R. falcatus* and *R. umatilla* were originally recognized as *Agosia* species by Gilbert & Evermann (1894). They subsequently remained as species due to putative morphological distinctiveness, while other recognized species were relegated to subspecific status (Schultz 1936; Hubbs et al. 1974; Deacon and Williams 1984). Four genetic studies have included both *R. falcatus* and *R. umatilla*, with an additional three evaluating only *R. falcatus*. Of these seven, four were congruent with our results (McPhail and Taylor 2009; Kim and Conway 2014; VanMeter 2017), whereas a fifth study was partially congruent (Smith et al. 2017), in that it related *R. falcatus* to the Middle Snake River whereas *R. umatilla* was more closely aligned with coastal drainages of the Pacific Northwest. The final two studies were indeterminate due to lack of locality data for *R. osculus* (Houston et al. 2012), or a deficit of samples from outside of the Pacific Northwest (Haas 2001).

Rhinichthys falcatus and *R. umatilla* are morphologically diagnosable, as are many subspecies of *R. osculus* found in or near their native range (Markle 2016). However, the former are scarcely separated from *R. osculus* of the Middle Snake River, based upon both nuclear and mitochondrial markers. Given these results, *R. falcatus* and *R. umatilla* should be considered as subspecies of *R. osculus*, at least until a comprehensive morphological study and re-description can be conducted.

Our results question the hybrid origin of *R. umatilla* (i.e., *R. falcatus* x *R. osculus*: Gilbert and Evermann 1894), as previously interpreted from genetic and morphological evidence that employed a mitochondrial gene (*Cytochrome B*) and one nuclear locus (internal transcribed spacer) (Haas 2001). From these results, Haas (2001) hypothesized that *R. umatilla* resulted from a hybridization event that occurred in large glacial lakes of the Snake and Columbia basins during the Wisconsin glaciation, then dispersed to its modern distribution. However, our analyses found just 7.7% of tests in HyDe were congruent with a hybrid origin hypothesis. Interestingly, just four populations from the Snake River provide statistically significant evidence for a hybrid origin of *R. umatilla*, while the remaining eight did not. The Columbia River did not show evidence of involvement in a hybrid origin for *R. umatilla* either. These results combined with *COI* data suggest *R. umatilla* is instead a morphological variant of *R. osculus*.

Tiaroga cobitis

Tiaroga is currently recognized by the American Fisheries Society (Page et al. 2013) as a *Rhinichthys* synonym. It was previously noted as "*Rhinichthys*-like" (Lee et al. 1980), but with synonymy based on results of a morphological study of *Rhinichthys*, *Tiaroga*, and *Agosia* (Woodman 1992). The phylogeny that supported *T. cobitis* as Rhinichthys was based upon just eleven morphological characters, of which two were uninformative (autapomorphic), two falsified the cladogram, and seven supported it as completely resolved (Woodman 1992). However, the inclusion of *Tiaroga* and *Agosia* within *Rhinichthys* was based primarily upon a single character: the adductor mandibulae muscle complex (Woodman 1992).

This inclusion also conflicts with subsequent genetic work. Schönhuth et al. (2012) applied three nuclear loci and a mitochondrial gene to allocate western North American minnows. Their results allocated *Agosia* as sister to *Algansea*, with *Tiaroga* displaying variable relationships. Nuclear loci recovered *Oregonichthys* as more closely related to *Rhinichthys* than *Tiaroga*, and the mitochondrial phylogeny recovered *Oregonichthys* as sister to *Tiaroga*. Schönhuth et al. (2018) later used two mitochondrial and two nuclear loci to evaluate Leuciscinae that reinforced the *Oregonichthys*/ *Tiaroga* relationship. The inclusion of *Tiaroga* to yield a monophyletic *Rhinichthys* would therefore necessitate a synonymization of at least *Oregonichthys*, and possibly several additional genera: *Agosia*, *Algansea*, *Erimonax*, *Exoglossum*, *Hybopsis*, *Platygobio*, and *Yuriria*. In this regard, our results agree with those of Schönhuth et al. (2012, 2018).

Patterson's D-statistic did recover statistically significant, although scant, evidence of historical introgression. Hybridization of these two species is unlikely due to differences in reproductive mode, with unique reproductive behaviors cancelling their early spring spawning periods (Minckley and Marsh 2009). For example, *Tiaroga* is the only cyprinid known to place egg clumps in cavities under rocks (Johnston and Page 1992), coupled with the fact that males display egg-guarding behavior (Vives and Minckley 1990). We thus recommend that *Tiaroga* be resurrected, as supported by molecular evidence, reproductive ecology, and the failure of morphological analyses to separate the species.

Evolutionary Significant Units

It is important to delineate ESUs (Ryder 1986; Waples 1991; Moritz 1994) as a mechanism to promote the conservation of *R. osculus*. ESUs provide a mechanistic explanation

for the distributions of endemic western fishes, and thus represent geographically isolated lineages for which a specific or subspecific name has not been applied. Although given its wide distribution, *R. osculus* seems at first to be an atypical target for conservation. Yet many of its composite forms are not only narrowly endemic but also imperiled due to anthropogenic water use, among others (Moyle et al. 2013). As evidence, relatively stable populations such as those in the Gila River, experienced a 16.5% decline when pre-1980 records of occurrence were compared to those of the late 1990s (Olden and Poff 2005).

A total of 26 ESUs are recognized: Two represent the Columbia River and Willamette/ Coastal Oregon groups in the Columbia clade. The latter is potentially worthy of further subdivision, given the morphological differences found between the Willamette and coastal drainages (Zirges 1973). Our samples included but a single Willamette locality, so we cannot confirm its monophyly. At least three ESUs should be designated for *R. o. robustus*: Humboldt River, Smoke Creek, and the Walker Subbasin. However, the complex relationships among these basins requires additional study in that several major rivers (i.e., Carson, Truckee, and Quinn) were not sampled.

Similarly, the undescribed Sacramento subspecies seemingly contains two ESUs (The Pit River and the remainder of the Sacramento), whereas the physical separation of the San Benito from the Sacramento drainage would represent a third. The Oregon Lakes region is represented by at least 4 ESUs: Lake Albert, Warner Lakes, Goose Lake, and Stinking Lake Spring within the Mahleur Basin. Thousand Springs Creek of the Bonneville Basin is also a distinct ESU, as are the Owyhee River and Salmon Falls creek of the Snake River drainage.

However, ESU designation is confounded within the Snake River clade, given its low phylogenetic resolution with respect to modern drainages. The Upper Snake, Bear, and Weber

rivers formed a monophyletic group and thus represent a single ESU. The remaining Middle Snake River sites (excluding Salmon Falls Creek) should be grouped as a second. The Deschutes River and its tributary (Crooked River) should represent a third. Since many of these ESUs do not reflect modern hydrology, it may be necessary to designate management units within these ESUs, pending population genetic studies.

The Colorado River clade itself manifests several ESUs, with three designated as the Little Colorado, Bill Williams, and Upper Colorado/ Grand Canyon. Within the Virgin River system, previous subspecific designations were recovered, whereas the mainstem Virgin River and its tributaries represent a fourth ESU. Although the mainstem Virgin River does not form a monophyletic group due to historical mixing and/ or incomplete lineage sorting, it should still be recognized as separate due to its contemporary isolation from the White River, Moapa, and Meadow Valley Wash drainages. The remainder of the Lower Colorado River Basin should be divided into four ESUs composed of the Gila, San Pedro, Salt, and Agua Fria/ Santa Cruz/ Verde River clades.

Conclusion

Introgression and hybridization have played important roles in shaping the observed patterns of genetic diversity in *Rhinichthys* across the western landscape. While much of this resulted from past events (i.e., the Bonneville Flood), more modern occurrences are visible in the analysis of hybrid *R. cataractae* x *R. osculus*. The widespread distribution and diversity of *R. osculus* promotes the inference that adaptation to a harsh western landscape has been promoted by hybridization, particularly given evidence that it has facilitated local adaptation in many species (Ellstrand and Schierenbeck 2000; Hovick and Whitney 2014; Mesgaran et al. 2016).

Thus, the propensity of this species to mix with conspecifics or congeners upon secondary contact, coupled with its recognized status as a headwater specialist, are major components of its evolutionary success and promoted rapid dispersal following stream capture events.

Our results substantiate the monophyly of most *R. osculus* subspecies, with exceptions being widespread forms such as *R. o. yarrowi* and *R. o. robustus* that reside in areas with complicated geological histories. In addition, most undescribed subspecies were recovered as monophyletic, yet remain as undescribed. Our re-analysis of *Tiaroga* indicates that it should be resurrected as well. Our evaluation of *R. falcatus* and *R. umatilla* implied that these taxa should be recognized as morphological subspecies of *R. osculus* from the Middle Snake River. The latter recommendations highlight the need for a revision of the genus *Rhinichthys* employing morphological as well as genomic data, to include formal descriptions of unique species and subspecies, as well as the conservation of narrowly endemic and endangered lineages.

References

Aalto K (2006) The Klamath peneplain: a review of J.S. Diller's classic erosion surface. In: Snoke AW, Barnes CG (eds) Geological Studies in the Klamath Mountains Province, California and Oregon: A Volume in Honor of William P. Irwin. The Geological Society of America, Boulder, CO, pp 451–464

Abbott R, Albach D, Ansell S, et al (2013) Hybridization and speciation. J Evol Biol 26:229–246. doi: 10.1111/j.1420-9101.2012.02599.x

Aberer AJ, Kobert K, Stamatakis A (2014) ExaBayes: massively parallel Bayesian tree inference for the whole-genome era. Mol Biol Evol 31:2553–2556. doi: 10.1093/molbev/msu236

Anderson TK (2008) Inferring bedrock uplift in the Klamath Mountains Province from river profile analysis and digital topography. Master's Thesis, Texas Tech University

April J, Mayden RL, Hanner RH, Bernatchez L (2011) Genetic calibration of species diversity among North America's freshwater fishes. Proc Natl Acad Sci U S A 108:10602–10607

Ardren WR, Baumsteiger J, Allen CS (2010) Genetic analysis and uncertain taxonomic status of threatened Foskett Spring speckled dace. Conserv Genet 11:1299–1315. doi: 10.1007/s10592-009-9959-0

Arnold ML (1992) Natural hybridization as an evolutionary process. Annu Rev Ecol Syst 23:237–261

Bangs MR, Douglas MR, Mussmann SM, Douglas ME (2018) Unraveling historical introgression and resolving phylogenetic discord within *Catostomus* (Osteichthys: Catostomidae). BMC Evol Biol 18:86. doi: 10.1186/s12862-018-1197-y

Barton NH (2001) The role of hybridization in evolution. Mol Ecol 10:551–568

Behnke R (1979) Monograph of the Native Trouts of the Genus *Salmo* of Western North America. US Fish and Wildlife Service, Denver, CO

Billman EJ, Lee JB, Young DO, et al (2010) Phylogenetic divergence in a desert fish: differentiation of speckled dace within the Bonneville, Lahontan, and upper Snake River basins. West North Am Nat 70:39–47

Blischak PD, Chifman J, Wolfe AD, Kubatko LS (2018) HyDe: a Python package for genome-scale hybridization detection. Syst Biol syy023–syy023. doi: 10.1093/sysbio/syy023

Borley K, White MM (2006) Mitochondrial DNA variation in the endangered Colorado pikeminnow: a comparison among hatchery stocks and historic specimens. North Am J Fish Manag 26:916–920. doi: 10.1577/M05-176.1

Broughton J, Smith G (2016) The Fishes of Lake Bonneville: Implications for Drainage History, Biogeography, and Lake Levels. In: Shroder JF (ed) Developments in Earth Surface Processes. Elsevier, pp 292–351

Buerkle CA (2005) Maximum-likelihood estimation of a hybrid index based on molecular markers. Mol Ecol Notes 5:684–687. doi: 10.1111/j.1471-8286.2005.01011.x

Case TJ, Taper ML (2000) Interspecific competition, environmental gradients, gene flow, and the coevolution of species' borders. Am Nat 155:583–605. doi: 10.1086/303351

Catchen J, Hohenlohe PA, Bassham S, et al (2013) Stacks: an analysis tool set for population genomics. Mol Ecol 22:3124–3140. doi: 10.1111/mec.12354

Cayan DR, Das T, Pierce DW, et al (2010) Future dryness in the southwest US and the hydrology of the early 21st century drought. Proc Natl Acad Sci U S A 107:21271–21276

Chafin TK, Martin BT, Mussmann SM, et al (2017) FRAGMATIC: *in silico* locus prediction and its utility in optimizing ddRADseq projects. Conserv Genet Resour 1–4

Clement M, Posada D, Crandall KA (2000) TCS: a computer program to estimate gene genealogies. Mol Ecol 9:1657–1659

Cornelius R (1969) The Systematics and Zoogeography of *Rhinichthys osculus* (Girard) in Southern California. Master's Thesis, California State University Fullerton

Coyne JA, Orr HA (1998) The evolutionary genetics of speciation. Philos Trans R Soc B Biol Sci 353:287–305

Dasmahapatra KK, Walters JR, Briscoe AD, et al (2012) Butterfly genome reveals promiscuous exchange of mimicry adaptations among species. Nature 487:94–98. doi: 10.1038/nature11041

Deacon JE, Williams JE (1984) Annotated list of the fishes of Nevada. Proc Biol Soc Wash 97:103–118

Demarais BD, Dowling TE, Douglas ME, et al (1992) Origin of *Gila seminuda* (Teleostei: Cyprinidae) through introgressive hybridization: implications for evolution and conservation. Proc Natl Acad Sci U S A 89:2747–2751

Douglas MR, Brunner PC, Douglas ME (2003) Drought in an evolutionary context: molecular variability in Flannelmouth Sucker (*Catostomus latipinnis*) from the Colorado River Basin of western North America. Freshw Biol 48:1254–1273. doi: 10.1046/j.1365-2427.2003.01088.x

Douglas MR, Davis MA, Amarello M, et al (2016) Anthropogenic impacts drive niche and conservation metrics of a cryptic rattlesnake on the Colorado Plateau of western North America. R Soc Open Sci 3:. doi: 10.1098/rsos.160047

Dowling TE, DeMarais BD (1993) Evolutionary significance of introgressive hybridization in cyprinid fishes. Nature 362:444–446

Dowling TE, Markle DF, Tranah GJ, et al (2016) Introgressive hybridization and the evolution of lake-adapted Catostomid fishes. PLoS ONE 11:1–27. doi: 10.1371/journal.pone.0149884

Dowling TE, Secor CL (1997) The role of hybridization and introgression in the diversification of animals. Annu Rev Ecol Syst 28:593–619

Durand EY, Patterson N, Reich D, Slatkin M (2011) Testing for ancient admixture between closely related populations. Mol Biol Evol 28:2239–2252. doi: 10.1093/molbev/msr048

Eaton DA (2014) PyRAD: assembly of *de novo* RADseq loci for phylogenetic analyses. Bioinformatics 30:1844–1849. doi: 10.1093/bioinformatics/btu121

Edmands S (2002) Does parental divergence predict reproductive compatibility? Trends Ecol Evol 17:520–527

Edwards SV, Xi Z, Janke A, et al (2016) Implementing and testing the multispecies coalescent model: a valuable paradigm for phylogenomics. Mol Phylogenet Evol 94:447–462. doi: 10.1016/j.ympev.2015.10.027

Ellstrand NC, Schierenbeck KA (2000) Hybridization as a stimulus for the evolution of invasiveness in plants? Proc Natl Acad Sci U S A 97:7043. doi: 10.1073/pnas.97.13.7043

Enzel Y, Wells SG, Lancaster N (2003) Late Pleistocene lakes along the Mojave River, southeast California. In: Enzel Y, Wells SG, Lancaster N (eds) Paleoenvironments and Paleohydrology of the Mojave and Southern Great Basin Deserts. Geological Society of America, Boulder, CO, pp 61–78

Gerber AS, Tibbets CA, Dowling TE (2001) The role of introgressive hybridization in the evolution of the *Gila robusta* complex (Teleostei: Cyprinidae). Evolution 55:2028–2039. doi: 10.1111/j.0014-3820.2001.tb01319.x

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 7:229–234

Gilbert CH, Evermann BW (1894) A report upon investigations in the Columbia River Basin: with descriptions of four new species of fishes. Bull US Fish Comm 14:169–207

Girard C (1856) Researches upon the cyprinoid fishes inhabiting the fresh waters of the United States, west of the Mississippi Valley, from specimens in the museum of the Smithsonian Institution. Proc Acad Nat Sci Phila 8:165–218

Gompert Z, Buerkle CA (2010) INTROGRESS: a software package for mapping components of isolation in hybrids. Mol Ecol Resour 10:378–384

Gompert Z, Lucas LK, Nice CC, et al (2012) Genomic regions with a history of divergent selection affect fitness of hybrids between two butterfly species: genomics of speciation. Evolution 66:2167–2181. doi: 10.1111/j.1558-5646.2012.01587.x

Grant PR, Grant BR (1992) Hybridization of bird species. Science 256:193–197

Green DM, Sharbel TF, Kearsley J, Kaiser H (1996) Postglacial range fluctuation, genetic subdivision and speciation in the western North American spotted frog complex, *Rana pretiosa*. Evolution 50:374–390. doi: 10.2307/2410808

Haas GR (2001) The evolution through natural hybridizations of the Umatilla Dace (Pisces: *Rhinichthys umatilla*), and their associated ecology and systematics

Harrison HB, Berumen ML, Saenz-Agudelo P, et al (2017) Widespread hybridization and bidirectional introgression in sympatric species of coral reef fish. Mol Ecol 26:5692–5704

Harrison RG, Larson EL (2014) Hybridization, introgression, and the nature of species boundaries. J Hered 105:795–809

Hershler R, Liu H-P (2004) A molecular phylogeny of aquatic gastropods provides a new perspective on biogeographic history of the Snake River region. Mol Phylogenet Evol 32:927–937

Hoang DT, Chernomor O, von Haeseler A, et al (2018) UFBoot2: improving the ultrafast bootstrap approximation. Mol Biol Evol 35:518–522

Hoekzema K, Sidlauskas BL (2014) Molecular phylogenetics and microsatellite analysis reveal cryptic species of speckled dace (Cyprinidae: *Rhinichthys osculus*) in Oregon's Great Basin. Mol Phylogenet Evol 77:238–250. doi: 10.1016/j.ympev.2014.04.027

Hopken MW, Douglas MR, Douglas ME (2013) Stream hierarchy defines riverscape genetics of a North American desert fish. Mol Ecol 22:956–971. doi: 10.1111/mec.12156

Houston DD (2009) Molecular systematics and phylogeography of the genus *Richardsonius*. Doctoral Dissertation, University of Nevada Las Vegas

Houston DD, Evans RP, Shiozawa DK (2012) Evaluating the genetic status of a Great Basin endemic minnow: the relict dace (*Relictus solitarius*). Conserv Genet 13:727–742. doi: 10.1007/s10592-012-0321-6

Hovick SM, Whitney KD (2014) Hybridisation is associated with increased fecundity and size in invasive taxa: meta-analytic support for the hybridisation-invasion hypothesis. Ecol Lett 17:1464–1477. doi: 10.1111/ele.12355

Hubbs CL (1979) List of the fishes of California. Occas Pap Calif Acad Sci 133:1–51

Hubbs CL, Miller RR (1948) The zoological evidence: correlation between fish distribution and hydrographic history in the desert basins of western United States. In: The Great Basin, with Emphasis on Glacial and Postglacial Times. University of Utah, Salt Lake City, UT, pp 17–166

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

Hubert N, Hanner R, Holm E, et al (2008) Identifying Canadian freshwater fishes through DNA barcodes. PLoS ONE 3:e2490

Hubert N, Renno J-F (2006) Historical biogeography of South American freshwater fishes. J Biogeogr 33:1414–1436. doi: 10.1111/j.1365-2699.2006.01518.x

Johnson JB (2002) Evolution after the flood: phylogeography of the desert fish Utah chub. Evolution 56:948–960

Johnston CE, Page LM (1992) The evolution of complex reproductive strategies in North American minnows (Cyprinidae). In: Mayden RL (ed) Systematics, Historical Ecology, and North American Freshwater Fishes. Stanford University Press, Stanford, CA, pp 600–621

Jordan DS (1889) Report of explorations in Colorado and Utah during the summer of 1889, with an account of the fishes found in each of the river basins examined. Bull US Fish Comm 9:1–63

Jordan DS, Evermann BW (1896) The fishes of North and Middle America. Bull U S Natl Mus 47:1–1236

Jordan DS, Evermann BW, Clark HW (1930) Check list of the fishes and fishlike vertebrates of North and Middle America north of the northern boundary of Venezuela and Colombia. Rep US Comm Fish Fisc Year 1928 2:1–670

Kim D, Conway KW (2014) Phylogeography of *Rhinichthys cataractae* (Teleostei: Cyprinidae): pre-glacial colonization across the Continental Divide and Pleistocene diversification within the Rio Grande drainage. Biol J Linn Soc 111:317–333

Kisel Y, Barraclough TG (2010) Speciation has a spatial scale that depends on levels of gene flow. Am Nat 175:316–334. doi: 10.1086/650369

Kozlov AM, Aberer AJ, Stamatakis A (2015) ExaML version 3: a tool for phylogenomic analyses on supercomputers. Bioinformatics 31:2577–2579

Lee DS, Gilbert CR, Hocutt CH, et al (1980) Atlas of North American Freshwater Fishes. North Carolina State Museum of Natural History, Raleigh, NC

Leigh JW, Bryant D (2015) POPART: full-feature software for haplotype network construction. Methods Ecol Evol 6:1110–1116

Lessios HA, Kessing BD, Pearse JS (2001) Population structure and speciation in tropical seas: global phylogeography of the sea urchin *Diadema*. Evolution 55:955–975. doi: 10.1554/0014-3820(2001)055[0955:PSASIT]2.0.CO;2

Link PK, Fanning CM, Beranek LP (2005) Reliability and longitudinal change of detrital-zircon age spectra in the Snake River system, Idaho and Wyoming: An example of reproducing the bumpy barcode. Sediment Geol 182:101–142

Link PK, Kaufman DS, Thackray GD (1999) Field guide to Pleistocene lakes Thatcher and Bonneville and the Bonneville flood, southeastern Idaho. In: Hughes SS, Thackray GD (eds) Guidebook to the Geology of Eastern Idaho. Idaho Museum of Natural History, Pocatello, ID, pp 251–266

Link PK, McDonald HG, Fanning C, Godfrey AE (2002) Detrital zircon evidence for Pleistocene drainage reversal at Hagerman Fossil Beds National Monument, central Snake River Plain, Idaho. In: Bonnichsen B, White CM, McCurry M (eds) Tectonic and Magmatic Evolution of the Snake River Plain Volcanic Province, Idaho Geological Survey Bulletin. pp 105–119

Loxterman JL, Keeley ER (2012) Watershed boundaries and geographic isolation: patterns of diversification in cutthroat trout from western North America. BMC Evol Biol 12:38–54

Maddison W, Knowles L (2006) Inferring phylogeny despite incomplete lineage sorting. Syst Biol 55:21–30. doi: 10.1080/10635150500354928

Maddison WP (1997) Gene trees in species trees. Syst Biol 46:523–536. doi: 10.2307/2413694

Magill J, Cox A (1981) Post-Oligocene tectonic rotation of the Oregon western Cascade Range and the Klamath Mountains. Geology 9:127–131

Malde HE (1968) The catastrophic late Pleistocene Bonneville flood in the Snake River plain, Idaho. In: Geological Survey Professional Paper 596. US Government Printing Office, Washington, DC, pp 1–69

Mallet J (2005) Hybridization as an invasion of the genome. Trends Ecol Evol 20:229–237. doi: 10.1016/j.tree.2005.02.010

Markle DF (2016) A Guide to Freshwater Fishes of Oregon. Oregon State University Press, Corvallis, OR

Mayr E (1942) Systematics and the Origin of Species, from the Viewpoint of a Zoologist. Harvard University Press, Cambridge, MA

McPhail J, Lindsey C (1986) Zoogeography of the freshwater fishes of Cascadia (the Columbia system and rivers north to the Stikine). In: Hocutt CH, Wiley EO (eds) The Zoogeography of North American Freshwater fishes. John Wiley & Sons, New York, pp 615–637

McPhail J, Taylor E (2009) Phylogeography of the longnose dace (*Rhinichthys cataractae*) species group in northwestern North America—the origin and evolution of the Umpqua and Millicoma dace. Can J Zool 87:491–497

McPhail JD (2007) The Freshwater Fishes of British Columbia. University of Alberta, Edmonton, AB

Mesgaran MB, Lewis MA, Ades PK, et al (2016) Hybridization can facilitate species invasions, even without enhancing local adaptation. Proc Natl Acad Sci U S A 113:10210. doi: 10.1073/pnas.1605626113

Miller RR (1958) Origin and affinities of the freshwater fish fauna of western North America. In: Zoogeography. American Society for the Advancement of Science, Washington, DC, pp 187–222

Miller RR (1984) *Rhinichthys deaconi*, a new species of dace (Pisces: Cyprinidae) from southern Nevada. Occas Pap Mus Zool Univ Mich 707:1–21

Miller RR, Miller RG (1948) The contribution of the Columbia River system to the fish fauna of Nevada: five species unrecorded from the state. Copeia 1948:174–187. doi: 10.2307/1438452

Miller RR, Smith GR (1981) Distribution and evolution of *Chasmistes* (Pisces: Catostomidae) in western North America. Occas Pap Mus Zool Univ Mich 696:1–46

Minckley W, Hendrickson D, Bond C (1986) Geography of western North American freshwater fishes: description and relationships to intracontinental tectonism. In: Hocutt CH, Wiley EO (eds) The Zoogeography of North American Freshwater Fishes. John Wiley & Sons, New York, pp 519–613

Minckley W, Marsh PC (2009) Inland Fishes of the Greater Southwest: Chronicle of a Vanishing Biota. University of Arizona Press, Tucson, AZ

Minckley WL (1973) Fishes of Arizona. Arizona Game and Fish Department, Phoenix, AZ

Moritz C (1994) Defining “evolutionarily significant units” for conservation. Trends Ecol Evol 9:373–374

Moyle PB (2002) Inland Fishes of California. University of California Press, Berkeley, CA

Moyle PB, Kiernan JD, Crain PK, Quiñones RM (2013) Climate change vulnerability of native and alien freshwater fishes of California: a systematic assessment approach. PLoS ONE 8:e63883. doi: 10.1371/journal.pone.0063883

Nguyen L-T, Schmidt HA, von Haeseler A, Minh BQ (2014) IQ-TREE: a fast and effective stochastic algorithm for estimating maximum-likelihood phylogenies. Mol Biol Evol 32:268–274

Nosil P, Vines TH, Funk DJ (2005) Reproductive isolation caused by natural selection against immigrants from divergent habitats. Evolution 59:705–719. doi: 10.1111/j.0014-3820.2005.tb01747.x

Oakey DD, Douglas ME, Douglas MR (2004) Small fish in a large landscape: diversification of *Rhinichthys osculus* (Cyprinidae) in western North America. Copeia 2004:207–221

Olden J, Poff N (2005) Long-term trends of native and non-native fish faunas in the American Southwest. Anim Biodivers Conserv 28:75–89

Oviatt CG (1997) Lake Bonneville fluctuations and global climate change. Geology 25:155–158

Page LM, Espinosa-Pérez H, Findley LT, et al (2013) Common and Scientific Names of Fishes from the United States, Canada, and Mexico. American Fisheries Society, Bethesda, MD

Peterson BK, Weber JN, Kay EH, et al (2012) Double digest RADseq: an inexpensive method for *de novo* SNP discovery and genotyping in model and non-model species. PLoS ONE 7:1–11. doi: 10.1371/journal.pone.0037135

Pfrender ME, Hicks J, Lynch M (2004) Biogeographic patterns and current distribution of molecular-genetic variation among populations of speckled dace, *Rhinichthys osculus* (Girard). Mol Phylogenet Evol 30:490–502. doi: 10.1016/S1055-7903(03)00242-2

Rambaut A, Drummond AJ, Xie D, et al (2018) Posterior summarization in Bayesian phylogenetics using Tracer 1.7. Syst Biol syy032–syy032. doi: 10.1093/sysbio/syy032

Räsänen K, Hendry AP (2008) Disentangling interactions between adaptive divergence and gene flow when ecology drives diversification. Ecol Lett 11:624–636

Reheis M, Morrison R (1997) High, old pluvial lakes of western Nevada. Brigh Young Univ Geol Stud 42:459–492

Repenning CA, Weasma TR, Scott GR (1995) The early Pleistocene (latest Blancan-earliest Irvingtonian) Froman Ferry fauna and history of the Glenns Ferry Formation, southwestern Idaho. US Government Printing Office, Washington, DC

Roskowski JA, Patchett PJ, Spencer JE, et al (2010) A late Miocene–early Pliocene chain of lakes fed by the Colorado River: evidence from Sr, C, and O isotopes of the Bouse Formation and related units between Grand Canyon and the Gulf of California. Geol Soc Am Bull 122:1625–1636

Ryder OA (1986) Species conservation and systematics: the dilemma of subspecies. Trends Ecol Evol 1:9–10

Sada DW, Britten HB, Brussard PF (1995) Desert aquatic ecosystems and the genetic and morphological diversity of Death Valley system speckled dace. Am Fish Soc Symp 17:350–359

Schönhuth S, Shiozawa DK, Dowling TE, Mayden RL (2012) Molecular systematics of western North American cyprinids (Cypriniformes: Cyprinidae). Zootaxa 3586:1–303

Schönhuth S, Vukić J, Šanda R, et al (2018) Phylogenetic relationships and classification of the Holarctic family Leuciscidae (Cypriniformes: Cyprinoidei). Mol Phylogenet Evol. doi: 10.1016/j.ympev.2018.06.026

Schrempf D, Minh BQ, De Maio N, et al (2016) Reversible polymorphism-aware phylogenetic models and their application to tree inference. J Theor Biol 407:362–370

Schultz LP (1936) Keys to the fishes of Washington, Oregon and closely adjoining regions. Umiversity Wash Publ Biol 2:105–228

Sigler WF, Sigler JW (1987) Fishes of the Great Basin: A Natural History. University of Nevada Press, Reno, NV

Smith G, Dowling T, Gobalet K, et al (2002) Biogeography and timing of evolutionary events among Great Basin fishes. Gt Basin Aquat Syst Hist 33:175–234

Smith GR (1978) Biogeography of intermountain fishes. Gt Basin Nat Mem 2:17–42

Smith GR (1975) Fishes of the Pliocene Glenns Ferry Formation, southwest Idaho. Univ Mich Mus Paleontol Pap Paleontol 14:1–68

Smith GR (1966) Distribution and evolution of the North American catostomid fishes of the subgenus *Pantosteus*, genus *Catostomus*. Misc Publ Mus Zool Univ Mich 129:1–137

Smith GR, Chow J, Unmack PJ, et al (2017) Evolution of the *Rhinichthys osculus* complex (Teleostei: Cyprinidae) in western North America. Misc Publ Mus Zool Univ Mich 204:1–83

Smith GR, Dowling TE (2008) Correlating hydrographic events and divergence times of speckled dace (*Rhinichthys*: Teleostei: Cyprinidae) in the Colorado River drainage. In: Reheis MC, Hershler R, Miller DM (eds) Special Paper 439: Late Cenozoic Drainage History of the Southwestern Great Basin and Lower Colorado River Region: Geologic and Biotic Perspectives. Geological Society of America, Boulder, CO, pp 301–317

Snyder JO (1905) Notes on the fishes of the streams flowing into the San Francisco Bay. Rept Bur Fish App 5:327–338

Springer MS, Gatesy J (2016) The gene tree delusion. Mol Phylogenet Evol 94:1–33. doi: 10.1016/j.ympev.2015.07.018

Stamatakis A (2014) RAxML version 8: a tool for phylogenetic analysis and post-analysis of large phylogenies. Bioinformatics 30:1312–1313

Stevens MI, Hogg ID (2003) Long-term isolation and recent range expansion from glacial refugia revealed for the endemic springtail *Gomphiocephalus hodgsoni* from Victoria Land, Antarctica. Mol Ecol 12:2357–2369. doi: 10.1046/j.1365-294X.2003.01907.x

Taylor D (1985) Evolution of freshwater drainages and molluscs in western North America. In: Smiley CJ (ed) Late Cenozoic History of the Pacific Northwest. American Association for the Advancement of Science, San Francisco, CA, pp 265–321

Unmack PJ, Dowling TE, Laitinen NJ, et al (2014) Influence of introgression and geological processes on phylogenetic relationships of western North American mountain suckers (*Pantosteus*, Catostomidae). PLoS ONE 9:e90061

VanMeter JJ (2017) The Santa Ana speckled dace (*Rhinichthys osculus*): phylogeography and molecular evolution of the mitochondrial DNA control region. Master's Thesis, California State University San Bernardino

Vives SP, Minckley W (1990) Autumn spawning and other reproductive notes on loach minnow, a threatened cyprinid fish of the American southwest. Southwest Nat 35:451–454

Wagner CE, Keller I, Wittwer S, et al (2013) Genome-wide RAD sequence data provide unprecedented resolution of species boundaries and relationships in the Lake Victoria cichlid adaptive radiation. Mol Ecol 22:787–798. doi: 10.1111/mec.12023

Waples RS (1991) Pacific salmon, *Oncorhynchus* spp., and the definition of "species" under the Endangered Species Act. Mar Fish Rev 53:11–22

Wiesenfeld JC, Goodman DH, Kinziger AP (2018) Riverscape genetics identifies speckled dace (*Rhinichthys osculus*) cryptic diversity in the Klamath–Trinity Basin. Conserv Genet 19:111–127. doi: 10.1007/s10592-017-1027-6

Williams JE (1978) Taxonomic Status of Rhinichthys osculus (Cyprinidae) in the Moapa River, Nevada. Southwest Nat 23:511. doi: 10.2307/3670257

Woodhouse CA, Meko DM, MacDonald GM, et al (2010) A 1,200-year perspective of 21st century drought in southwestern North America. Proc Natl Acad Sci U S A 107:21283–21288. doi: 10.1073/pnas.0911197107

Woodman DA (1992) Systematic relationships within the cyprinid genus *Rhinichthys*. In: Mayden RL (ed) Systematics, Historical Ecology, and North American Freshwater Fishes. Stanford University Press, Stanford, CA, pp 374–391

Zirges MH (1973) Morphological and meristic characteristics of ten populations of blackside dace, *Rhinichthys osculus nubilus* (Girard), from western Oregon. Master's Thesis, Oregon State University

Appendix

Table 1. Results from program HyDe testing the putative hybrid origin of *Rhinichthys umatilla* (i.e., *R. falcatus* x *R. osculus*). Only 11 tests presented below (out of 142) were significant. Basin = drainage from which the parental *R. osculus* population was sampled. SPD = parental *R. osculus* population. Z = Z-score test statistic. *p*(Z) = p-value associated with the Z-score.

Basin	SPD	Z	*p* (Z)
Bonneville	Mink Creek	6.53	3.3×10^{-11}
Snake River	Bruneau River	6.53	3.4×10^{-11}
	Little Wood River	5.69	6.6×10^{-9}
	Raft River	6.10	5.4×10^{-10}
	South Fork Rock Creek	7.56	2×10^{-14}
Colorado River	Big Sandy River	6.58	2.3×10^{-11}
	Desolation Canyon	6.06	6.9×10^{-10}
	Dolores River	5.53	1.6×10^{-8}
Southern California Coastal	San Gabriel River	6.23	2.4×10^{-10}
Virgin River	Eagle Valley Reservoir	6.10	5.4×10^{-10}
	Washington Fields Diversion	5.67	7.2×10^{-9}

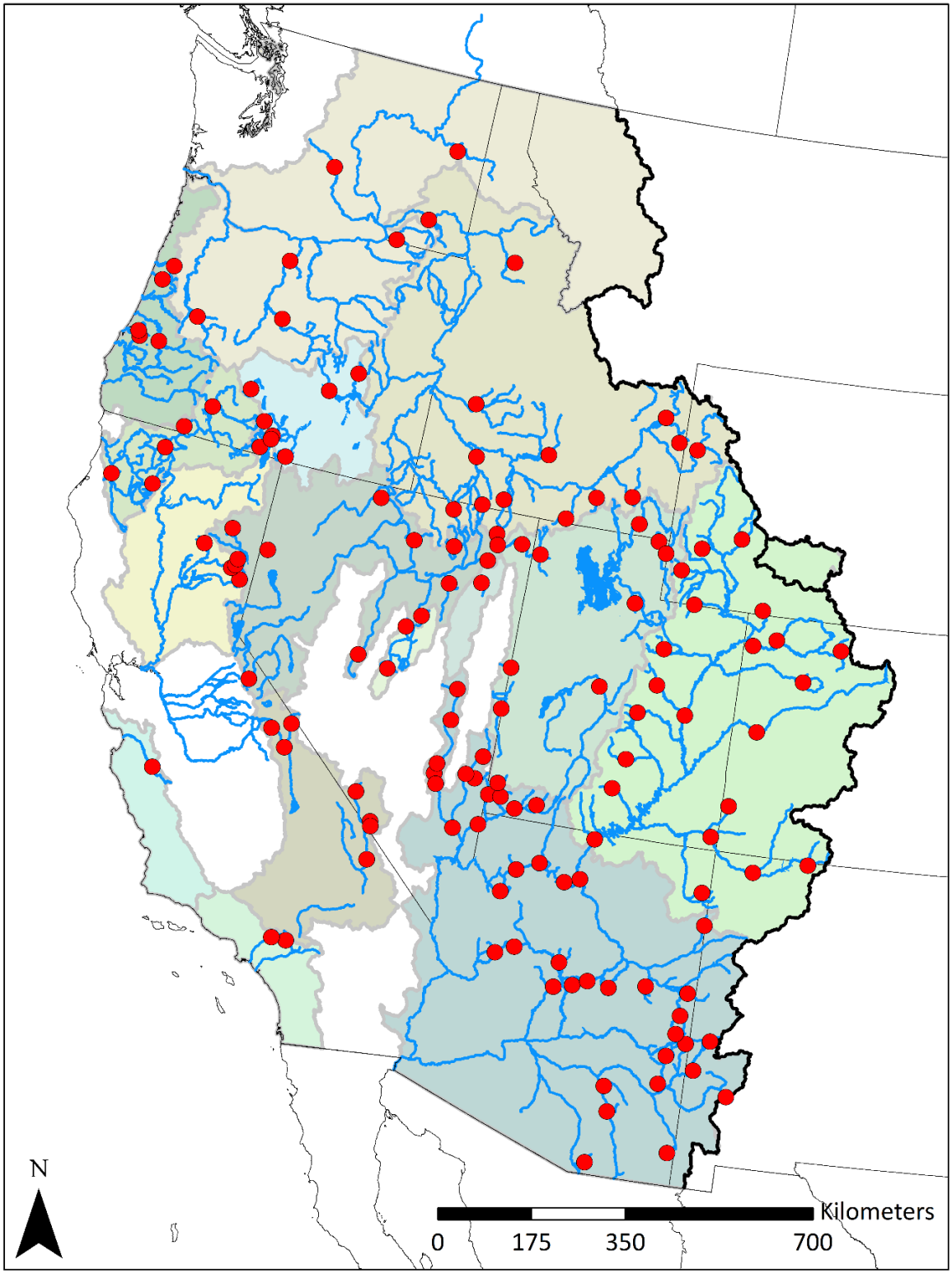

Figure 1. Speckled Dace sampled from 141 localities distributed west of the Continental Divide (designated by a thick black line). Shaded areas represent drainage basins. Blue lines represent rivers and streams. Red circles indicate sampling localities.

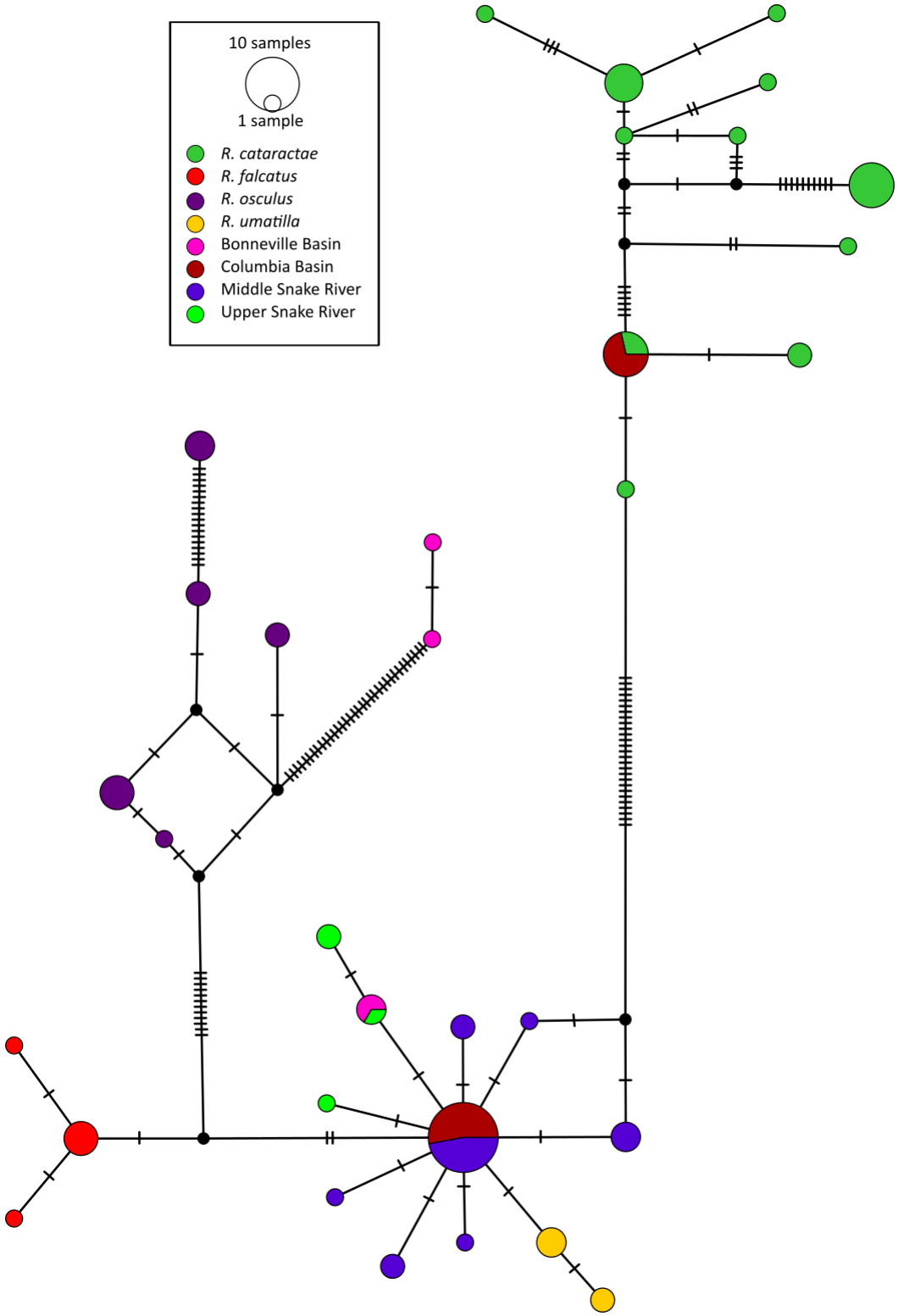

Figure 2. A haplotype network derived from 648 base pairs of the *cytochrome c oxidase I* (*COI*) mitochondrial gene showing relationships among *Rhinichthys* collected from the Bonneville, Columbia, and Snake River Basins. Colored nodes represent observed haplotypes, with their number indicated by the size of each node. Black nodes represent unobserved haplotypes. Black hash marks between nodes represent the number of mutations separating haplotypes.

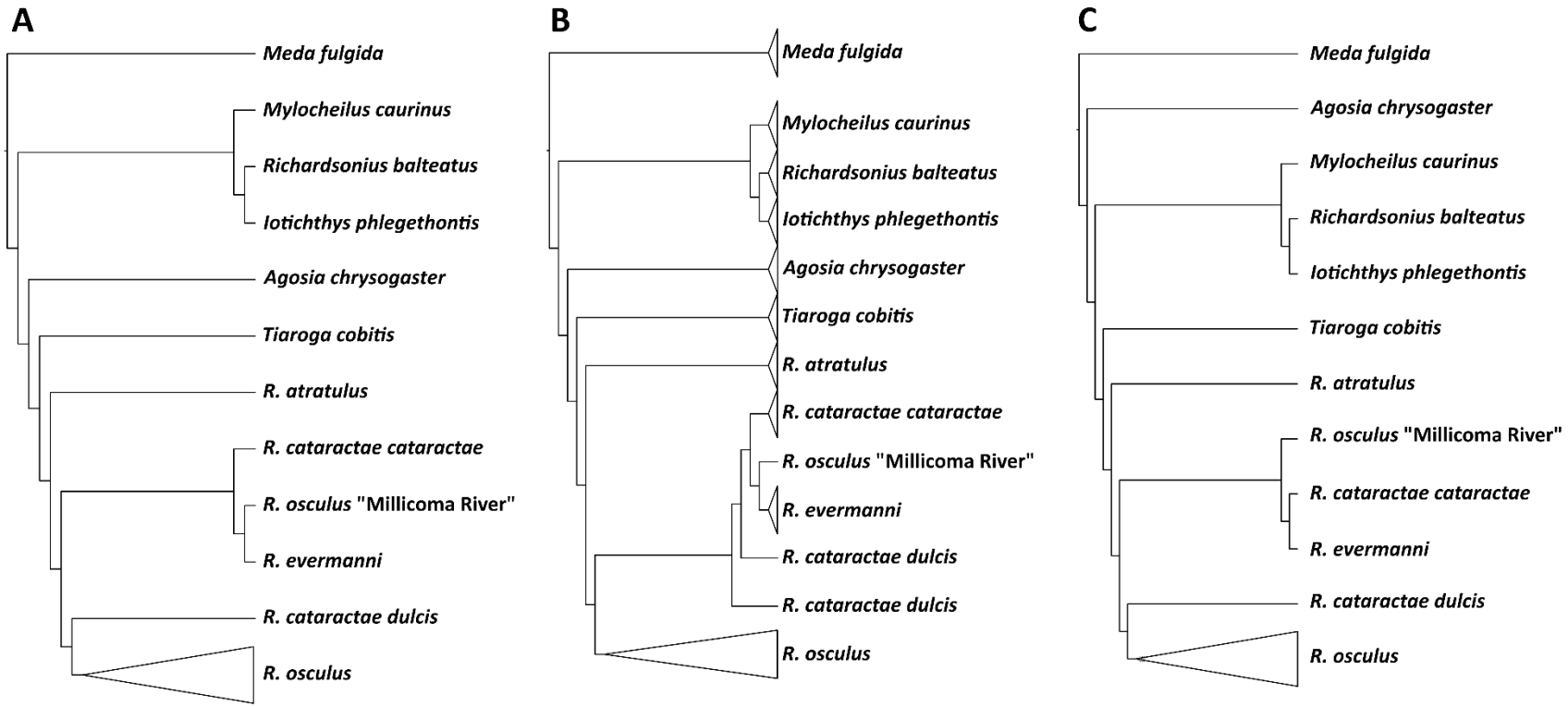

Figure 3. Phylogenetic trees presented as results from three separate methods that evaluated relationships within *Rhinichthys*: A = ExaBayes, B = ExaML, and C = IQ-TREE (PoMo). Similar relationships were recovered among non-*Rhinichthys* taxa for most analytical methods, however IQ-TREE differed in its relationship of *Agosia* to other taxa.

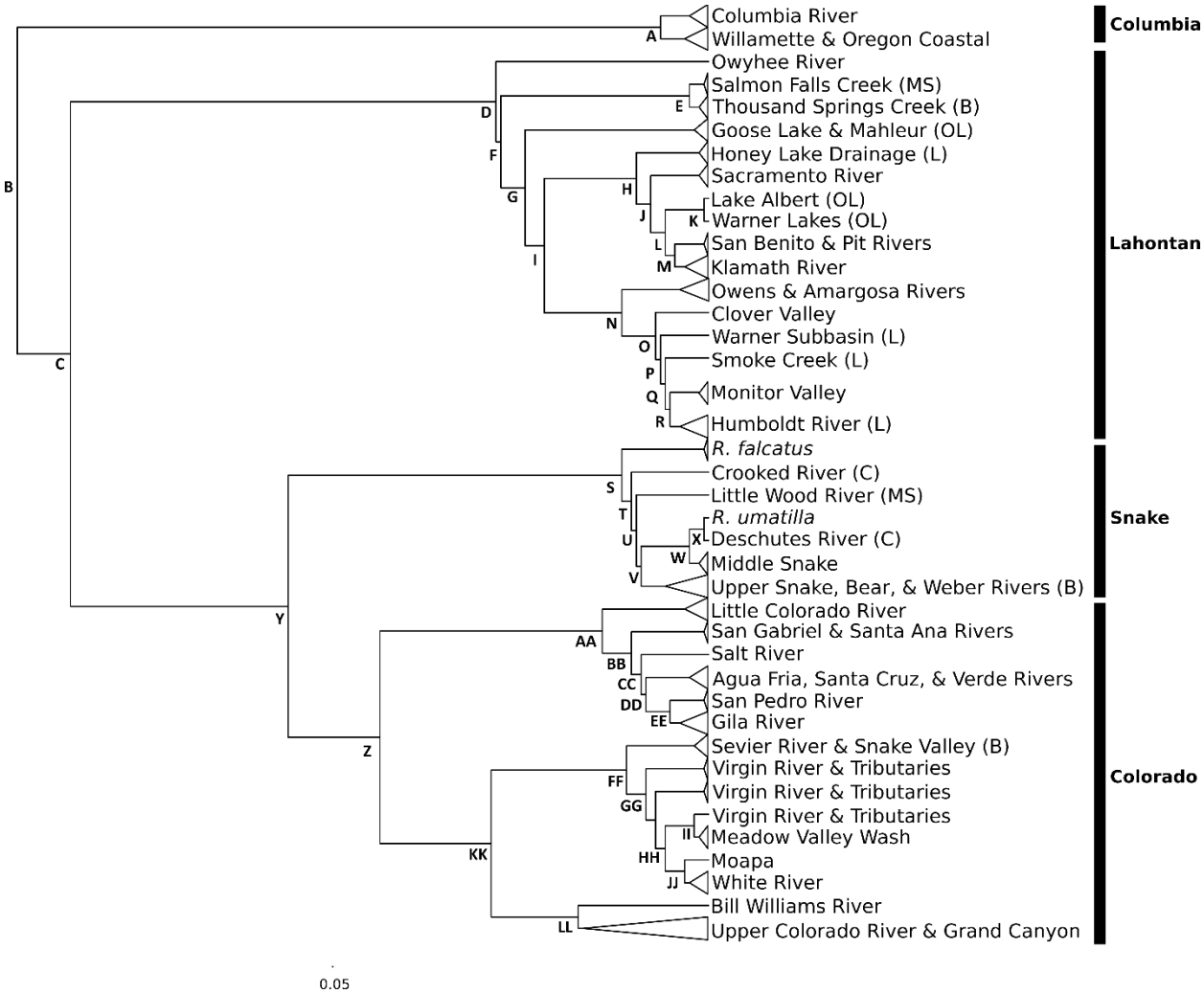

Figure 4. EXABAYES result depicting relationships within *Rhinichthys osculus*. Four major clades broadly represent the Columbia River, Lahontan Basin, Snake River, and Colorado River. All nodes were supported with a Bayesian posterior probability of 1. Labels for nodes correspond to those depicted in Figure 5.

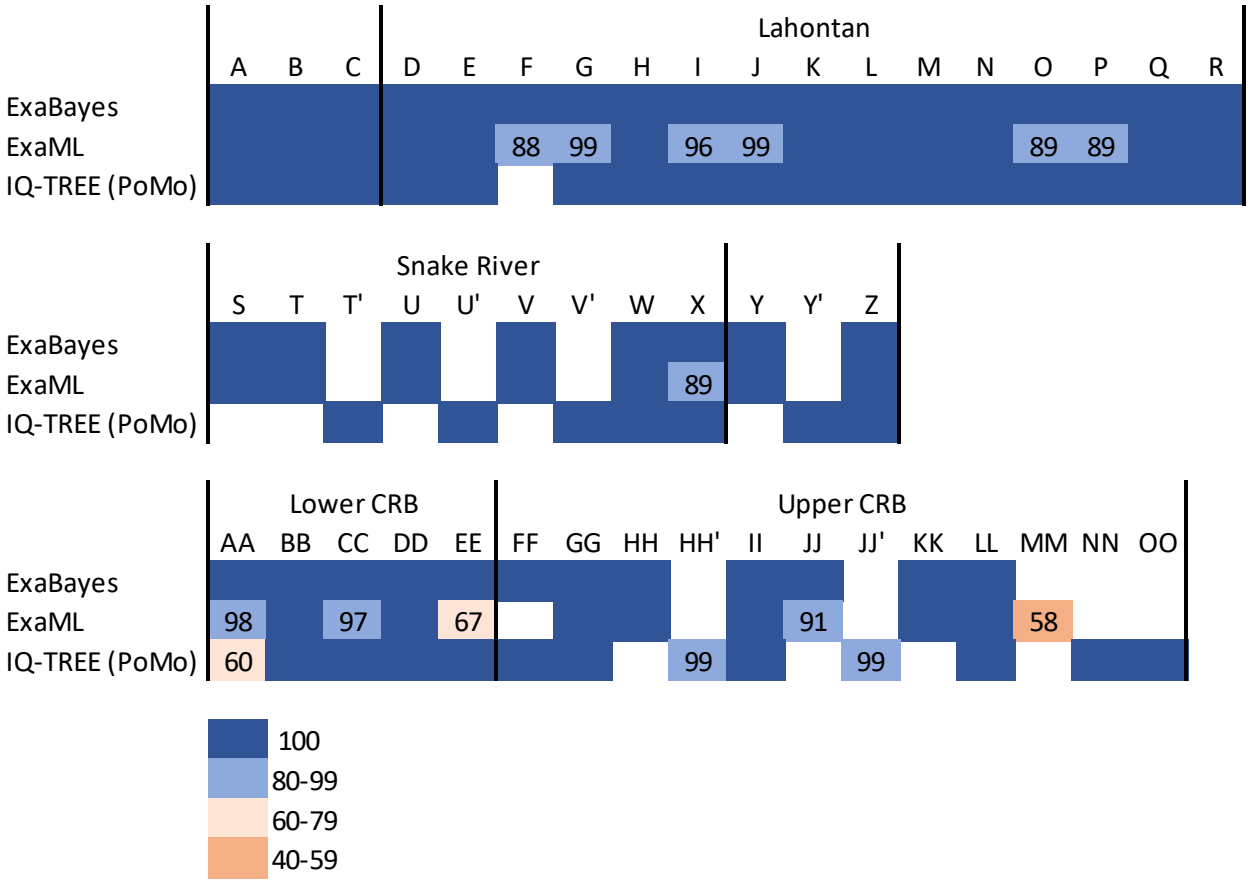

Figure 5. Diagram depicting concordance among three phylogenetic reconstructions of relationships among *Rhinichthys osculus*. IQ-TREE and ExaML showed the greatest number of nodes with low support (bootstrap support < 70). Nodes A, B, C, Y and Z represent deeper relationships within the tree that were mostly conserved among methods. Node labels correspond to the tree in Figure 4 and online supplemental material. Numbers represent bootstrap support values for ExaML and IQ-TREE. All nodes in the ExaBayes analysis were supported by Bayesian posterior probability = 1.

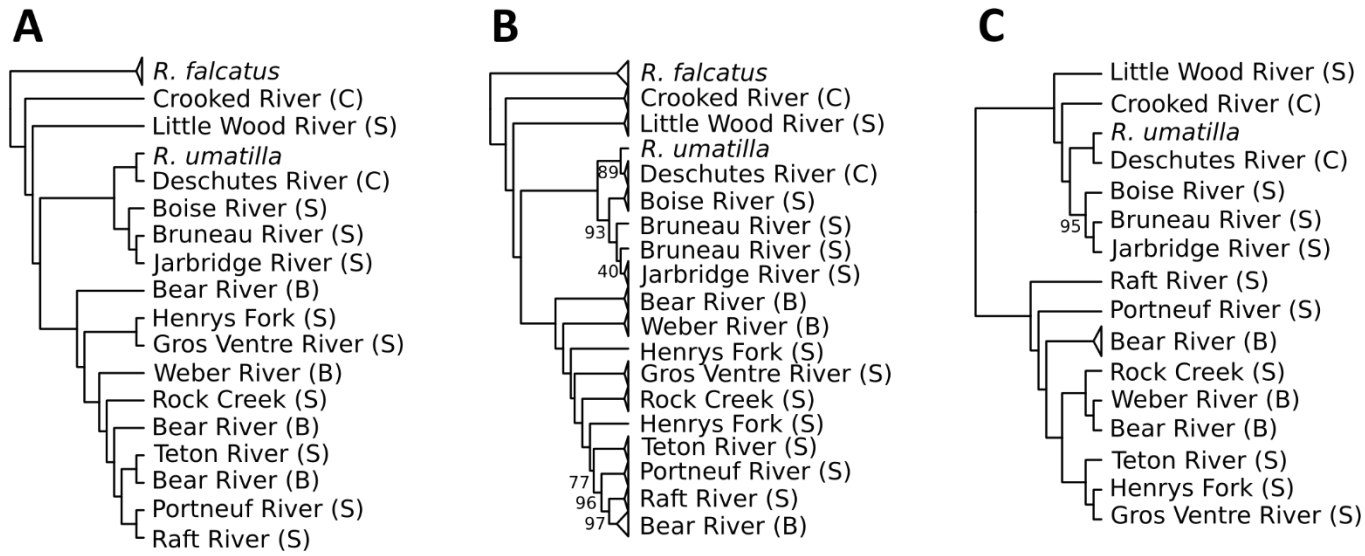

Figure 6. Phylogenetic relationships of *Rhinichthys osculus* found among rivers within the Snake River clade as depicted by three separate methodologies: A = ExaBayes, B = ExaML, and C = IQ-TREE (PoMo). *Rhinichthys falcatus* is absent from the IQ-TREE in that it was recovered as sister to the Colorado River Basin. Letters in parentheses following river names indicate the basin within which the river is found: Bonneville (B), Columbia (C), or Snake (S). All nodes in the ExaBayes result were supported by Bayesian posterior probability = 1. Numbers at nodes for the ExaML and PoMo results represent bootstrap support values. Nodes without numbers had bootstrap support of 100.

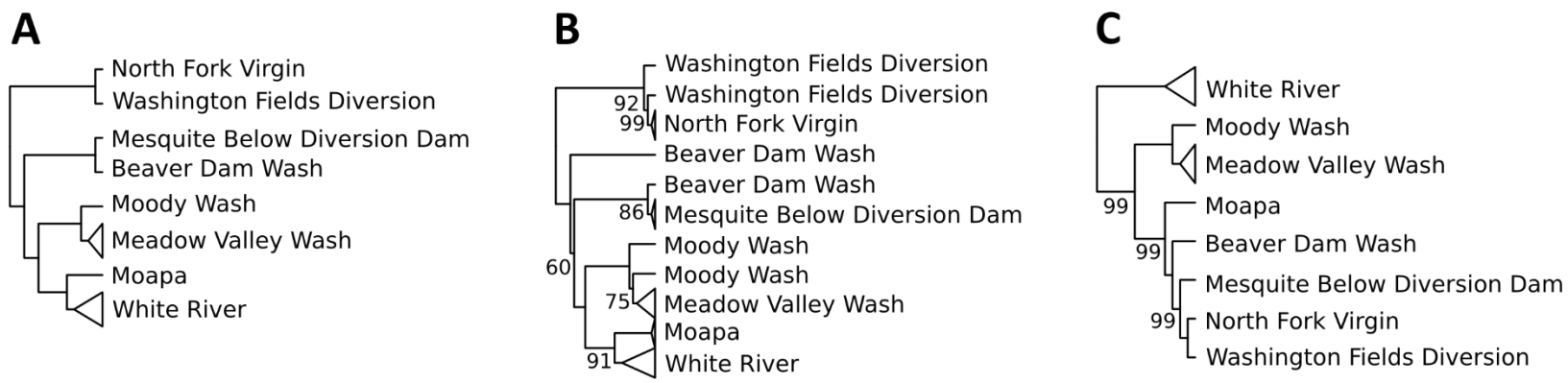

Figure 7. Phylogenetic relationships of *Rhinichthys osculus* found within the Virgin River clade. Four methods were utilized to evaluate relationships: A = ExaBayes, B = ExaML, and C = IQ-TREE (PoMo). All nodes in the ExaBayes result were supported by Bayesian posterior probability = 1. Numbers at nodes for the ExaML and PoMo results represent bootstrap support values. Nodes without numbers had bootstrap support of 100.

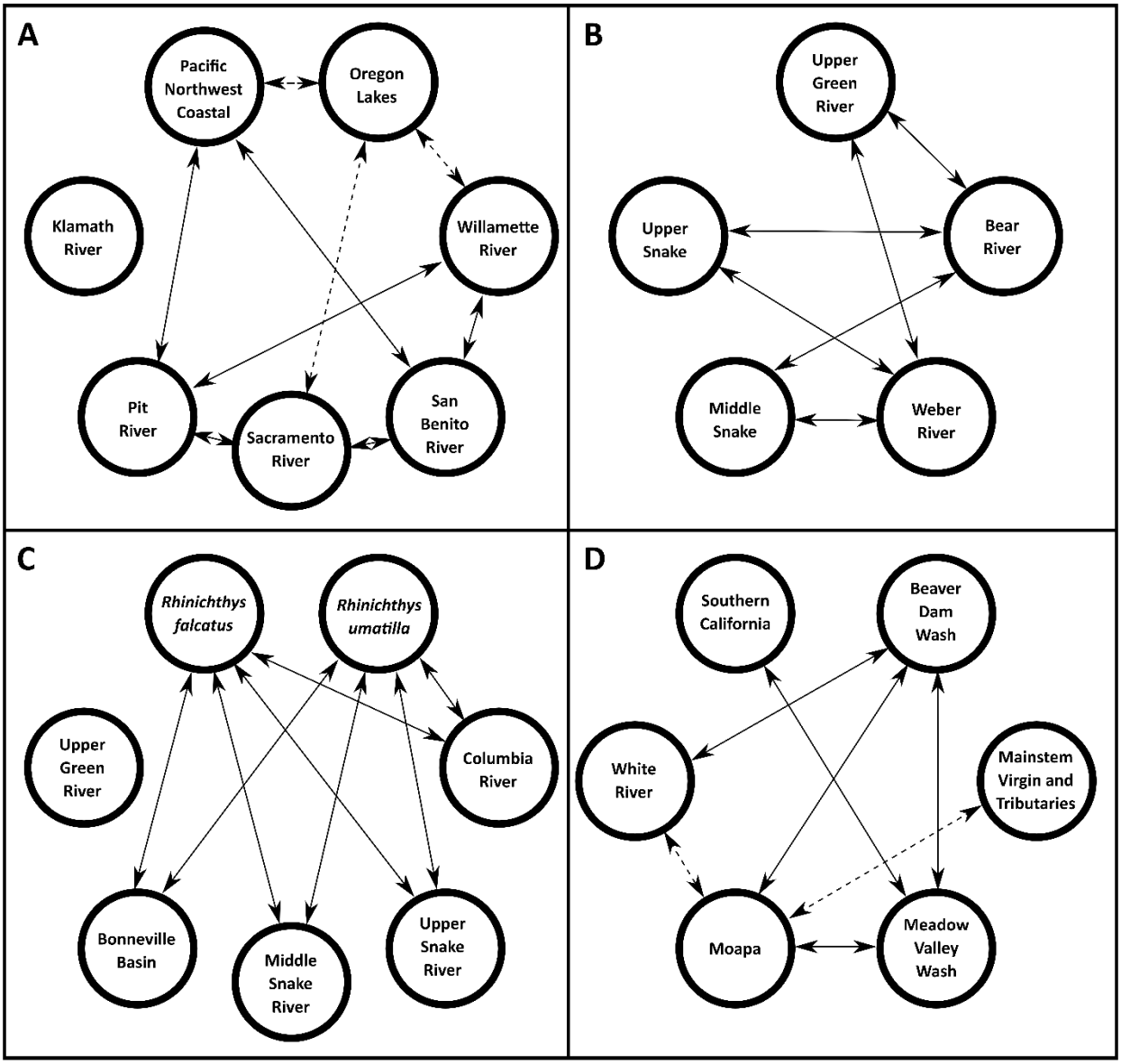

Figure 8. Patterns of introgression that involve (A) the Klamath River, (B) the Snake River, (C) *R. falcatus* and *R. umatilla*, and (D) the Virgin River. Arrows with solid lines connecting rivers indicate significant D-statistic values for >33% of tests. The dashed lines in (A) connecting the Oregon Lakes region to other basins (Pacific Coastal, Willamette, and Sacramento Rivers) indicate that significance related only to the Rattlesnake Creek drainage of the Mahleur basin. Dashed lines in (D) connecting Moapa to the White River and mainstem Virgin River represents very low levels of mixing among these rivers (< 5% of tests were significant).

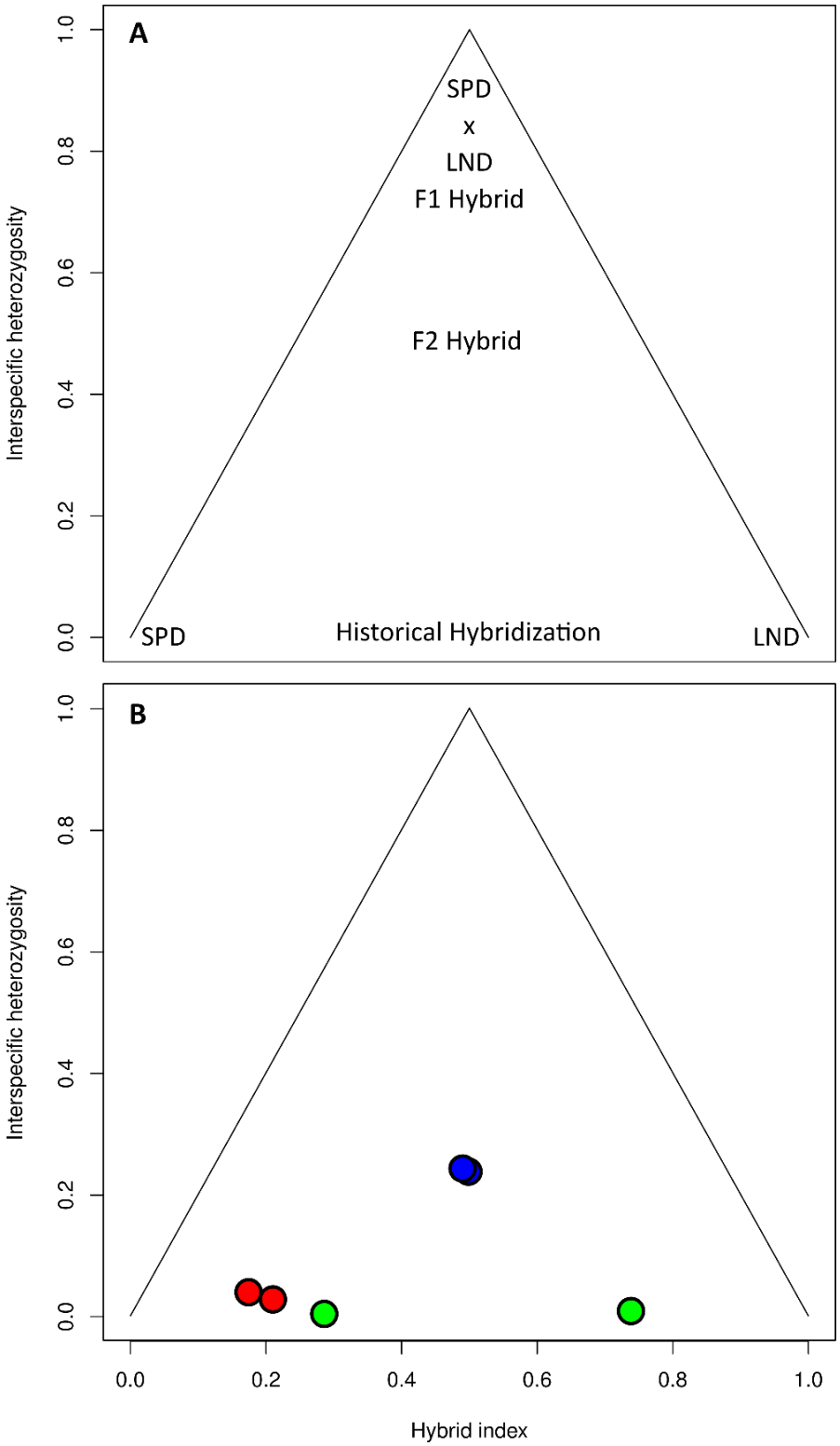

Figure 9. Hybrid index calculated for *Rhinichthys cataractae dulcis* (Blue), *R. osculus* from the Millicoma River (Green), and *R. osculus* from the Deschutes River (Red). The parental genotypes were represented by *R. c. cataractae* from the South Platte River, Colorado and *R. osculus* from the Verde River drainage, Arizona.

III. Evolution of Speckled Dace as gauged within two large, divergent Basins of western North America: The Great Basin and Colorado River

Introduction

The physical structure of watersheds (i.e., their linkages, distances, and structure) not only impact life histories of riverine fishes and drive their evolution (Hughes et al. 2009; Hopken et al. 2013) but also serve as necessary benchmarks for the Anthropocene (Gleick 2010; Vörösmarty et al. 2010). In this sense, highly dendritic stream networks often promote asymmetrical gene flow and induce unique community structure (Morrissey and de Kerckhove 2009; Brown and Swan 2010), whereas the opposite is true for those systems more linear and unimpeded where gene flow and ephemeral metapopulations predominate (Barson et al. 2009; Díez-del-Molino et al. 2013; Davis et al. 2018).

Riverine systems are frequently disrupted, with the pluvial connectivity or stream capture provided by natural events (Smith et al. 1983; Della Croce et al. 2014; Houston et al. 2015) often provoking unique faunal assemblages (Minckley et al. 1986; Craw et al. 2007). However, anthropogenic modifications are also influential in that they not only restrict movements of endemic species but also create favorable conditions for exotics (Rahel 2010; Osmundson 2011; McCauley et al. 2015). Dams and impoundments, for example, modify riverine dynamics (Poff et al. 2007) and impact gene flow (Raeymaekers et al. 2008; Fluker et al. 2014), with larger, more migratory fishes of commercial value being impacted (Waples et al. 2008), but less so for those smaller and non-game (Alò and Turner 2005).

These processes are best represented in fishes of western North America, and have been depicted via two zoogeographical models (Meffe and Vrijenhoek 1988): The ‘Stream Hierarchy

Model' (SHM) explains geospatial distribution of genetic diversity through stream distance and stream network branching patterns. It predicts lower genetic diversity among headwater populations relative to downstream counterparts (Paz-Vinas et al. 2015; Thomaz et al. 2016), with dendritic pattern serving as barriers to gene flow (Della Croce et al. 2014). In contrast, the 'Death Valley Model' (DVM) depicts genetic differentiation among fishes in highly fractured landscapes such as the Basin and Range Province of western North America. Here, genetic differentiation is a function of isolation, with high levels of divergence and scant gene flow predicted among populations (Whiteley et al. 2010). Each model was applied initially to fishes with narrow geographic ranges: SHM for the Sonoran Topminnow (*Poeciliopsis occidentalis*); DVM for Pupfishes (*Cyprinodon* spp.) (Meffe and Vrijenhoek 1988), and the potential for species-specific life history attributes may impact model results.

Rhinichthys osculus, or Speckled Dace (SPD), is a small, widely distributed and abundant cyprinid fish with an evolutionary history that seemingly reflects both zoogeographic models. As such, it provides an opportunity for each model to not only be tested, but also replicated. For example, SPD is found within the well-connected and exorheic Colorado River Basin (CRB) (i.e., that drains directly into the ocean or to another water body so connected). As such, it represents a good test of the SHM model. SPD is also found within the Great Basin that is equally as dispersed spatially as the CRB but is endorheic (i.e., without connectivity with other water bodies). It would serve as a test of the DVM model. Hence, both models are tested within a single species distributed across a spectrum of habitat conditions, rather than across a pair of species and consequently removing life history divergence as a potential issue.

Here we employ SPD to test the following hypotheses: 1) Populations are structured according to contemporary riverine connectivity; 2) Geographic distance serves as a proxy for

genetic distance; 3) The spatial patterning of a dendritic network drives the distribution of genetic diversity; 4) The geospatial distribution of genetic diversity conforms to the particular geographic regions which it inhabits. Implications for the conservation of narrow endemics and SPD as a whole will be also explored.

Methods

Sampling

Samples (N=1,130) were collected throughout the GB and CRB between 1989-2016 by state, federal, and tribal agencies, with additional contributions from researchers at Arizona State University, Idaho State University, and the University of Nevada, Las Vegas. They represent 39 sampling locations within the GB (N=386; μ=9.9 samples/ site), and 77 CRB localities (N=744; μ=9.6 samples/ site).

Data Collection

Whole genomic DNA extractions were performed using one of four methods: Gentra Puregene DNA Purification Tissue kit, QIAGEN DNeasy Blood and Tissue Kit, QIAamp Fast DNA Tissue Kit, or CsCl-gradient. DNA quantity was measured using a Qubit 2.0 fluorometer (Thermo Fisher Scientific, Inc.), whereas quality was assessed visually on a 2.0% agarose gel. Preparation of double digest Restriction-Site Associated DNA (ddRAD) libraries followed Peterson et al. (2012). Restriction digest of 1µg genomic DNA/sample was performed in 50µl reactions containing 5µl New England BioLabs CutSmart Buffer and 20 units each *Pst*I, and *Msp*I. Samples were digested at 37ºC for 24 hours then purified using Agencourt AMPure XP beads (Beckman Coulter, Inc.).

Barcoded samples (100 ng DNA each) were pooled in sets of 48 following Illumina adapter ligation, then size-selected to retrieve DNA fragments between 375 and 425 bp in length (Chafin et al. 2017) using the Pippin Prep System (Sage Science). Size-selected DNA was subjected to 12 cycles of PCR amplification using Phusion high-fidelity DNA polymerase (New England Bioscience), according to manufacturer protocols. Subsequent quality checks to confirm successful amplification of DNA were performed via Agilent 2200 TapeStation and qPCR. Three indexed libraries (144 samples) were pooled per lane for 100bp single-end sequencing on either Illumina HiSeq 2000 (University of Wisconsin Biotechnology Center), or HiSeq 4000 (University of Oregon Genomics & Cell Characterization Core Facility).

Alignment

Samples were grouped by respective geographic regions then subdivided according to historic connectivity and inferred phylogenetic relationships. This was done to facilitate retention of informative loci within each region, in that highly conserved loci can emerge when divergent populations are grouped in ddRAD analyses (Eaton et al. 2015). The CRB was divided into five subregions: Upper Colorado (N=221), Virgin River (N=205), Grand Canyon (N=149), Gila (N=105), and Little Colorado (N=64). The GB was divided into three, per Pleistocene hydrology (Grasso 1996; Oviatt 1997; Adams and Wesnousky 1998). These are: Former Lake Lahontan, including Monitor and Clover Valleys (N=170), Death Valley ecosystem (Amargosa and Owens River drainages; N=112), and the former Lake Bonneville (N=104).

Data were de-multiplexed and filtered in STACKS (Catchen et al. 2013) to discard reads with uncalled bases or low Phred quality scores (<10), while simultaneously attempting to recover those reads with ambiguous barcodes (≤1 mismatched nucleotide). The *de novo*

assembly of ddRAD loci within each region (N=2) and subregion (N=8) was accomplished in PYRAD (Eaton 2014) with a clustering threshold at 0.85. All reads containing >4 bases with Phred quality scores <20 were removed from analysis. A minimum of 15 reads was required for an individual to retain a locus. Putative paralogs were eliminated by discarding loci with heterozygosity >0.6. Loci were retained if present in >50% of samples from each subregion, a procedure accomplished by filtering in step 7 of PYRAD.

Population Structure

The input, filtering, and execution of a Maximum Likelihood (ML) approach to assess population structure (ADMIXTURE v1.3: Alexander et al. 2009) was facilitated by a custom pipeline (https://github.com/stevemussmann/AdmixturePipeline). A .vcf file containing all SNPs was first converted to .ped format (PLINK 1.9: Purcell et al. 2007), with one SNP per locus sampled prior to running ADMIXTURE. Clustering values (K=1 to 20) were evaluated for each subregion, with 20 replicates for each K. Cross-validation (CV) values were calculated in ADMIXTURE, per program specifications.

Multiple independent runs were evaluated (CLUMPAK; Kopelman et al. 2015) so as to identify different ADMIXTURE modes within a single K-value. Major clusters were identified in CLUMPAK using a similarity score of 0.9, and a custom script (https://github.com/stevemussmann/ADMIXTURE_cv_sum) summarized variability in CV values. K-values associated with lowest CV values represented the best estimate of population structure within each subregion. Samples were grouped according to ADMIXTURE assignment and pairwise F_{ST} values were calculated (ARLEQUIN 3.5; Excoffier and Lischer 2010). Pairwise significance (Bonferroni-adjusted $\alpha = 0.05$) compensated for multiple comparisons.

Stream Hierarchy Model

Multiple methods were used to assess gene flow within the CRB and evaluate the fit of the SHM. Isolation by distance (IBD) was evaluated in ARLEQUIN via Mantel test by correlating a pairwise distance matrix of linearized F_{ST} values ($LinF_{ST}$) with linear river distance (measured in KM) between each pair of sites. Significance was assessed using 16,000 permutations. Geographic distance was calculated using the Network Analyst Origin-Destination Cost Matrix tool in ARCGIS 10.4 (ESRI, Redlands, CA).

Linearized F_{ST} values were also evaluated in STREAMTREE (Kalinowski et al. 2008), under the assumption that river network branching patterns approximate a tree of distances among sampling localities. These results also represent a second test of the SHM (with input for STREAMTREE derived using software at https://github.com/stevemussmann/STREAMTREE_arcpy). Models fit (i.e., IBD vs. SHM) was compared using coefficients of determination (R^2).

Death Valley Model

An analysis of molecular variance (AMOVA: Excoffier et al. 1992) was performed in ARLEQUIN for GB and CRB populations, with groupings per ADMIXTURE results and significance derived by permutation (N=16,000). Both regions were then contrasted according to values derived for among-group variance (F_{CT}). Here, the GB should display greater among-group but reduced within-population variation relative to the CRB. This would be consistent with the disparity in connectivity between the two (i.e., greater potential for gene flow among populations in the CRB vs. GB), thus establishing the DVM as the model for GB isolation.

Migration

We calculated migration rates in the CRB using BA3-SNPS (a modification of BAYESASS 3.0.4; Wilson and Rannala 2003, available at https://github.com/stevemussmann/BAYESASS3-SNPs). Computational issues make it impractical to evaluate migration among all 77 CRB localities at once (Faubet et al. 2007). This was remedied by using a sliding window approach that estimates migration rates among triplets of neighboring sampling locations (N=618 groups). Input files were again generated by repeating Step 7 of PYRAD to isolate loci present in ≥ 50% of samples for each triad of localities. Each BA3-SNPS instance was allowed to run for 40 million Markov Chain Monte Carlo (MCMC) generations, the first 25% of which were discarded as burn-in. Results were summarized using a mixture of Perl and Python code to assess variability of independent migration rate assessments for the same population pairs. An elliptic envelope method (Rousseeuw and Driessen 1999) was used to detect outliers by gauging migration rates against pairwise stream distances between localities (available at https://github.com/stevemussmann/bayesass_sum). Adjustment of mixing parameters for BA3-SNPS was automated using a binary search algorithm (https://github.com/stevemussmann/BA3-SNPS-autotune).

Results

Alignment

The number of loci recovered per region was variable ($\bar{x}$ = 10,620, σ = 2,588) and ranged from a maximum of 15,120 in the Little Colorado River to a minimum of 6,569 in the Gila River Basin (Table 1). This also correlated with the amount of missing data for each subregion, with the Little Colorado River having the lowest percentage (15.34%) whereas the Gila River had the

highest (38.32%). This reflects differential sequencing coverage as well, with the Little Colorado having the second highest mean value ($\bar{x}$ = 67.60) whereas the Gila River had the second lowest ($\bar{x}$ = 48.87).

Population Structure

ADMIXTURE identified 30 populations in the CRB, six of which are in the upper basin (Figure 2A): Chinle Wash, Green River, Paria River, San Juan River, Smiths Fork River, and Vermillion Creek. The San Rafael and Fremont rivers share a close affinity with the San Juan population. Some individuals from Smiths Fork were found to share ancestry with the Bear River of the Bonneville Basin (data not shown).

The Virgin River and its tributaries represent an additional eight populations (Figure 2B), three of which correspond to the pluvial White River drainage of Nevada (the habitat of *R. o. velifer*). One White River population is represented by a single individual, with partial assignments to nine other fish sampled from Moapa, the White River, and Meadow Valley Wash. The Moapa Speckled Dace (*R. o. moapae*) is recovered as a fourth population while the Meadow Valley Wash system represents a fifth. The headwaters of Beaver Dam Wash represent a sixth population, while the upper Virgin River is the seventh. The mainstem Virgin River represents the eighth population and shows evidence of introgression from tributary populations.

The remaining 16 populations are distributed throughout the lower CRB. These include five distributed linearly along the length of the Grand Canyon (Figure 2C), a sixth representing the Bill Williams River (Figure 2D), and four tributaries of the Little Colorado River (Figure 2E). The remaining six are found within the Gila River system, specifically the Agua Fria River,

Eagle Creek, San Pedro River, San Simon River, Upper Gila River, and the Verde River (Figure 2D). The Upper Gila River population shows mixed ancestry with the Eagle Creek population.

A total of 17 populations were recovered in the Great Basin, mostly corresponding to known geographic breaks, while the Lahontan Basin was divided into seven (Figure 3A), one of which represents the endangered Clover Valley SPD (*R. o. oligoporus*). Monitor Valley SPD (*R. o. ssp.*) was also recovered as unique, but surprisingly, clustered with samples from Coils Creek despite the lack of a contemporary connection. Samples of *R. o. robustus* represent four unique populations, the largest of which is distributed throughout the Humboldt River drainage. The remaining three are split among Canyon Creek (Upper Quinn River), the Walker Lake Subbasin, the Smoke Creek system, and Honey and Eagle lakes of northern California. Honey Lake also shows evidence of introgression from Smoke Creek.

In the Death Valley region, all proposed SPD subspecies were recovered as unique (Figure 3B). The only formally described subspecies from this region, the endangered *R. o nevadensis*, inhabits springs within Ash Meadows National Wildlife Refuge. The remaining undescribed subspecies are distributed among Oasis Valley and Amargosa Canyon in the Amargosa drainage, with two populations in the Owens River drainage (i.e., Long Valley Speckled Dace, *R. o. ssp.*, and the Owens River itself). A locality from Benton Valley was recovered as a sixth population.

The Bonneville Basin was divided into four populations (Figure 3C). The first of these corresponds to the Bear and Weber rivers, while the second encompasses the Thousand Springs Creek area in the northwestern part of former Lake Bonneville. Most of the southern Bonneville Basin (i.e., Sevier River and Snake Valley) comprised a single population, with the exception of Shoal Creek in the Escalante Desert.

AMOVA and F statistics

The results of the AMOVA (Table 2) revealed greater among-group variation in the GB (50.11%; F_{CT} = 0.501; p<0.0001) relative to the CRB (27.42%; F_{CT} = 0.274; p<0.00001). Within-population variation in the CRB (72.53%; F_{ST} = 0.275; p<0.00001) was greater than the GB (43.61%; F_{ST} = 0.564; p<0.0001). Pairwise F_{ST} values comparing ADMIXTURE-defined populations within the CRB and GB are provided in Tables 3 and 4. All values in the GB were significant (p < 0.0001). Nearly all were significant in the CRB, save for some involving the San Pedro River (21/406: 5.2%). F_{IS} values for each sampling site are found in Table 5. Most F_{IS} values were low (<0.1), but the GB yielded a greater proportion >0.1 (33%; 13/39) when compared to the CRB (3.9%; 3/77).

IBD and the StreamTree Model

The correlation of genetic and geographic distances was low but significant (R^2=0.159, p < 0.0001) (Figure 4A). In contrast, a more resolved fit for the data was provided by the STREAMTREE model (R^2=0.787 p < 0.0001) (Figure 4B). Figure 5 highlights the genetic distance explained by each stream segment in STREAMTREE. High values ($LinF_{ST}$ >0.36) were frequently observed in the Lower CRB. Examples include the stream segments leading to the Agua Fria ($LinF_{ST}$=1.20) and San Pedro ($LinF_{ST}$=0.53) rivers, as well as Fossil and Eagle creeks ($LinF_{ST}$=0.48 and 0.36, respectively). The Bill Williams River ($LinF_{ST}$=1.19) shows a high level of divergence, as does the connection between the Gila River and the remainder of the Lower CRB ($LinF_{ST}$=0.10). The connection between the Grand Canyon and the Virgin River is elevated ($LinF_{ST}$=0.09), as is that between the Little Colorado and the Grand Canyon ($LinF_{ST}$=0.12). Very

few connections in the Upper CRB showed elevated levels of genetic distance, save the North Fork of Vermillion Creek in Wyoming ($LinF_{ST}$=0.59).

Migration

The BA3-SNPs analyses identified those locations linked via migration (Table 6), with source-and-sink populations identified (Figure 6). Not surprisingly, many elevated migration rates detected by the outlier analysis involve sites either in close geographic proximity or within the same CRB subdrainage. For example, many were detected in the Virgin River system, which may also be a function of sampling density. This area contained 205/744 (27.6%) of all CRB samples, rivaled only by the upper basin (221, or 29.7%). However, upper basin samples are distributed over a greater area. A similar trend was observed in the Grand Canyon, where several sites are in close proximity. One surprise was the identification of Desolation Canyon (DES: lower Green River) as a sink population involving many distant sources (μ=634 km), to include sites from the lower basin (Grand Canyon). Other surprising connections occurred between the Virgin River (CCN), Gila (GLN), and Verde (SPG) sites (μ CCN distance=1,533 km). However, these results may be spurious due to the admixed nature of the CCN site, which can produce bizarre allele frequencies that may coincidentally match far away localities. In all cases (N=5), at least one site involved in an elevated migration rate to/from a distant site (>1,000 km) was admixed. Furthermore, migration outliers between subdrainages most commonly involved the Upper CRB and Grand Canyon (72.7%; 8/11 examples).

Discussion

Here we evaluated the effects of isolation and connectivity on the evolution of the widely-distributed Speckled Dace, and to test zoogeographical models regarding its geospatial distribution of genetic diversity. We were particularly interested in the manner by which: 1) A pioneer species evolves within a large, dendritic stream network, and 2) the manner by which it reacts to broad vicariant events of a geomorphic nature. We then juxtaposed our results against two general themes that resonated in the literature for SPD: 1) A variety of natural and anthropogenic processes have influenced its distribution and diversity, and 2) undocumented diversity is greater than expected.

Contrasts within and between Stream Models

The GB and CRB followed expectations of the DVM and SHM, respectively. Much of the CRB diversity is contained within populations rather than among sites (Table 2). This is intuitive in that SPD populations are connected by riverine habitats throughout much of the upper and lower CRB. Those in the Virgin River system are less so, but with pluvial connections providing occasional opportunities for gene flow. In contrast, a greater portion of the diversity in the GB (50.11%) was found among sampling sites as opposed to within populations. This separation is largely geographic in that connectivity previously occurred end-of-Pleistocene (Grasso 1996; Oviatt 1997; Adams and Wesnousky 1998).

F_{IS} values for GB and CRB sites also fit predictions of the SHM and DV models (i.e., lower values in well-connected areas but greater in isolated populations). F_{IS} values for CRB sites were generally low (μ=-0.004), with values >0.1 only being found at Maynard Springs (dmay: 0.437), Smith's Fork (SMF: 0.275), and Chevelon Creek (dchv: 0.1598). The common

theme is that each site contains subpopulation structure, indicating presence of a Wahlund effect that increases F_{IS} estimates (Wahlund 1928). Maynard Springs has a known mixed ancestry (USFWS 1998) with SPD from two additional springs being introduced in 1991.

Sites throughout the GB have comparatively higher F_{IS} estimates, indicating greater deviation from Hardy-Weinberg Equilibrium (HWE). Many with F_{IS} >0.1 are found in the Lahontan and surrounding areas, with a site of mixed ancestry near Lassen Volcanic National Park being greatest (dlva: F_{IS}=0.259). Again, this underscores the isolated nature of these sites, and sustains expectations of the DVM in that reduced gene flow drives the evolution and divergence of these populations (Meffe and Vrijenhoek 1988).

Dendritic Networks

The CRB reflects gene flow typical of dendritic stream networks (Hughes et al. 2009), as seen in the close proximity of distinct genetic populations to one another. The Grand Canyon typifies this situation, where its tributary streams are separated by short segments of the Colorado River, such that populations are isolated without obvious physical barriers (i.e., presence of 'soft' barriers such as water currents or distance that disrupt adult movements and/or larval dispersal). A second example is the Little Colorado River, where populations sampled near tributary headwaters show little evidence of mixing with one another.

The confluences of streams in both examples (above) represent unsuitable habitats for SPD (Turner and Robison 2006). In this sense, SPD prefers 2nd and 3rd order streams (Moyle 2002), and these are readily available in the CRB. However, SPD is still found in atypical habitat such as Lake Tahoe (Sigler and Sigler 1987), but larger, mainstem rivers or anthropogenic reservoirs are considered less suitable (Minckley 1973). Additionally, the modern Colorado

River formed ~5.3 million years ago (House et al. 2008) and lakes disappeared once lava dams eroded within the Grand Canyon (mid-Pleistocene: Crow et al. 2008). In this context, these results make sense because selective pressure for survival in large rivers or lacustrine habitats has been absent for hundreds of thousands of years (Dalrymple and Hamblin 1998).

Asymmetric gene flow among sites is another common trend observed in dendritic networks (Morrissey and de Kerckhove 2009). Examples include genotypes in Chinle Wash and Vermillion Creek (San Juan River) that are also found in the Green River (upper CRB). Eagle Creek genotypes also show evidence of mixing with the upper Gila River, as well as the Virgin River system. Here, mainstem Virgin River populations are also introgressed by individuals from the Upper Virgin, Clover Creek, and Moapa populations. The latter two are aberrant in this regard, given the lack of stable contemporary connectivity, and may thus represent echoes of past connectivity.

The influence of past connections on the spatial distribution of diversity within the CRB is apparent when IBD and STREAMTREE models are evaluated. The IBD model is a poor fit to the CRB, due to the relatively short riverine distances separating sites with elevated genetic differentiation, and vice versa. The STREAMTREE model is a much closer fit, under the assumption that the riverine network approximates a distance-based phylogenetic tree. Yet it too falls short of the high R^2 values found in studies at much smaller geographic scales (Kalinowski et al. 2008). Again, these results reflect the fluctuating connectivity among river segments coupled with potential extirpations/ recolonizations in others.

These discrepancies in the distribution of genetic diversity in the CRB are again illustrated in the Little Colorado River and Grand Canyon. In a phylogenetic analysis, the Little Colorado is recovered as the sister clade of the Gila River/ Southern California coastal drainages

(See Chapter 1), whereas in STREAMTREE it reflects a close relationship with the Grand Canyon and the Upper CRB. The latter scenario is plausible in a morphological perspective since Little Colorado SPD resemble the Upper CRB subspecies (*R. o. yarrowi*), supporting a connection with the Grand Canyon and Upper Basin (Minckley et al. 1986). However, Minckley (1973) also recognized that the Gila River subspecies (*R. o. osculus*) and *R. o. yarrowi* "intergrade chaotically" along the Mogollon Rim, suggesting a history of repeated stream capture across this geographic feature.

An alternate explanation for this discordance involves the natural impoundments formed by Pleistocene lava dams in the Grand Canyon (Hamblin 1994). These formed large, steep-shored impoundments that held more water than Lakes Powell and Mead combined (Dalrymple and Hamblin 1998). Such habitats would have been unfavorable for a small minnow whose preferred niche involved 2nd or 3rd order streams. Thus, the genetic signatures in the Little Colorado are potential vestiges of a once more widely-distributed form subsequently replaced by an Upper CRB form. The Little Colorado River was isolated from Grand Canyon by the development of a vicariant lava flow ~20,000 years ago (Duffield et al. 2006).

Other locations within the CRB do not fit the SHM, a result of various hydrologic events. Smiths Fork (Upper CRB) contains many samples that align more closely with the Bear and Snake river drainages (data not shown), likely due to stream capture during the Pleistocene (Smith et al. 2002; Loxterman and Keeley 2012). Some of these fishes are assigned almost entirely to the Bear and Snake river population, indicating an event more recent than the Pleistocene.

Few surprises emerge when STREAMTREE results are examined for the Upper Basin. A slightly elevated value is apparent for Whisky Creek in Chinle Wash, but it remains closely

related to others in the area (F_{ST}= -0.23-0.16). The stream segment leading to the North Fork of Vermillion Creek shows the greatest discrepancy between observed and fitted genetic distance and may indicate a longer period of isolation or rather, a closer affinity with an SPD that remains unsampled. However, exceptions within an otherwise genetically homogenous Upper CRB remain sparse. Population structure and migration analyses also corroborate the apparent homogeneity of SPD in this region.

Migration

Results for the STREAMTREE and migration analyses reflect complementary aspects of the same processes. While STREAMTREE identified differences among populations, the migration analysis underscored the potential for sources and sinks within the CRB (Kawecki and Holt 2002). Evidence for migration is abundant in the Virgin River, due largely to the density of our sampling. Fifteen of the 40 outlier migration rates reflect movement among sites in the Virgin River System. Two populations in particular – the Virgin River above Santa Clara (VSC) and Beaver Dam Wash above Motoqua (BDM) – showed rates of immigration that were higher than surrounding localities. The ADMIXTURE analysis also reflects a high level of gene flow into the mainstem Virgin River population. These results indicate that upstream Virgin River serves as source, whereas the mainstem river is a sink, and underscores in turn the presence of degraded habitat along the Virgin River.

River systems with reduced habitat quality due to urbanization also reflect similar trends. Fishes in relatively pristine reaches are mostly impervious to downstream problems, however degraded downstream habitats may persist only by receiving migrants from pristine areas (Waits et al. 2008). This also describes the situation in the mainstem Virgin River, much of which

transects small towns and agricultural fields that contain diversion dams, such as Washington Fields and Quail Creek. Fish barriers along the Virgin River also prevent upstream migration of non-native fishes from Lake Mead. The headwaters of Virgin River tributaries often persist on state or federally protected lands, meaning they are relatively unaffected by these anthropogenic issues. Both ADMIXTURE and migration analyses indicate that headwaters of the Virgin River provide input for the mixed ancestry of the mainstem. For example, the locality with migrants from the greatest number of sources (i.e., VSC) is located on the Santa Clara River below Gunlock Dam (Table 6). Likewise, nearly every locality along the Virgin River mainstem shows admixture from upstream populations (Figure 2B). The combined results of these analyses are consistent with the negative impacts of degraded habitats on SPD populations in the mainstem Virgin River.

Desolation Canyon (DES) of the Green River represents another major sink population. Yet it is somewhat surprising that high rates of migration are found in this region, given the relative genetic homogeneity of the Green River. DES is unique in experiencing a high rate of immigration from several distant sources, to include the Grand Canyon (pairwise distance >800 km). Contemporary movements upstream from the Grand Canyon are now blocked by Glen Canyon Dam, and subsequently Lake Powell. Other sources would require movement downstream from above Flaming Gorge Dam. Pairwise F_{ST} values are low among Green River sites, possibly indicative of large population sizes that counterbalance the effects of genetic drift (Frankham 1996), and which maintain allelic frequencies that reflect spurious migration into Desolation Canyon.

A more plausible explanation for the homogeneity of the Green River may emerge when other Upper CRB native fishes are compared. Colorado Pikeminnow (*Ptychocheilus lucius*) and

Flannelmouth Sucker (*Catostomus latipinnis*) exhibit little diversity with regards to mtDNA, suggesting a post-Pleistocene bottleneck throughout the Upper CRB (Douglas et al. 2003; Borley and White 2006). Upper CRB *Gila* also exhibit scant mtDNA variation, so much so that *G. cypha* and *G. robusta* share mtDNA that is also commonly found within *G. elegans* (Gerber et al. 2001). This has also been interpreted as the result of a post-Pleistocene drying that forced fishes into refugia, possibly on numerous occasions, with subsequent genetic homogenization as a result (Douglas and Douglas 2007).

In addition, microsatellite DNA studies of Bluehead Sucker (*C. discobolus*) revealed little evidence of population structure in the Upper CRB, suggesting the Green River as a potential management unit (MU: Hopken et al. 2013). Extreme and prolonged drought in this area were frequent at end-of-Pleistocene, particularly within the past few centuries (Woodhouse et al. 2010). Patterns of genetic diversity for SPD match those of other species (above), thus sustaining the hypothesis that a region-wide event may have negatively impacted not only mainstem fishes, but also those inhabiting tributaries throughout the region.

Undocumented Diversity

This study adds to the literature that documents unidentified diversity within and among SPD populations. *Rhinichthys* has long been recognized as containing a great deal of geographically localized morphological diversity (Jordan and Evermann 1896; Jordan et al. 1930; Deacon and Williams 1984), with additional diversity identified genetically (Hoekzema and Sidlauskas 2014; Wiesenfeld et al. 2018). This study also highlights as distinct multiple populations of *R. o. robustus* throughout the former Lahontan basin. Several distinct *R. osculus* populations were identified from the Virgin River, as well as multiple distinct populations in

Meadow Valley Wash (*R. o. ssp*) and the pluvial White River (*R. o. velifer*). Although most are found in highly isolated areas, unique populations also arise in well-connected river systems such as Vermillion Creek (a Green River tributary).

Unfortunately, some of this diversity has only been recognized following a decline in abundance or outright extinction. The most extreme example is Marble Creek within the Owens River Valley of California. Genetic analyses demonstrate uniqueness on par with five other populations of the Death Valley region, all of which have been proposed as subspecies of *R. osculus*. However, this particular population was eliminated by a flood in 1989 (Moyle et al. 2015). Another example is the pluvial White River drainage of Nevada. SPD was once the most abundant fish in the Pahranagat Valley (USFWS 1998) but is now restricted to isolated springs throughout its former range (Guadalupe 2015). Fish representing three distinct genetic populations were identified from this area, with at least two still extant (Figure 2B: ISP, PAH, and SUN). However, the source of the third population could not be definitively identified in this study, as it appeared at low frequency across multiple sampling localities.

A surprising lack of population structure was also found in central Nevada localities [i.e. Coils Creek (COC) and Monitor Valley (STN, DPB)] (Figure 3). These localities were unambiguously assigned to a single genetic population, with high gene flow among them (per low F_{ST} estimates). The two systems lack connectivity, even in wet years, potentially due to recent drying over the past two centuries (Jeff Petersen, Nevada Department of Wildlife, pers. comm.). Anthropogenic transfer between systems may offer one explanation, or the presence of large population sizes that negate the effects of genetic drift. F_{IS} values were relatively high for Monitor Valley sample sites (DPB: 0.095; STN: 0.116) compared to that for Coils Creek (0.062). The high inbreeding coefficient for the Monitor Valley population suggests isolation, but without

significant differentiation from Coils Creek, possibly due to contemporary transfer during wetter times.

The identification of hidden diversity underscores the importance of quantifying and monitoring genetic diversity in wide-ranging species such as SPD. Numerous recent studies have identified "species-level divergence" as a basis for revising SPD (Pfrender et al. 2004; Hoekzema and Sidlauskas 2014; Wiesenfeld et al. 2018). Data presented herein support this argument, but with the caveat that a holistic approach is required (Padial et al. 2010; Fujita et al. 2012). Microsatellites are quickly evolving markers that can accurately estimate population divergence (Oliveira et al. 2006), yet are limited with regard to a phylogenetic signal (Petren et al. 1999; Ochieng et al. 2007). The pattern of gene flow derived from ddRAD allele frequencies in this study also provides accurate insights into the recent past (Davey and Blaxter 2010). Likewise, mtDNA data are gleaned from a single molecule that can be variable within-species depending upon the evolutionary rate of the locus examined (Thomaz et al. 1996). It may also be discordant with nuclear data due to lateral gene transfer (Funk and Omland 2003; Chan and Levin 2005), as noted several times among western fishes (Gerber et al. 2001; Dowling et al. 2016), to include *Rhinichthys* (see Chapter 1). Any forthcoming revisions should also take into consideration the morphological variation found in SPD (Smith et al. 2017), as well as employing the latest genetic tools that will allow adaptations to be more formally assayed through genomic studies (Harrisson et al. 2014; Hoban et al. 2016).

The results presented here are consistent with the delineation of evolutionary significant units (ESUs) in Chapter 1 (Waples 1991). Genetic populations within these ESUs should be designated as management units (MUs), pending taxonomic revision or studies of ecological adaptations (Palsbøll et al. 2006). However, distinct populations have already been lost due to

natural disasters (Benton Valley, California), and more will follow due to the groundwater pumping of isolated springs in the more arid regions of the west (Sigler and Sigler 1987; Minckley and Unmack 2000; Moyle 2002). Thus, a proactive management approach is necessary (Brooks et al. 2006). Additional data to complete designation of ESUs and MUs for the Lahontan and surrounding areas is also needed. Samples for several areas (Big Smoky Valley, Independence Valley, Carson River, and the Truckee River) were not obtained in time to be included in this study. Additional sampling within the Quinn river system would also be beneficial.

Conclusion

Conservation implications are apparent in the riverscape analysis of SPD provided herein. Many dynamic forces have acted upon SPD throughout its range, including homogenizing vs asymmetrical gene flow, and isolation by vicariance. Additional forces have impacted the evolution of this species in other river basins (i.e., the Columbia and Snake Rivers), where sympatry with congeners is more prevalent and lateral gene transfer among species is apparent (Pfrender et al. 2004; Smith et al. 2017; Wiesenfeld et al. 2018).

SPD is viewed as an atypical target of conservation, in that it is widespread and abundant save for a few narrow endemics. However, its conservation value becomes magnified when it is viewed as a surrogate for other stream fishes of similar size, habitat preference, or life history that must also respond to anthropogenic modifications (Caro 2010). Many endangered western fishes share at least one of these traits, and the impacts identified for SPD in this study will similarly impact them. In addition, the near-ubiquity of SPD in western North America also

promotes a variety of comparisons with threatened and endangered species of limited distribution, as demonstrated for the Upper CRB.

A core principle of conservation genetics is to quantify the natural processes promoting the evolution of life, and utilize them for the management of species and their habitats (Frankham 1995). Past and ongoing natural processes that promoted lineage diversification have been identified herein, with implications not only for management of SPD but also other native fishes throughout western North America. A trenchant example is the status of fishes in the Grand Canyon. If indeed they represent recolonization from the Upper Basin following extirpation, then anthropogenic modifications will further isolate populations and negatively impact upstream habitats as well (Pringle 1997).

Implications of climate change must also be considered. Genomic tools now exist to identify the manner by which fishes adapt to changing conditions (Cure et al. 2017), and these should be a primary focus of future conservation genomic studies. The beneficial effects of gene flow are now controversial in that a population of maladapted individuals may persist as a sink despite receiving individuals from a source (Moore and Hendry 2009). Future studies should focus on the geographic regions where anthropogenic activities are expected to admix populations, and to quantify the genomic rearrangements that occur in response to a changing environment.

References

Adams KD, Wesnousky SG (1998) Shoreline processes and the age of the Lake Lahontan highstand in the Jessup embayment, Nevada. Geol Soc Am Bull 110:1318–1332. doi: 10.1130/0016-7606(1998)110<1318:SPATAO>2.3.CO;2

Alexander DH, Novembre J, Lange K (2009) Fast model-based estimation of ancestry in unrelated individuals. Genome Res 19:1655–1664. doi: 10.1101/gr.094052.109

Alò D, Turner TF (2005) Effects of habitat fragmentation on effective population size in the endangered Rio Grande silvery minnow: genetic effects of river fragmentation. Conserv Biol 19:1138–1148. doi: 10.1111/j.1523-1739.2005.00081.x

Barson NJ, Cable J, van Oosterhout C (2009) Population genetic analysis of microsatellite variation of guppies (*Poecilia reticulata*) in Trinidad and Tobago: evidence for a dynamic source–sink metapopulation structure, founder events and population bottlenecks. J Evol Biol 22:485–497. doi: 10.1111/j.1420-9101.2008.01675.x

Borley K, White MM (2006) Mitochondrial DNA variation in the endangered Colorado pikeminnow: a comparison among hatchery stocks and historic specimens. North Am J Fish Manag 26:916–920. doi: 10.1577/M05-176.1

Brooks TM, Mittermeier RA, da Fonseca GAB, et al (2006) Global biodiversity conservation priorities. Science 313:58. doi: 10.1126/science.1127609

Brown BL, Swan CM (2010) Dendritic network structure constrains metacommunity properties in riverine ecosystems. J Anim Ecol 79:571–580. doi: 10.1111/j.1365-2656.2010.01668.x

Caro T (2010) Conservation by Proxy: Indicator, Umbrella, Keystone, Flagship, and Other Surrogate Species. Island Press, Washington, DC

Catchen J, Hohenlohe PA, Bassham S, et al (2013) Stacks: an analysis tool set for population genomics. Mol Ecol 22:3124–3140. doi: 10.1111/mec.12354

Chafin TK, Martin BT, Mussmann SM, et al (2017) FRAGMATIC: *in silico* locus prediction and its utility in optimizing ddRADseq projects. Conserv Genet Resour 1–4

Chan KMA, Levin SA (2005) Leaky prezygotic isolation and porous genomes: rapid introgression of maternally inherited DNA. Evolution 59:720–729. doi: 10.1111/j.0014-3820.2005.tb01748.x

Craw D, Burridge C, Anderson L, Waters JM (2007) Late Quaternary river drainage and fish evolution, Southland, New Zealand. Geomorphology 84:98–110. doi: 10.1016/j.geomorph.2006.07.008

Crow R, Karlstrom KE, McIntosh W, et al (2008) History of Quaternary volcanism and lava dams in western Grand Canyon based on lidar analysis, 40Ar/39Ar dating, and field

studies: Implications for flow stratigraphy, timing of volcanic events, and lava dams. Geosphere 4:183–206. doi: 10.1130/GES00133.1

Cure K, Thomas L, Hobbs J-PA, et al (2017) Genomic signatures of local adaptation reveal source-sink dynamics in a high gene flow fish species. Sci Rep 7:8618. doi: 10.1038/s41598-017-09224-y

Dalrymple GB, Hamblin WK (1998) K-Ar ages of Pleistocene lava dams in the Grand Canyon in Arizona. Proc Natl Acad Sci U S A 95:9744–9749

Davey JW, Blaxter ML (2010) RADSeq: next-generation population genetics. Brief Funct Genomics 9:416–423. doi: 10.1093/bfgp/elq031

Davis CD, Epps CW, Flitcroft RL, Banks MA (2018) Refining and defining riverscape genetics: How rivers influence population genetic structure. Wiley Interdiscip Rev Water 5:e1269. doi: 10.1002/wat2.1269

Deacon JE, Williams JE (1984) Annotated list of the fishes of Nevada. Proc Biol Soc Wash 97:103–118

Della Croce P, Poole GC, Payn RA, Izurieta C (2014) Simulating the effects of stream network topology on the spread of introgressive hybridization across fish populations. Ecol Model 279:68–77. doi: 10.1016/j.ecolmodel.2014.02.014

Díez-del-Molino D, Carmona-Catot G, Araguas R-M, et al (2013) Gene flow and maintenance of genetic diversity in invasive mosquitofish (*Gambusia holbrooki*). PLoS ONE 8:e82501. doi: 10.1371/journal.pone.0082501

Douglas MR, Brunner PC, Douglas ME (2003) Drought in an evolutionary context: molecular variability in Flannelmouth Sucker (*Catostomus latipinnis*) from the Colorado River Basin of western North America. Freshw Biol 48:1254–1273. doi: 10.1046/j.1365-2427.2003.01088.x

Douglas MR, Douglas ME (2007) Genetic structure of humpback chub *Gila cypha* and roundtail chub *G. robusta* in the Colorado River ecosystem. US Geological Survey, Grand Canyon Monitoring and Research Center, Flagstaff, AZ

Dowling TE, Markle DF, Tranah GJ, et al (2016) Introgressive hybridization and the evolution of lake-adapted Catostomid fishes. PLoS ONE 11:1–27. doi: 10.1371/journal.pone.0149884

Duffield W, Riggs N, Kaufman D, et al (2006) Multiple constraints on the age of a Pleistocene lava dam across the Little Colorado River at Grand Falls, Arizona. Geol Soc Am Bull 118:421–429. doi: 10.1130/B25814.1

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

Excoffier L, Smouse PE, Quattro JM (1992) Analysis of molecular variance inferred from metric distances among DNA haplotypes: application to human mitochondrial DNA restriction data. Genetics 131:479–491

Faubet P, Waples RS, Gaggiotti O (2007) Evaluating the performance of a multilocus Bayesian method for the estimation of migration rates. Mol Ecol 16:1149–1166. doi: 10.1111/j.1365-294X.2007.03218.x

Fluker BL, Kuhajda BR, Harris PM (2014) The effects of riverine impoundment on genetic structure and gene flow in two stream fishes in the Mobile River basin. Freshw Biol 59:526–543. doi: 10.1111/fwb.12283

Frankham R (1995) Conservation genetics. Annu Rev Genet 29:305–327. doi: 10.1146/annurev.ge.29.120195.001513

Frankham R (1996) Relationship of genetic variation to population size in wildlife. Conserv Biol 10:1500–1508. doi: 10.1046/j.1523-1739.1996.10061500.x

Fujita MK, Leaché AD, Burbrink FT, et al (2012) Coalescent-based species delimitation in an integrative taxonomy. Trends Ecol Evol 27:480–488. doi: 10.1016/j.tree.2012.04.012

Funk DJ, Omland KE (2003) Species-level paraphyly and polyphyly: frequency, causes, and consequences, with insights from animal mitochondrial DNA. Annu Rev Ecol Evol Syst 34:397–423. doi: 10.1146/annurev.ecolsys.34.011802.132421

Gerber AS, Tibbets CA, Dowling TE (2001) The role of introgressive hybridization in the evolution of the *Gila robusta* complex (Teleostei: Cyprinidae). Evolution 55:2028–2039. doi: 10.1111/j.0014-3820.2001.tb01319.x

Gleick PH (2010) Roadmap for sustainable water resources in southwestern North America. Proc Natl Acad Sci U S A 107:21300–21305. doi: 10.1073/pnas.1005473107

Grasso DN (1996) Hydrology of modern and late Holocene lakes, Death Valley, California. US Geological Survey, Denver, CO

Guadalupe K (2015) Nevada Department of Wildlife Native Fish and Amphibians Field Trip Report. Nevada Department of Wildlife, Las Vegas, NV

Hamblin WK (1994) Late Cenozoic lava dams in the western Grand Canyon. Geol Soc Am Mem 183:1–139

Harrisson KA, Pavlova A, Telonis-Scott M, Sunnucks P (2014) Using genomics to characterize evolutionary potential for conservation of wild populations. Evol Appl 7:1008–1025. doi: 10.1111/eva.12149

Hoban S, Kelley JL, Lotterhos KE, et al (2016) Finding the genomic basis of local adaptation: pitfalls, practical solutions, and future directions. Am Nat 188:379–397. doi: 10.1086/688018

Hoekzema K, Sidlauskas BL (2014) Molecular phylogenetics and microsatellite analysis reveal cryptic species of speckled dace (Cyprinidae: *Rhinichthys osculus*) in Oregon's Great Basin. Mol Phylogenet Evol 77:238–250. doi: 10.1016/j.ympev.2014.04.027

Hopken MW, Douglas MR, Douglas ME (2013) Stream hierarchy defines riverscape genetics of a North American desert fish. Mol Ecol 22:956–971. doi: 10.1111/mec.12156

House PK, Pearthree PA, Perkins ME (2008) Stratigraphic evidence for the role of lake spillover in the inception of the lower Colorado River in southern Nevada and western Arizona. In: Reheis MC, Hershler R, Miller DM (eds) Special Paper 439: Late Cenozoic Drainage History of the Southwestern Great Basin and Lower Colorado River Region: Geologic and Biotic Perspectives. Geological Society of America, Boulder, CO

Houston DD, Evans RP, Shiozawa DK (2015) Pluvial drainage patterns and Holocene desiccation influenced the genetic architecture of relict dace, *Relictus solitarius* (Teleostei: Cyprinidae). PLoS ONE 10:1–12. doi: 10.1371/journal.pone.0138433

Hughes JM, Schmidt DJ, Finn DS (2009) Genes in streams: using DNA to understand the movement of freshwater fauna and their riverine habitat. BioScience 59:573–583. doi: 10.1525/bio.2009.59.7.8

Jordan DS, Evermann BW (1896) The fishes of North and Middle America. Bull U S Natl Mus 47:1–1236

Jordan DS, Evermann BW, Clark HW (1930) Check list of the fishes and fishlike vertebrates of North and Middle America north of the northern boundary of Venezuela and Colombia. Rep US Comm Fish Fisc Year 1928 2:1–670

Kalinowski ST, Meeuwig MH, Narum SR, Taper ML (2008) Stream trees: a statistical method for mapping genetic differences between populations of freshwater organisms to the sections of streams that connect them. Can J Fish Aquat Sci 65:2752–2760

Kawecki TJ, Holt RD (2002) Evolutionary consequences of asymmetric dispersal rates. Am Nat 160:333–347. doi: 10.1086/341519

Kopelman NM, Mayzel J, Jakobsson M, et al (2015) CLUMPAK: a program for identifying clustering modes and packaging population structure inferences across K. Mol Ecol Resour 15:1179–1191. doi: 10.1111/1755-0998.12387

Loxterman JL, Keeley ER (2012) Watershed boundaries and geographic isolation: patterns of diversification in cutthroat trout from western North America. BMC Evol Biol 12:38–54

McCauley LA, Anteau MJ, van der Burg MP, Wiltermuth MT (2015) Land use and wetland drainage affect water levels and dynamics of remaining wetlands. Ecosphere 6:1–22

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

Minckley W, Hendrickson D, Bond C (1986) Geography of western North American freshwater fishes: description and relationships to intracontinental tectonism. In: Hocutt CH, Wiley EO (eds) The Zoogeography of North American Freshwater Fishes. John Wiley & Sons, New York, pp 519–613

Minckley W, Unmack PJ (2000) Western springs: their faunas, and threats to their existence. In: Freshwater Ecoregions of North America: A Conservation Assessment. Island Press, Washington, DC, pp 52–53

Minckley WL (1973) Fishes of Arizona. Arizona Game and Fish Department, Phoenix, AZ

Moore J-S, Hendry AP (2009) Can gene flow have negative demographic consequences? Mixed evidence from stream threespine stickleback. Philos Trans R Soc B Biol Sci 364:1533–1542. doi: 10.1098/rstb.2009.0007

Morrissey MB, de Kerckhove DT (2009) The maintenance of genetic variation due to asymmetric gene flow in dendritic metapopulations. Am Nat 174:875–889. doi: 10.1086/648311

Moyle PB (2002) Inland Fishes of California. University of California Press, Berkeley, CA

Ochieng JW, Steane DA, Ladiges PY, et al (2007) Microsatellites retain phylogenetic signals across genera in eucalypts (Myrtaceae). Genet Mol Biol 30:1125–1134

Oliveira EJ, Pádua JG, Zucchi MI, et al (2006) Origin, evolution and genome distribution of microsatellites. Genet Mol Biol 29:294–307

Osmundson DB (2011) Thermal regime suitability: assessment of upstream range restoration potential for Colorado pikeminnow, a warmwater endangered fish. River Res Appl 27:706–722. doi: 10.1002/rra.1387

Oviatt CG (1997) Lake Bonneville fluctuations and global climate change. Geology 25:155–158

Padial JM, Miralles A, De la Riva I, Vences M (2010) The integrative future of taxonomy. Front Zool 7:1–14. doi: 10.1186/1742-9994-7-16

Palsbøll PJ, Bérubé M, Allendorf FW (2006) Identification of management units using population genetic data. Trends Ecol Evol 22:11–16. doi: 10.1016/j.tree.2006.09.003

Paz-Vinas I, Loot G, Stevens VM, Blanchet S (2015) Evolutionary processes driving spatial patterns of intraspecific genetic diversity in river ecosystems. Mol Ecol 24:4586–4604. doi: 10.1111/mec.13345

Peterson BK, Weber JN, Kay EH, et al (2012) Double digest RADseq: an inexpensive method for *de novo* SNP discovery and genotyping in model and non-model species. PLoS ONE 7:1–11. doi: 10.1371/journal.pone.0037135

Petren K, Grant BR, Grant PR (1999) A phylogeny of Darwin's finches based on microsatellite DNA length variation. Proc R Soc Lond B Biol Sci 266:321. doi: 10.1098/rspb.1999.0641

Pfrender ME, Hicks J, Lynch M (2004) Biogeographic patterns and current distribution of molecular-genetic variation among populations of speckled dace, *Rhinichthys osculus* (Girard). Mol Phylogenet Evol 30:490–502. doi: 10.1016/S1055-7903(03)00242-2

Poff NL, Olden JD, Merritt DM, Pepin DM (2007) Homogenization of regional river dynamics by dams and global biodiversity implications. Proc Natl Acad Sci U S A 104:5732–5737

Pringle CM (1997) Exploring how disturbance is transmitted upstream: going against the flow. J North Am Benthol Soc 16:425–438

Purcell S, Neale B, Todd-Brown K, et al (2007) PLINK: a tool set for whole-genome association and population-based linkage analyses. Am J Hum Genet 81:559–575

Raeymaekers JAM, Maes GE, Geldof S, et al (2008) Modeling genetic connectivity in sticklebacks as a guideline for river restoration. Evol Appl 1:475–488. doi: 10.1111/j.1752-4571.2008.00019.x

Rahel FJ (2010) Homogenization, differentiation, and the widespread alteration of fish faunas. Am Fish Soc Symp 73:311–326

Rousseeuw PJ, Driessen KV (1999) A fast algorithm for the minimum covariance determinant estimator. Technometrics 41:212–223. doi: 10.1080/00401706.1999.10485670

Sigler WF, Sigler JW (1987) Fishes of the Great Basin: A Natural History. University of Nevada Press, Reno, NV

Smith G, Dowling T, Gobalet K, et al (2002) Biogeography and timing of evolutionary events among Great Basin fishes. Gt Basin Aquat Syst Hist 33:175–234

Smith GR, Chow J, Unmack PJ, et al (2017) Evolution of the *Rhinichthys osculus* complex (Teleostei: Cyprinidae) in western North America. Misc Publ Mus Zool Univ Mich 204:1–83

Smith GR, Hall JG, Koehn RK, Innes DJ (1983) Taxonomic relationships of the Zuni Mountain sucker, *Catostomus discobolus yarrowi*. Copeia 1983:37–48. doi: 10.2307/1444696

Thomaz AT, Christie MR, Knowles LL (2016) The architecture of river networks can drive the evolutionary dynamics of aquatic populations. Evolution 70:731–739. doi: 10.1111/evo.12883

Thomaz D, Guiller A, Clarke B (1996) Extreme divergence of mitochondrial DNA within species of pulmonate land snails. Proc R Soc Lond B Biol Sci 263:363. doi: 10.1098/rspb.1996.0056

Turner TF, Robison HW (2006) Genetic diversity of the Caddo madtom, *Noturus taylori*, with comments on factors that promote genetic divergence in fishes endemic to the Ouachita Highlands. Southwest Nat 51:338–345

USFWS (1998) Recovery Plan for the Aquatic and Riparian Species of Pahranagat Valley. US Fish and Wildlife Service, Portland, OR

Vörösmarty CJ, McIntyre PB, Gessner MO, et al (2010) Global threats to human water security and river biodiversity. Nature 467:555–561. doi: 10.1038/nature09440

Wahlund S (1928) Zusammensetzung von populationen und korrelationserscheinungen vom standpunkt der vererbungslehre aus betrachtet. Hereditas 11:65–106. doi: 10.1111/j.1601-5223.1928.tb02483.x

Waits ER, Bagley MJ, Blum MJ, et al (2008) Source–sink dynamics sustain central stonerollers (*Campostoma anomalum*) in a heavily urbanized catchment. Freshw Biol 53:2061–2075. doi: 10.1111/j.1365-2427.2008.02030.x

Waples RS (1991) Pacific salmon, *Oncorhynchus* spp., and the definition of "species" under the Endangered Species Act. Mar Fish Rev 53:11–22

Waples RS, Zabel RW, Scheuerell MD, Sanderson BL (2008) Evolutionary responses by native species to major anthropogenic changes to their ecosystems: Pacific salmon in the Columbia River hydropower system. Mol Ecol 17:84–96. doi: 10.1111/j.1365-294X.2007.03510.x

Whiteley AR, Hastings K, Wenburg JK, et al (2010) Genetic variation and effective population size in isolated populations of coastal cutthroat trout. Conserv Genet 11:1929–1943. doi: 10.1007/s10592-010-0083-y

Wiesenfeld JC, Goodman DH, Kinziger AP (2018) Riverscape genetics identifies speckled dace (*Rhinichthys osculus*) cryptic diversity in the Klamath–Trinity Basin. Conserv Genet 19:111–127. doi: 10.1007/s10592-017-1027-6

Wilson GA, Rannala B (2003) Bayesian inference of recent migration rates using multilocus genotypes. Genetics 163:1177–1191

Woodhouse CA, Meko DM, MacDonald GM, et al (2010) A 1,200-year perspective of 21st century drought in southwestern North America. Proc Natl Acad Sci U S A 107:21283–21288. doi: 10.1073/pnas.0911197107

Appendix

Table 1. The number of loci recovered by PYRAD for regions and subregions. Loci = the number of loci recovered [a locus to be present in at least 50% of all individuals within a subregion (italicized names)]. % Missing = actual percentage of missing data for each region or subregion. Mean Coverage = average sequencing depth of each locus for each subregion.

Region	Loci	% Missing	Mean Coverage
Colorado River Basin	9725	30.57	52.80
Grand Canyon	11529	23.01	45.73
Gila River	6569	38.32	48.87
Little Colorado River	15120	15.34	67.60
Upper Colorado River Basin	11005	26.94	56.57
Virgin River	8926	32.11	51.26
Great Basin	9479	27.32	62.23
Bonneville	9052	29.35	69.07
Death Valley	12556	20.16	61.07
Lahontan	10204	25.09	58.85

Table 2. Results of an Analysis of Molecular Variance (AMOVA). Columns correspond to the amount of genetic variance explained by the F-statistics derived in the AMOVA. The Colorado River Basin (CRB) follows the expectations of the stream hierarchy model due to a lower among-group and greater within-population genetic variance. In contrast, the GB follows expectations of the Death Valley model by exhibiting greater among-group relative to within-population variance.

	Among Groups	**Among populations within groups**	**Within Populations**
Colorado River	27.42%	0.05%	72.53%
Great Basin	50.11%	6.28%	43.61%

Table 3. Pairwise F_{ST} values comparing ADMIXTURE-defined genetic populations in the Colorado River Basin (CRB). All values were significant ($\alpha<0.0001$) except those appearing in bold. Populations in the table are arranged according to geographic subregion. Pairwise comparisons among populations from the same subregion are color-coded. **Upper CRB (grey)**: PR = Paria River; CW = Chinle Wash; SJR = San Juan River; GR = Green River; SF = Smiths Fork; NFVC = North Fork Vermillion Creek. **Grand Canyon (red)**: LCtK = Lava-Chuar to Kwagunt; ECtBA = Elves Chasm to Bright Angel; WWtK = Whitmore Wash to Kanab; HC = Havasu Creek; SCtD = Surprise Canyon to Diamond. **Little Colorado (yellow)**: EF = East Fork; ECC = East Clear Creek; CC = Chevelon Creek; SC = Silver Creek. **Virgin River (blue)**: UV = Upper Virgin; BDW = Beaver Dam Wash; P = Pahranagat; M = Moapa; VM = Virgin Mainstem; MVW = Meadow Valley Wash; WR = White River. **Lower Colorado River (green)**: SP = San Pedro; UGR = Upper Gila River; EC = Eagle Creek; BWR = Bill Williams River; VR = Verde River; SSR = San Simon River; AF = Agua Fria.

	AF	SSR	VR	BWR	EC	UGR	SP	WR	MVW	VM	M	P	BDW	UV	SC	CC	ECC	EF	SCtD	HC	WWtK	ECtBA	LCtK	NFVC	SF	GR	SJR	CW
PR	0.63	0.55	0.55	0.54	0.63	0.39	**-0.18**	0.44	0.48	0.36	0.44	0.46	0.42	0.43	0.55	0.50	0.51	0.48	0.26	0.44	0.16	0.24	0.17	0.41	0.29	0.19	0.15	0.22
CW	0.69	0.63	0.59	0.58	0.67	0.47	0.20	0.39	0.43	0.31	0.42	0.46	0.45	0.53	0.59	0.52	0.49	0.45	0.23	0.52	0.06	0.20	0.12	0.47	0.15	0.09	0.13	
SJR	0.61	0.55	0.51	0.51	0.62	0.37	**0.01**	0.40	0.42	0.28	0.40	0.44	0.38	0.41	0.54	0.48	0.47	0.43	0.15	0.41	**-0.02**	0.13	0.06	0.33	0.15	0.02		
GR	0.55	0.48	0.44	0.46	0.59	0.27	**-0.11**	0.44	0.45	0.28	0.43	0.46	0.33	0.32	0.53	0.48	0.48	0.48	0.19	0.35	0.07	0.15	0.10	0.24	0.24			
SF	0.50	0.39	0.40	0.50	0.53	0.00	**-1.47**	0.32	0.39	0.22	0.38	0.32	0.29	0.26	0.45	0.40	0.45	0.41	0.33	0.45	0.26	0.29	0.19	0.37				
NFVC	0.79	0.73	0.70	0.69	0.74	0.62	0.44	0.52	0.55	0.47	0.53	0.56	0.59	0.66	0.67	0.63	0.61	0.56	0.42	0.66	0.27	0.41	0.33					
LCtK	0.55	0.47	0.44	0.48	0.56	0.27	**-0.18**	0.37	0.40	0.24	0.37	0.39	0.31	0.34	0.47	0.41	0.42	0.38	0.15	0.35	0.04	0.07						
ECtBA	0.63	0.57	0.53	0.55	0.64	0.40	0.04	0.45	0.48	0.33	0.46	0.49	0.40	0.45	0.57	0.52	0.52	0.49	0.17	0.39	0.04							
WWtK	0.54	0.47	0.43	0.45	0.57	0.26	**-0.19**	0.42	0.44	0.27	0.40	0.42	0.31	0.31	0.50	0.44	0.46	0.44	0.12	0.29								
HC	0.81	0.77	0.73	0.73	0.75	0.67	0.55	0.55	0.58	0.51	0.57	0.60	0.62	0.71	0.69	0.65	0.63	0.57	0.41									
SCtD	0.64	0.58	0.57	0.56	0.66	0.43	**-0.01**	0.47	0.51	0.38	0.48	0.50	0.45	0.46	0.58	0.55	0.55	0.52										
EF	0.54	0.47	0.46	0.62	0.57	0.25	**-0.30**	0.48	0.51	0.40	0.48	0.49	0.44	0.43	0.40	0.36	0.39											
ECC	0.63	0.57	0.54	0.66	0.65	0.40	**-0.05**	0.50	0.54	0.44	0.51	0.52	0.49	0.49	0.40	0.13												
CC	0.64	0.56	0.55	0.67	0.62	0.40	**-0.11**	0.48	0.52	0.43	0.49	0.49	0.50	0.53	0.39													
SC	0.69	0.61	0.61	0.71	0.67	0.49	**0.02**	0.52	0.56	0.49	0.54	0.53	0.56	0.58														
UV	0.74	0.67	0.62	0.68	0.66	0.51	0.34	0.33	0.37	0.22	0.33	0.39	0.41															
BDW	0.66	0.59	0.56	0.60	0.63	0.42	**0.05**	0.25	0.27	0.12	0.17	0.22																
P	0.58	0.50	0.52	0.58	0.59	0.30	**-0.48**	0.11	0.28	0.16	0.17																	
M	0.56	0.49	0.48	0.56	0.59	0.29	**-0.23**	0.25	0.29	0.09																		
VM	0.53	0.44	0.42	0.51	0.55	0.21	**-0.17**	0.19	0.19																			
MVW	0.56	0.49	0.48	0.58	0.60	0.29	**-0.16**	0.27																				
WR	0.52	0.45	0.44	0.54	0.56	0.24	**-0.24**																					
SP	0.37	0.20	**0.01**	0.42	**-0.12**	**0.02**																						
UGR	0.56	0.24	0.39	0.65	0.11																							
EC	0.66	0.32	0.56	0.76																								
BWR	0.79	0.75	0.73																									
VR	0.48	0.56																										
SSR	0.70																											

Table 4. Pairwise F_{ST} values comparing ADMIXTURE-defined genetic populations in the Great Basin (GB). All values were significant ($\alpha<0.0001$). Populations in the table are arranged according to geographic subregion. Pairwise comparisons among intra-subregion populations are color-coded. **Death Valley (yellow):** OV = Oasis Valley; AC = Amargosa Canyon; LV = Long Valley; AM = Ash Meadows; BV = Owens River (Benton Valley); OR = Owens River. **Bonneville Basin (blue):** SC = Shoal Creek; NWB = Northwest Bonneville; SEV = Sevier River; BR = Bear River. **Lahontan Basin (green):** CC = Canyon Creek; CV = Clover Valley; WR = Walker River; PC = Pine Creek; HR = Humboldt River; MV = Monitor Valley; SC = Smoke Creek.

	SMO	**MV**	**HR**	**PC**	**WR**	**CV**	**CC**	**BR**	**SEV**	**NWB**	**SC**	**OR**	**BV**	**AM**	**LV**	**AC**
OV	0.52	0.55	0.49	0.66	0.49	0.65	0.74	0.62	0.68	0.66	0.71	0.28	0.48	0.35	0.69	0.27
AC	0.48	0.52	0.47	0.64	0.42	0.62	0.75	0.63	0.69	0.65	0.72	0.23	0.48	0.27	0.72	
LV	0.60	0.63	0.56	0.75	0.59	0.73	0.84	0.71	0.76	0.72	0.82	0.57	0.74	0.73		
AM	0.53	0.56	0.49	0.68	0.50	0.66	0.77	0.65	0.70	0.67	0.74	0.30	0.53			
BV	0.48	0.52	0.44	0.69	0.44	0.63	0.82	0.64	0.71	0.64	0.79	0.26				
OR	0.36	0.42	0.36	0.52	0.32	0.51	0.62	0.50	0.56	0.54	0.57					
SC	0.50	0.55	0.48	0.68	0.50	0.65	0.73	0.51	0.24	0.58						
NWB	0.46	0.51	0.47	0.59	0.46	0.60	0.63	0.46	0.55							
SEV	0.47	0.53	0.43	0.63	0.51	0.64	0.62	0.42								
BR	0.39	0.46	0.37	0.56	0.42	0.59	0.58									
CC	0.53	0.61	0.53	0.71	0.56	0.69										
CV	0.42	0.48	0.42	0.59	0.41											
WR	0.12	0.26	0.17	0.36												
PC	0.29	0.44	0.38													
HR	0.13	0.11														
MV	0.20															

Table 5. F_{IS} and sample size (N) for 77 sampling localities in the Colorado River Basin (CRB: columns 1 and 2)) and 39 in the Great Basin (GB: column 3). The mean CRB F_{IS} = 0.004 (Standard Deviation = 0.09), whereas that for the GB = 0.066 (Standard Deviation = 0.07).

Locality	N	F_{IS}	Locality	N	F_{IS}	Locality	N	F_{IS}
182	10	0.0467	LAV	5	-0.0039	AMA	32	0.0435
ARU	12	-0.0621	LCR	10	0.0892	AMC	10	-0.011
BAC	10	-0.0036	LIT	7	0.085	ASH	10	-0.0121
BBC	9	0.029	LYC	10	0.0179	ASR	10	-0.0801
BBL	9	0.0335	MAT	10	0.0584	BRL	3	0.0328
BDM	6	-0.0904	MBD	8	0.0353	CAN	8	-0.0141
BDS	10	-0.024	MCE	10	-0.0974	CLV	10	0.0803
BDW	10	0.0335	MOA	22	0.0242	COC	10	0.0621
BFT	10	0.0222	MOD	8	0.0862	DAI	4	-0.0671
BIG	10	-0.0297	MUC	10	-0.0203	dbox	7	0.1073
BOU	10	-0.008	MVW	10	0.036	dbrr	11	0.0702
CCN	4	-0.0019	NCC	10	-0.0785	dbsc	10	0.0484
CHE	12	0.1598	NFV	10	-0.1543	dhar	10	-0.0333
CHU	9	-0.134	NTH	10	-0.0397	dlva	11	0.2594
CLC	10	0.0401	NVC	10	-0.1953	dnfh	11	0.1156
CYC	10	-0.144	PAH	10	0.047	dorb	10	0.1239
danr	6	-0.0107	PAR	12	0.0343	DPB	11	0.0953
DES	7	-0.0306	PIP	10	-0.017	dpin	11	0.0746
DIA	10	0.031	RYC	10	0.0195	dres	7	0.1331
dmay	7	0.4366	SCG	9	0.0567	dsfk	11	0.1
DOL	12	-0.0392	SHN	10	-0.0555	dths	7	0.0932
dvel	7	0.02	SIC	10	0.0635	ECN	4	0.0085
EAG	12	-0.0134	SMF	7	0.2751	EFH	10	0.1743
ECC	12	0.0711	SPG	10	-0.1176	FRZ	12	0.0287
EFL	10	-0.0153	SUN	10	0.0453	GAN	8	0.0924
ELK	10	-0.0267	SUR	10	-0.0372	HCR	10	0.1104
ELV	10	0.0008	SYC	10	-0.0884	IVH	10	0.158
EVR	7	-0.0233	TSA	10	0.0133	LMR	5	0.0009
FER	8	-0.021	TUL	10	0.0086	LOO	10	0.0686
FOS	10	-0.0437	VSC	10	-0.0121	LVD	20	-0.018
FRA	12	-0.1113	WEN	10	0.0578	MIK	5	0.001
FRE	10	0.0056	WFA	8	0.0021	PRC	10	0.1348
GLN	9	-0.0573	WFG	10	-0.0636	RUP	10	0.1349
HAM	10	0.0276	WHY	10	-0.1338	SER	7	0.0903
HAV	10	-0.2056	WIL	10	0.0351	SEV	16	0.0753
HFK	10	0.0492	WIT	10	0.0218	SHO	10	-0.0277
ISP	8	0.0249	WPC	10	0.0086	SMO	10	0.1083
KAN	10	-0.0257	YAM	10	-0.1334	STN	10	0.1157
KWA	10	-0.0766				WCU	5	0.0931

Table 6. Outlier migration rates calculated in BA3-SNPS. Each triplet of neighboring sites was evaluated for contemporary migration, resulting in multiple observations (N) of the mean migration rate (Rate) for some site pairs. Migration rate standard deviation (StDev) was calculated when multiple observations were made. The stream distance between sites (KM) is provided in kilometers. Drainage = the drainage(s) within which each pair of sites exists. Sink = the recipient population. Source = the source population of migrants.

Drainage	Sink	Source	N	Rate	StDev	KM
Gila River	ARU	TUL	1	0.170	-	461
	GLN	TUL	15	0.165	0.043	106
Virgin River	BDM	BDW	5	0.119	0.047	28
	BDM	BDS	1	0.148	-	59
	BDM	VSC	2	0.130	0.051	82
	CCN	MVW	16	0.127	0.052	40
	CCN	MBD	1	0.125	-	257
	dmay	dvel	1	0.133	-	43
	LAV	WFA	1	0.167	-	19
	MBD	LIT	19	0.147	0.046	24
	MVW	CLC	16	0.142	0.043	12
	MVW	EVR	1	0.119	-	52
	SUN	ISP	1	0.180	-	73
	VSC	SCG	1	0.205	-	44
	VSC	MOD	1	0.154	-	47
	VSC	MBD	1	0.128	-	81
	VSC	BDW	1	0.128	-	110
Grand Canyon	BOU	182	15	0.158	0.038	139
	CHU	LCR	6	0.139	0.046	12
	CHU	182	15	0.164	0.045	192
	KWA	182	15	0.120	0.040	206
	PIP	182	15	0.185	0.043	151
	WIT	182	21	0.162	0.038	9
Green River	DES	WPC	26	0.182	0.045	319
	DES	BBC	21	0.187	0.045	486
	DES	BFT	1	0.139	-	605
	DES	BIG	1	0.167	-	723
	SMF	HFK	1	0.133	-	364
	YAM	ELK	1	0.128	-	287
Virgin - Gila	CCN	SPG	1	0.125	-	1419
	CCN	GLN	1	0.148	-	1648
Grand Canyon - Green	CHU	WPC	1	0.139	-	965
	DES	182	17	0.168	0.043	833
	DES	WIT	1	0.195	-	842
San Juan - Green	danr	BBC	1	0.133	-	1329
	MCE	WPC	1	0.205	-	1020
San Juan - Grand Canyon	DOL	182	15	0.122	0.037	749
	MCE	182	15	0.171	0.043	666
	MCE	WIT	1	0.205	-	674
Grand Canyon - Gila	PIP	SPG	1	0.128	-	1555

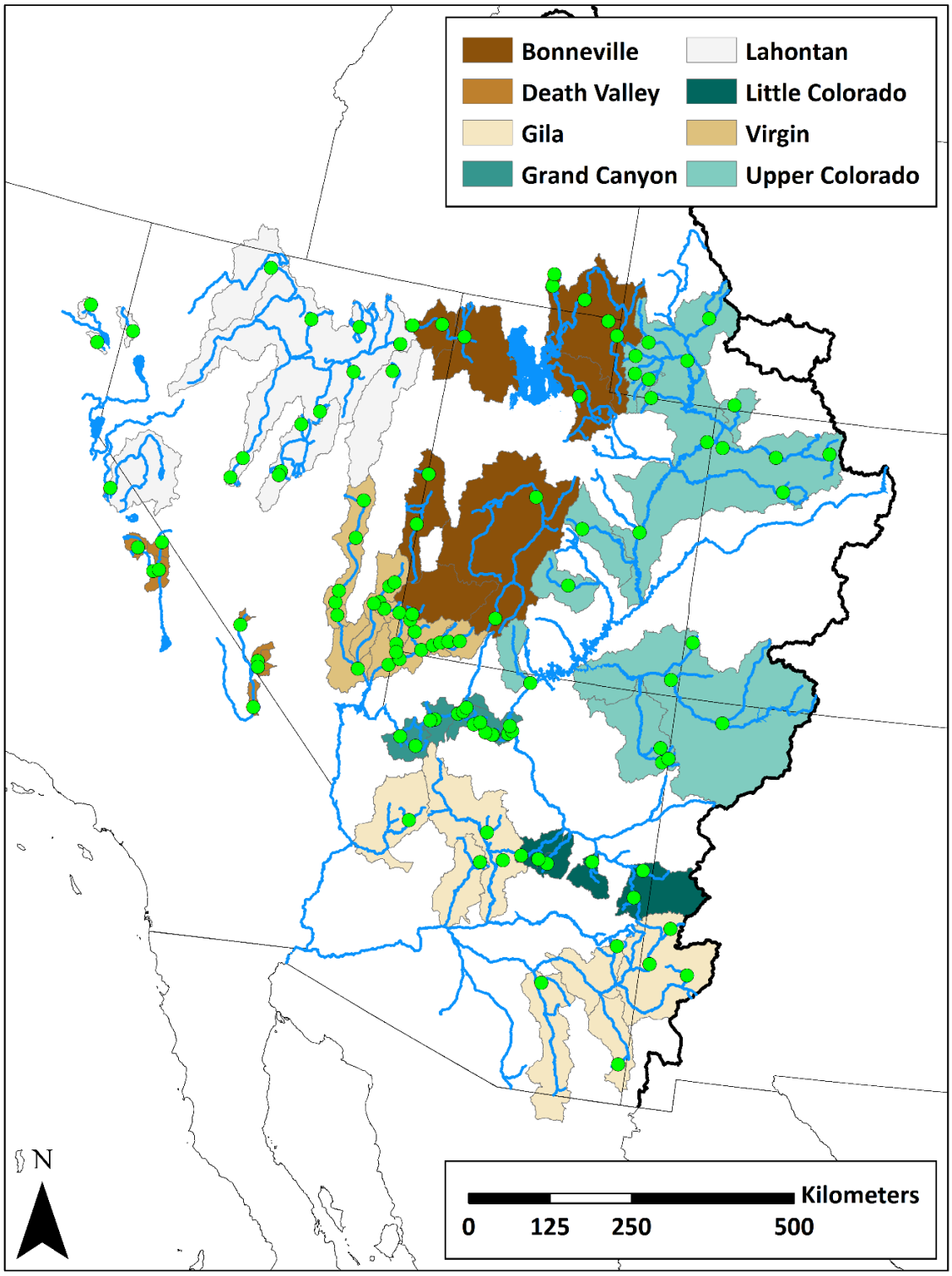

Figure 1. Speckled Dace (SPD) samples collected from 116 sites distributed throughout the Colorado River Basin (CRB) and Great Basin (GB). Colored areas of the map represent subregions that were evaluated for population structure. The Bonneville, Death Valley, and Lahontan areas comprise the GB, while the remaining five subregions form the CRB.

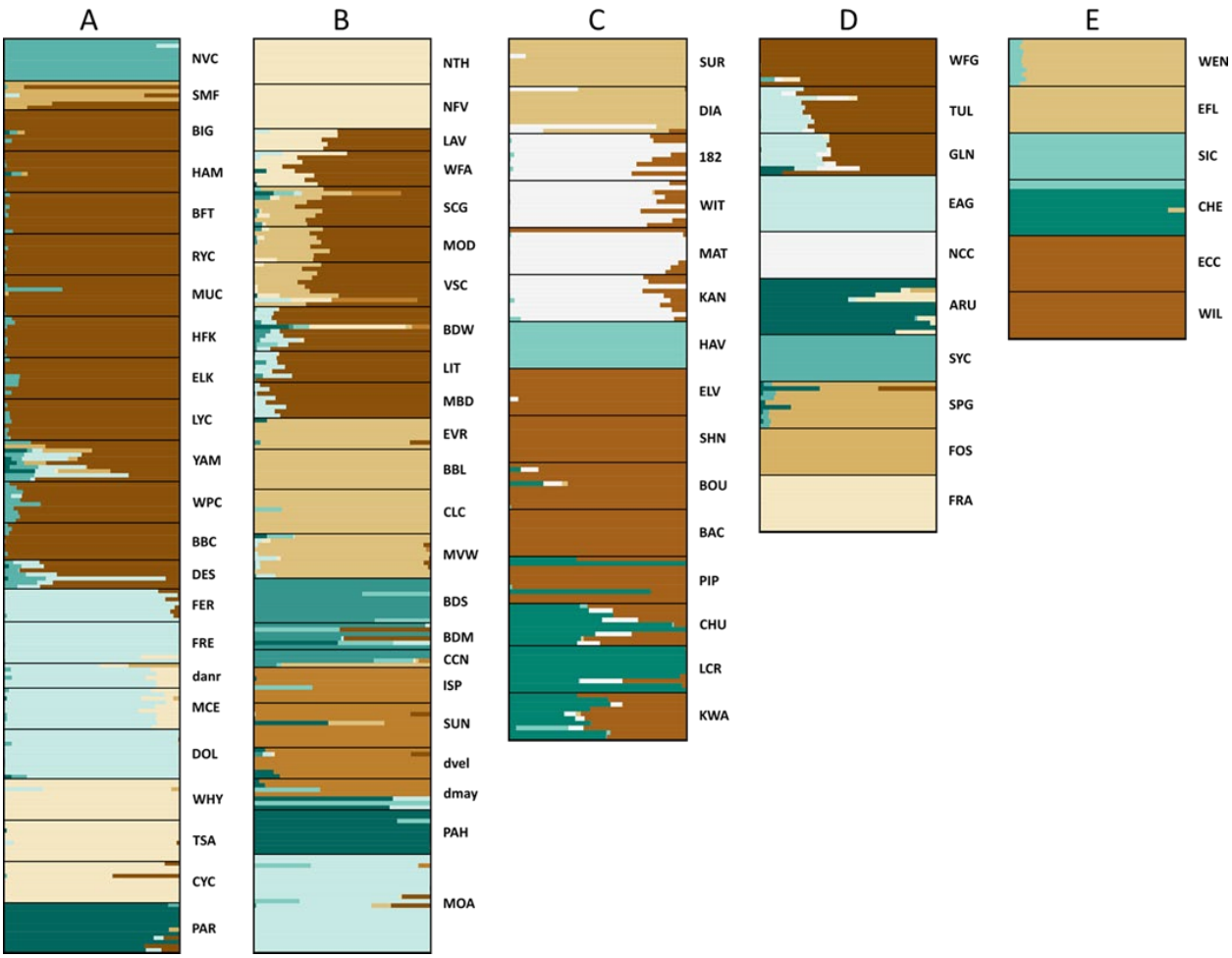

Figure 2. Population structure in the Colorado River Basin (CRB). Upper CRB populations (**A:** N=6) are: Vermillion Creek (NVC), Smiths Fork (SMF), Green River (BIG to DES), San Juan, San Rafael, and Fremont rivers (FER-DOL), Chinle Wash (WHY-CYC) and the Paria River (PAR). Virgin River populations **(B: N=8)** include the upper Virgin River (NTH-NFV), Mainstem Virgin River (LAV-MBD), Meadow Valley Wash (EVR-MVW), Beaver Dam Wash (BDS-CCN), White River (ISP-dmay), Pahranagat (PAH), and Moapa (MOA). An eight population was fractionally assigned to multiple localities in the pluvial White River. Grand Canyon populations **(C: N=5)** were delimited by river miles (RM) with the exception of Havasu Creek (HAV). The remainder include RM 225-249 (SUR-DIA), RM 143-189 (182-KAN), RM 87-117 (ELV-PIP), and RM 225-249 (CHU-KWA). Gila and Bill Williams drainages **(D: N=7)** include the Upper Gila (WFG-GLN), Eagle Creek (EAG), San Simon (NCC), San Pedro (ARU), Agua Fria (SYC), Verde (SPG-FOS), and Bill Williams (FRA). The Little Colorado **(E: N=4)** encompasses the East Fork (WEN-EFL), Silver Creek (SIC), Chevelon Creek (CHE), and East Clear Creek (ECC-WIL).

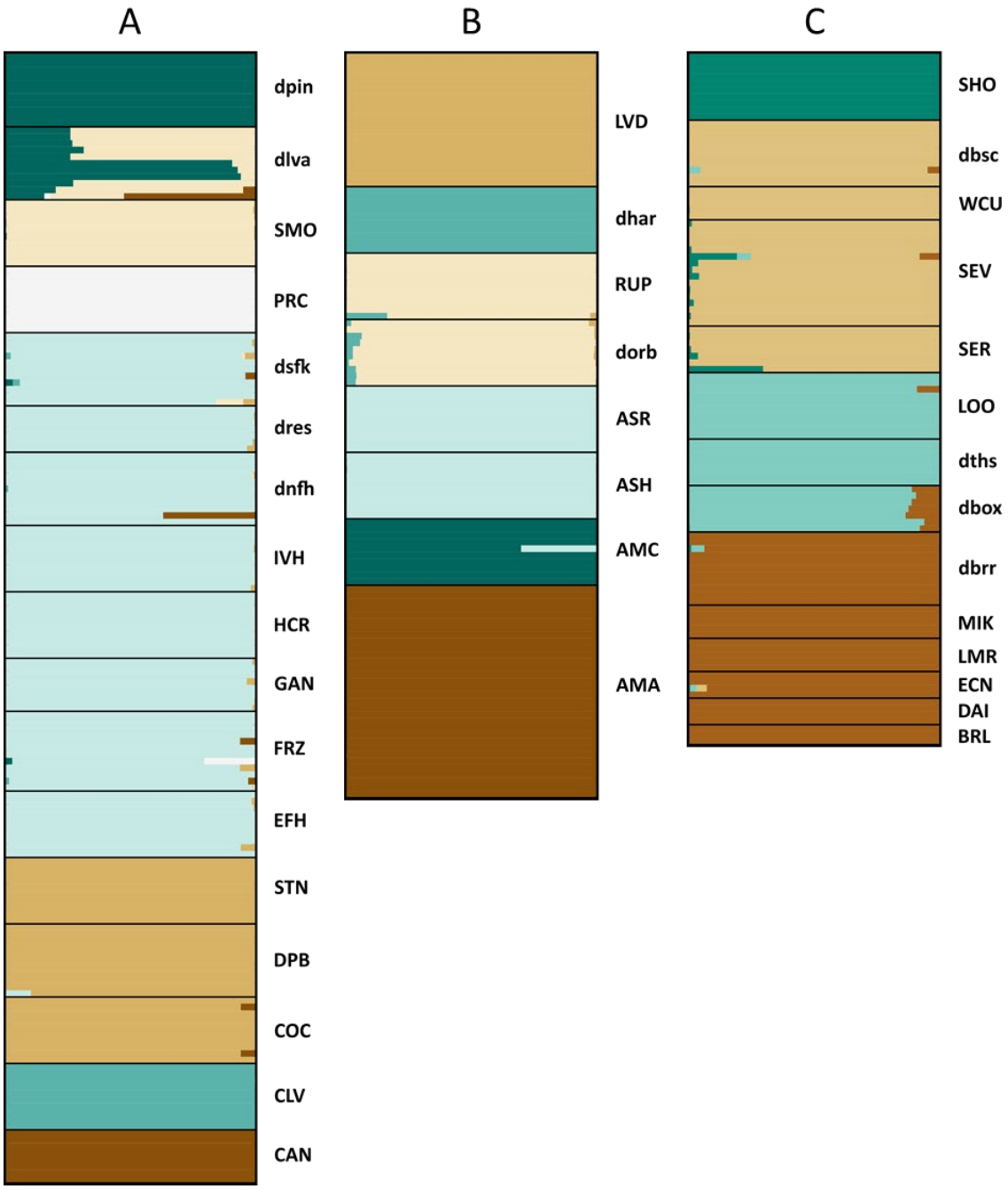

Figure 3. Population structure in the Great Basin (GB). The Lahontan basin **(A: N=7)** includes Honey and Eagle Lakes (dpin-dlva), Smoke Creek (SMO), Walker Subbasin (PRC), Humboldt River (dsfk-EFH), Monitor Valley (STN-COC), Clover Valley (CLV), and Quinn River (CAN). Death Valley region **(B: N=6)** depicts Long Valley (LVD), Benton Valley (dhar), Owens River (RUP-dorb), Ash Meadows (ASR-ASH), Amargosa Canyon (AMC), and Oasis Valley (AMA). The Bonneville basin **(C: N=4)** encompasses Escalante Desert (SHO), Snake Valley/ Sevier River (dbsc-SER), Thousand Springs Creek (LOO-dbox), and Bear/ Weber rivers (dbrr-BRL).

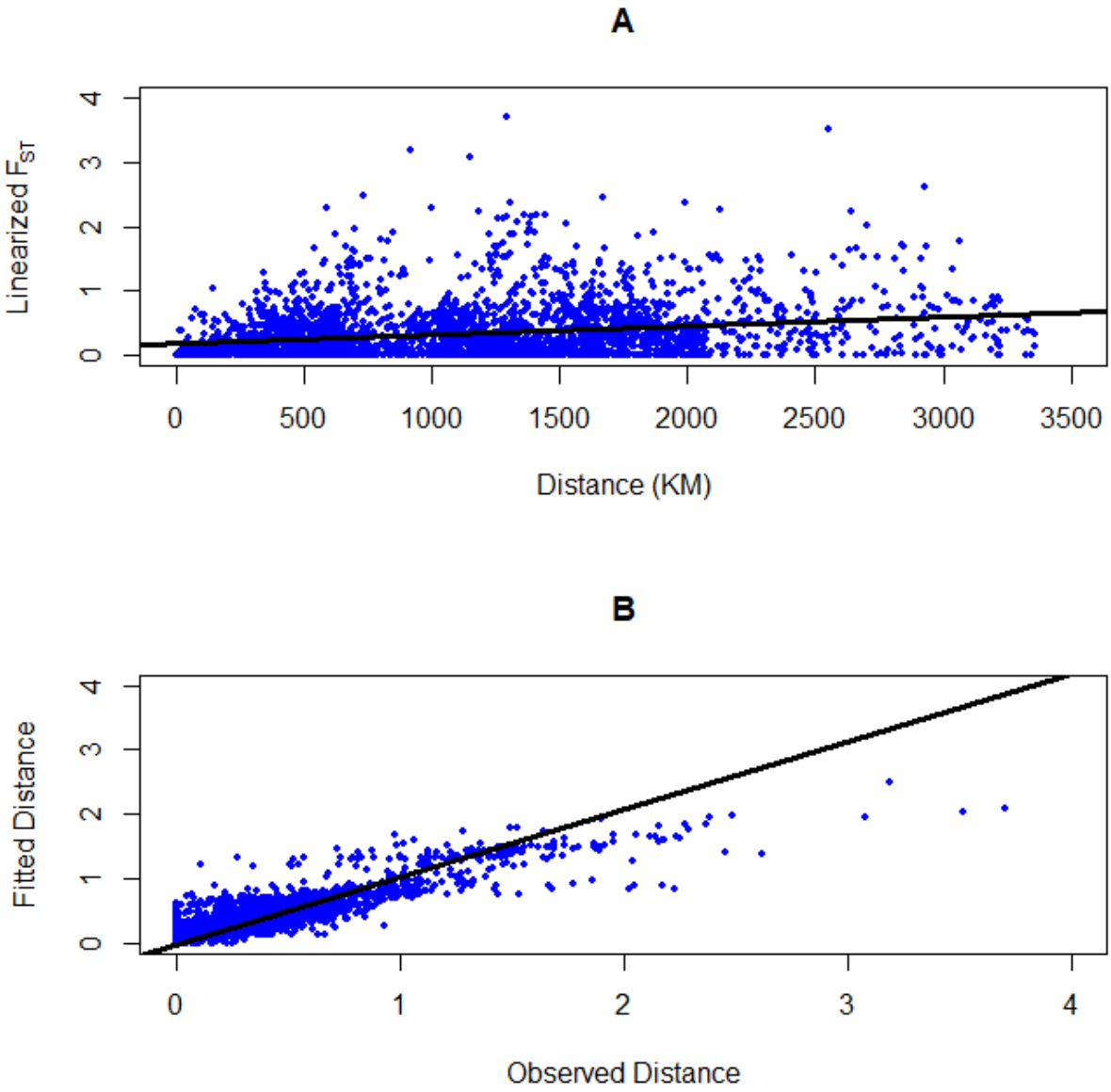

Figure 4. Results of the Stream Hierarchy Model (SHM) fitted to the Colorado River Basin (CRB). (A) = a test of isolation by distance (IBD) for sampling localities in the CRB. Riverine distances were measured using ArcGIS 10.4 (ESRI, Redlands, CA). Results are significant (R^2=0.159; $p < 0.001$). (B) = expected versus observed genetic distance as calculated by StreamTree and displays a tighter fit ($R^2 = 0.787$; $p < 0.0001$) with regard to the distribution of genetic diversity within the CRB relative to the IBD model.

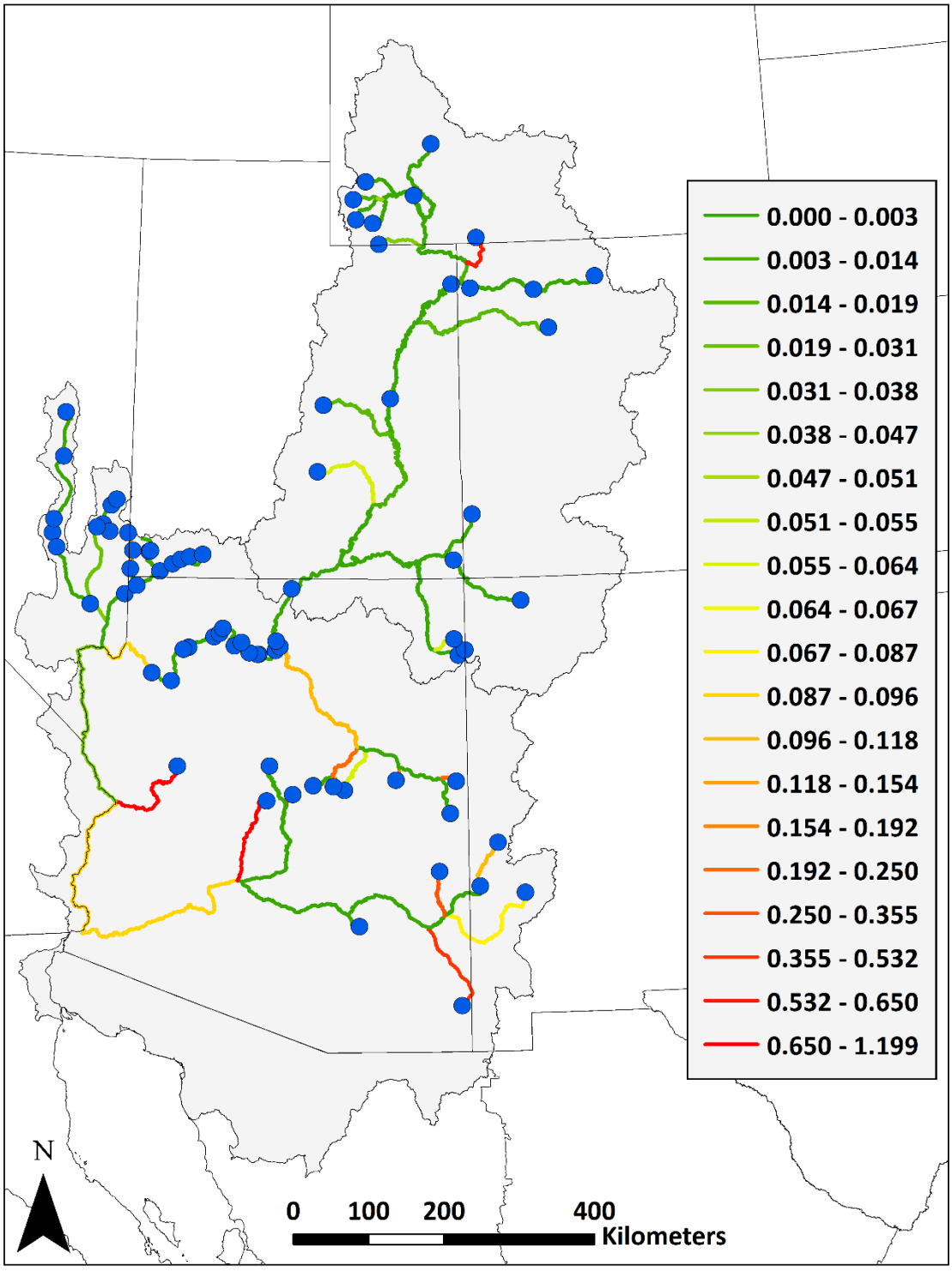

Figure 5. Genetic distance (=Linearized F_{ST}) explained by stream segments in the Colorado River Basin (CRB), as calculated by STREAMTREE. Red segments indicate elevated genetic distance separating populations whereas green depicts very low genetic distance. Elevated values largely equate to the lower CRB, whereas low genetic distances separate upper CRB sites.

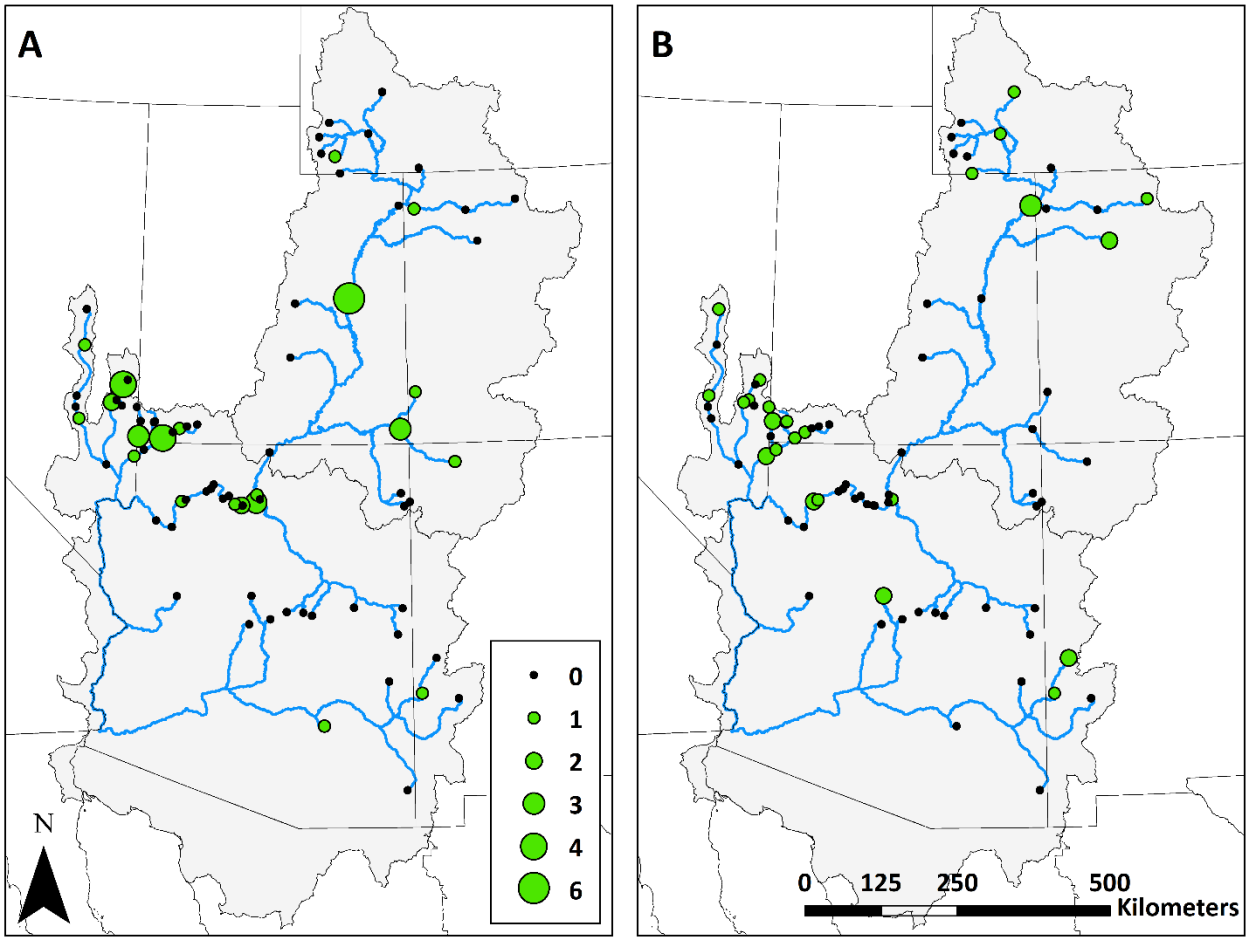

Figure 6. Sink (A) and source (B) populations detected by analysis of migration using BAYESASS3-SNPS. Populations as migration rate outliers are depicted in green. The size of each green circle represents the number of times it was detected as an outlier. Larger circles indicate a sink (A) or source (B) was detected multiple times (per numerical score).

IV. Molecular Dating of Speckled Dace Lineages in the Death Valley Ecosystem

Introduction

The juxtaposition of molecular data with the fossil record not only provides taxonomic groups with an estimate of time since divergence (Pisani and Liu 2015), but also a mechanism by which to gauge the process (Bell 2015). This approach has been employed numerous times as a means of bookmarking significant events in the evolution of global biodiversity, to include the origins of animals (dos Reis et al. 2015), the shift in arthropods from water to land (Lozano-Fernandez et al. 2016), the tempo of speciation in primates (Yoder and Yang 2000), the timing of adaptive radiation in Darwin's finches (Sato et al. 2001), and the rapid evolution of viruses (Leitner and Albert 1999).

While a powerful technique, molecular dating also has pitfalls. For example, it is only as accurate as the calibrations that generate the estimates (Graur and Martin 2004; Sauquet et al. 2012). Concomitant with this is the fact that many lineages suffer from a dearth of accurately dated fossils (Norell and Novacek 1992) and this, in turn, requires the implementation of other, more ancillary approaches that calibrate the clock (Ho et al. 2015). However, these must be well justified in that they may not correlate with the vicariant events that drive speciation (De Baets et al. 2016). Such asides obviously loom large when molecular data are employed as a surrogate for time since diversification.

The Death Valley (DV) ecosystem in eastern California/ western Nevada is a harsh and arid landscape that encompasses an endemic and presumably archaic fish fauna (Sigler & Sigler 1987). As such, it provides an ideal setting to ascertain the impact of geomorphic events on the evolution of aquatic organisms. Not surprisingly, it has also become a contemporary arena for

competing hypotheses that attempt to pinpoint the origin and evolutionary trajectory of an endemic fish, the Devil's Hole Pupfish (DHP; *Cyprinodon diabolis*). A key aspect is that results in support of each hypothesis must be congruent with the mechanisms that underpin molecular clocks. These are: The variation implied in fossil dates, and the manner by which biogeographic clocks are calibrated.

Controversy

The study organism, a small (<25mm TL) Cyprinodontiform fish, is endemic to an aquifer-fed pool (i.e., Devil's Hole) within a limestone cavern at Ash Meadows National Wildlife Refuge in the Mojave Desert east of Death Valley. The species has already gained notoriety in that it was declared as endangered in 1967 (32 FR 4001), but subsequent anthropogenic use of groundwater in the 1980s reduced the water level at Devil's Hole and impacted its survival, thus galvanizing intense legal battles (Deacon and Williams 1991). Given this, it has been described as the rarest fish on Earth (https://www.economist.com/science-and-technology/2013/01/19/in-a-hole; https://www.cnet.com/features/the-deep-dark-quest-to-save-the-pupfish/).

With regard to the controversy, one hypothesis argues that DHP diversified in lockstep with the origin of Devil's Hole at ~60 ka (thousand years before present) (Sağlam et al. 2016, 2018). This age is in agreement with a Pliocene-age *Cyprinodon* fossil from Death Valley (Miller 1945; Martin and Wainwright 2011; Martin et al. 2016). A second hypothesis advocates a more recent diversification (6.5-2.5 ka; Martin et al. 2016; Martin and Höhna 2017) based upon the formation (at ~8 ka) of Lake Chichancanab (northern Yucatan Peninsula, Mexico). However, this estimate conflicts with our current understanding of palaeohydrology in western North

American, as well as proposed ages of all North American *Cyprinodon* (Knott et al. 2018). Also of interest is the fact that this estimate requires an "overland dispersal" of pupfish as a mechanism of compliance (Martin and Turner 2018).

An Alternative Test

Two issues must be established so as to calibrate a molecular clock for the region: The veracity of geological events used as anchor points, as well as an independent test of dispersal mechanisms for fishes. The only other fish distributed throughout DV is the Speckled Dace (*Rhinichthys osculus*; SPD), a small cyprinid with considerable among-population variability (Sada et al. 1995; Oakey et al. 2004; Furiness 2012). Five subspecies of 'special concern' are recognized in the region (Moyle et al. 2015), none of which are formally described save one (Gilbert 1893; La Rivers 1962; Williams et al. 1982; Deacon and Williams 1984). Two reside within the Owens Basin (Long Valley and Owens River), and three in the Amargosa Basin [Ash Meadows (*R. o. nevadensis*), Amargosa Canyon, and Oasis Valley].

SPD would represent a model system for assessing palaeohydrological connections in the region, and thus, as an alternative test for the diversification of DHP. Both genera overlap in the region (Moyle 2002), and each has inhabited the Owens and Amargosa basins since at least late Pleistocene, possibly longer (Miller 1945), but with the understanding that differential colonization may have occurred. Thus, a parallel test to compare against the fit of an historic versus contemporary model for Devil's Hole would be of interest not only for DHP but also SPD.

Pupfish in the Amargosa Basin share a closer affinity with congeners in the Colorado River Basin than the Owens River (Echelle et al. 2005; Echelle 2008), although this conflicts

with known palaeohydrology (Knott et al. 2008). A similar origin has been proposed for SPD, but genetic evidence (Oakey et al. 2004) suggests that Owens and Amargosa SPD are instead sister taxa with a close relationship to the Lahontan. Regardless of origin, the co-distribution of both genera is supported by numerous studies: Molecular ecology (Smith and Dowling 2008), palaeohydrology (Jayko et al. 2008; Knott et al. 2008), fossil evidence (Miller 1945; Smith et al. 2017).

This study provides a comprehensive population-level evaluation of DV SPD in the context of basin evolution. The timing of divergence was evaluated for all populations, with known geological events serving as anchor points. Population structure and historic introgression were evaluated to gauge contemporary and historical gene flow within and among basins. Nuclear loci under selection were identified and evaluated as a means of testing for adaptive divergence in isolated populations. Finally, genetic distinction of all narrowly endemic DV populations was statistically contrasted, and management implications discussed.

Methods

Sampling

Eight locations were selected for population genetic analysis, representing all five putative subspecies (N=50 samples/ 4 localities in the Owens; N=62 samples/ 4 localities in the Amargosa) (Figure 1). Sampling spanned 1989-2016, with Long Valley sampled twice (1989, 2016) and Oasis Valley three times (1993, 2004, 2016). Five Eastern Blacknose Dace (*R. atratulus*) from the Rogue River in Michigan served as outgroup.

Data Collection

Whole genomic DNA extractions were performed using one of four methods: Gentra Puregene DNA Purification Tissue kit, QIAGEN DNeasy Blood and Tissue Kit, QIAamp Fast DNA Tissue Kit, or CsCl-gradient. Extracted DNA was visualized on a 2.0% agarose gel and quantified with a Qubit 2.0 fluorometer (Thermo Fisher Scientific, Inc.). Double digest Restriction-Site Associated DNA (ddRAD) library preparation followed the methods of Peterson et al. (2012). Barcoded samples (100 ng DNA each) were pooled in sets of 48 following Illumina adapter ligation, then size selected (at 375-425 bp; Chafin et al. 2017) using the Pippin Prep System (Sage Science). Size-selected DNA was subjected to 12 cycles of PCR amplification using Phusion high-fidelity DNA polymerase (New England Bioscience) according to manufacturer protocols. Subsequent quality checks were performed via Agilent 2200 TapeStation and qPCR to confirm successful library amplification. Three indexed libraries (N=144 samples) were pooled per lane for 100bp single-end (SE) sequencing (Illumina HiSeq 2000, University of Wisconsin Biotechnology Center; HiSeq 4000, University of Oregon Genomics & Cell Characterization Core Facility).

Thirty-six samples were selected for a 250bp paired-end (PE) library (Illumina HiSeq 2500, University of Wisconsin Biotechnology Center). This included range-wide sampling of *R. osculus* (N=22) and DV SPD (N=5). Several other *Rhinichthys* species (N=1 each: *R. atratulus*, *R. cataractae*, *R. evermanni*, *R. falcatus*, and *R. umatilla*) were included, as were outgroup taxa (N=1 each: *Tiaroga cobitis*, *Richardsonius balteatus*, *Mylocheilus caurinus*, and *Iotichthys phlegethontis*).

Both SE and PE libraries were de-multiplexed and filtered for quality using the process_radtags module of STACKS (Catchen et al. 2013). All reads with uncalled bases or Phred

quality scores <10 were discarded. Reads that passed quality filtering but had ambiguous barcodes were recovered when possible (≤ 1 mismatched nucleotide). Overlapping PE reads were merged (PEAR; Zhang et al. 2014), whereas those non-overlapping (≤ 5% of reads per sample) were discarded.

We employed a clustering threshold of 0.85 (PYRAD; Eaton 2014) for our separate *de novo* assemblies of ddRAD loci for SE and PE libraries. Reads with > four low quality bases (Phred quality score < 20) were removed from analysis. A minimum of 15 reads was required for a locus to be called for an individual. A filter to remove putative paralogs was applied by discarding loci with heterozygosity > 0.6. Those loci present in at least 50% of ingroup samples (N=56) were retained in the SE data, whereas for the PE assembly, at least 33 samples (92% of total) were required. These stipulations reduced the effects of missing data on phylogenetic tree branch lengths (Zheng and Wiens 2015). Loci containing >10 heterozygous sites were also removed from the final PE assembly.

Loci Under Selection

All loci recovered for the SE dataset were subjected to F_{ST} outlier analysis, with unlinked SNPs from pyRAD first analyzed in BAYESCAN v2.1 (Foll and Gaggiotti 2008). The recommended program settings were utilized, i.e. 20 pilot runs (5,000 generations each) followed by 100,000 Markov Chain Monte Carlo (MCMC) generations (including 50,000 burn-in). Data were thinned by the retention of every 10^{th} sample only, equating to 5,000 total MCMC samples. Outlier status was determined by a false discovery rate of 0.05.

BAYESCAN has the lowest Type I and II error rates among comparable software, yet a single outlier-detection method elicits some level of uncertainty (Narum and Hess 2011). Thus,

cross-validation of results was conducted using the hierarchical island model (ARLEQUIN; Excoffier and Lischer 2010) and the FDIST2 method (LOSITAN; Beaumont and Nichols 1996; Antao et al. 2008). Conditions for ARLEQUIN included 50 simulated groups, 100 simulated demes per group, and 20,000 coalescent simulations, with significance at $p<0.05$, whereas LOSITAN included 100,000 total simulations, assuming a "neutral" forced mean F_{ST}, 95% confidence interval (CI), and false discovery rate of 0.1. SNPs determined to be under selection by BAYESCAN and at least one other method were extracted from the SE dataset for downstream analysis.

Population Structure Analysis

An analysis of molecular variance (AMOVA) was calculated for the full SE SNP dataset (ARLEQUIN; Excoffier and Lischer 2010) and again for SNPs under selection (SE-select). In both cases, pairwise F_{ST} values were used to evaluate genetic differentiation among localities, and among events within localities.

A Maximum Likelihood (ML) approach (ADMIXTURE v1.3; Alexander et al. 2009) was utilized to assess population structure in both SE and SE-select datasets. A custom pipeline (https://github.com/stevemussmann/AdmixturePipeline) sampled one SNP per locus before sending data to ADMIXTURE. Clustering (K) values of 1 to 12 were evaluated, each with 20 replicates. Cross-validation (CV) values were calculated in ADMIXTURE following program instructions.

Output was evaluated (CLUMPAK; Kopelman et al. 2015) so as to automate the process of summarizing multiple independent ADMIXTURE runs. A Markov clustering algorithm identified different modes calculated by ADMIXTURE within a single K value, with a threshold at 0.9

identifying major clusters that were summarized using a custom script that visualized variability in CV values (https://github.com/stevemussmann/ADMIXTURE_cv_sum). The best explanation of population structure was that K value associated with the lowest CV score.

Phylogenetic Analysis

A phylogenetic tree of sampling localities was calculated using the SE dataset. The reversible polymorphism-aware phylogenetic model (POMO: Schrempf et al. 2016) was applied (IQ-TREE; Nguyen et al. 2014). This model accounts for incomplete lineage sorting by allowing polymorphic states within populations rather than following the traditional assumption in DNA substitution models that taxa are fixed for a specific nucleotide at a locus. A virtual population size of 19 was assumed to model genetic drift. Mutations followed a GTR substitution model, and rate heterogeneity was modeled at four rate categories. An ultrafast bootstrap algorithm (Hoang et al. 2018) performed 1,000 bootstrap replicates.

Introgression & Hybridization

Tests for introgression were performed within and among the Amargosa and Owens basins using Patterson's D-statistic (Durand et al. 2011). A Z-score was calculated from 1,000 bootstrap replicates for each test, with statistical significance evaluated at a Holm-Bonferroni adjusted threshold (α=0.0001). Preliminary evaluations indicated high levels of admixture in Amargosa Canyon, so the potential for hybrid origin was evaluated (HYDE; Blischak et al. 2018). The model employs a coalescent-based approach that compares a putative hybrid species against two candidate parental species and an outgroup taxon. Significance was evaluated by a Bonferroni adjusted threshold ($\alpha=8.3 \times 10^{-4}$). *Rhinichthys atratulus* was used as the outgroup

taxon for all tests of hybridization and introgression. To assess the extent and relative age of the Amargosa Canyon hybridization event, we calculated a hybrid index (Buerkle 2005) using the est.h function in the INTROGRESS R package (Gompert and Buerkle 2010). Data were filtered to use only fixed, unlinked SNPs among parental populations. Interspecific heterozygosity was calculated using the calc.intersp.het function in INTROGRESS, and results were visualized using the triangle.plot function.

Molecular Clock Analysis

The PE dataset was partitioned (IQ-TREE v1.5.5; Nguyen et al. 2014), with the top 10% of partition schemes searched with relaxed clustering and greedy algorithms (Lanfear et al. 2012, 2014). The minimum branch length was set to $1x10^{-6}$ and the safe numerical mode was utilized to prevent "numerical underflow" errors. We then determined an appropriate number of independent clock models for each partition (CLOCKSTAR2; Duchêne et al. 2014). A calculated tree topology (EXAML; Kozlov et al. 2015) served as input for branch length optimization. To select the optimal number of clock models, we implemented the PAM clustering algorithm with 500 bootstrap replicates.

A time-calibrated tree was calculated in BEAST v2.4.7 (Bouckaert et al. 2014). Data were divided into six partitions, with 365 of 372 loci (98.12%) as input for the TPM3u model. Two loci each (0.54%) were modeled by K2P and TPM2u. One locus each (0.27%) was modeled by K3Pu, JC, and F81. Invariant sites and gamma rate parameters were also applied to each partition, and a single relaxed log normal clock model was implemented. A birth-death process was used to describe the tree model. Two independent replicates of 100 million MCMC generations were conducted, with samples drawn every 5,000 generations. The first ten million

generations were discarded as burn-in following visual examination (Tracer v1.7; Rambaut et al. 2018).

In the BEAST analysis, five fossil calibrations were applied with the first representing the oldest known *Rhinichthys* fossil (8.4 Ma; Drewsey Formation: Smith et al. 2017). The earliest *R. osculus* fossil (at 4 Ma, Glenns Ferry Formation; Smith and Dowling 2008) was placed at the split between *R. osculus* and *R. cataractae*. The earliest *Richardsonius* fossil (3.5 Ma; Glenns Ferry Formation: Smith 1975; Neville et al. 1979; Kimmel 1982) was used to estimate the minimum age of the most recent common ancestor (MRCA) for *Iotichthys* and *Richardsonius*. The earliest *Mylocheilus* fossil (7.0 Ma; Chalk Hills Formation, Dowling et al. 2002; Smith et al. 2002) was placed as the MRCA between *Mylocheilus* and the *Iotichthys/ Richardsonius* clade. An *R. osculus* fossil from Owens Valley (45 ka) calibrated the split between Long Valley SPD and the remainder of the Owens and Amargosa subspecies (Smith et al. 2017). All calibration priors were described by a log normal distribution with a standard deviation of 1.5 Ma, save the Owens Basin fossil which was provided a 0.35 Ma standard deviation.

Bayes Factor Delimitation

A multispecies coalescent (MSC) based species delimitation method (BFD*: Leaché et al. 2014) was applied (SNAPP; Bryant et al. 2012) to test which subspecies, populations, or geographic divisions most appropriately divided observed genetic diversity into discrete, well-supported units. Data were filtered in the Phrynomics R package (Leaché et al. 2015) to remove invariant sites, non-binary SNPs, and loci appearing in >95% of individuals. Ten samples representing the most recent field collections from each of six ADMIXTURE-defined clusters were retained, and yielded 601 SNPS from 60 individuals. The prior value for the population mutation

rate (Θ) was estimated using mean pairwise sequence divergence (7.99 x 10^{-3}) between *R. osculus* and *R. cataractae*. This value was set as the mean of a gamma-distributed prior. The lineage birth rate (λ) of the Yule model was fixed using calculations from pyule (https://github.com/joaks1/pyule). A λ-value of 181.49 assumed tree height to be half of the maximum observed pairwise sequence divergence. Path sampling was set to 48 steps of 500,000 MCMC generations with 100,000 discarded as burn-in. Bayes factors (BF) were calculated from normalized marginal likelihoods and compared according to Leaché et al. (2014). Models were evaluated for statistical significance via BF (Kass and Raftery 1995).

Results

Alignment

The SE dataset yielded 12,556 loci ($\bar{x}$=10,024.6; σ=1,613.3) present in at least 50% of individuals (N=56). The actual proportion of missing data was 20.16%. Average sequencing depth was 61.07x (σ=22.48x). The PE dataset yielded 372 loci ($\bar{x}$=345.7; σ=65.9) in at least 91.67% of individuals (N=33), with the actual proportion of missing data at 8.66%. Just 23 loci were recovered for the *M. caurinus* sample, and was judged as a statistical outlier due to sample degradation. *Tiaroga cobitis* had the next fewest loci at 227. Average PE sequencing depth was 69.4x (σ=24.92x).

F_{ST} Outlier

BAYESCAN recovered 223 total loci under selection, with 46 under positive selection and 177 under balancing selection (Figure 2). These results were cross-referenced against the 410 loci recovered by ARLEQUIN, and the 100 loci found by LOSITAN. The three methods did not

converge on a consensus, with BAYESCAN and LOSITAN having no loci in common. However, 137 loci were congruent among BAYESCAN and ARLEQUIN, all of which were under balancing selection.

Each ADMIXTURE-identified cluster contained at least one monomorphic F_{ST} outlier locus. The Long Valley population exhibited the most, with 59 of 137 (43.07%) being fixed. The population with the next greatest number was Benton Valley with 18 (13.14%). Amargosa Canyon had the third highest, with seven (5.11%). Oasis Valley (N=2; 1.46%), Ash Meadows (N=1; 0.73%), and Owens Valley (N=1; 0.73%) had fewest. No private alleles were detected at any F_{ST} outlier locus for any population.

Population Structure

Genetic diversity within the Owens and Amargosa basins was best represented by six genetic clusters for SE (Figure 3A) and SE-select datasets (Figure 3B). All proposed subspecies were recovered as unique populations, but with Owens River divided into Bishop, CA (RUP and dorb) and Benton Valley (dhar). The SE-select results showed a similar trend but with greater admixture. Most admixture occurs among Amargosa Basin populations, although some introgression into the Owens Basin was also evident. Loci under selection indicated unique genetic signatures corresponding to the same six populations recovered by the full SE dataset.

AMOVA results for both datasets revealed strong genetic differentiation among sampling localities (SE F_{ST}=0.52; SE-select F_{ST}=0.40; $p<0.001$). Genetic differences among ADMIXTURE-defined clusters were also high (SE F_{CT}=0.49; SE-select F_{CT}=0.39; $p<0.001$), while the proportion of among-locality variation within clusters was low (SE F_{SC}=0.06; SE-select F_{SC}=0.02). The proportion of genetic variance distributed among ADMIXTURE-identified clusters

was greatest for the SE dataset (SE=49.18%; SE-select=39.11%) whereas the greatest source for the SE-select dataset was found within sampling localities (SE=47.79%; SE-select=59.6%3). The proportion of variation distributed among localities within clusters was very low for both datasets (SE=3.04%; SE-select=1.26%).

Pairwise F_{ST} values for both datasets demonstrated significant differentiation among sampling localities (Table 1). All for the SE dataset were significant save those comparing 1993 and 2016 collections of Oasis Valley, and the two Owens Valley (dorb/ RUP) populations. The greatest pairwise values were comparisons of Long Valley with other localities (F_{ST}=0.508-0.783), with Benton Valley the second greatest (F_{ST}=0.298-0.685). Similar trends were observed for SE-select, with the most notable exception a lack of significant pairwise F_{ST} values when within-population sampling events were compared. Although pairwise F_{ST} values were lower overall, Long Valley (F_{ST}=0.536-0.770) and Benton Valley (F_{ST}=0.139-0.770) again exhibited higher levels of divergence relative to other localities.

Introgression & Hybridization

HyDe revealed Amargosa Canyon as a putative hybrid between Oasis Valley and Ash Meadows (p=0.0004). A hybrid index was calculated for this scenario (Figure 4) based upon 29 SNPs representing fixed differences between Ash Meadows and Oasis Valley. The genomic composition of each Amargosa Canyon sample is an approximate 50/50 representation of each parent (Figure 4B), and ranged from recent to historical hybridization (Figure 4A). Amargosa Canyon samples were collected ten months after a significant flood event that temporarily reconnected Amargosa Canyon to upstream populations (i.e., Ash Meadows and Oasis Valley).

Tests of introgression via the four-taxon D-statistic test failed to indicate significant introgression of Oasis Valley or Ash Meadows alleles into Amargosa Canyon (Figure 5). This is explained by the nearly equal contribution of each subspecies to the Amargosa Canyon genome, which homogenizes the distribution of ABBA and BABA sites among species involved in these tests. Only 169 of 5,000 tests (3%, Oasis Valley) and 79 of 5,000 (2%, Ash Meadows) showed significant introgression into Amargosa Canyon.

Oasis Valley did show significant historical mixing with Owens Basin subspecies, with 1,069 of 5,000 tests (21%, Benton Valley), 3,469 of 10,000 (35%, Owens River), and 2,311 of 5,000 (46%, Long Valley) showing evidence of mixing with Oasis Valley. Benton Valley was introgressed by Long Valley (5,865 of 10,000 tests: 59%). The Owens River showed a high level of mixing with Benton Valley (8,117 of 10,000 tests: 81%), the highest recorded for any group in either basin. This may explain the elevated F_{ST} values that separate this population from others in the area. Given the rarity of hydrological connections (Knott et al. 2008), many of the inter-basin mixing events may be the result of incomplete lineage sorting (Maddison and Knowles 2006) or ancient hybridization events.

Phylogenetic Analysis

The POMO analysis revealed that Owens Basin localities are paraphyletic with respect to the Amargosa Basin (Figure 6). Owens River SPD from a locality near Bishop, CA and a private pond formed sister taxa, and were more closely related to the three Amargosa River subspecies than to Benton Valley. Relationships were consistent with expected patterns that resulted from transfer of SPD from the Owens to the Amargosa Basin as a result of hydrological connections.

All nodes were well supported, and relationships were congruent with those recovered in the fossil-calibrated tree.

Molecular Clock Analysis

The root of the fossil-calibrated tree (mean=16.09 Ma; 95% CI=14.21-18.38 Ma) coincides with a Miocene origin of modern western fish fauna (Smith 1981; Minckley et al. 1986). Divergence dates within the outgroup clade are older than recorded in earlier studies. Mean age for the *Iotichthys*/ *Richardsonius* split was previously estimated at 3.5 Ma (Houston et al. 2010), in contrast to the 6.09 Ma (95% CI=5.7-7.11 Ma) derived herein. The split between this clade and *Mylocheilus* was historically deeper at 12.1 Ma (95% C =11.5-13.69 Ma), compared to the previous value of 6.2 Ma (Houston et al. 2010). However, this could be an artifact of inflated branch lengths resulting from a high proportion of missing data in *Mylocheilus*.

Smith et al. (2017) recovered a mean age of 7.8 Ma for the *R. osculus*/ *R. cataractae* split, which is coincident with the age observed here (6.45 Ma; 95% CI=4.25-8.66 Ma). The MRCA of *Rhinichthys* (13.92 Ma; 95% CI=13.7-14.41 Ma) was also consistent with their origin date of 13.7-15 Ma (Smith et al. 2017). However, discrepancies appear in both topology and divergence time within the *R. osculus* clade. For instance, Smith et al. recovered *R. falcatus* as sister to Willamette River *R. osculus*, and *R. Umatilla* as sister to Oregon coastal drainage *R. osculus*. They also recovered *R. umatilla* as being older than *R. falcatus*. In contrast, ddRAD data recovered *R. falcatus* (2.87 Ma) as older than *R. umatilla* (1.11 Ma), with each being part of a Snake River/ Northeastern Bonneville clade.

Mean ages within the DV clade were generally older than anticipated, but with confidence intervals that overlapped with hydrological events. A mean age of 1.06 Ma (95% CI=0.44=1.76 Ma) was recovered for the split between Long Valley and the other DV taxa. This range encompasses the Long Valley Caldera formation which is credited with the isolation of Long Valley SPD (Bailey et al. 1976). The Owens and Amargosa Basin split is also older than anticipated (0.83 Ma; 95% CI=0.35-1.41 Ma), as are splits within the Amargosa Basin (Oasis Valley vs. Amargosa Canyon and Ash Meadows: 0.58 Ma, 95% CI=0.23-1.0 Ma; Amargosa Canyon vs. Ash Meadows: 0.38 Ma, 95% CI=0.13-0.69 Ma).

Bayes Factor Delimitation

The BFD analysis (Table 2) decisively favored splitting the Owens and Amargosa basins into six unique lineages (BF=307.08). In this regard, a BF of 10 is considered very strong support (Kass and Raftery 1995). The six groups correspond to those populations identified by ADMIXTURE (i.e., the five subspecies plus Benton Valley). The next two models most highly ranked served to collapse subspecies within the Amargosa Basin: one model grouped Oasis Valley and Amargosa Canyon while the other grouped Ash Meadows and Amargosa Canyon.

Discussion

A Molecular Clock for Speckled Dace

An overview of DV palaeohydrology (Figure 1) provides a baseline for subsequent interpretation of the time-calibrated tree. The headwaters of the Owens and Amargosa rivers lacked a continuous connection during the Pleistocene, and given this, DV and its surrounding basins must have become linked in a stepwise fashion (Knott et al. 2008). Lake Russell

(Pleistocene precursor to Mono Lake) formed a connection to the Lahontan Basin through the Walker River Basin prior to 1.6 Ma, and this served as a potential entry for SPD into the Owens Basin (Reheis et al. 2002a, b). From 1.2-0.6 Ma, there were at least four separate connections between Lakes Searles and Panamint (Jannik et al. 1991), allowing potential movement among these basins. During this approximate time, Death Valley seemingly contained several small, disconnected lakes (Morrison 1991; Reheis et al. 2002a), and was isolated from the Amargosa River until at least 0.6 Ma, possibly until 0.16 Ma (Hillhouse 1987). The connection between the Owens and Amargosa rivers occurred from 0.18-0.13 Ma (Morrison 1991; Lowenstein et al. 1999; Larsen et al. 2003).

This sequential pattern of connectivity is reflected in SPD relationships. Speciation events within the Owens/ Amargosa clade are consistent with the order of geological events, indicating that hydrology was influential in deriving relationships. Furthermore, dates associated with the oldest geological and molecular events overlap, with the MRCA of the Owens/ Amargosa/ Lahontan clade, dated at 3.93-1.54 Ma and coincident with the separation of the Walker River from Lake Russel. The formation of the Long Valley Caldera also corresponded with the split between Long Valley and all other Owens/ Amargosa SPD (95% CI=0.44-1.76 Ma). However, shallow nodes appear shifted from their putatively associated geological events, likely due to a bias associated with choice of sample and fossil calibration.

Multiple fossil calibrations, as employed herein, improve the reliability of divergence times (Soltis et al. 2002; Smith and Peterson 2002; Conroy and van Tuinen 2003). However, additional important factors must be considered when relaxed clock methods are utilized (Benton and Donoghue 2007; Ho and Phillips 2009; Roquet et al. 2013). These center on distance to calibrated nodes (Rutschmann et al. 2007; Marshall 2008), and their balanced distribution across

the tree (Saladin et al. 2017). In this study, two (of five) occur at nodes defining outgroup relationships, while another two are found deep within the *Rhinichthys* clade. The only recent calibration node is found within the DV clade. These are unevenly distributed in the tree, with many clades somewhat distant. This, in turn, may also provide a more historic bias for divergence times, as evidenced by the Moapa/ Pahranagat clade. The Moapa and White rivers connected to the Virgin River during the late Pleistocene/ early Holocene (Hubbs and Miller 1948; Smith 1978), whereas the divergence estimated herein is older (95% CI=1.22-0.22 Ma). Additionally, those that overlap with corresponding geological events in the DV clade consistently do so in the youngest extent of their CI estimates.

Despite this caveat, palaeohydrology remains the most likely explanation for divergent SPD lineages in the Owens and Amargosa basins. Phylogenetic analysis of the SE dataset revealed paraphyly of the Owens Basin with respect to the Amargosa. Samples from the Owens River near Bishop were more closely related to the Amargosa Basin than those from Benton Valley, which split from the Amargosa at 1.76-0.44 Ma. This predates the Owens/ Amargosa split, and the 0.18-0.13 Ma connection of the Owens Basin to the Amargosa. Logic dictates that the Owens/ Amargosa split is younger than that for Amargosa/ Benton Valley, and could be consistent with the 0.18-0.13 Ma Owens/ Amargosa hydrological connection. Thus, results remain consistent with a hypothesis of geological events as a driver of divergence among SPD.

The strongest contradiction to the palaeohydrology hypothesis is found in Amargosa lineage diversification. The alternative hypothesis requires diversification to have occurred external to the modern Amargosa basin. Allopatric speciation was possible in that Panamint Valley separated into northern and southern components when water levels fluctuated during the Pleistocene (Jayko et al. 2008; Knott et al. 2008), and the molecular date for the Ash Meadows/

Oasis Valley split (95% CI=1.0-0.23 Ma) precedes the 0.18-0.13 Ma Owens/ Amargosa hydrological connection. The separation of Amargosa Canyon from Ash Meadows (95% CI=0.69-0.13 Ma) is also coincident with Lake Panamint joining with the Amargosa Basin. This would allow the Amargosa Canyon SPD hybridization event to occur during late-Pleistocene pluvial periods. However, SPD readily hybridize upon secondary contact, as evidenced by the apparent recent mixing of Amargosa Canyon with conspecifics following the October 2015 flood. It therefore seems unlikely that Oasis Valley and Ash Meadows forms would have maintained isolation in a lacustrine environment during the Panamint to Amargosa transfer, and this alternative hypothesis is rejected in favor of an historic bias in molecular ages.

A Molecular Clock for Desert Pupfish

These results pose two important questions for divergence dating of pupfish in the region: 1) For how long have pupfish and SPD been contemporary in these watersheds, and 2) Is *Cyprinodon* evolution also coincident with sequential geological events? Although answers to both are inconclusive, results herein provide guidance in evaluating pupfish divergence. A controversial Pliocene-aged *Cyprinodon* fossil was recorded from Death Valley (Miller 1945), however both its age and relationship to modern *Cyprinodon* have been questioned (Martin et al. 2016). If the age and relationship is accurate, then *Cyprinodon* has been subjected to the same hydrological events as have *Rhinichthys*.

This second question is complicated by incomplete taxon sampling in recent *Cyprinodon* phylogenies. Martin et al. (2016) excluded Owens Pupfish (*C. radiosus*) in their analyses, whereas Sağlam et al. (2016) did not, yet still employed a narrow focus on *C. diabolis* and *C. nevadensis mionectes*. The most comprehensive *Cyprinodon* phylogeny (Echelle 2008) indicated

that Colorado River Basin (CRB) and Chihuahuan Desert species shared a closer relationship with Amargosa Basin pupfishes than did the Owens Pupfish, indicating a possible CRB origin for the former. Echelle (2008) relied upon three mtDNA markers, and several unresolved or poorly supported nodes emerged from those analyses. However, analyses using modern nuclear DNA markers most often yield increased resolution and a capacity to allocate problematic taxa (Eaton and Ree 2013; Díaz-Arce et al. 2016). Thus, a comprehensive pupfish phylogeny based on modern techniques is a necessity, particularly given the lack of a verifiable connection between DV and the CRB for the past 3 million years (Brown and Rosen 1995; Enzel et al. 2002; Knott et al. 2008). If the tree topology for *Cyprinodon* matched known hydrological connections, then colonization via river systems clearly offers a more parsimonious explanation than overland dispersal (Knott et al. 2018; Martin and Turner 2018).

Subspecific Divisions

MSC-based species delimitation methods outperform their distance-based counterparts (Yu et al. 2017), and have successfully delineated taxonomic units in several problematic groups (Hedin 2014; Hedin et al. 2015; Herrera and Shank 2016). However, a cautious interpretation is important, not only because over-splitting of taxa can occur under certain conditions (Barley et al. 2018; Leaché et al. 2018), but also due to the a propensity for MSC-based methods to parse populations rather than speciation events (Sukumaran and Knowles 2017). This stems from the many assumptions implicit to MSC, such as random mating, neutral markers, a lack of post-speciation gene flow, no within-locus recombination, and no linkage disequilibrium (Degnan and Rosenberg 2009). Despite this, species delimitation tools remain useful as a mechanism to improve our understanding of biodiversity (Leaché et al. 2018), particularly when non-genetic

properties of taxonomic groups are also considered (Barley et al. 2018). Six lineages identified herein were considered as evolutionarily significant units (ESU; Ryder 1986; Waples 1991; Moritz 1994).

While statistically significant differentiation of SPD lineages was demonstrated, morphological, ecological, and life history information must also be considered. Sada (1995) provided the most comprehensive overview of putative DV subspecies. He focused on broader regional morphological trends, yet also found "highly significant differences among all populations for all meristic and mensural characters" (Sada et al. 1995). Ordination revealed unique body shapes (i.e., slender and elongate versus short and deep-bodied) that are typical responses to stream flow (Brinsmead and Fox 2005; Collin and Fumagalli 2011). Long Valley and Ash Meadows SPD occur in spring habitats and their outflows, whereas others in this study are found in stream habitats. Benton Valley and Owens SPD are known from cold-water streams and irrigation ditches, whereas Oasis Valley and Amargosa Canyon SPD are native to the Amargosa River. Thus, environmental response to habitat is a contributing factor when body shape differences are considered.

Taxon-specific meristic counts must also be interpreted carefully. Ranges of counts frequently overlap, and additional details for putative subspecies are lacking or conflated with other taxa (Moyle et al. 2015). This is especially true of Oasis Valley SPD, initially lumped with Ash Meadows SPD (Gilbert 1893; La Rivers 1962), but with a lack of morphological details when first noted as a distinct subspecies (Williams et al. 1982; Deacon and Williams 1984). Likewise, few details are available for Amargosa Canyon SPD (Scoppettone et al. 2011), other than a qualitative description that details a smaller head depth, shorter snout-to-nostril length, greater length between anal and caudal fins, greater numbers of pectoral rays, and fewer

vertebrae when compared with other SPD subspecies (Moyle et al. 2015). Details for Ash Meadows SPD are often qualitative as well, with a diagnosis indicating an incomplete lateral line, seven anal fin rays, a relatively large head, short and deep body with a dark stripe running its entire length, and a small eye (Gilbert 1893).

The Owens River SPD is recorded as being locally variable. Meristic counts (Moyle et al. 2015) are summarized for four populations, including Benton Valley, which prevents most *ad hoc* within-basin comparisons. However, it does have maxillary barbels as a characteristic distinguishing it from conspecifics in surrounding basins. Furthermore, Benton Valley populations have qualitatively longer pelvic fins, and lower lateral line and pore scale counts relative to others within-basin (Moyle et al. 2015). In contrast, Long Valley SPD has a higher pectoral and pelvic fin ray count, elevated lateral line scale count, and fewer lateral line pores.

Ecological, life history, and morphological data are not conclusive enough to define taxonomic units in DV. However, this represents a deficiency in data rather than homogeneity of traits, and there are indications that subspecies may segregate morphologically. In contrast, multiple lines of genetic data provide clear signals of differentiation, and inferences from modern genomic techniques define gaps in traditional ecological knowledge (Crandall et al. 2000; Funk et al. 2012). F_{ST} outlier loci were used in this context to evaluate potential ecological adaptation of SPD lineages, as loci under selection have the potential to reveal cryptic signals of adaptive divergence (Tigano et al. 2017). While this method does not replace traditional field observations (Funk et al. 2012), it does reveal that neutral variation among the DV subspecies aligns with adaptive variation.

Benton Valley was last sampled in 1989. A previous phylogeny constructed from restriction-site mapping of mitochondrial genomes failed to distinguish it from Owens River

SPD (Oakey et al. 2004). However, nuclear data reported herein show a clear distinction between the two. Unfortunately this area was impacted by a flood in 1989 (Moyle et al. 2015), and wildlife managers have been unable to access the area, generating uncertainty over the persistence of this population (Steve Parmenter, CDFW, personal communication).

Genomic tools also demonstrated the previously unknown hybrid status of Amargosa Canyon SPD. Hybridization has been commonplace among desert fishes (Bangs et al. 2018) and has acted as a mechanism of speciation (Gerber et al. 2001). Additionally, anthropogenic climate change has induced hybridization of divergent species (Muhlfeld et al. 2014; Canestrelli et al. 2017) and is therefore an evolutionary mechanism inherent to the post-Pleistocene desiccation of western North America (Woodhouse et al. 2010). Hybridization has long been a contentious conservation topic (Fitzpatrick et al. 2015), particularly given that a policy for hybrids is not contained within Endangered Species Act (vonHoldt et al. 2017). However, the historical and ongoing lineage mixing that generated Amargosa Canyon SPD has all appearances of a natural occurrence that should not preclude its protection (Allendorf et al. 2001).

Conclusion

Divergence dates of six distinct SPD lineages within the DV region conform to documented Pleistocene hydrological connections among basins, as demonstrated from reduced representation genomic analyses. This calls into question the contemporary estimates for *Cyprinodon* species, especially if those patterns of speciation also conform to known historical hydrology. The DV SPD lineages are narrowly endemic relicts of a Pleistocene ecosystem that now persists in small desert oases. Despite academic recognition of their precarious existence,

federal protection is lacking, save one (*R. o. nevadensis*). Results of this study validate these lineages as ESUs eligible for protection under existing conservation laws.

References

Alexander DH, Novembre J, Lange K (2009) Fast model-based estimation of ancestry in unrelated individuals. Genome Res 19:1655–1664. doi: 10.1101/gr.094052.109

Allendorf FW, Leary RF, Spruell P, Wenburg JK (2001) The problems with hybrids: setting conservation guidelines. Trends Ecol Evol 16:613–622. doi: 10.1016/S0169-5347(01)02290-X

Antao T, Lopes A, Lopes RJ, et al (2008) LOSITAN: A workbench to detect molecular adaptation based on a F_{ST}-outlier method. BMC Bioinformatics 9:323. doi: 10.1186/1471-2105-9-323

Bailey RA, Dalrymple GB, Lanphere MA (1976) Volcanism, structure, and geochronology of Long Valley Caldera, Mono County, California. J Geophys Res 81:725–744

Bangs MR, Douglas MR, Mussmann SM, Douglas ME (2018) Unraveling historical introgression and resolving phylogenetic discord within *Catostomus* (Osteichthys: Catostomidae). BMC Evol Biol 18:86. doi: 10.1186/s12862-018-1197-y

Barley AJ, Brown JM, Thomson RC (2018) Impact of model violations on the inference of species boundaries under the multispecies coalescent. Syst Biol 67:269–284. doi: 10.1093/sysbio/syx073

Beaumont MA, Nichols RA (1996) Evaluating loci for use in the genetic analysis of population structure. P Roy Soc Lond B Bio 263:1619. doi: 10.1098/rspb.1996.0237

Bell CD (2015) Between a rock and a hard place: applications of the "molecular clock" in systematic biology. Syst Biol 40:6–13. doi: 10.1600/036364415X686297

Benton MJ, Donoghue PCJ (2007) Paleontological evidence to date the tree of life. Mol Biol Evol 24:26–53. doi: 10.1093/molbev/msl150

Blischak PD, Chifman J, Wolfe AD, Kubatko LS (2018) HyDe: a Python package for genome-scale hybridization detection. Syst Biol syy023–syy023. doi: 10.1093/sysbio/syy023

Bouckaert R, Heled J, Kühnert D, et al (2014) BEAST 2: a software platform for Bayesian evolutionary analysis. PLoS Comput Biol 10:1–6. doi: 10.1371/journal.pcbi.1003537

Brinsmead J, Fox MG (2005) Morphological variation between lake- and stream-dwelling rock bass and pumpkinseed populations. J Fish Biol 61:1619–1638. doi: 10.1111/j.1095-8649.2002.tb02502.x

Brown WJ, Rosen MR (1995) Was there a Pliocene-Pleistocene fluvial-lacustrine connection between Death Valley and the Colorado River? Quaternary Res 43:286–296. doi: 10.1006/qres.1995.1035

Bryant D, Bouckaert R, Felsenstein J, et al (2012) Inferring species trees directly from biallelic genetic markers: bypassing gene trees in a full coalescent analysis. Mol Biol Evol 29:1917–1932

Buerkle CA (2005) Maximum-likelihood estimation of a hybrid index based on molecular markers. Mol Ecol Notes 5:684–687. doi: 10.1111/j.1471-8286.2005.01011.x

Canestrelli D, Bisconti R, Chiocchio A, et al (2017) Climate change promotes hybridisation between deeply divergent species. PeerJ 5:e3072. doi: 10.7717/peerj.3072

Catchen J, Hohenlohe PA, Bassham S, et al (2013) Stacks: an analysis tool set for population genomics. Mol Ecol 22:3124–3140. doi: 10.1111/mec.12354

Chafin TK, Martin BT, Mussmann SM, et al (2017) FRAGMATIC: *in silico* locus prediction and its utility in optimizing ddRADseq projects. Conserv Genet Resour 1–4

Collin H, Fumagalli L (2011) Evidence for morphological and adaptive genetic divergence between lake and stream habitats in European minnows (*Phoxinus phoxinus*, Cyprinidae). Mol Ecol 20:4490–4502. doi: 10.1111/j.1365-294X.2011.05284.x

Conroy CJ, van Tuinen M (2003) Extracting time from phylogenies: positive interplay between fossil and genetic data. J Mammal 84:444–455. doi: 10.1644/1545-1542(2003)084<0444:ETFPPI>2.0.CO;2

Crandall KA, Bininda-Emonds ORP, Mace GM, Wayne RK (2000) Considering evolutionary processes in conservation biology. Trends Ecol Evol 15:290–295. doi: 10.1016/S0169-5347(00)01876-0

De Baets K, Antonelli A, Donoghue PCJ (2016) Tectonic blocks and molecular clocks. Philos T Roy Soc B 371:. doi: 10.1098/rstb.2016.0098

Deacon JE, Williams CD (1991) Ash Meadows and the legacy of the Devils Hole pupfish. In: Minckley WL, Deacon JE (eds) Battle Against Extinction: Native Fish Management in the American West. University of Arizona Press, Tucson, AZ, pp 69–87

Deacon JE, Williams JE (1984) Annotated list of the fishes of Nevada. P Biol Soc Wash 97:103–118

Degnan JH, Rosenberg NA (2009) Gene tree discordance, phylogenetic inference and the multispecies coalescent. Trends Ecol Evol 24:332–340. doi: 10.1016/j.tree.2009.01.009

Díaz-Arce N, Arrizabalaga H, Murua H, et al (2016) RAD-seq derived genome-wide nuclear markers resolve the phylogeny of tunas. Mol Phylogenet Evol 102:202–207. doi: 10.1016/j.ympev.2016.06.002

dos Reis M, Thawornwattana Y, Angelis K, et al (2015) Uncertainty in the timing of origin of animals and the limits of precision in molecular timescales. Curr Biol 25:2939–2950. doi: 10.1016/j.cub.2015.09.066

Dowling TE, Tibbets CA, Minckley W, et al (2002) Evolutionary relationships of the Plagopterins (Teleostei: Cyprinidae) from cytochrome b sequences. Copeia 2002:665–678

Duchêne S, Molak M, Ho SYW (2014) ClockstaR: choosing the number of relaxed-clock models in molecular phylogenetic analysis. Bioinformatics 30:1017–1019. doi: 10.1093/bioinformatics/btt665

Durand EY, Patterson N, Reich D, Slatkin M (2011) Testing for ancient admixture between closely related populations. Mol Biol Evol 28:2239–2252. doi: 10.1093/molbev/msr048

Eaton DA (2014) PyRAD: assembly of *de novo* RADseq loci for phylogenetic analyses. Bioinformatics 30:1844–1849. doi: 10.1093/bioinformatics/btu121

Eaton DA, Ree RH (2013) Inferring phylogeny and introgression using RADseq data: an example from flowering plants (*Pedicularis*: Orobanchaceae). Syst Biol 62:689–706. doi: 10.1093/sysbio/syt032

Echelle AA (2008) The western North American pupfish clade (Cyprinodontidae: *Cyprinodon*): mitochondrial DNA divergence and drainage history. In: Reheis MC, Hershler R, Miller DM (eds) Special Paper 439: Late Cenozoic Drainage History of the Southwestern Great Basin and Lower Colorado River Region: Geologic and Biotic Perspectives. Geological Society of America, Boulder, CO, pp 27–38

Echelle AA, Carson EW, Echelle AF, et al (2005) Historical biogeography of the new-world pupfish genus *Cyprinodon* (Teleostei: Cyprinodontidae). Copeia 2005:320–339. doi: 10.1643/CG-03-093R3

Enzel Y, Knott JR, Anderson K, et al (2002) Is there any Eeidence of mega-Lake Manly in the eastern Mojave Desert during oxygen isotope stage 5e/6? - A comment on Hooke, R.L. (1999), Lake Manly (?) shorelines in the eastern Mojave Desert, California, Quaternary Research, v. 52, p. 328–336. Quaternary Res 57:173–176. doi: 10.1006/qres.2001.2300

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

Fitzpatrick BM, Ryan ME, Johnson JR, et al (2015) Hybridization and the species problem in conservation. Current Zoology 61:206–216. doi: 10.1093/czoolo/61.1.206

Foll M, Gaggiotti O (2008) A genome-scan method to identify selected loci appropriate for both dominant and codominant markers: a Bayesian perspective. Genetics 180:977–993. doi: 10.1534/genetics.108.092221

Funk WC, McKay JK, Hohenlohe PA, Allendorf FW (2012) Harnessing genomics for delineating conservation units. Trends Ecol Evol 27:489–496. doi: 10.1016/j.tree.2012.05.012

Furiness SJ (2012) Population Structure of Death Valley System Speckled Dace (*Rhinichthys osculus*). MS, Texas A&M University Corpus Christi

Gerber AS, Tibbets CA, Dowling TE (2001) The role of introgressive hybridization in the evolution of the *Gila robusta* complex (Teleostei: Cyprinidae). Evolution 55:2028–2039. doi: 10.1111/j.0014-3820.2001.tb01319.x

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 American Fauna 7:229–234

Gompert Z, Buerkle CA (2010) INTROGRESS: a software package for mapping components of isolation in hybrids. Mol Ecol Resour 10:378–384

Graur D, Martin W (2004) Reading the entrails of chickens: molecular timescales of evolution and the illusion of precision. Trends Genet 20:80–86. doi: 10.1016/j.tig.2003.12.003

Hedin M (2014) High-stakes species delimitation in eyeless cave spiders (*Cicurina*, Dictynidae, Araneae) from central Texas. Mol Ecol 24:346–361. doi: 10.1111/mec.13036

Hedin M, Carlson D, Coyle F (2015) Sky island diversification meets the multispecies coalescent – divergence in the spruce-fir moss spider (*Microhexura montivaga*, Araneae, Mygalomorphae) on the highest peaks of southern Appalachia. Mol Ecol 24:3467–3484. doi: 10.1111/mec.13248

Herrera S, Shank TM (2016) RAD sequencing enables unprecedented phylogenetic resolution and objective species delimitation in recalcitrant divergent taxa. Mol Phylogenet Evol 100:70–79. doi: 10.1016/j.ympev.2016.03.010

Hillhouse JW (1987) Late Tertiary and Quaternary geology of the Tecopa Basin, southeastern California. U.S. Geological Survey Miscellaneous Investigations Map I-1728, scale 1:48,000

Ho SYW, Phillips MJ (2009) Accounting for calibration uncertainty in phylogenetic estimation of evolutionary divergence times. Syst Biol 58:367–380. doi: 10.1093/sysbio/syp035

Ho SYW, Tong KJ, Foster CSP, et al (2015) Biogeographic calibrations for the molecular clock. Biol Lett 11:20150194. doi: 10.1098/rsbl.2015.0194

Hoang DT, Chernomor O, von Haeseler A, et al (2018) UFBoot2: improving the ultrafast bootstrap approximation. Mol Biol Evol 35:518–522

Houston DD, Shiozawa DK, Riddle BR (2010) Phylogenetic relationships of the western North American cyprinid genus *Richardsonius*, with an overview of phylogeographic structure. Mol Phylogenet Evol 55:259–273. doi: 10.1016/j.ympev.2009.10.017

Hubbs CL, Miller RR (1948) The zoological evidence: correlation between fish distribution and hydrographic history in the desert basins of western United States. In: The Great Basin,

with Emphasis on Glacial and Postglacial Times. University of Utah, Salt Lake City, UT, pp 17–166

Jannik NO, Phillips FM, Smith GI, Elmore D (1991) A ^{36}Cl chronology of lacustrine sedimentation in the Pleistocene Owens River system. Geol Soc Am Bull 103:1146–1159. doi: 10.1130/0016-7606(1991)103<1146:ACCOLS>2.3.CO;2

Jayko AS, Forester RM, Kaufman DS, et al (2008) Late Pleistocene lakes and wetlands, Panamint Valley, Inyo County, California. In: Reheis MC, Hershler R, Miller DM (eds) Special Paper 439: Late Cenozoic Drainage History of the Southwestern Great Basin and Lower Colorado River Region: Geologic and Biotic Perspectives. Geological Society of America, Boulder, CO

Kass RE, Raftery AE (1995) Bayes factors. J Am Stat Assoc 90:773–795

Kimmel PG (1982) Stratigraphy, age, and tectonic setting of the Miocene-Pliocene lacustrine sediments of the western Snake River Plain, Oregon and Idaho. Cenozoic geology of Idaho: Idaho Bureau of Mines and Geology Bulletin 26:559–578

Knott JR, Machette MN, Klinger RE, et al (2008) Reconstructing late Pliocene to middle Pleistocene Death Valley lakes and river systems as a test of pupfish (Cyprinodontidae) dispersal hypotheses. In: Reheis MC, Hershler R, Miller DM (eds) Special Paper 439: Late Cenozoic Drainage History of the Southwestern Great Basin and Lower Colorado River Region: Geologic and Biotic Perspectives. Geological Society of America, Boulder, CO

Knott JR, Phillips FM, Reheis MC, et al (2018) Geologic and hydrologic concerns about pupfish divergence during the last glacial maximum. P Roy Soc Lond B Bio 285:. doi: 10.1098/rspb.2017.1648

Kopelman NM, Mayzel J, Jakobsson M, et al (2015) CLUMPAK: a program for identifying clustering modes and packaging population structure inferences across K. Mol Ecol Resour 15:1179–1191. doi: 10.1111/1755-0998.12387

Kozlov AM, Aberer AJ, Stamatakis A (2015) ExaML version 3: a tool for phylogenomic analyses on supercomputers. Bioinformatics 31:2577–2579

La Rivers I (1962) Fishes and Fisheries of Nevada. Nevada Fish and Game Commission, Carson City, NV

Lanfear R, Calcott B, Ho SY, Guindon S (2012) PartitionFinder: combined selection of partitioning schemes and substitution models for phylogenetic analyses. Mol Biol Evol 29:1695–1701

Lanfear R, Calcott B, Kainer D, et al (2014) Selecting optimal partitioning schemes for phylogenomic datasets. BMC Evol Biol 14:1–14

Larsen D, Knott JR, Jayko AS (2003) Comparison of middle Pleistocene records in Death, Panamint, and Tecopa Valleys, California: implications for regional paleoclimate. Geological Society of America Abstracts with Programs 33:333

Leaché AD, Banbury BL, Felsenstein J, et al (2015) Short tree, long tree, right tree, wrong tree: new acquisition bias corrections for inferring SNP phylogenies. Syst Biol 64:1032–1047. doi: 10.1093/sysbio/syv053

Leaché AD, Fujita MK, Minin VN, Bouckaert RR (2014) Species delimitation using genome-wide SNP data. Syst Biol 63:534–542. doi: 10.1093/sysbio/syu018

Leaché AD, Zhu T, Rannala B, Yang Z (2018) The spectre of too many species. Syst Biol syy051–syy051. doi: 10.1093/sysbio/syy051

Leitner T, Albert J (1999) The molecular clock of HIV-1 unveiled through analysis of a known transmission history. P Natl Acad Sci USA 96:10752–10757. doi: 10.1073/pnas.96.19.10752

Lowenstein TK, Li J, Brown C, et al (1999) 200 k.y. paleoclimate record from Death Valley salt core. Geology 27:3–6. doi: 10.1130/0091-7613(1999)027<0003:KYPRFD>2.3.CO;2

Lozano-Fernandez J, Carton R, Tanner AR, et al (2016) A molecular palaeobiological exploration of arthropod terrestrialization. Philos T Roy Soc B 371:. doi: 10.1098/rstb.2015.0133

Maddison W, Knowles L (2006) Inferring phylogeny despite incomplete lineage sorting. Syst Biol 55:21–30. doi: 10.1080/10635150500354928

Marshall CR (2008) A simple method for bracketing absolute divergence times on molecular phylogenies using multiple fossil calibration points. Am Nat 171:726–742. doi: 10.1086/587523

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. P Roy Soc Lond B Bio 283:1–10. doi: 10.1098/rspb.2015.2334

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. doi: 10.1111/mec.14404

Martin CH, Turner BJ (2018) Long-distance dispersal over land by fishes: extremely rare ecological events become probable over millennial timescales. P Roy Soc Lond B Bio 285:. doi: 10.1098/rspb.2017.2436

Martin CH, Wainwright PC (2011) Trophic novelty is linked to exceptional rates of morphological diversification in two adaptive radiations of *Cyprinodon* pupfish. Evolution 65:2197–2212. doi: 10.1111/j.1558-5646.2011.01294.x

Miller RR (1945) Four new species of fossil cyprinodont fishes from eastern California. Journal of the Washington Academy of Sciences 35:315–321

Miller RR (1946) Correlation between fish distribution and Pleistocene hydrography in eastern California and southwestern Nevada, with a map of the Pleistocene waters. J Geol 54:43–53

Minckley W, Hendrickson D, Bond C (1986) Geography of western North American freshwater fishes: description and relationships to intracontinental tectonism. In: Hocutt CH, Wiley EO (eds) The Zoogeography of North American Freshwater Fishes. John Wiley & Sons, New York, pp 519–613

Moritz C (1994) Defining "evolutionarily significant units" for conservation. Trends Ecol Evol 9:373–374

Morrison RB (1991) Quaternary stratigraphic, hydrologic, and climatic history of the Great Basin, with emphasis on Lakes Lahontan, Bonneville, and Tecopa. In: Morrison RB (ed) Quaternary Nonglacial Geology. Geological Society of America

Moyle P, Quiñones RM, Katz J, Weaver J (2015) Fish species of special concern of California. California Department of Fish and Wildlife

Moyle PB (2002) Inland Fishes of California. University of California Press, Berkeley, CA

Muhlfeld CC, Kovach RP, Jones LA, et al (2014) Invasive hybridization in a threatened species is accelerated by climate change. Nat Clim Change 4:620

Narum SR, Hess JE (2011) Comparison of F_{ST} outlier tests for SNP loci under selection. Mol Ecol Resour 11:184–194. doi: 10.1111/j.1755-0998.2011.02987.x

Neville C, Opdyke ND, Lindsay EH, Johnson NM (1979) Magnetic stratigraphy of Pliocene deposits of the Glenns Ferry Formation, Idaho, and its implications for North American mammalian biostratigraphy. Am J Sci 279:503–526

Nguyen L-T, Schmidt HA, von Haeseler A, Minh BQ (2014) IQ-TREE: a fast and effective stochastic algorithm for estimating maximum-likelihood phylogenies. Mol Biol Evol 32:268–274

Norell MA, Novacek MJ (1992) The fossil record and evolution: comparing cladistic and paleontologic evidence for vertebrate history. Science 255:1690–1693

Oakey DD, Douglas ME, Douglas MR (2004) Small fish in a large landscape: diversification of *Rhinichthys osculus* (Cyprinidae) in western North America. Copeia 2004:207–221

Peterson BK, Weber JN, Kay EH, et al (2012) Double digest RADseq: an inexpensive method for *de novo* SNP discovery and genotyping in model and non-model species. PLoS ONE 7:1–11. doi: 10.1371/journal.pone.0037135

Pisani D, Liu AG (2015) Animal evolution: only rocks can set the clock. Curr Biol 25:R1079–R1081. doi: 10.1016/j.cub.2015.10.015

Rambaut A, Drummond AJ, Xie D, et al (2018) Posterior summarization in Bayesian phylogenetics using Tracer 1.7. Syst Biol syy032–syy032. doi: 10.1093/sysbio/syy032

Reheis M, Sarna-Wojcicki A, Reynolds R, et al (2002a) Pliocene to middle Pleistocene lakes in the western Great Basin: ages and connections. In: Hershler R, Madsen DB, Currey DR (eds) Great Basin Aquatic Systems History: Smithsonian Contributions to the Earth Sciences. pp 53–108

Reheis MC, Stine S, Sarna-Wojcicki AM (2002b) Drainage reversals in Mono Basin during the late Pliocene and Pleistocene. Geol Soc Am Bull 114:991–1006. doi: 10.1130/0016-7606(2002)114<0991:DRIMBD>2.0.CO;2

Reich D, Green RE, Kircher M, et al (2010) Genetic history of an archaic hominin group from Denisova Cave in Siberia. Nature 468:1053

Roquet C, Thuiller W, Lavergne S (2013) Building megaphylogenies for macroecology: taking up the challenge. Ecography 36:13–26. doi: 10.1111/j.1600-0587.2012.07773.x

Rutschmann F, Eriksson T, Abu Salim K, Conti E (2007) Assessing calibration uncertainty in molecular dating: The assignment of fossils to alternative calibration points. Syst Biol 56:. doi: 10.1080/10635150701491156

Ryder OA (1986) Species conservation and systematics: the dilemma of subspecies. Trends Ecol Evol 1:9–10

Sada DW, Britten HB, Brussard PF (1995) Desert aquatic ecosystems and the genetic and morphological diversity of Death Valley system speckled dace. Am Fish S S 17:350–359

Sağlam İK, Baumsteiger J, Smith MJ, et al (2016) Phylogenetics support an ancient common origin of two scientific icons: Devils Hole and Devils Hole pupfish. Mol Ecol 25:3962–3973. doi: 10.1111/mec.13732

Sağlam İK, Baumsteiger J, Smith MJ, et al (2018) Best available science still supports an ancient common origin of Devils Hole and Devils Hole pupfish. Mol Ecol 27:839–842. doi: 10.1111/mec.14502

Saladin B, Leslie AB, Wüest RO, et al (2017) Fossils matter: improved estimates of divergence times in *Pinus* reveal older diversification. BMC Evol Biol 17:95. doi: 10.1186/s12862-017-0941-z

Sato A, Tichy H, O'hUigin C, et al (2001) On the origin of Darwin's finches. Mol Biol Evol 18:299–311. doi: 10.1093/oxfordjournals.molbev.a003806

Sauquet H, Ho SYW, Gandolfo MA, et al (2012) Testing the impact of calibration on molecular divergence times using a fossil-rich group: the case of *Nothofagus* (Fagales). Syst Biol 61:289–313

Schrempf D, Minh BQ, De Maio N, et al (2016) Reversible polymorphism-aware phylogenetic models and their application to tree inference. J Theor Biol 407:362–370

Scoppettone GG, Hereford ME, Rissler PH, et al (2011) Relative abundance and distribution of fishes within an established Area of Critical Environmental Concern, of the Amargosa River Canyon and Willow Creek, Inyo and San Bernardino Counties, California. US Geological Survey

Smith AB, Peterson KJ (2002) Dating the time of origin of major clades: molecular clocks and the fossil record. Annu Rev Earth Pl Sc 30:65–88. doi: 10.1146/annurev.earth.30.091201.140057

Smith G, Dowling T, Gobalet K, et al (2002) Biogeography and timing of evolutionary events among Great Basin fishes. Great Basin Aquatic Systems History 33:175–234

Smith GR (1975) Fishes of the Pliocene Glenns Ferry Formation, southwest Idaho. University of Michigan Museum of Paleontology, Papers on Paleontology 14:1–68

Smith GR (1981) Late Cenozoic freshwater fishes of North America. Annu Rev Ecol Syst 12:163–193

Smith GR (1978) Biogeography of intermountain fishes. Great Basin Naturalist Memoirs 2:17–42

Smith GR, Chow J, Unmack PJ, et al (2017) Evolution of the *Rhinichthys osculus* complex (Teleostei: Cyprinidae) in western North America. Miscellaneous Publications Museum of Zoology University of Michigan 204:1–83

Smith GR, Dowling TE (2008) Correlating hydrographic events and divergence times of speckled dace (*Rhinichthys*: Teleostei: Cyprinidae) in the Colorado River drainage. In: Reheis MC, Hershler R, Miller DM (eds) Special Paper 439: Late Cenozoic Drainage History of the Southwestern Great Basin and Lower Colorado River Region: Geologic and Biotic Perspectives. Geological Society of America, Boulder, CO, pp 301–317

Soltis PS, Soltis DE, Savolainen V, et al (2002) Rate heterogeneity among lineages of tracheophytes: integration of molecular and fossil data and evidence for molecular living fossils. P Natl Acad Sci USA 99:4430. doi: 10.1073/pnas.032087199

Sukumaran J, Knowles LL (2017) Multispecies coalescent delimits structure, not species. P Natl Acad Sci USA 114:1607. doi: 10.1073/pnas.1607921114

Tigano A, Shultz AJ, Edwards SV, et al (2017) Outlier analyses to test for local adaptation to breeding grounds in a migratory arctic seabird. Ecol Evol 7:2370–2381. doi: 10.1002/ece3.2819

vonHoldt BM, Brzeski KE, Wilcove DS, Rutledge LY (2017) Redefining the role of admixture and genomics in species conservation. Conserv Lett 11:e12371. doi: 10.1111/conl.12371

Waples RS (1991) Pacific salmon, *Oncorhynchus* spp., and the definition of "species" under the Endangered Species Act. Mar Fish Rev 53:11–22

Williams CD, Hardy TP, Deacon JE (1982) Distribution and status of fishes of the Amargosa River Canyon, California. Unpublished Report submitted to US Fish and WIldlife Service Endangered Species Office, Sacramento, CA

Woodhouse CA, Meko DM, MacDonald GM, et al (2010) A 1,200-year perspective of 21st century drought in southwestern North America. P Natl Acad Sci USA 107:21283–21288. doi: 10.1073/pnas.0911197107

Yoder AD, Yang Z (2000) Estimation of primate speciation dates using local molecular clocks. Mol Biol Evol 17:1081–1090. doi: 10.1093/oxfordjournals.molbev.a026389

Yu G, Rao D, Matsui M, Yang J (2017) Coalescent-based delimitation outperforms distance-based methods for delineating less divergent species: the case of *Kurixalus odontotarsus* species group. Sci Rep-UK 7:16124. doi: 10.1038/s41598-017-16309-1

Zhang J, Kobert K, Flouri T, Stamatakis A (2014) PEAR: a fast and accurate Illumina Paired-End reAd mergeR. Bioinformatics 30:614–620. doi: 10.1093/bioinformatics/btt593

Zheng Y, Wiens JJ (2015) Do missing data influence the accuracy of divergence-time estimation with BEAST? Mol Phylogenet Evol 85:41–49. doi: 10.1016/j.ympev.2015.02.002

Appendix

Table 1. Pairwise F_{ST} values calculated via AMOVA in Arlequin for Amargosa and Owens River basin Speckled Dace (*Rhinichthys osculus*). Cells shaded in blue represent comparisons among Amargosa Basin localities, whereas those in green compare localities within the Owens River Basin. Values below the diagonal were calculated for the full dataset of 12,556 loci. Values above the diagonal were calculated from 137 loci determined to be under selection. Bold F_{ST} values lack statistical significance, whereas all others are significant at $p<0.001$. Sites for which there were multiple sampling events (AMA, LVD) also reflect the year when collections occurred. SPD locations are: AMA=Oasis Valley; AMC=Amargosa Canyon; ASH and ASR=Ash Meadows (*Rhinichthys osculus nevadensis*); dhar=Benton Valley; dorb and RUP=Owens Valley; LVD=Long Valley. Numbers next to sampling locality names represent years during which repeated collections occurred.

	AMA-1993	AMA-2004	AMA-2016	AMC	ASH	ASR	RUP	dorb	dhar	LVD-1989	LVD-2016
AMA-1993	*	**-0.002**	**-0.011**	0.186	0.226	0.242	0.183	0.190	0.280	0.546	0.615
AMA-2004	0.026	*	**0.007**	0.253	0.374	0.380	0.229	0.259	0.383	0.643	0.721
AMA-2016	**-0.004**	0.027	*	0.209	0.256	0.300	0.201	0.223	0.327	0.607	0.683
AMC	0.251	0.259	0.241	*	0.149	0.152	0.105	0.104	0.259	0.536	0.607
ASH	0.403	0.438	0.403	0.311	*	0.084	0.217	0.273	0.337	0.605	0.669
ASR	0.401	0.430	0.402	0.300	0.219	*	0.225	0.271	0.309	0.603	0.663
RUP	0.318	0.296	0.297	0.272	0.369	0.375	*	**-0.003**	0.139	0.549	0.590
dorb	0.319	0.323	0.313	0.282	0.387	0.385	**0.019**	*	0.169	0.552	0.601
dhar	0.519	0.558	0.525	0.488	0.619	0.602	0.298	0.302	*	0.741	0.770
LVD-1989	0.681	0.716	0.688	0.658	0.755	0.743	0.518	0.508	0.603	*	**0.061**
LVD-2016	0.718	0.750	0.727	0.701	0.783	0.775	0.592	0.582	0.685	0.056	*

Table 2. Bayes Factor Delimitation for six populations of Speckled Dace (*Rhinichthys osculus*) recovered by ADMIXTURE analysis (Figure 3). Fourteen models were tested and ordered by model preference (=Rank) based upon Bayes factors (=BF) calculated by comparing Marginal Likelihood (=MarL) values computed in SNAPP. Models tested a maximum of six divisions (=Splits) of populations based upon current geography and known historical connections among basins. Letters below population names indicate the manner by which locations were grouped for each model (=Rank).

Oasis Valley	Ash Meadows	Amargosa Canyon	Owens River	Benton Valley	Long Valley	Splits	MarL	BF	Rank
A	B	C	D	E	F	6	-9741.865121	0	1
A	B	A	C	D	E	5	-10048.94228	-307.0771568	2
A	B	B	C	D	E	5	-10097.7503	-355.8851795	3
A	A	A	B	C	D	4	-10558.95611	-817.0909871	4
A	B	C	D	D	E	5	-10626.96155	-885.096432	5
A	B	A	C	C	D	4	-10933.87547	-1192.010345	6
A	A	B	C	C	D	4	-11161.8967	-1420.03158	7
A	A	A	B	B	C	3	-11444.03214	-1702.16702	8
A	A	A	A	B	C	3	-11864.85644	-2122.991321	9
A	B	B	C	C	D	4	-12470.13991	-2728.274787	10
A	A	A	A	A	B	2	-13185.04573	-3443.180606	11
A	B	B	C	C	C	3	-14343.6593	-4601.794181	12
A	A	A	A	B	B	2	-14480.36624	-4738.501122	13
A	A	A	B	B	B	2	-14515.77419	-4773.90907	14

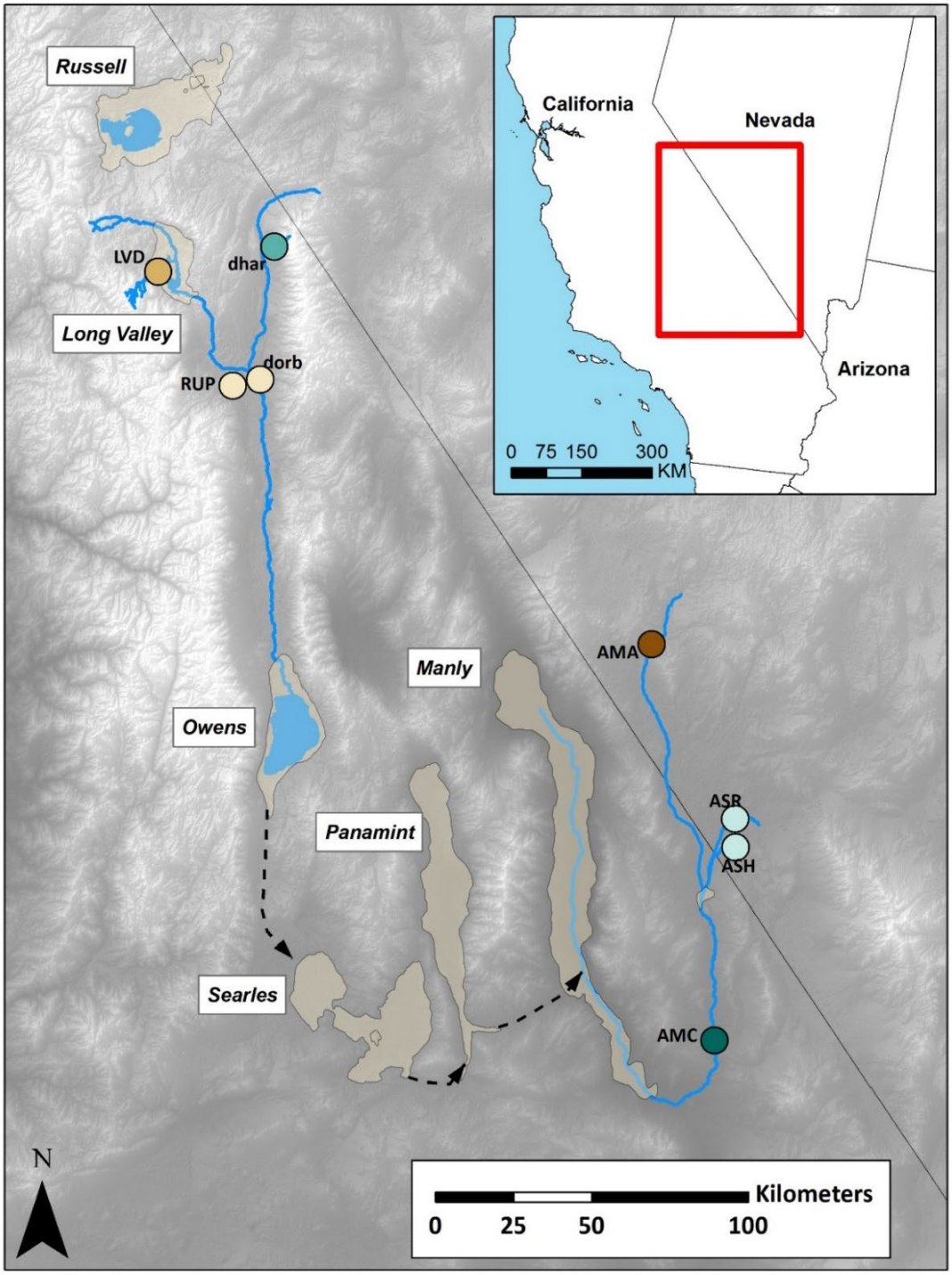

Figure 1. Sampling sites for Speckled Dace (*Rhinichthys osculus*; SPD) within the Owens and Amargosa river basins. Locality abbreviations and colors correspond with population structure bar-plots (Figure 3). Tan outlines represent the extent of Pleistocene lakes, as named in white boxes. Dashed arrows represent Pleistocene connections among basins. SPD locations are: AMA=Oasis Valley; AMC=Amargosa Canyon; ASH and ASR=Ash Meadows (*Rhinichthys osculus nevadensis*); dhar=Benton Valley; dorb and RUP=Owens Valley; LVD=Long Valley.

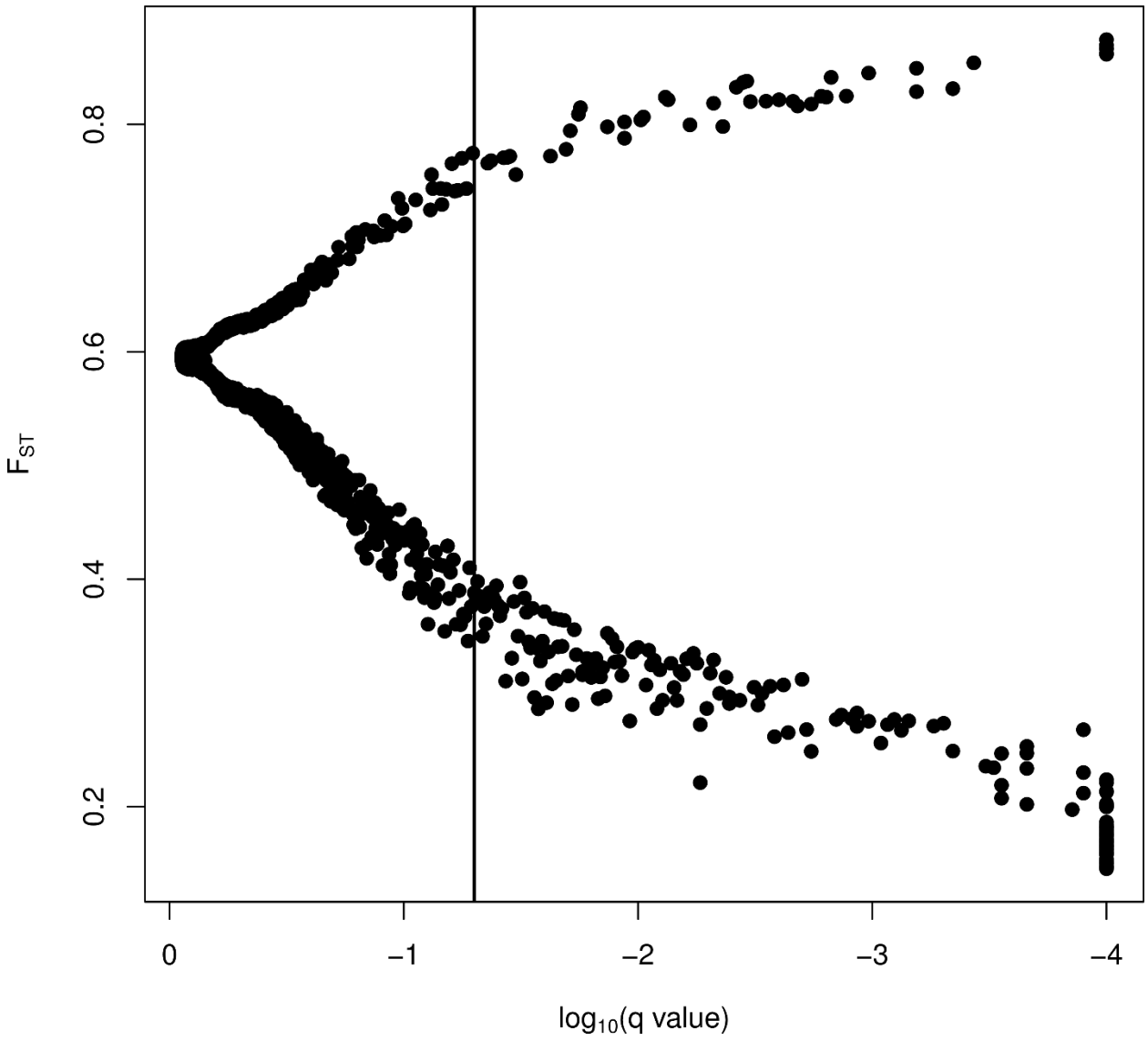

Figure 2. Results of BAYESCAN analysis for 12,556 loci recovered from 112 Speckled Dace (*Rhinichthys osculus*) samples from six populations. The F_{ST} value for each ddRAD locus is plotted against the q-value that represents the chance that a locus is under selection. The vertical bar shows the log-transformed critical q-value [$\log_{10}(q)$=-1.3] that signifies a locus under selection. Loci with a high F_{ST} value plotted to the right of the line are under positive selection, whereas those with lower F_{ST} values are indicative of balancing selection. For visualization purposes, those markers with a $\log_{10}(q)$ <-4 are shown at -4 on the plot.

Figure 3. Results of ADMIXTURE analyses showing population structure among Speckled Dace (*Rhinichthys osculus*; SPD) sampling sites. (A) Population structure recovered through analysis of the full single-end sequencing dataset (12,556 loci). (B) Structure recovered from 137 loci determined to be under selection by F_{ST} outlier analyses. SPD locations are: AMA=Oasis Valley; AMC=Amargosa Canyon; ASH and ASR=Ash Meadows (*Rhinichthys osculus nevadensis*); dhar=Benton Valley; dorb and RUP=Owens Valley; LVD=Long Valley. Numbers next to locality names represent the year when collection occurred.

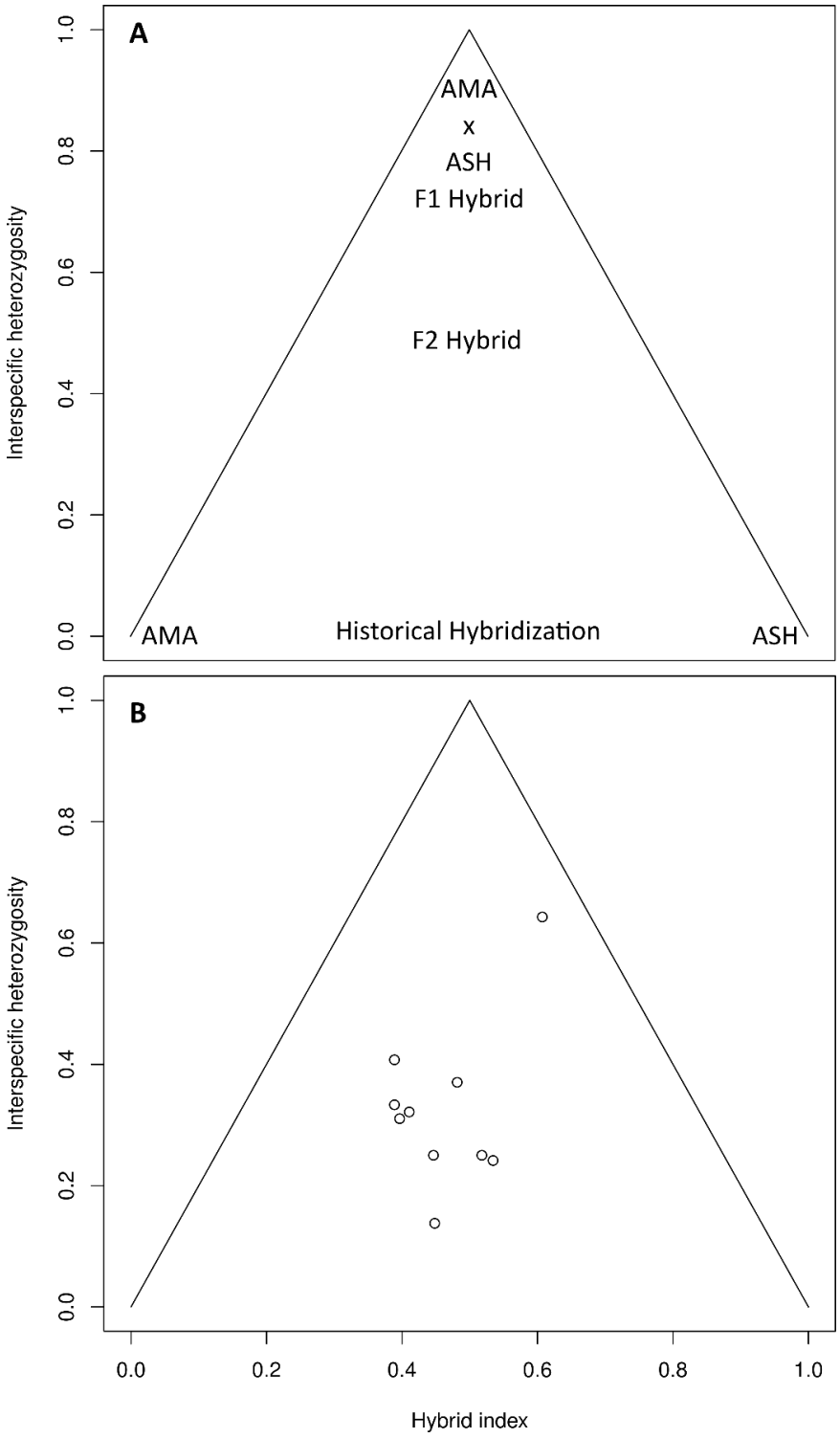

Figure 4. Hybrid index results for Amargosa Canyon Speckled Dace (*Rhinichthys osculus;* SPD). (A) Interpretation of hybrid index results. (B) Hybrid index calculated for ten Amargosa Canyon SPD samples. Parental genotypes are represented by Oasis Valley (AMA) and Ash Meadows (ASH).

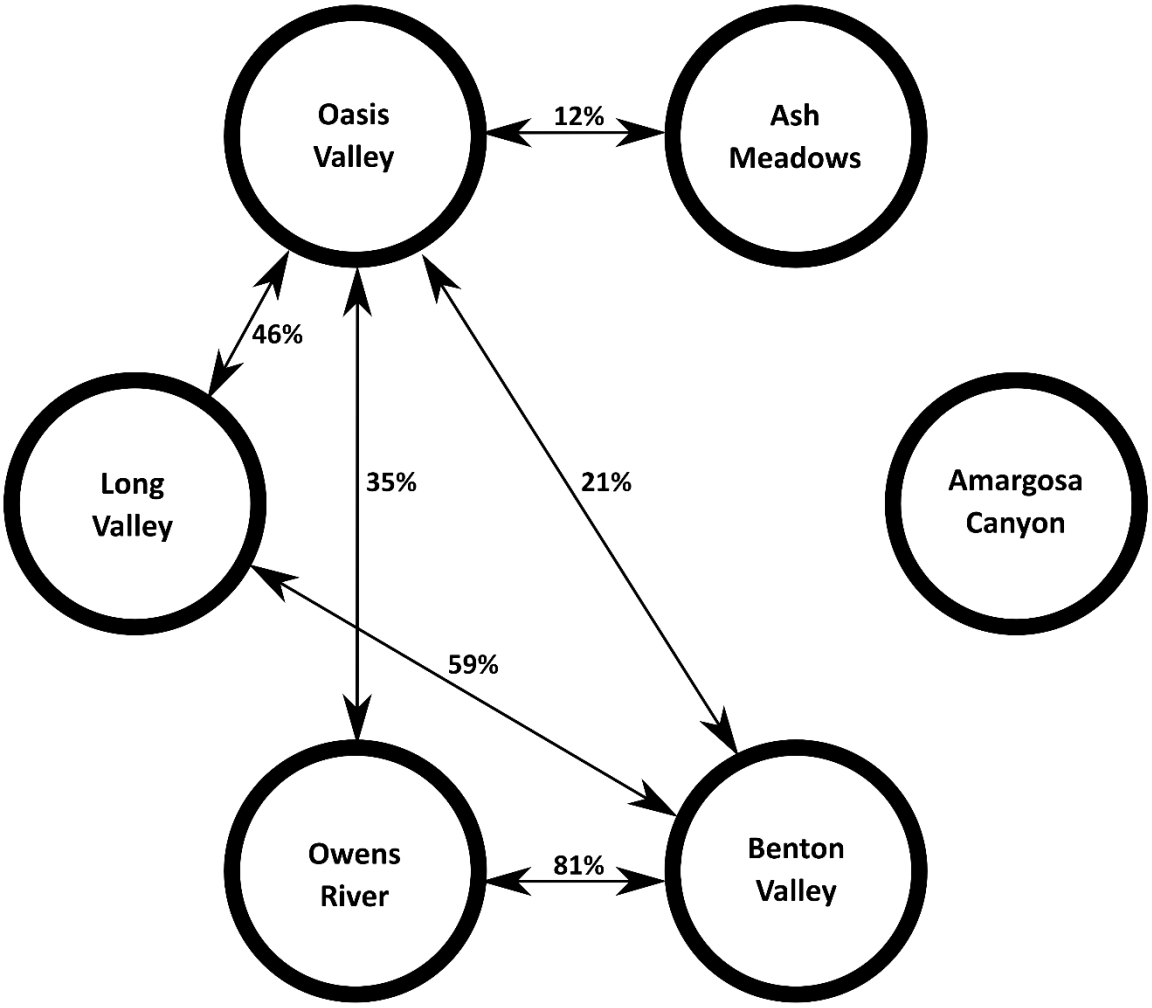

Figure 5. Patterns of introgression among Owens and Amargosa River basin Speckled Dace (*Rhinichthys osculus*). Lines connecting circles represent significant D-statistic values among lineages. Numbers represent the percent of tests found to be statistically significant ($p<0.0001$). Comparisons <10% significant are not shown.

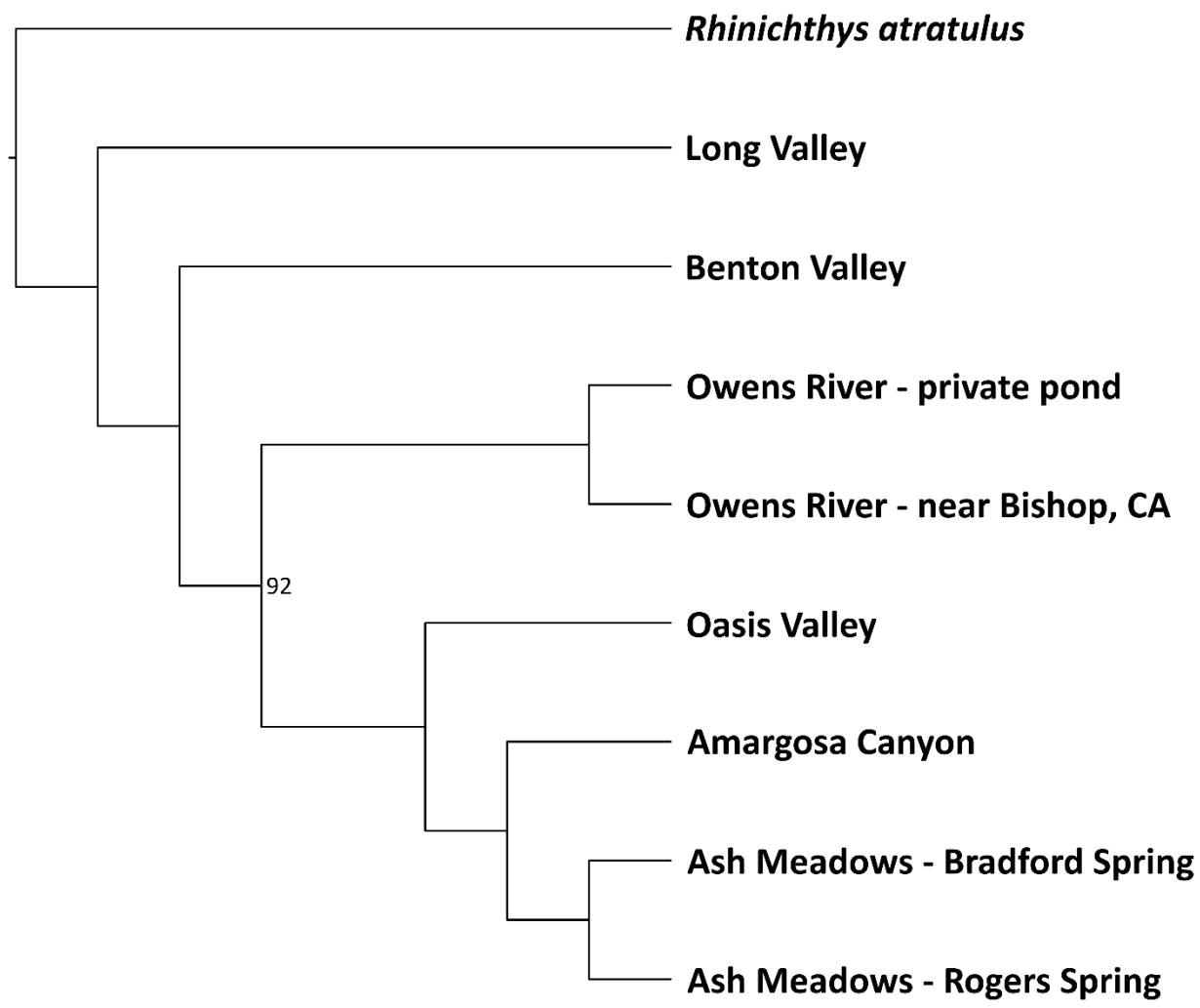

Figure 6. Phylogenetic tree of Owens and Amargosa basin Speckled Dace (*Rhinichthys osculus*) sampling localities. The tree was produced using polymorphism-aware models (POMO) in IQ-TREE. Bootstrap support values are displayed for those nodes with support <100.

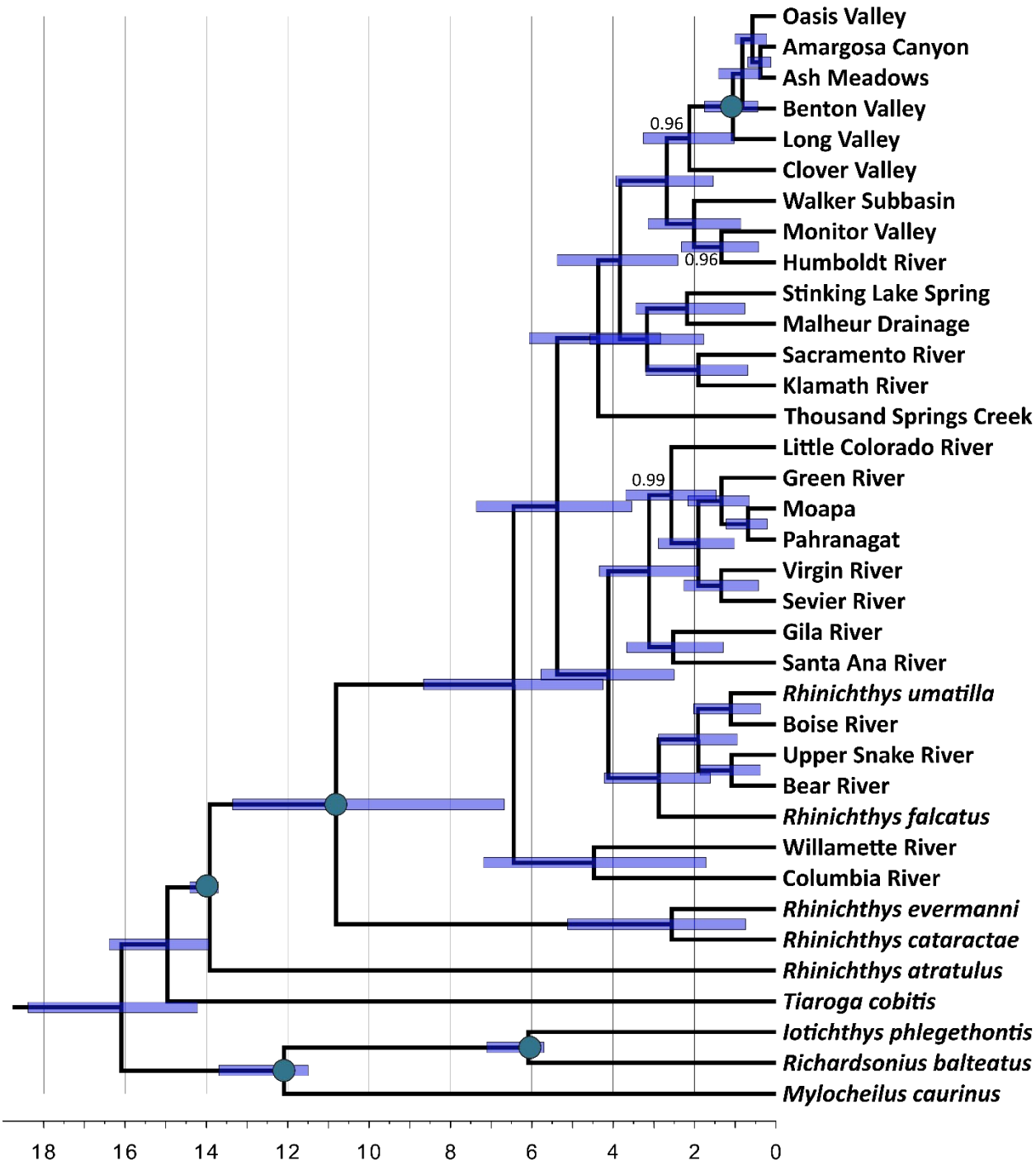

Figure 7. Time-calibrated tree calculated in BEAST 2.4.7 for populations of Speckled Dace (*Rhinichthys osculus*) and seven outgroups. Support values are Bayesian posterior probability (BPP). Only BPP values <1 are reported. The x-axis scale indicates node ages in millions of years. Blue bars represent 95% confidence intervals for divergence dates. Nodes receiving fossil calibration priors are indicated by green circles.

V. COMP-D: A Program for Comprehensive Computation of D-statistics and Population Summaries of Reticulated Evolution

Introduction

Historical introgression among closely related species (Hou et al. 2015; Zhang et al. 2016; Zheng & Janke 2018) can be successfully analyzed using Patterson's D-statistic (Green et al. 2010; Durand et al. 2011) and its enhancements (Eaton & Ree 2013; Pease & Hahn 2015). However, a variety of computational issues limit their accessibility and constrain applicability. For example, one program (pyRAD; Eaton 2014) requires that D-statistic calculations be tethered to a pipeline-specific format. In addition, other packages only compute the four-taxon test (ANGSD: Korneliussen et al. 2014; EvobiR: Blackmon & Adams 2015), while another implements Partitioned-D analyses in addition (ADMIXTOOLS: Patterson et al. 2012). A stand-alone program to calculate the various D-statistics (DFOIL: Pease & Hahn 2015) is available but offers limited options for assessing significance.

The evaluation of reticulated evolution would be greatly improved if a stand-alone, user-friendly program not only coalesced the multiple D-statistic programs but also compiled their results in an efficient manner. Importantly, the statistics in these programs verify the signature of introgression, its relative timing (Partitioned-D: Eaton et al. 2015), and its direction (D-foil: Árnason et al. 2018). Their independent analyses also help distinguish historical versus contemporary introgression, a necessity when anthropogenic drivers of introgression are suspected (Malukiewicz et al. 2015). These considerations are vital for developing conservation policies informed by the latest genomic techniques, and highlight their importance in the field of conservation genetics (Allendorf et al 2001; Bohling 2016).

We remedy these deficiencies by developing a software package (COMP-D) that not only calculates Patterson's D-statistic, but also its various modifications. The program is computationally efficient, employs file formats common to conservation genetic applications, and outputs files that are easily parsed either manually or by machine. Statistical options are also available for statistical evaluation of individual tests or populations as a whole.

Program Description

COMP-D recursively identifies biallelic loci for each unique combination of taxa within a whitespace-delimited file of taxon names. Single nucleotide polymorphism (SNP) data are input in PHYLIP- or STRUCTURE-format. Site patterns (i.e., ABBA, BABA, etc.) are calculated for each locus using a specified outgroup that represents the ancestral genotype. Three options are available for scoring site patterns at heterozygous loci: 1) random draw of an allele from a binomial distribution; 2) ignored if at least one individual is heterozygous; 3) as derived via SNP frequency formulas (Durand et al. 2011; Eaton et al. 2015). These provide an advantage over existing software (pyRAD) that ignores heterozygous loci when the four-taxon test is implemented.

Once site patterns are determined, D-statistics are calculated for one of three user-specified tests: The four-taxon test (Durand et al. 2011), Partitioned-D (Eaton & Ree 2013), or D-foil (Pease & Hahn 2015). Statistical confidence is assessed using two methods available in separate software packages. The first is equivalent to the Z-score produced by pyRAD (Eaton 2014). Here, nonparametric bootstrapping of loci is parallelized by Message Passing Interface (MPI) implementation to estimate the standard deviation of each test statistic, with comparison to a normal distribution $N(0,1)$. COMP-D also offers a $\chi 2$ goodness of fit test with one degree of

freedom as an alternative method to assess statistical significance (previously implemented in the DFOIL program; Pease & Hahn 2015).

A second improvement assesses introgression at the population level by aggregating results derived from multiple tests involving the same populations. COMP-D accomplishes this by calculating a Z-score derived from the mean and standard deviation of all individual tests performed as a single pass of the program. This follows Eaton (2014), but uses the D-statistics calculated from individual tests to derive a population mean and standard deviation, rather than relying upon the nonparametric bootstrap procedure as implemented for individual significance tests. A population-level quantification of introgression is calculated, and pseudoreplication is removed from the equation as this may generate standard deviation estimates lower than the true sample standard deviation (Efron 1981). This population-centric approach also reduces the risk of Type-I error associated with performing multiple comparisons among taxa, and also avoids a reduction in statistical power that can stem from Bonferroni procedures (Holm 1979, Perneger 1998, Rice 1989).

Methods

The D-statistic function of pyRAD offers the closest functionality to COMP-D and was thus selected to benchmark and validate our results. On the other hand, DFOIL (Pease & Hahn 2015) was not selected in that it is not multithreaded and does not employ a bootstrap to assess significance of individual tests.

Hybridization and introgression data for western North American catostomid fishes (Bangs et al. 2018) were utilized for verification and benchmarking. We evaluated 27 analyses encompassing 6,300 ABBA-BABA tests. Although Bangs et al. (2018) did not perform any five-

taxon tests, their data were used here to derive benchmarks for Partitioned-D and D-foil tests by evaluating introgression among *Catostomus discobolus* and *C. clarkii*. All tests were performed on a computer with dual Intel Xeon E5-4627 3.30GHz processors, 265GB RAM, and within a 64-bit Linux environment.

Results

Computational speed increased significantly in COMP-D (Figure 1), with four-taxon tests completed on average 3.93x more rapidly than in pyRAD (σ=1.82x). Substantial increases were also observed when fewer than 16 cpus were employed ($\overline{x}$=4.43x, σ=1.78x, max=9.78x). Performance gains for Partitioned-D were reflected by a 3.01x increase (σ=1.07x), improving average performance for 8 or fewer cpus by 3.42x (σ=0.90x). However, pyRAD-based D-foil calculations failed to complete within 48 runtime hours, indicating a potential software bug. For COMP-D, the longest-running D-foil calculations required approximately 4 hours (i.e., 750 tests/single cpu). Results from Partitioned-D and D-foil were compatible with the four-taxon tests of Bangs et al. (2018) for introgression among *C. discobolus* and *C. clarkii*, thus verifying the statistics accuracy of the calculations.

Although both programs produced similar four-taxon results, COMP-D required considerably less computational time. All four-taxon evaluations matched those in Bangs et al. (2018), but the numbers of significant tests were variable. This was influenced by four factors, with the first reflecting the options for assessing significance. The frequency-based χ2-test was a more conservative estimate, with fewer significant results than the Z-score (Table I). Random treatment of heterozygous sites revealed the number of significant χ2 results were less than or

equal to those detected via Z-score (24/27 result groups; 88.9%). Both inclusion and exclusion of heterozygous sites yielded similar trends (26/27; 96.3%).

The second source of variability was the manner by which heterozygous loci were processed. Ignoring them resulted in a 26.85% (σ=3.05%) decrease (Table II) in the number of usable loci on average. This highlights the risk of losing valuable information by excluding heterozygous sites (Zheng & Janke 2018), but also underscores a real biological phenomenon. Examining variable treatment of heterozygous sites through repeated analyses with different parameters in COMP-D can promote detection of recent hybridization. Ignoring heterozygous sites in an analysis of contemporary hybrids dramatically reduces the number of significant results (Table I: Lines 12-13). This alone does not demonstrate recent hybridization but indicates that additional analyses should be conducted for the evaluation of contemporary hybridity. This has emerged as an important consideration in other ABBA-BABA tests (Martin et al. 2015; Ottenburghs et al. 2017).

The final two factors that bear on the number of significant results are rounding error and the inherent randomness of bootstrapping. Bootstrapping is utilized to calculate the standard deviation for each test which becomes the denominator of the Z-score calculation. The standard deviation in pyRAD is rounded to three decimal places, whereas in COMP-D it is four. PyRAD thus inflates standard deviations somewhat and this, in turn, deflates Z-scores. Standard deviation values were compared for the three result groups with the greatest discrepancy in significant results between pyRAD and COMP-D (Table I: Lines 24-26). In 92.9% of the tests (884/952), the standard deviation calculated by pyRAD was greater than in COMP-D ($\overline{x}$ =0.027, σ=0.024). These subsequently pushed below that threshold many values that were at the cusp of significance (α=0.001).

Conclusion

The computationally more efficient and adaptable COMP-D can significantly improve studies of reticulated evolution, with results now extended to the population level as well. Standardized input formats are promoted whereas pipeline-specific file formats have been eliminated. The expansion of compatible input formats will facilitate analyses of historical introgression in conservation-related fields where non-model organisms often lack prior genomic information (i.e., RADseq studies). In addition, the reduction in computational time allows these algorithms to be implemented by researchers with limited (or no) access to high-performance computing resources.

References

Allendorf FW, Leary RF, Spruell P, Wenburg JK (2001) The problems with hybrids: setting conservation guidelines. Trends Ecol Evol 16:613–622. doi: 10.1016/S0169-5347(01)02290-X

Árnason Ú, Lammers F, Kumar V, et al (2018) Whole-genome sequencing of the blue whale and other rorquals finds signatures for introgressive gene flow. Sci Adv 4:. doi: 10.1126/sciadv.aap9873

Bangs MR, Douglas MR, Mussmann SM, Douglas ME (2018) Unraveling historical introgression and resolving phylogenetic discord within *Catostomus* (Osteichthys: Catostomidae). BMC Evol Biol 18:86. doi: 10.1186/s12862-018-1197-y

Blackmon H, Adams RA (2015) EvobiR: Tools for comparative analyses and teaching evolutionary biology

Bohling JH (2016) Strategies to address the conservation threats posed by hybridization and genetic introgression. Biol Conserv 203:321–327. doi: 10.1016/j.biocon.2016.10.011

Durand EY, Patterson N, Reich D, Slatkin M (2011) Testing for ancient admixture between closely related populations. Mol Biol Evol 28:2239–2252. doi: 10.1093/molbev/msr048

Eaton DA (2014) PyRAD: assembly of *de novo* RADseq loci for phylogenetic analyses. Bioinformatics 30:1844–1849. doi: 10.1093/bioinformatics/btu121

Eaton DA, Hipp AL, González-Rodríguez A, Cavender-Bares J (2015) Historical introgression among the American live oaks and the comparative nature of tests for introgression. Evolution 69:2587–2601. doi: 10.1111/evo.12758

Eaton DA, Ree RH (2013) Inferring phylogeny and introgression using RADseq data: an example from flowering plants (*Pedicularis*: Orobanchaceae). Syst Biol 62:689–706. doi: 10.1093/sysbio/syt032

Efron B (1981) Nonparametric estimates of standard error: The jackknife, the bootstrap and other methods. Biometrika 68:589–599. doi: 10.1093/biomet/68.3.589

Green RE, Krause J, Briggs AW, et al (2010) A draft sequence of the Neandertal genome. Science 328:. doi: 10.1126/science.1188021

Holm S (1979) A simple sequentially rejective multiple test procedure. Scand J Stat 6:65–70

Hou Y, Nowak MD, Mirré V, et al (2015) Thousands of RAD-seq loci fully resolve the phylogeny of the highly disjunct arctic-alpine genus *Diapensia* (Diapensiaceae). PLoS ONE 10:e0140175

Korneliussen TS, Albrechtsen A, Nielsen R (2014) ANGSD: Analysis of next generation sequencing data. BMC Bioinformatics 15:356. doi: 10.1186/s12859-014-0356-4

Malukiewicz J, Boere V, Fuzessy LF, et al (2015) Natural and anthropogenic hybridization in two species of eastern Brazilian marmosets (*Callithrix jacchus* and *C. penicillata*). PLoS ONE 10:e0127268. doi: 10.1371/journal.pone.0127268

Martin SH, Davey JW, Jiggins CD (2015) Evaluating the use of ABBA–BABA statistics to locate introgressed loci. Mol Biol Evol 32:244–257

Ottenburghs J, Megens H-J, Kraus RHS, et al (2017) A history of hybrids? Genomic patterns of introgression in the true geese. BMC Evol Biol 17:201. doi: 10.1186/s12862-017-1048-2

Patterson N, Moorjani P, Luo Y, et al (2012) Ancient admixture in human history. Genetics 192:1065–1093. doi: 10.1534/genetics.112.145037

Pease JB, Hahn MW (2015) Detection and polarization of introgression in a five-taxon phylogeny. Syst Biol 64:651–662. doi: 10.1093/sysbio/syv023

Perneger TV (1998) What's wrong with Bonferroni adjustments. Br Med J 316:1236. doi: 10.1136/bmj.316.7139.1236

Rice WR (1989) Analyzing tables of statistical tests. Evolution 43:223–225. doi: 10.2307/2409177

Zhang W, Dasmahapatra KK, Mallet J, et al (2016) Genome-wide introgression among distantly related *Heliconius* butterfly species. Genome Biol 17:1–15. doi: 10.1186/s13059-016-0889-0

Zheng Y, Janke A (2018) Gene flow analysis method, the D-statistic, is robust in a wide parameter space. BMC Bioinformatics 19:10. doi: 10.1186/s12859-017-2002-4

Appendix

Table 1. Results of four-taxon D-statistic tests comparing methods for handling heterozygous loci in COMP-D with results obtained from pyRAD. Each column shows the number of statistically significant tests (α=0.001) for each method of assessing significance in each treatment. COMP-D offers two methods of assessing statistical significance (Z-scores and Chi-square tests) whereas pyRAD offers only Z-scores. Two treatments (Random and HetInclude) considered all heterozygous loci in D-statistic calculations, but differed by either randomly picking an allele to represent an individual (Random) or using SNP frequency calculations (HetInclude). The HetIgnore method removed all heterozygous loci from calculations. All tests were performed upon catostomid fishes of the western United States. Abbreviations for taxon names (P1, P2, P3, and O columns) are as follows: BBS = Bonneville Bluehead Sucker, BLS = Bridgelip Sucker, FMS = Flannelmouth Sucker, LNS = Longnose Sucker, MTS = Mountain Sucker, RBS = Razorback Sucker, SOS = Sonora Sucker, THS = Tahoe Sucker, WTS = White Sucker. Abbreviations in parentheses next to species abbreviations represent different populations. BB = Bonneville Basin, CB = Columbia River Basin, GC = Grand Canyon of the Colorado River, LB = Lahontan Basin, LC = Little Colorado River, UC = Upper Colorado River Basin, VR = Virgin River, wen = Wenima Wildlife Area of the Little Colorado River.

				Random		HetInclude		HetExclude			
P1	P2	P3	O	Z	χ^2	Z	χ^2	Z	χ^2	pyRAD	Total tests
THS	BLS	BBS	LNS	4	0	5	0	0	0	0	90
THS	BLS	MTS(CB)	LNS	0	0	0	0	1	0	1	36
THS	BLS	MTS(LB)	LNS	0	0	0	0	4	0	6	90
THS	BLS	MTS(BB)	LNS	1	0	1	0	4	0	4	108
FMS(UC)	FMS(GC)	SOS	WTS	28	27	18	7	17	5	5	360
FMS(UC)	FMS(VR)	SOS	WTS	419	419	420	420	419	418	405	420
FMS(GC)	FMS(VR)	SOS	WTS	420	420	420	420	419	418	417	420
FMS(LC)	FMS(VR)	SOS	WTS	341	342	346	340	348	346	337	350
FMS(wen)	FMS(VR)	SOS	WTS	12	11	13	11	63	51	52	210
FMS(LC)	FMS(UC)	SOS	WTS	0	0	0	0	0	0	0	300
FMS(LC)	FMS(GC)	SOS	WTS	0	0	0	0	0	0	1	300
FMS(UC)	FMS(wen)	SOS	WTS	169	167	174	173	117	117	96	180
FMS(GC)	FMS(wen)	SOS	WTS	171	170	173	165	119	115	100	180
FMS(UC)	SOS	RBS	WTS	239	238	239	239	240	240	238	240
FMS(GC)	SOS	RBS	WTS	238	237	239	239	234	233	232	240
FMS(VR)	SOS	RBS	WTS	239	236	255	247	279	278	277	280
FMS(LC)	SOS	RBS	WTS	175	168	174	171	185	186	178	200
FMS(wen)	SOS	RBS	WTS	93	89	97	93	112	106	100	120
FMS(GC)	FMS(UC)	RBS	WTS	1	3	0	0	5	0	20	288
FMS(UC)	FMS(LC)	RBS	WTS	1	0	0	0	4	0	0	240
FMS(GC)	FMS(LC)	RBS	WTS	0	0	0	0	6	1	2	240
FMS(UC)	FMS(wen)	RBS	WTS	9	7	7	1	7	3	1	144
FMS(GC)	FMS(wen)	RBS	WTS	11	10	7	1	8	3	2	144
FMS(UC)	FMS(VR)	RBS	WTS	173	160	126	153	227	192	155	336
FMS(GC)	FMS(VR)	RBS	WTS	176	165	164	132	193	150	128	336
FMS(LC)	FMS(VR)	RBS	WTS	61	58	76	58	132	107	100	280
FMS(wen)	FMS(VR)	RBS	WTS	0	1	0	0	19	9	8	168

Table 2. The number of biallelic loci recovered when allowing for heterozygous loci (Het. Included) and considering fixed loci only (Het. Excluded). Both the mean number of loci (Avg. Loci) and standard deviation (StDev) are presented for each treatment. The % decrease indicates the percentage of loci lost by considering only fixed differences among taxa. All tests were performed upon catostomid fishes of the western United States. Abbreviations for taxon names (P1, P2, P3, and O columns) are as follows: BBS = Bonneville Bluehead Sucker, BLS = Bridgelip Sucker, FMS = Flannelmouth Sucker, LNS = Longnose Sucker, MTS = Mountain Sucker, RBS = Razorback Sucker, SOS = Sonora Sucker, THS = Tahoe Sucker, WTS = White Sucker. Abbreviations in parentheses next to species abbreviations represent different populations. BB = Bonneville Basin, CB = Columbia River Basin, GC = Grand Canyon of the Colorado River, LB = Lahontan Basin, LC = Little Colorado River, UC = Upper Colorado River Basin, VR = Virgin River, wen = Wenima Wildlife Area of the Little Colorado River.

				Het. Included		Het. Excluded		
P1	**P2**	**P3**	**O**	**Avg. Loci**	**StDev Loci**	**Avg. Loci**	**StDev Loci**	**% Decrease**
THS	BLS	BBS	LNS	4006.40	4131.82	3203.97	3349.67	20.03%
THS	BLS	MTS(CB)	LNS	3069.14	2892.97	2450.78	2357.50	20.15%
THS	BLS	MTS(LB)	LNS	4095.22	3956.30	3286.28	3207.14	19.75%
THS	BLS	MTS(BB)	LNS	4309.40	4169.65	3421.12	3368.17	20.61%
FMS(UC)	FMS(GC)	SOS	WTS	8944.57	1369.22	6417.83	944.23	28.25%
FMS(UC)	FMS(VR)	SOS	WTS	9384.35	859.25	6759.10	605.67	27.97%
FMS(GC)	FMS(VR)	SOS	WTS	9384.35	859.25	6759.10	605.67	27.97%
FMS(LC)	FMS(VR)	SOS	WTS	6603.40	1955.18	4886.15	1458.37	26.01%
FMS(wen)	FMS(VR)	SOS	WTS	6215.60	1523.61	4434.45	1057.05	28.66%
FMS(LC)	FMS(UC)	SOS	WTS	6486.53	1899.93	4760.41	1408.06	26.61%
FMS(LC)	FMS(GC)	SOS	WTS	6486.53	1899.93	4760.41	1408.06	26.61%
FMS(UC)	FMS(wen)	SOS	WTS	6075.96	1473.55	4292.14	1014.08	29.36%
FMS(GC)	FMS(wen)	SOS	WTS	5911.59	1600.66	4178.88	1098.16	29.31%
FMS(UC)	SOS	RBS	WTS	8657.01	2376.47	6304.40	1696.69	27.18%
FMS(GC)	SOS	RBS	WTS	8657.01	2376.47	6141.78	1821.63	29.05%
FMS(VR)	SOS	RBS	WTS	8777.53	2422.55	6433.48	1750.48	26.71%
FMS(LC)	SOS	RBS	WTS	6117.02	2315.09	4555.35	1731.03	25.53%
FMS(wen)	SOS	RBS	WTS	5788.27	1926.93	4170.90	1363.35	27.94%
FMS(GC)	FMS(UC)	RBS	WTS	7701.15	4150.56	5478.28	2925.39	28.86%
FMS(UC)	FMS(LC)	RBS	WTS	6431.23	2471.50	4661.05	1778.71	27.52%
FMS(GC)	FMS(LC)	RBS	WTS	6431.23	2471.50	4528.60	1806.82	29.58%
FMS(UC)	FMS(wen)	RBS	WTS	6048.72	2046.89	4276.94	1406.54	29.29%
FMS(GC)	FMS(wen)	RBS	WTS	6048.72	2046.89	4276.94	1406.54	29.29%
FMS(UC)	FMS(VR)	RBS	WTS	9413.68	2572.11	6710.90	1782.78	28.71%
FMS(GC)	FMS(VR)	RBS	WTS	9156.07	2776.91	6531.51	1933.88	28.66%
FMS(LC)	FMS(VR)	RBS	WTS	6540.95	2529.87	4792.64	1842.15	26.73%
FMS(wen)	FMS(VR)	RBS	WTS	6179.17	2109.21	4417.12	1470.13	28.52%

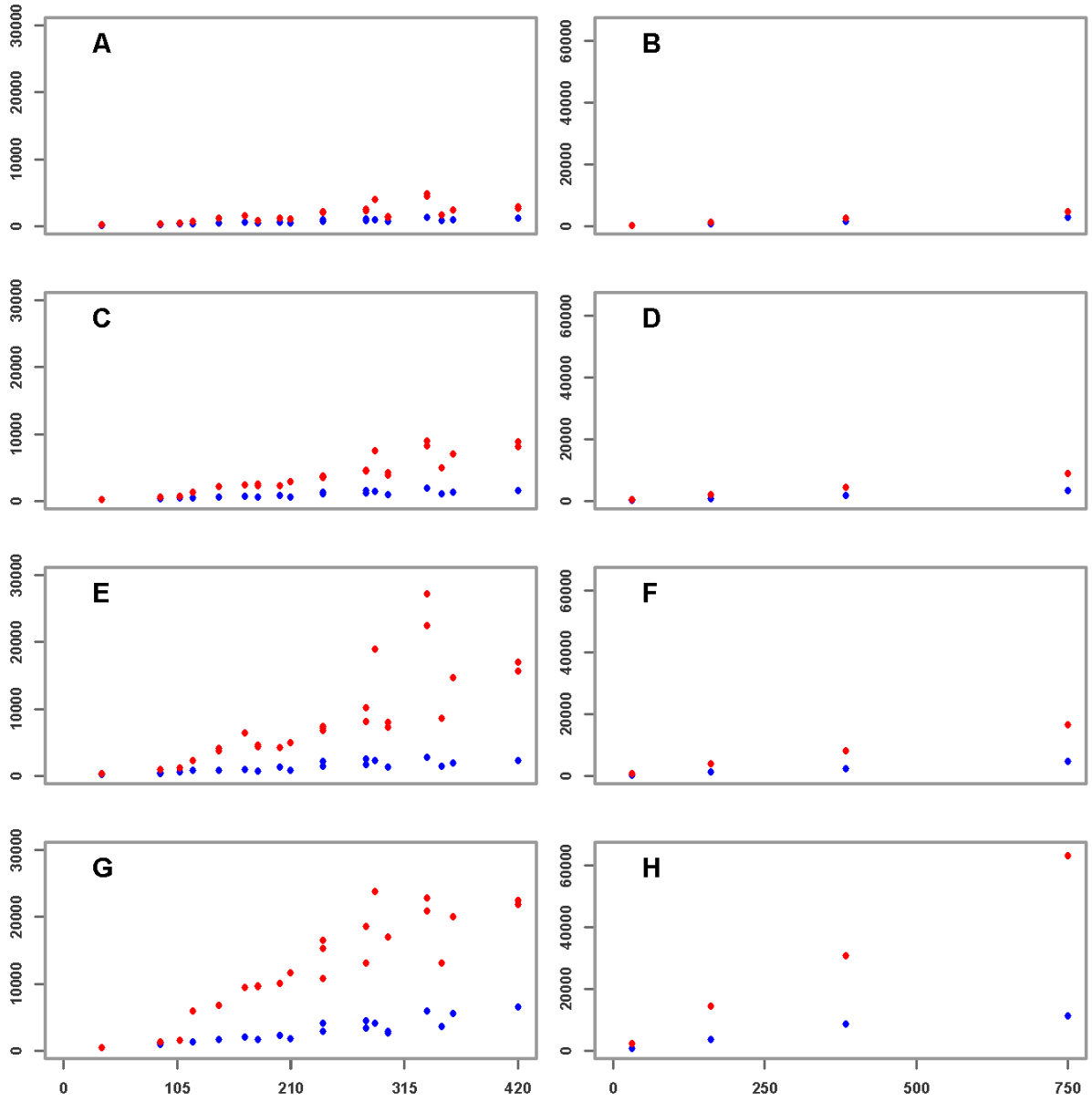

Figure 1. The relative performance of COMP-D (blue points) compared to D-statistic calculations in pyRAD (red points). X axes represent the number of tests performed in one instance of each program. Y-axes represent the number of seconds required to run each instance. All plots in the left column (A, C, E, and G) represent four-taxon test calculations. Plots in the right column (B, D, F, and H) represent Partitioned-D calculations. D-foil calculations could not be compared because the equivalent calculations in pyRAD never completed. Plots in the first row (A and B) represent tests performed with 16 processor cores (cpus). Tests for plots C and D required 8 cpus. E and F utilized 4 cpus, while G and H were run on a single cpu each. Figure was generated using R version 3.4.4 (R Core Team 2018).

VI. BA3-SNPs: Contemporary Migration Reconfigured in BayesAss for Next-Generation Sequence Data

Introduction

Accelerated loss of biodiversity is a fundamental issue in the Anthropocene (Lewis and Maslin 2015), with direct and potentially irreversible consequences for the well-being of humankind. To slow this erosion, scientists strive to ascertain at which level biodiversity should be conserved (Douglas and Brunner 2002). Yet these considerations are complicated both philosophically and methodologically as they reflect the underlying tension between taxonomy, conservation, and the immediacy of adaptive management (Frankham 2010).

This dilemma is further compounded by the spatial and temporal scales within which these disciplines are framed. The taxonomy of species, for example, is viewed as an historical process operating over geologic time, whereas conservation deals with environmental modifications of a contemporary nature that impact local populations. This gap is effectively bridged by conservation genetics (Allendorf et al. 2010), a multi-disciplinary science that employs molecular data to delineate biodiversity. Such data are often used to distinguish populations and gauge levels of differentiation at very fine spatial scales (e.g., Hopken et al. 2013; Mussmann et al. 2017). They can also provide a deep historical perspective by deriving genetic patterns that effectively differentiate species (e.g., Douglas et al. 2006, 2009). Both approaches effectively promote biodiversity management by facilitating the integration of conservation policies at the regional level (Willis and Birks 2006).

Recent advances in conservation genomic approaches (Allendorf 2016) allow for the development of quantitative metrics that can define ‘conservation units’ (Holycross and Douglas

2007; Sullivan et al. 2014), and identify adaptively divergent gene pools in non-model organisms (Funk et al. 2012). A common approach is the application of RAD (restriction-site associated DNA) sequencing, where targeted fragmentation using restriction enzymes to cut genomic DNA at predictable sequence motifs effectively reduces genomic complexity. This ensures that data remain consistent across individuals, a consideration particularly applicable for organisms that lack an *a priori* baseline regarding genome composition (Andrews and Luikart 2014; Puritz et al. 2014). Methodological iterations of the original RAD protocol (Baird et al. 2008) have been remarkably prolific in generating multiple derivatives, with applicability often extended towards particular research foci (e.g. Peterson et al. 2012; Ali et al. 2016).

However, the analysis of genome-scale data presents new computational challenges when placed within an hypotheses-testing framework. Furthermore, this consideration is often magnified by the limited bioinformatics skills found among conservation practitioners. These issues necessitate the development of user- and computationally-friendly algorithms that are flexible enough to parse next generation sequence data. This issue is a focal point of this study and serves to recognize that molecular methods often outpace analytical methods. As a result, programs once at the forefront of analysis may fall out of favor when data generated by newer molecular methods are in such amounts that more legacy analytical programs cannot parse.

One such example is the program BAYESASS (Wilson and Rannala 2003), designed as a method to diagnose demographic independence within and among populations (Palumbi 2003), a critical step in establishing conservation units. Given this, its use promoted the quantification of 'management units' (MUs - those reflecting <10% immigration; Palsbøll et al. 2006). A practical example is the delineation of hatchery broodstocks as being genetically equivalent to populations

within a species, a necessary endeavor to avoid outbreeding depression or reduced fitness due to the mixing locally adapted lines (Frankham et al. 2011).

Here, we provide a modification to this program that accepts modern SNP datasets as input, such as those produced via RADseq methods. We also implement a new tool for this approach that will streamline its applicability with multiple datasets.

Program Description

The source code for BAYESASS version 3.0.4 was downloaded and modified to provide several enhancements that facilitate the analysis of SNP datasets containing thousands of loci. This was accomplished by removing the hard-coded upper limit that constrained the amount of memory allocated for locus storage (N=420). Instead, we implemented dynamic memory allocation to promote the number of loci processed by BA3-SNPS, with restrictions manifested only by hardware and operating system limitations. Additional changes (e.g., an object-oriented programming approach, and streamlining of command line input parsing) are not immediately apparent but serve to promote future extension of the code by other researchers.

An additional tool (BA3-SNPS-AUTOTUNE) was also developed to accommodate larger datasets. Similar to its parent program, BA3-SNPS requires the adjustment of mixing parameters for migration rates, allele frequencies, and inbreeding coefficients. To accomplish this, a binary search algorithm was implemented in Python that conducts short exploratory runs, with the acceptance rates for each recorded, and parameters updated in subsequent runs until optimal values are achieved. Preferred targets for acceptance rates lay between 0.35 and 0.45, with those between 0.2 and 0.6 considered acceptable (Wilson and Rannala 2003). The algorithm

significantly accelerates the process of tuning the mixing parameters for each unique input by automating this procedure and removing the previously required "guess-and-check" process.

Methods

Performance benchmarking was conducted for the above programs by analyzing 618 files containing empirical data produced via ddRAD sequencing methods (Peterson et al. 2012). Each file contained from 21 to 43 samples ($\bar{x}$ = 30.5, σ = 3.1) genotyped at 632 to 12,408 SNPs ($\bar{x}$ = 8,366.8, σ = 1,918.04). Forty-million Markov Chain Monte Carlo (MCMC) generations were calculated for each run. The time needed to analyze each file was recorded. Linear models were evaluated to explain the runtime as a function of the number of samples, the number of loci, and the number of samples and loci (R Development Core Team 2018).

BA3-SNPS-AUTOTUNE was used to find optimal mixing parameters for each run, with exploratory analyses employing 10,000 MCMC generations in length. A maximum of 10 exploratory analyses were conducted for each data file. The number of repetitions required to find optimal mixing parameters was recorded for each, and mixing parameters verified to produce adequate MCMC acceptance rates (i.e., 0.2 < acceptance rate < 0.6). All tests were performed on a computer equipped with dual Intel Xeon E5-4627 3.30GHz processors, 265GB RAM, and with a 64-bit Linux environment. Since neither program is multithreaded, a single processor core was used per analysis.

Results

BA3-SNPS completed the analysis of all files in 7.7 to 374.03 hours ($\bar{x}$ = 124.44, σ = 47.39). The computational speed was evaluated relative to the number of individuals per data file

(Figure 1A), and the number of loci (Figure 1B). A weak yet significant correlation was found between sample size and run time ($R^2 = 0.015$, $p = 0.00134$). However, it appears the number of loci being evaluated by the program is a more accurate predictor of runtime ($R^2 = 0.307$, $p < 2.2 \times 10^{-16}$). Modest gains in model fit were achieved by couching runtime as a function of both samples and loci produced ($R^2 = 0.3346$, $p < 2.2 \times 10^{-16}$).

To find optimal mixing parameters, BA3-SNPS-AUTOTUNE required between three and ten rounds of optimization per data file ($\bar{x} = 4.15$, $\sigma = 0.72$) (Figure 2). The four data files that required 10 rounds of optimization were deemed statistical outliers. BA3-SNPS runs associated with these four files ultimately had acceptance rates for each parameter that fit within the recommended values specified in the original BAYESASS user manual, despite requiring ten rounds of optimization.

Conclusion

The enhancements to BAYESASS allowing for the analysis of contemporary SNP datasets provide a necessary upgrade to a valuable algorithm. Validation using empirical data indicated the suitability of the updated version for either high-performance computing clusters or modern desktop computers. The 40 million generations used in this study was deliberately excessive so as to push the limits of computing performance. Peak memory usage for the largest data file was measured at just under 3.5 GB, indicating that the program will be remain useful for those researchers that lack access to a computer cluster.

Furthermore, BA3-SNPS-AUTOTUNE successfully streamlined the tedious process of determining appropriate mixing parameters. We also verified that these parameters could be identified using low numbers of MCMC generations per exploratory run. The program rarely

required more than 5 rounds of optimization to obtain suitable parameters. This will greatly aid researchers who must run BA3-SNPS on multiple unique files. These programs not only modernized migration analyses to accommodate next-generation sequence data, but also increased the accessibility of these methods to a broad range of conservation geneticists.

References

Ali OA, O'Rourke SM, Amish SJ, et al (2016) RAD capture (rapture): flexible and efficient sequence-based genotyping. Genetics 202:389. doi: 10.1534/genetics.115.183665

Allendorf FW (2016) Genetics and the conservation of natural populations: allozymes to genomes. Mol Ecol 26:420–430. doi: 10.1111/mec.13948

Allendorf FW, Hohenlohe PA, Luikart G (2010) Genomics and the future of conservation genetics. Nat Rev Genet 11:697

Andrews KR, Luikart G (2014) Recent novel approaches for population genomics data analysis. Mol Ecol 23:1661–1667

Baird NA, Etter PD, Atwood TS, et al (2008) Rapid SNP discovery and genetic mapping using sequenced RAD markers. PLoS ONE 3:1–7. doi: 10.1371/journal.pone.0003376

Douglas ME, Douglas MR, Schuett GW, Porras LW (2006) Evolution of rattlesnakes (Viperidae; *Crotalus*) in the warm deserts of western North America shaped by Neogene vicariance and Quaternary climate change. Mol Ecol 15:3353–3374. doi: 10.1111/j.1365-294X.2006.03007.x

Douglas ME, Douglas MR, Schuett GW, Porras LW (2009) Climate change and evolution of the New World pitviper genus *Agkistrodon* (Viperidae). J Biogeogr 36:1164–1180. doi: 10.1111/j.1365-2699.2008.02075.x

Douglas MR, Brunner PC (2002) Biodiversity of central alpine *Coregonus* (Salmoniformes): Impact of one-hundred years of management. Ecol Appl 12:154–172. doi: 10.2307/3061143

Frankham R (2010) Challenges and opportunities of genetic approaches to biological conservation. Biol Conserv 143:1919–1927. doi: 10.1016/j.biocon.2010.05.011

Frankham R, Ballou JD, Eldridge MDB, et al (2011) Predicting the probability of outbreeding depression. Conserv Biol 25:465–475. doi: 10.1111/j.1523-1739.2011.01662.x

Funk WC, McKay JK, Hohenlohe PA, Allendorf FW (2012) Harnessing genomics for delineating conservation units. Trends Ecol Evol 27:489–496. doi: 10.1016/j.tree.2012.05.012

Holycross AT, Douglas ME (2007) Geographic isolation, genetic divergence, and ecological non-exchangeability define ESUs in a threatened sky-island rattlesnake. Biol Conserv 134:142–154. doi: 10.1016/j.biocon.2006.07.020

Hopken MW, Douglas MR, Douglas ME (2013) Stream hierarchy defines riverscape genetics of a North American desert fish. Mol Ecol 22:956–971. doi: 10.1111/mec.12156

Lewis SL, Maslin MA (2015) Defining the Anthropocene. Nature 519:171

Mussmann SM, Douglas MR, Anthonysamy WJB, et al (2017) Genetic rescue, the greater prairie chicken and the problem of conservation reliance in the Anthropocene. R Soc Open Sci 4:. doi: 10.1098/rsos.160736

Palsbøll PJ, Bérubé M, Allendorf FW (2006) Identification of management units using population genetic data. Trends Ecol Evol 22:11–16. doi: 10.1016/j.tree.2006.09.003

Palumbi SR (2003) Population genetics, demographic connectivity, and the design of marine reserves. Ecol Appl 13:146–158. doi: 10.1890/1051-0761(2003)013[0146:PGDCAT]2.0.CO;2

Peterson BK, Weber JN, Kay EH, et al (2012) Double digest RADseq: an inexpensive method for *de novo* SNP discovery and genotyping in model and non-model species. PLoS ONE 7:1–11. doi: 10.1371/journal.pone.0037135

Puritz JB, Matz MV, Toonen RJ, et al (2014) Demystifying the RAD fad. Mol Ecol 23:5937–5942

Sullivan BK, Douglas MR, Walker JM, et al (2014) Conservation and management of polytypic species: The little striped whiptail complex (*Aspidoscelis inornata*) as a case study. Copeia 2014:519–529. doi: 10.1643/CG-13-140

Willis KJ, Birks HJB (2006) What Is natural? The need for a long-term perspective in biodiversity conservation. Science 314:1261. doi: 10.1126/science.1122667

Wilson GA, Rannala B (2003) Bayesian inference of recent migration rates using multilocus genotypes. Genetics 163:1177–1191

Appendix

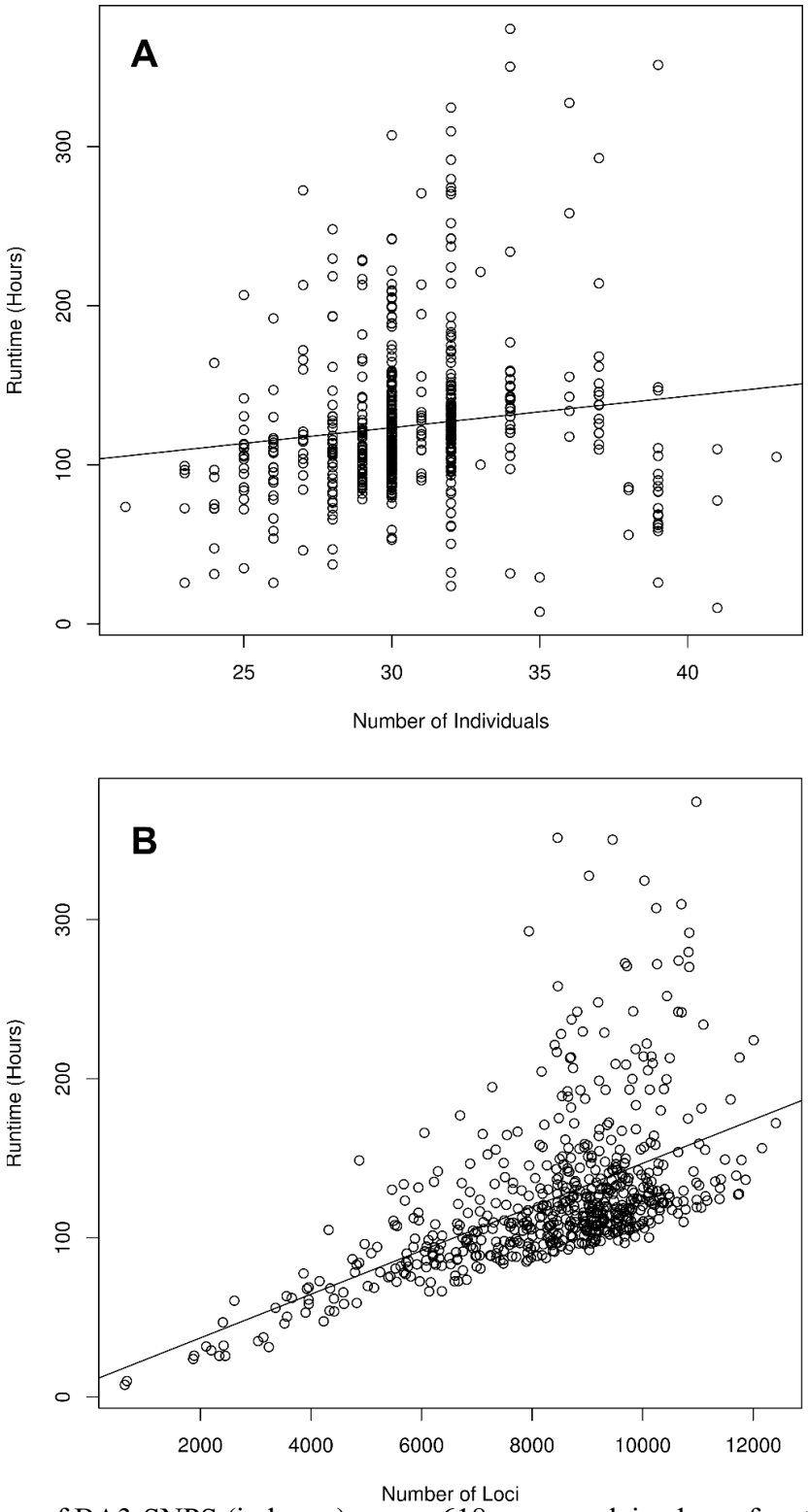

Figure 1. Runtime of BA3-SNPS (in hours) across 618 runs, explained as a function of (A) the number of individuals and (B) the number of loci. Note that "Number of Individuals" does not scale as predictably as "Number of Loci" - likely a consequence of runtime by variation among the groups compared, rather than the number of individuals within those populations.

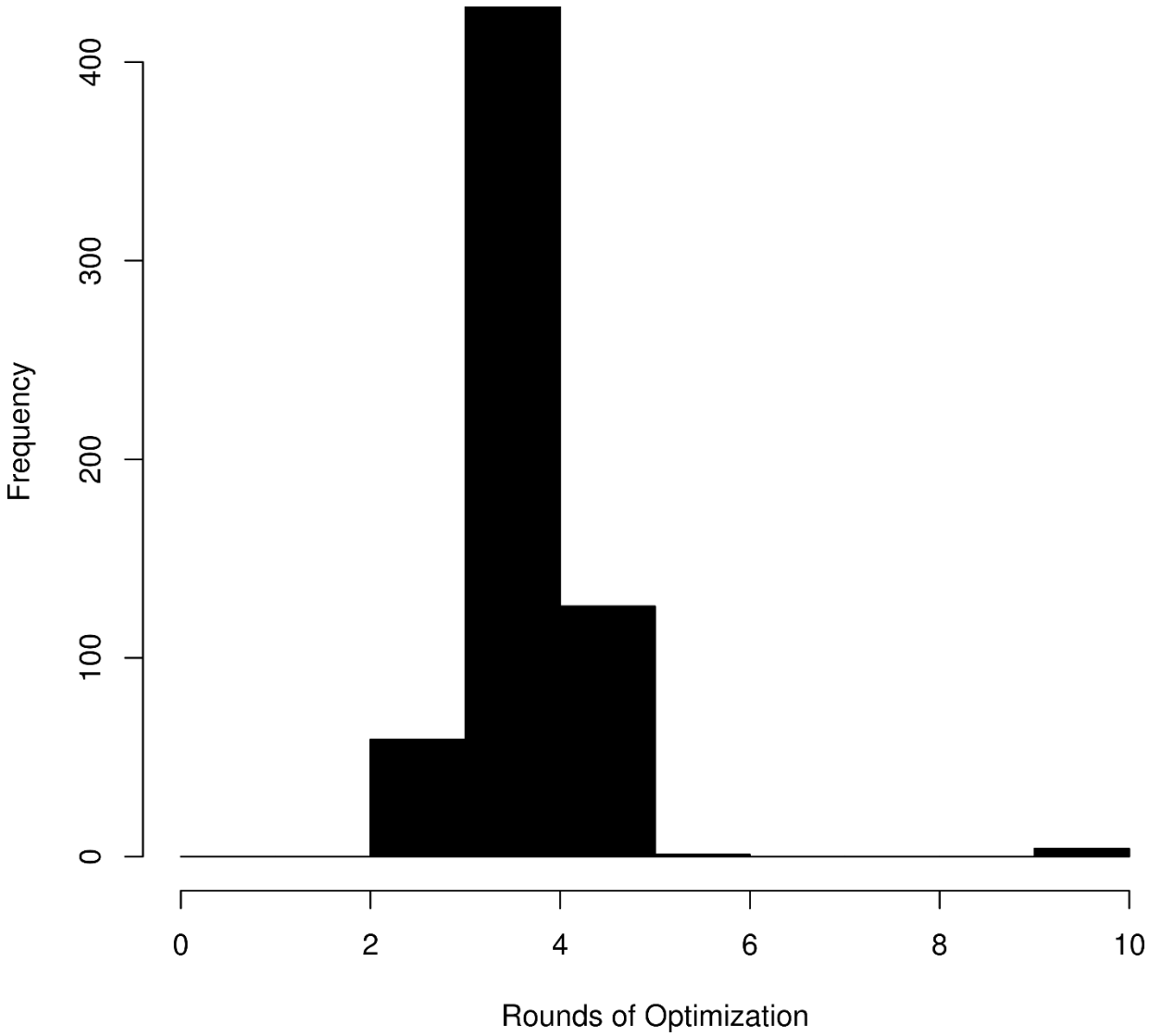

Figure 2. A histogram depicting the number of rounds of optimization required by BA3-SNPS-AUTOTUNE to find optimal mixing parameters for BA3-SNPS. The program was tested on 618 data files, and all but five completed optimizations in five rounds or less. All produced optimal mixing parameters that allowed MCMC rates to fall within acceptable thresholds for BA3-SNPS.

VII. Conclusion

Changes in the landscape of western North America since the Miocene have had powerful impacts on the inter- and intraspecific diversity of western fish fauna (Minckley et al. 1986). Natural forces such as plate tectonics, volcanism, and climate change have profoundly altered the landscape and impacted the evolution of the organisms that inhabit it (Spencer et al. 2008; Chamberlain et al. 2012). In this dissertation, mechanisms generating this diversity were evaluated on both geological and contemporary time scales. The geospatial distribution of diversity was also examined within and among drainage basins. These mechanisms are challenging to study at these scales due to the depauperate nature of western aquatic fauna, and the endangered status shared by many western fishes. In this sense, *Rhinichthys* has served as a proxy to identify relationships among drainage basins, identify areas in which localized diversity is realized, and provide another data point supporting trends that concern endangered aquatic species of the west (Gerber et al. 2001; Douglas et al. 2003; Borley and White 2006).

Chapter II elucidated the important role played by introgression and hybridization in *Rhinichthys* evolution, substantiated the monophyly of most *R. osculus* subspecies, and revealed that *Rhinichthys* is in dire need of taxonomic revision. Both historic events (e.g., the Bonneville Flood), and modern occurrences (e.g., hybrid *R. cataractae* x *R. osculus*) highlight the importance of this frequently undervalued evolutionary mechanism throughout the species' range. The propensity of *Rhinichthys* to mix with conspecifics or congeners upon secondary contact, coupled with its status as a headwater specialist, is a major component of its evolutionary success that promotes rapid dispersal and adaptation to unique environments (Ellstrand and Schierenbeck 2000; Hovick and Whitney 2014; Mesgaran et al. 2016).

Chapter III demonstrated the importance of dynamic riverine forces that act upon fishes, including homogenizing vs asymmetrical gene flow, and isolation by vicariance. Although *Rhinichthys* as a whole is an atypical target of conservation, this study verified the uniqueness of its many narrowly endemic subspecies, and elucidated trends that have conservation implications for endangered aquatic fauna of the region. Its conservation value is magnified when viewed as a surrogate for stream fishes of similar size, habitat preference, or life history that also contend with anthropogenic modifications (Caro 2010). Many endangered western fishes share at least one of these traits with *Rhinichthys*, meaning conservation issues relevant to *Rhinichthys* will similarly impact other species. The ubiquity of western *Rhinichthys* promotes a variety of comparisons with threatened and endangered species of limited distribution, as demonstrated for the Upper Colorado River Basin (Douglas and Douglas 2007; Hopken et al. 2013).

Chapter IV evaluated divergence dates of six distinct *Rhinichthys* lineages within the Owens and Amargosa drainages of eastern California/ western Nevada. *Rhinichthys* again demonstrated its value as a proxy for other native western fishes, while simultaneously validating the genetic distinctness of its own endangered, narrowly endemic forms. The molecular clock employed for this purpose revealed that divergence dates for Speckled Dace mostly conform to documented Pleistocene interbasin hydrological connections. This calls contemporary estimates for *Cyprinodon* species into question (Knott et al. 2018; Martin and Turner 2018), especially if those patterns of speciation also conform to known historical hydrology. The *Rhinichthys* lineages in this region are narrowly endemic relicts of a Pleistocene ecosystem that now persists in small desert oases. Results demonstrate that federal protections would greatly advance conservation actions to preserve these subspecies.

Chapters V and VI provide new and updated computational programs, respectively, for analyzing genome-scale SNP data. COMP-D (Chapter V) is a C++ program that takes advantage of high-performance computing resources to calculate Patterson's D-statistic (Durand et al. 2011) and its derivatives (Eaton and Ree 2013; Pease and Hahn 2015). This is an important asset for a changing world in which distinguishing among natural and human-mediated introgression is important for interpreting hybridization in the context of conservation laws (Allendorf et al. 2001; Fitzpatrick et al. 2015; vonHoldt et al. 2017). Chapter VI describes BA3-SNPS: an update to BAYESASS 3 (Wilson and Rannala 2003) that assesses contemporary migration using SNP data. This update breathes new life into an important algorithm that had previously been unusable for genome-scale datasets.

These five studies taken together use *Rhinichthys* as a proxy to study evolutionary mechanisms in a dynamic landscape, and provide new analytical tools for analyzing data in a conservation genetic framework. These resources will be important in the context of global change. Furthermore, they provide a step towards a better understanding of adaptation to the western landscape. Conservation genomic studies must make the investigation of genomic adaptation to changing conditions a primary focus (Cure et al. 2017). Future studies should evaluate the geographic regions where anthropogenic activities are expected to admix populations, quantify genomic rearrangements that occur in response to changing environments, and provide a bridge between these phenomena.

References

Allendorf FW, Leary RF, Spruell P, Wenburg JK (2001) The problems with hybrids: setting conservation guidelines. Trends Ecol Evol 16:613–622. doi: 10.1016/S0169-5347(01)02290-X

Borley K, White MM (2006) Mitochondrial DNA variation in the endangered Colorado pikeminnow: a comparison among hatchery stocks and historic specimens. North Am J Fish Manag 26:916–920. doi: 10.1577/M05-176.1

Caro T (2010) Conservation by Proxy: Indicator, Umbrella, Keystone, Flagship, and Other Surrogate Species. Island Press, Washington, DC

Chamberlain CP, Mix HT, Mulch A, et al (2012) The Cenozoic climatic and topographic evolution of the western North American Cordillera. Am J Sci 312:213–262. doi: 10.2475/02.2012.05

Cure K, Thomas L, Hobbs J-PA, et al (2017) Genomic signatures of local adaptation reveal source-sink dynamics in a high gene flow fish species. Sci Rep 7:8618. doi: 10.1038/s41598-017-09224-y

Douglas MR, Brunner PC, Douglas ME (2003) Drought in an evolutionary context: molecular variability in Flannelmouth Sucker (*Catostomus latipinnis*) from the Colorado River Basin of western North America. Freshw Biol 48:1254–1273. doi: 10.1046/j.1365-2427.2003.01088.x

Douglas MR, Douglas ME (2007) Genetic structure of humpback chub *Gila cypha* and roundtail chub *G. robusta* in the Colorado River ecosystem

Durand EY, Patterson N, Reich D, Slatkin M (2011) Testing for ancient admixture between closely related populations. Mol Biol Evol 28:2239–2252. doi: 10.1093/molbev/msr048

Eaton DAR, Ree RH (2013) Inferring Phylogeny and Introgression using RADseq Data: An Example from Flowering Plants (Pedicularis: Orobanchaceae). Syst Biol 62:689–706. doi: 10.1093/sysbio/syt032

Ellstrand NC, Schierenbeck KA (2000) Hybridization as a stimulus for the evolution of invasiveness in plants? Proc Natl Acad Sci U S A 97:7043. doi: 10.1073/pnas.97.13.7043

Fitzpatrick BM, Ryan ME, Johnson JR, et al (2015) Hybridization and the species problem in conservation. Curr Zool 61:206–216. doi: 10.1093/czoolo/61.1.206

Gerber AS, Tibbets CA, Dowling TE (2001) The role of introgressive hybridization in the evolution of the *Gila robusta* complex (Teleostei: Cyprinidae). Evolution 55:2028–2039. doi: 10.1111/j.0014-3820.2001.tb01319.x

Hopken MW, Douglas MR, Douglas ME (2013) Stream hierarchy defines riverscape genetics of a North American desert fish. Mol Ecol 22:956–971. doi: 10.1111/mec.12156

Hovick SM, Whitney KD (2014) Hybridisation is associated with increased fecundity and size in invasive taxa: meta-analytic support for the hybridisation-invasion hypothesis. Ecol Lett 17:1464–1477. doi: 10.1111/ele.12355

Knott JR, Phillips FM, Reheis MC, et al (2018) Geologic and hydrologic concerns about pupfish divergence during the last glacial maximum. Proc R Soc Lond B Biol Sci 285:. doi: 10.1098/rspb.2017.1648

Martin CH, Turner BJ (2018) Long-distance dispersal over land by fishes: extremely rare ecological events become probable over millennial timescales. Proc R Soc Lond B Biol Sci 285:. doi: 10.1098/rspb.2017.2436

Mesgaran MB, Lewis MA, Ades PK, et al (2016) Hybridization can facilitate species invasions, even without enhancing local adaptation. Proc Natl Acad Sci U S A 113:10210. doi: 10.1073/pnas.1605626113

Minckley W, Hendrickson D, Bond C (1986) Geography of western North American freshwater fishes: description and relationships to intracontinental tectonism. In: Hocutt CH, Wiley EO (eds) The Zoogeography of North American Freshwater Fishes. John Wiley & Sons, New York, pp 519–613

Pease JB, Hahn MW (2015) Detection and polarization of introgression in a five-taxon phylogeny. Syst Biol 64:651–662. doi: 10.1093/sysbio/syv023

Spencer JE, Smith GR, Dowling TE (2008) Middle to late Cenozoic geology, hydrography, and fish evolution in the American Southwest. In: Reheis MC, Hershler R, Miller DM (eds) Special Paper 439: Late Cenozoic Drainage History of the Southwestern Great Basin and Lower Colorado River Region: Geologic and Biotic Perspectives. Geological Society of America, pp 279–299

vonHoldt BM, Brzeski KE, Wilcove DS, Rutledge LY (2017) Redefining the role of admixture and genomics in species conservation. Conserv Lett 11:e12371. doi: 10.1111/conl.12371

Wilson GA, Rannala B (2003) Bayesian inference of recent migration rates using multilocus genotypes. Genetics 163:1177–1191