CONSERVATION GENOMIC ASSESSMENT OF TWO IMPERILED FRESHWATER FISHES, LEON SPRINGS PUPFISH (CYPRINODON BOVINUS) AND PECOS GAMBUSIA (GAMBUSIA NOBILIS)

A Thesis

by

ROBERT J. BRETZING-TUNGATE

BS, Texas A&M University-Corpus Christi, 2020

Submitted in Partial Fulfillment of the Requirements for the Degree of

MASTER OF SCIENCE

in

FISHERIES AND MARICULTURE

Texas A&M University-Corpus Christi
Corpus Christi, Texas

December 2023

December 2023

CONSERVATION GENOMIC ASSESSMENT OF TWO IMPERILED FRESHWATER FISHES, LEON SPRINGS PUPFISH (CYPRINODON BOVINUS) AND PECOS GAMBUSIA (GAMBUSIA NOBILIS)

A Thesis

by

ROBERT JAMES BRETZING-TUNGATE

This thesis meets the standards for scope and quality of
Texas A&M University-Corpus Christi and is hereby approved.

David Portnoy, PhD
Chair

Kevin Conway, PhD
Committee Member

Christopher Hollenbeck, PhD
Committee Member

December 2023

ABSTRACT

Freshwater fishes are increasingly recognized as one of the most imperiled groups of vertebrates, with a growing body of research highlighting the significant threat posed to their biodiversity by human activities. Anthropogenic actions, such as habitat modification and destruction, pollution, overexploitation, and the introduction of invasive species, have led to a decline in the number of freshwater fish species worldwide. Addressing this imminent crisis requires comprehensive conservation efforts, stricter regulatory frameworks, habitat restoration, and heightened public awareness. This project aimed to provide data important for the conservation of two imperiled desert freshwater fishes, *Cyprinodon bovinus* and *Gambusia nobilis*, both of which are federally listed as endangered and have undergone range contractions throughout the western United States due to anthropogenic activity. Using genomic techniques, I assessed patterns of genetic diversity within and between populations of both species and screened for evidence of hybridization with introduced congeners. No evidence of contemporary hybridization was found between *C. bovinus* and *C. variegatus*, but admixture was detected among *G. nobilis* and its respective invasive congeners (*G. geiseri* and *G. affinis*). Fine-scale population structure was evident for both species of interest and estimates of effective population sizes were low for both species. The results of the study will help update conservation management plans to help mitigate the threat of extinction for both species.

ACKNOWLEDGEMENTS

I would like to express my gratitude to all the collaborators who have contributed to the completion of this research project. Your support and assistance were invaluable in shaping the direction and quality of the work.

First and foremost, I would like to extend my appreciation to my advisor, David S. Portnoy whose guidance and support played a large role in the development and completion of this research. His mentorship and support have been instrumental in guiding me as a researcher and aspiring ichthyologist.

I am also thankful to my collaborators in the Marine Genomics lab who provided valuable support, feedback, and direction throughout the course of this study. Their contributions greatly helped guide me through the methods and results outlined within this manuscript as well as provided numerous supplementary materials to help me along the way. I would like to thank Andrew Fields and Lizz Hunt especially for their invaluable mentorship during this project. Without their help the timeline and course of this project would have developed very differently.

I would like to acknowledge the organization that provided financial support for this project, The Nature Conservancy. Without their funding, this research would not have been possible.

I would like extend gratitude to the participants of our study, whose willingness to engage and share their experiences was essential to the data collection process. Kevin Conway of Texas A&M University for provided support during field work as well as collecting extra samples when needed. I would like to greatly thank Kevin for also providing all the high-quality photographs used for this study. Also, I would like to thank Megan Bean for her support out of

Texas Parks and Wildlife. Without her aid, field sampling for this project could not have been completed.

Lastly, I want to express my appreciation to my family and friends for their patience, encouragement, and understanding during these three years. Their unwavering support sustained me throughout this journey, and without them I could not have made it to the end of my graduate degree. This research paper stands as a collective effort, and the contributions of these individuals and organizations have been instrumental in its completion. I am sincerely grateful for their support and collaboration.

TABLE OF CONTENTS

Page

LIST OF FIGURES

Page

LIST OF TABLES

Page

CHAPTER I

INTRODUCTION

Conservation Genomics

Conservation genomics can provide a comprehensive understanding of the genetic makeup, population structure, and evolutionary history of species allowing for more effective conservation strategies (Moran 2002). By analyzing genetic data, researchers can assess the genetic health of endangered species, detect inbreeding and hybridization, and identify individuals for captive breeding or reintroduction programs, while enabling the identification of genetically distinct populations, even at fine-scale levels, all of which are essential for developing targeted conservation plans. Conservation genomics can also help to identify genes and genetic variants associated with adaptive traits, providing insights into the potential resilience of species in the face of environmental heterogeneity and decline (Fagan et al. 2002). Additionally, conservation genomics can aid in understanding the impacts of habitat loss, fragmentation, and climate change on genetic diversity and population viability (Moran 2002; Avise 2010). The advancement of genetic techniques and tools has had a great effect on conservation by increasing the amount of information, i.e. loci that can be sampled at a relatively low cost (Peterson et al. 2012), which allows for greater resolution to detect genetic change within populations (e.g., genetic diversity, effective population size, drift, or hybridization and introgression) and between populations (e.g., populations differentiation) over time. Temporal conservation genomic studies are another tool critical in tracking threatened and critically endangered freshwater fish species as rates of freshwater resource quality and availability continues to decrease in North America (Leidy and Moyle 2021a).

Temporal Genetic Studies

Conducting temporal genomic studies on endangered fishes is important for several reasons. Within small populations and species with short generation times, monitoring genetic diversity and population fragmentation is a key factor in building effective management plans. This is especially the case for species that are understudied and exhibit many of the characteristics associated with extinction, e.g., small and fragmented populations with reduced genetic diversity (Pavlova et al. 2017b). For species with small populations, inbreeding and genetic drift can have detrimental effects on genetic diversity and population persistence. Inbreeding occurs when individuals mate with close relatives and often occurs in small populations where there are few potential mates. Inbreeding increasing the likelihood of inheriting identical alleles from both parents leading to a reduction in heterozygosity (Todesco et al. 2016; Chan et al. 2019). As a result, small populations become more homozygous, increasing the chance of expression of deleterious recessive genetic traits and decreasing relative fitness. Furthermore, in small populations, genetic drift can lead to the rapid loss of and fixation of alleles, leaving populations with limited adaptive potential to respond to environmental changes or challenges. As populations continue to decrease in size, these two processes above work in conjunction and eventually populations may lose the adaptive capacity to deal with environmental changes (Fagan et al. 2002). Lastly, introductions of congeneric species can lead to increased competition for limited freshwater resources as well as hybridization, with the latter potentially causing outbreeding depression, decreases in genetic variation, and loss of local adaptative variation (Black et al. 2017; Barker et al. 2019). Conducting genomic studies on endangered fishes across multiple time points provides valuable information for conservation and management, including the ability to monitor genetic diversity, assess population dynamics,

identify changes in inbreeding, and track hybridization events. This knowledge is crucial for developing effective conservation strategies and ensuring the long-term survival of endangered fish species (Allendorf 1989; Vrijenhoek 1994).

Desert Fishes

Desert fishes exhibit higher frequencies of endangerment and require conservation efforts due to their unique ecological and evolutionary circumstances, as well as the numerous threats they face in arid climates, including limited habitat and high levels of endemism (Jaeger et al. 2014). Desert fishes inhabit specialized aquatic ecosystems such as desert springs, small streams, and other isolated water bodies, all of which are subject to rapid deterioration due to human and climatological stressors. These habitats are typically small, isolated and limited in distribution, making them highly vulnerable to water extraction, pollution, and habitat degradation from human activities. Coupling these risks with small population sizes and, commonly, short generation times, these fishes are at great risk of large-scale and small-scale genetic and ecological changes at multiple scales (Laub and Budy 2015).

Genomic Studies for Conservation Management

The preservation of North American freshwater fish biodiversity is an important priority due to the high susceptibility of freshwater fishes to both natural and anthropogenic disturbances. Threats to freshwater biodiversity include non-native species introductions, human induced changes to water flow, pollution, overexploitation, and climate driven processes (Jaeger et al. 2014). As suitable habitat changes and introduced species spread, hybridization of formerly isolated taxa becomes more frequent (Allendorf et al. 2001). As a response, several endemic freshwater fishes of Texas have been chosen to establish refuge populations as a response to the threats these fishes face. These include one of the focal species of this project (*Cyprinodon*

bovinus) and other pupfishes including *C. elegans* (Comanche springs pupfish) and *C. pecosensis* (Pecos pupfish), and *C. eremus* (Sonyota pupfish) (Baugh and Deacon 1988; Rodríguez-Ramírez et al. 2023). The goal of these refuge programs is to maintain refuge populations of imperiled animals and repatriate those refuge organisms back to their habitat when conditions are more favorable. In instances where refuges are necessary but haven't been implemented, some state agencies have reclaimed land and build refugia for fish that have been impacted by anthropogenic activity. Of these fishes, the *C. elegans* and *Gambusia nobilis* were relocated to a manmade Cienega in 1993 in Jeff Davis County, Texas and have been kept there since as a refuge populations for each species (Winemiller and Anderson 1997). Across taxa that make up a large portion on north American desert fishes, refuges have proven to be invaluable tools for management, allowing for small scale tracking of genomic variation within refuge populations that will ideally lead to the stocking and enhancement of wild populations.

References

Allendorf FW (1989) Oxford Surveys in Evolutionary Biology. Evolution 43:1338–1339. https://doi.org/10.1111/j.1558-5646.1989.tb02584.x

Allendorf FW, Leary RF, Spruell P, Wenburg JK (2001) The Problems with Hybrids: Setting Conservation Guidelines. Trends in Ecology & Evolution 16:613–622. https://doi.org/10.1016/S0169-5347(01)02290-X

Avise JC (2010) Perspective: Conservation Genetics Enters the Genomics Era. Conserv Genet 11:665–669. https://doi.org/10.1007/s10592-009-0006-y

Barker AM, Adams DH, Driggers WB, et al (2019) Hybridization Between Sympatric Hammerhead Sharks in the Western North Atlantic Ocean. Biol Lett 15:20190004. https://doi.org/10.1098/rsbl.2019.0004

Baugh TM, Deacon JE (1988) Evaluation of the Role of Refugia in Conservation Efforts for the Devils Hole pupfish, *Cyprinodon diabolis* Wales. Zoo Biology 7:351–358. https://doi.org/10.1002/zoo.1430070406

Black AN, Seears HA, Hollenbeck CM, Samollow PB (2017) Rapid Genetic and Morphologic Divergence Between Captive and Wild Populations of the Endangered Leon Springs pupfish, *Cyprinodon bovinus*. Mol Ecol 26:2237–2256. https://doi.org/10.1111/mec.14028

Chan WY, Hoffmann AA, van Oppen MJH (2019) Hybridization as a Conservation Management Tool. Conservation Letters 12:e12652. https://doi.org/10.1111/conl.12652

Fagan WF, Unmack PJ, Burgess C, Minckley WL (2002) Rarity, Fragmentation, and Extinction Risk in Desert Fishes. Ecology 83:3250–3256. https://doi.org/10.1890/0012-9658(2002)083[3250:RFAERI]2.0.CO;2

Jaeger KL, Olden JD, Pelland NA (2014) Climate Change Poised to Threaten Hydrologic Connectivity and Endemic Fishes in Dryland Streams. Proceedings of the National Academy of Sciences 111:13894–13899. https://doi.org/10.1073/pnas.1320890111

Laub BG, Budy P (2015) Assessing the Likely Effectiveness of Multispecies Management for Imperiled Desert Fishes with Niche Overlap Analysis. Conservation Biology 29:1153–1163. https://doi.org/10.1111/cobi.12457

Leidy RA, Moyle PB (2021) Keeping Up with the Status of Freshwater Fishes: A California (USA) Perspective. Conservation Science and Practice 3:e474. https://doi.org/10.1111/csp2.474

Moran P (2002) Current Conservation Genetics: Building an Ecological Approach to the Synthesis of Molecular and Quantitative Genetic Methods. Ecology of Freshwater Fish 11:30–55. https://doi.org/10.1034/j.1600-0633.2002.110105.x

Pavlova A, Gan HM, Lee YP, et al (2017) Purifying Selection and Genetic Drift Shaped Pleistocene Evolution of the Mitochondrial Genome in an Endangered Australian Freshwater Fish. Heredity 118:466–476. https://doi.org/10.1038/hdy.2016.120

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:e37135. https://doi.org/10.1371/journal.pone.0037135

Rodríguez-Ramírez R, Echelle AA, Varela-Romero A, et al (2023) Genetic Evaluation of the Refuge Program for Sonoyta Pupfish *Cyprinodon eremus* (Cyprinodontidae) in Mexico. J Ichthyol. https://doi.org/10.1134/S0032945223060140

Todesco M, Pascual MA, Owens GL, et al (2016) Hybridization and Extinction. Evolutionary Applications 9:892–908. https://doi.org/10.1111/eva.12367

Vrijenhoek RC (1994) Genetic Diversity and Fitness in Small Populations. In: Loeschcke V, Jain SK, Tomiuk J (eds) Conservation Genetics. Birkhäuser, Basel, pp 37–53

Winemiller KO, Anderson AA (1997) Response of Endangered Desert Fish Populations to a Constructed Refuge. Restoration Ecology 5:204–213. https://doi.org/10.1046/j.1526-100X.1997.09725.x

CHAPTER II

CONSERVATION GENOMIC ASESSMENT OF THE IMPERILED LEON SPRINGS PUPFISH (*CYPRINODON BOVINUS*) ACROSS TEMPORAL AND SPATIAL SCALES

This chapter has been prepared for publication in the journal of Conservation Genomics. Modifications to formatting have been made to comply with thesis formatting requirements.

Abstract

The Leon Springs pupfish, *Cyprinodon bovinus,* is federally listed as critically endangered and is confined to an extremely narrow range in the southwestern United States. The only known contemporary locations for the species are spring fed habitat in the Diamond Y Spring Preserve in Pecos County, Texas and a reserve population is maintained by U.S. Fish and Wildlife Service in Dexter, New Mexico. To inform continued conservation and management of the species, a comprehensive conservation genomics study was performed. No evidence of contemporary hybridization or historical introgression between *C. bovinus* and the invasive congener *C. variegatus* was found. The reserve population was found to be significantly differentiated from the two wild samples, but the two wild samples (located less than 1 km from each other) also exhibited significant heterogeneity, providing evidence of population structure at a very small spatial scale. Estimates of contemporary diversity and effective population size were smaller for both wild samples than the reserve population and temporal comparisons suggest an increased magnitude of drift acting on the wild populations relative to the reserve population.

Introduction

Genetic monitoring programs have become important tools for conservation and management of endangered freshwater fishes (Schwartz et al. 2007; Bernos et al. 2020). Many of these species have short generation times, small population sizes and fragmented distributions,

making them at risk of rapidly losing genetic diversity due to genetic drift and inbreeding (Kennedy 1977). Furthermore, the widespread introduction of non-native species poses threats in terms of competition and predation (Liss et al. 2016). In the case of introduced congeners, hybridization can result in the introduction of maladaptive variation and in a worst-case scenario the genetic swamping and loss of native species (Allendorf et al. 2001; Todesco et al. 2016). To combat these problems, refuge populations are often established which can be used to augment or replace the wild populations if necessary (Meretsky et al. 2006). However, refuge populations must be managed carefully to ensure that they maintain sufficient levels of standing genetic variation and remain genetically similar to wild populations (Love Stowell et al. 2017; Novak et al. 2020). Genomic approaches can provide insight into all of these processes by allowing researchers to assess the adaptive potential of endangered species, detect inbreeding and hybridization, and monitor captive breeding or reintroduction programs, while enabling the identification of genetically distinct populations, even at fine-scales, which is essential for developing targeted conservation plans (Meffe and Vrijenhoek 1988).

One group of freshwater fishes of particular conservation concern in southwestern North America (swNA) are pupfishes in the genus *Cyprinodon* (Family Cyprinodontidae). The genus contains 44 species, including 34 species found in the swNA (Texas, New Mexico, Arizona, Nevada, California, and Mexico; (Echelle et al. 2005; Echelle 2008). Phylogenetic analyses suggest that a common ancestor to the groups was widespread across a wetter swNA during the Pliocene (~3 MYA), but as the climate became arid, populations became isolated in pockets of habitat that remained and began to speciate (Eschellle et al. 2005). Many pupfishes in swNA show extremely limited ranges, with fragmented distributions and small population sizes (Echelle 2008). Because swNA is an arid region, human water usage has become a major threat

to pupfishes (Baugh and Deacon 1988; Lewis et al. 2013a; Black et al. 2017). To deal with these issues, pupfish habitat has been protected by state or federal agencies or acquired in land purchase by non-government organizations when possible (Minckley and Deacon 2017). Furthermore, example species that are now maintained in refuge populations include *C. diabola* (Devils Hole pupfish) and *C. bovinus* (Leon Springs pupfish; Bough and Deacon, 1988; Black et al. 2017).

The focal species of this study, *C. bovinus*, is currently found only in a series of spring fed pools on the Diamond Y Spring Preserve (DY) in Pecos County, Texas. Like many other species of pupfish, *C. bovinus* is sexually dimorphic with males typically reaching lengths of around 2.5 to 3.8 centimeters and females reaching slightly smaller sizes (Kennedy 1977). The species is also dichromatic, with large breeding male *C. bovinus* exhibiting yellow dorsally with flecks of blue and yellow on the body and a dark margin on the caudal, while females have visible dark barring and are drabber (Fig. 2.1; male *C. bovinus*). Breeding usually occurs during the springs and early summer months and involves large males guarding territory around the water's edge, with females selecting males and laying eggs in territories to be fertilized (Al-Shaer et al. 2016; Bernos et al. 2020). *C. bovinus* has a short generation time, reaching maturity between 4-6 months and living on average ~1 year (Kennedy 1977).

Historically, *C. bovinus* ranged from Leon Springs (the location of their first description) to the lower reaches of the Pecos River drainage in West Texas, found in clear, cool, spring-fed waters. However, due to issues with water usage, the species was extirpated from much of its range and thought to be extinct (Hubbs 1957) until it was rediscovered in the area around DY (Black et al. 2017). In 1980, the species was listed as endangered, and a recovery plan was adopted in 1985 (Black et al. 2016). Prior to this, in the 1970's a refuge population of *C. bovinus*

was established at the Southwestern Natural Resources and Rehabilitation Centre (REF) in Dexter, New Mexico, using 80 wild fish from DY, as a response to the species nearly being extirpated from DY because of the introduction of an invasive congener *C. variegatus* (sheepshead minnow). Besides posing a threat in terms of competition, experimental and empirical research have demonstrated that *C. bovinus* readily hybridizes with *C. variegatus* (Garret 1980), threatening the integrity of the *C. bovinus* gene pool (Hubbs 1980). To deal with the issue, all *C. variegatus, C. bovinus*, and admixed individuals from DY were culled, followed by introduction of *C. bovinus* from the refuge population in 1976 (Echelle et al. 2004). The second restoration effort from 1998-2001 added an additional 7,755 to 8,364 to DY. Population declines were seen again beginning in 2001, hypothesized to be due to decreases in habitat quality and quantity due to proximal oil and gas extractions. Another round of augmentation from the refuge population was coupled with habitat restoration in a lower reach water body (Monsanto Pool), but by 2013 that population had collapsed (Black et al. 2017). In 2015, an additional 500 pupfish from the refuge were released into Monsanto Pool (Al-Shaer et al. 2018), but a survey in 2019 found no fish. Despite all of the effort expended, the only habitats with contemporary *C. bovinus* populations are in DY.

Black et al. (2017) examined the degree of contemporary hybridization between *C. variegatus* and *C. bovinus* as well as morphological and genetic divergence between the refuge and wild populations of *C. bovinus* in DY using samples collected in 2013. They found significant differences in morphology and genetic variation between the refuge and the wild, but no significant evidence of contemporary hybridization between the *C. bovinus* and *C. variegatus*. The study provided critical data for a first pass genomic assessment of *C. bovinus* but was a snapshot in time of processes (drift and hybridization) that are dynamic. Furthermore, Black et

al. (2017) treated all individuals sampled from the wild as a single population. Though historically heavy rains have allowed for temporary connections to form between spring fed pools in DY, there have been local and state-wide droughts that have left these pools completely disconnected since 2018 (Ryan Smith, Per. Com), which is potentially enough time for significant divergence to occur, given the species short generation time and potentially small population sizes. Therefore, this study utilized the data from Black et al. (2017) and a fine-scale sampling in the wild to look for evidence of hybridization and assess contemporary genetic diversity within space and across time to better understand processes causing the wild and refuge populations to move apart, as well as assess for population structure within the wild

Methods

Field Sampling

Fin clips were collected from 30 to 40 *C. bovinus* at DY and from the United States Fish and Wildlife Service (USFWS) refuge population (hereafter REF) at the Southwestern Native Aquatic Resource and Recovery Centre (SNARCC) in Dexter, New Mexico. Four discrete sampling locations in DY were identified in advance where *C. bovinus* were known to be present, but one site had unexpectedly desiccated, leaving no habitat, and the species was absent at a second site. Thus, only two locations were sampled on DY: Karges Springs (KGS) and the Headwater Spring (HEAD), which are not directly connected, though only separated by ~0.5 km. Fin clips were also collected from *C. variegatus* from four locations across the species' range in Texas. All tissues were immersed in 20% salt-saturated DMSO buffer (Seutin et al. 1991) upon collection in the field and stored at room temperature until time of extraction. Tissues were collected from DY in March of 2020, while tissues from REF and *C. variegatus* were collected between June – December of 2021. To capture individuals, a seine net was used when sites were

large enough and dip nets at all other times. A sterile razor blade was used to remove the upper 25% of the caudal fin from each captured fish and all tools were cleaned using de-ionized water and 10% bleach solutions between fish. After handling, individuals were held in a bucket filled with water from their habitat, which was oxygenated using a bubbler and placed in the shade to ensure survival after release. Three whole *C. bovinus* were euthanized in total, preserved in 95% ethanol and are stored as voucher specimens in the Biodiversity, Research, and Teaching Collection in College Station. Whole individuals of *C. variegatus* were collected using dip nets, euthanized using clove oil, and preserved in 95% ethanol. Fin clips were later taken in the lab and immersed in 20% salt-saturated DMSO until time of extraction, while the rest of the tissue samples were stored at -20ºF in 95% ethanol at the Marine Genomics Lab at Texas A&M University-Corpus Christi. The distribution of sampling sites is shown in Figure 2.2, including the total count of individuals sampled at each location. Sampling for *C. bovinus* and *C. variegatus* within Pecos County was conducted with TPWD and TNCT biologists under permit number TE814933, and *C. variegatus* sampled outside of Pecos County were collected under permit number SPR-0614-111. All animal use and care followed IACUC animal protocol TAMU IACUC 2021-0001.

Genomic Library Preparation and Sequencing

Genomic DNA was extracted from fin clips using MagBind Blood and Tissue HDQ DNA extraction kits (Omega Bio-tek). Extractions were then electrophoresed through a 1% agarose gel and quantified with AccuBlue High Sensitivity dsDNA Quantitation Kits (Biotium). Reduced representation libraries were assembled following a modified version of the double-digestion restriction-site associated DNA sequencing (ddRADseq) protocol of Peterson et al. (2012). Briefly, libraries were assembled using ~500 ng of genomic DNA which was digested

using restriction enzymes *Eco*RI and *Msp*I (New England Biolabs). After digestion, DNA was purified using Mag-Bind TotalPure NGS (Omega Bio-tek), quantified, and standardized to 100ng/μL. For each digestion, one of 48 barcoded adapters was ligated to the *Eco*RI site and a common adaptor was ligated to the *Msp*I site. Ligation reactions were then PCR tested using a two-step PCR protocol (amplification for 18 & 32 cycles) and checked using gel electrophoresis. Successful ligation reactions were pooled into one of four indexed libraries (Table S.2.1), purified using PEG, and quantified again. Fragments were size selected between 313-437bp using Pippin Prep (Sage Science) and PCR amplified for 14 cycles to incorporate P2 adapters containing index sequences. Libraries were then purified one more time and quantified. To ensure proper size selection, pooled fragments were run on a fragment analyzer. Libraries were paired end sequenced on a single lane of an Illumina HiSeq 4000 (2x150bp).

COI Sequencing

A subset of individuals from both species was selected haphazardly for sequencing of the mitochondrially encoded (mtDNA) cytochrome oxidase subunit I gene (COI). DNA was extracted using MagBind Blood and Tissue HDQ DNA extraction kits (Omega Bio-tek) and a 698 bp fragment of CO1 was PCR amplified using universal F1/F2 & R1/R2 fish primers (Ward et al. 2005). PCR reactions (30 μL) contained 5 x GoTaq buffer, 1.5 mM magnesium chloride, 1% Tween, 2.5 mM dNTPs each, 0.25 mM of each primer, 0.03 units of *Taq* polymerase, 1 μL of template DNA, and water. Cycling was performed as follows: denaturation at 95 ºC for 2 minutes, followed by 35 cycles of denaturation at 95 ºC for 60 seconds, annealing at 52 ºC for 60 seconds, and extension at 72 ºC for 90 seconds, a final extension was done at 72 ºC for 10 minutes. Amplicons were purified using Mag-Bind TotalPure NGS. To prepare the samples for sequencing, amplicons were standardized to 10ng/μL in 10 μL of water and Sanger sequenced on

an ABI3730 XL DNA Analyzer at the Core Lab at Texas A&M University-Corpus Christi. Raw Sanger sequences were quality trimmed and edited by using TRIMMOMATIC v8.22 (Free Software Foundation, Inc.) and compared in bulk to the NCBI nucleotide collection using BLASTn (Altschul et al. 1990) to obtain individual species IDs.

Bioinformatics and Filtering

For downstream analysis, three datasets were produced. One dataset was created for hybrid analysis and contained all contemporary *C. variegatus* and *C. bovinus*. The second dataset included only contemporary *C. bovinus* individuals and was used to look at patterns of contemporary genetic structure and diversity. Finally, the third dataset combined data from contemporary individuals of *C. bovinus* with sequence data from historical samples of *C. bovinus* (Black et al. 2017) to assess temporal changes in diversity.

To create the three datasets, raw Illumina reads for *C. bovinus and C. variegatus* and reads from Black et al. (2017) were demultiplexed using a custom Perl script and processed using the DDOCENT v2.9.1 pipeline (Puritz et al.). For each dataset, appropriate demultiplexed sequences were quality trimmed and stacked into orthologous scaffolds and made into a de novo reference. The reference was then optimized for mapping using custom BASH scripts. Quality trimmed sequences were then mapped back onto the reference and genomic variants were scored and compiled into a VCF file. Variants were then quality filtered using a combination of VCFTOOLS (Danecek et al. 2011) and various BASH and Perl scripts. Filtering followed guidance of O'Leary et al. (2018) to remove low quality or artefactual SNPs as well as potentially paralogs and low-quality individuals. Genotypes with quality < 15 and called from < 10 reads were coded as missing, retaining loci with quality > 15, genotype call rate > 50%, and mean depth 10. Additionally, loci were filtered for allele balance, mapping quality ratios, strand

balance, paired status, depth/quality ratio, and heterozygosity. Individuals with > 20% missing data were removed. Using a custom Perl script, SNP variants found in the same contig were phased into microhaplotypes (hereafter loci) following Willis et al. (2017). Pair-wise relatedness (Wang 2014) was used to screen for duplicate pairs and identify potentially related individuals. To test for putative loci under selection the FDIST (Antao et al. 2008) method was implemented in ARLEQUIN v3.5.2.2 (Excoffier and Lischer 2010) using 20,000 coalescent simulations and an island model. A Bayesian approach for outlier detection was also implemented, using BAYESCAN (Foll and Gaggiotti 2008) with 30 pilot runs of 5,000 iterations, flowed by a burn in of 50,000 iterations and 500,000 iterations sampled 5,000 times, and a *q*-value of 0.05. While outlier loci provide evidence for localized adaptation, they may cause bias in neutral data sets (Fagan et al. 2002). Therefore, outliers were removed from downstream analyses because the focus of this study is understanding patterns of population structure and drift.

Hybrid Analysis

Using the hybrid dataset, reads were mapped to a multispecies reference and loci that were specific to each species group (diagnostic loci) were used to identify pure individuals of *C. bovinus* and pure *C. variegatus*. NEWHYBRIDS v2.0 (Anderson and Thompson 2002) was then used to assign individuals back to the parent species or one of three hybrid classes (F1, F1 x *C. bovinus*, and F1 x *C. variegatus*). The program uses a Bayesian clustering algorithm that maximizes the posterior probability of each individual's assignment, using an estimated proportion of hybrid individuals and the assumed genetic model. A secondary approach for hybrid detection was implemented in *Adegenet* v2.5.1 (Jombart 2008) in R. For this approach, 30 F1 hybrids, 30 F1 x *C. bovinus* backcrosses, and 30 F1 x *C. vareigatus* backcrosses were simulated. Hybrids individuals and empirical individuals were then plotted using PCA to see

how empirical samples grouped with simulated individuals. Finally, species identities based on COI were compared to species identities based on nuclear data to validate hybrid analyses and assesses for historical mtDNA introgression.

Population Genetic Analyses

For the contemporary *C. bovinus* dataset, a single-level analysis of molecular variance (AMOVA) was carried out in ARLEQUIN v3.5.2.2, using a locus-by-locus framework to account for uneven levels of missing data across loci (Weir and Cockerham 1984). Significance was determined at an α-level of 0.05 by permuting individuals among locations 10,000 times and 95% confidence intervals determined using 20,000 bootstrap replicates. Post-hoc pairwise F_{ST} was then estimated in ARLEQUIN with significance determined as above and 95% confidence intervals calculated in *hierfstat* (Goudet 2005) in R. To assess the number of genetic groups present in the data, discriminant analysis of principal components (DAPC; Jombart et al. 2010) was implemented in *adegenet* using *K*-means clustering (K = 2-6) with the number of clusters selected by comparing Bayesian information criterion values (BIC). The optimal number of PCs to retain was then determined using cross validation and membership probabilities for each individual to the inferred clusters were calculated. Effective population size (N_E) was estimated for each location using the linkage disequilibrium approach with an allele frequency cut-off of 0.1 and 95% confidence intervals determined using 1,000 jackknife replicates, as implemented in NEESTIMATOR v2.1 (Do et al. 2014). To account for physical linkage, all contemporary N_E estimates were adjusted according to (Waples et al. 2016). Mean expected heterozygosity (H_E; Nei 1973), and rarefied allelic richness (A_R; El Mousadik and Petit 1996) were estimating using *hierfstat*. Friedman's tests were conducted on both metrics to test for homogeneity among

locations, and post-hoc Wilcoxon tests conducted to assess pairwise differences using the R package *coin* v.1.3.1 (Hothorn et al. 2008).

Temporal Analysis

For the contemporary/historical dataset, F_{ST} was estimated between samples over time: contemporary refuge (CREF) vs historical refuge (HREF) and contemporary wild (CWILD) vs historical wild (HWILD), and across samples within time: CREF vs CWILD and HREF vs HWILD. F_{ST} was calculated using ARLEQUIN with significance determined as above and 95% confidence intervals calculated in *hierfstat*. Structure was also visualized using PCA as implemented in *adegenet*. Lastly, estimates of variance N_E (Waples 1989) and confidence intervals were made for the refuge (across seven generations) and wild (across seven generations) using Pollack's F (Pollak 1983) and a minimum allele frequency of 0.1 in NEESTIMATOR.

Results

Hybrid Detection

After filtering, the final hybrid dataset contained 113 individuals genotyped at 4,734 loci, with an average of 6.7 alleles per locus. No hybrids were detected using NEWHYBRIDS and no empirical samples grouped with simulated hybrids or backcrosses using PCA (Figure S1). Therefore, no individuals were removed from the contemporary *C. bovinus* dataset. Individual species identities inferred from COI and nuclear data corresponded, providing no evidence of historical mtDNA introgression.

Population Genetic Analyses

After filtering, the contemporary *C. bovinus* data contained 72 individuals genotyped at 4,598 loci, with an average of 2.34 alleles per locus. Thirty loci putatively under selection were

found and removed, leaving 4,568 neutral loci. One pair of individuals had a relatedness value of 0.98, leading to the removal of one individual. Therefore, the final dataset contained 71 individuals genotyped across 4,568 loci. The component of genetic variation attributable to differences among samples was highly significant (%V = 5.49, $P < 0.0001$; Table 2.1). Post-hoc estimate of pairwise F_{ST} were significant between all samples but were an order of magnitude greater in comparisons involving REF relative to the comparison involving the two wild populations (Table 2.2). The minimum BIC value was obtained for K=2 (Figure S2), concordant with a refuge group and a wild group (Figure S3), and 100% of individuals assigned back to their group of origin. Estimated N_E for KGS, HEAD, and REF were 252, 126, and 601 respectively (Table 2.3). Estimates of H_E were 0.32, 0.32 and 0.35, and estimates of A_R were 2.05, 2.08, and 2.21, for KGS, HEAD, and REF respectively (Table 2.4). Friedman's tests for within population diversity were significant for H_E (; H_E: $X^2(2) = 84.4$, p-value = 2.2×10^{-16}) but not A_R. Post-hoc Wilcoxon tests for H_E were only significant between REF and wild populations, with REF having significantly higher diversity (Table 2.5).

Temporal Analysis

The final contemporary/historical dataset contained 117 individuals genotyped at 3,281 loci, with an average of 2.27 alleles per locus. Pairwise F_{ST} was nearly an order of magnitude greater for spatial comparisons (HWILD vs HREF; CWILD vs CREF) than temporal comparisons (HREF vs CREF; HWILD vs CWILD; Fig. 2.3). The temporal comparison between refuge samples was ~5x smaller than the temporal comparison between wild populations, and the former was not significant. Variance estimates of effective population size were larger for the refuge population than the wild (Fig. 2.4).

Discussion

No contemporary hybrids or evidence of introgression were detected in this study, despite the presence of *C. variegatus* in habitat close to DY (Shepta 2022). These results agree with the findings in Black et al. (2017) and suggest that extensive culling in DY effectively removed all individuals carrying *C. variegatus* genetic material from native *C. bovinus* populations. Furthermore, no *C. variegatus* were encountered in DY, indicating that introduced *C. variegatus* to the surrounding area might not have made it into DY pools to this point. It is important to note that the level of interactions between potentially hybridizing species is dynamic and that current apparent lack of gene flow does not guarantee a lack of gene flow in the future, so continued monitoring will be necessary (Perry et al, 2002; Todesco et al. 2016). Another point of concern is the lack of evidence of post-zygotic barriers in among species in the genus *Cyprinodon*. Laboratory studies have documented decreases in male hybrid fertility, survivorship as well as a female preference for true species over hybrid males species (Rosenfield et al. 2004; Tech 2006). Lastly, *C. variegatus* for this project were sampled at locations far from DY to ensure that pure *C. variegatus* were used to look for introgression. While the results of mtDNA analysis are clear, there could be components of *C. variegatus* nuclear variation specific to populations in Pecos County present in *C. bovinus* that were not detected.

Population structure and differentiation were significant in the contemporary analysis between the wild and refuge populations, concordant with Black et al. (2017). While Black et al. (2017) did not look at fine scale differences within DY, the present study found discrete population structure within the wild samples despite the short geographic distance (~0.5 km) between the two wild sampling locations. The results indicate that there has been little to no gene flow between KGS and HEAD in the recent past, consistent with the observation that the two

sites sampled have been disconnected due to drought since 2018. Given this species' short generation time and the suspected small number of individuals left in the wild, little time would be required for genetic drift to create significant differentiation between populations (Blažek et al. 2013).

This study found that the refuge was more diverse than either of the wild populations. Consistent with this, the refuge population had large estimated contemporary N_E relative to both wild populations. Small contemporary N_E in the wild population further reflects the notion that both KGS and HEAD contain small numbers of breeding individuals and do not receive contribution via gene flow (Meretsky et al. 2006). Although KGS and HEAD had similar expected heterozygosity, HEAD had higher allelic richness despite smaller contemporary N_E, perhaps indicating differences in recent past demography between the two populations (Leberg 2002). Selection could also be acting at a fine-scale population level contributing to differentiation between the two habitats in DY, but the majority of detected outliers emphasized differences between REF and the two wild populations, rather than difference between KGS and HEAD.

Finally, variance N_E of the wild was smaller than the refuge and closer to the estimate of contemporary N_E for HEAD than KGS. This could be due to the effect of combining two differentiated populations (HEAD and KGS) for temporal analysis, or it could simply reflect that the realized population size of HEAD is smaller than KGS (Debouzie 1980). Furthermore, the historical and contemporary wild samples were significantly differentiated, indicating changes in allele frequencies over the time period between the two studies. On the other hand, the refuge population variance N_E was only slightly lower than the estimated contemporary N_E and confidence intervals for estimates overlapped. This finding, along with the presence of only

slight levels of differentiation measured between the historical and contemporary refuge samples, suggest that refuge populations is being maintained in a way that is minimizing the effects of genetic drift and maintaining genetic diversity (Meretsky et al. 2006; Laub and Budy 2015). Taken as a whole, it appears that increased divergence between the wild and refuge populations seen in this study, relative to Black et al. (2017), is likely attributable to drift in the wild, with wild populations experiencing reductions in genetic diversity and the refuge remaining relatively stable.

With the significant difference in genetic diversity and differentiation between the two wild sites and the refuge, genetic rescue via assisted migration from the refuge may be necessary. If the wild populations remain isolated, they will likely continue to diverge (Booy et al. 2000; Whiteley et al. 2015) and with contemporary estimate of N_E in the low hundreds they may be at risk of extirpation/extinction (Frankham et al. 2014). Lack of connectivity associated with drought likely led to the formations of distinct *C. bovinus* populations in discrete locations and may have also led to decreases in the size of the gene pools of these populations (Jaeger et al. 2014). This cycle of decreasing population sizes and decreasing genetic diversity can create an extinction vortex, a phenomenon that is seen in desert fishes (Booy et al. 2000; Todesco et al. 2016). One example of note is in the Macquarie perch (*Macquaria australasica*) from Australia. Multiple fragmentation events led to dire reductions in genetic diversity between sub populations and decreases in effective population sizes that were below the threshold to retain adaptive potential. The study found that assisted gene flow among wild populations or from a refuge population was necessary if the sub populations were to persist into the future (Pavlova et al. 2017a). Assisted migration between wild populations might also be a strategy for increasing genetic diversity (Maehr et al. 2002). Given the information from this study's tandem analysis of

the refuge and wild populations with historical data from Black et al. (2017), genetic rescue of the wild populations might be plausible, but caution will be needed as the current study is unable to address effects of potentially deleterious alleles being introduced from a separately maintained genetic unit. Future studies should maintain focus on contemporary genetic diversity and hybridization while habitat conditions are becoming more unfavorable. Secondly, it has been shown that after the culling and introduction of refuge individuals of *C. bovinus* in the late 1900s and the early 2000s, little time is needed for the wild populations to completely diverge from the refuge. This cycle is likely to continue if the same restoration tactic is used again. Bi-directional transplant of individuals (i.e., transplant wild *C. bovinus* to the refuge then refuge *C. bovinus* to the wild after a number of generations have passed) might be more beneficial to introduce new variation to the refuge instead of continuously homogenizing the wild using refuge individuals, as well as assisted migration between discrete wild populations to increase the effective population sizes in the wild, and increase heterozygosity the ensure the goal of management isn't just short term survivability, but prioritizing long-term population persistence (Frankham et al. 2014a, b).

References

Allendorf FW, Leary RF, Spruell P, Wenburg JK (2001) The Problems with Hybrids: Setting Conservation Guidelines. Trends in Ecology & Evolution 16:613–622. https://doi.org/10.1016/S0169-5347(01)02290-X

Al-Shaer L, Bloch A, Little K, Itzkowitz M (2018) Monitoring Social Behaviour as an Assessment of Translocation Success in a Reintroduced Population of the Endangered Leon Springs pupfish (*Cyprinodon bovinus*). Aquatic Conservation: Marine and Freshwater Ecosystems 28:. https://doi.org/10.1002/aqc.2889

Al-Shaer L, Bloch A, Paciorek T, et al (2016) Renovated Breeding Habitat Use in Wild & Captive-bred Populations of an Endangered Desert Pupfish. Journal of Biodiversity & Endangered Species 4:. https://doi.org/10.4172/2332-2543.1000156

Altschul SF, Gish W, Miller W, et al (1990) Basic Local Alignment Search Tool. Journal of Molecular Biology 215:403–410. https://doi.org/10.1016/S0022-2836(05)80360-2

Anderson EC, Thompson EA (2002) A Model-Based Method for Identifying Species Hybrids Using Multilocus Genetic Data. Genetics 160:1217–1229

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. https://doi.org/10.1186/1471-2105-9-323

Baugh TM, Deacon JE (1988) Evaluation of the Role of Refugia in Conservation Efforts for the Devils Hole Pupfish, *Cyprinodon diabolis* Wales. Zoo Biology 7:351–358. https://doi.org/10.1002/zoo.1430070406

Bernos TA, Jeffries KM, Mandrak NE (2020) Linking Genomics and Fish Conservation Decision Making: A Review. Rev Fish Biol Fisheries 30:587–604. https://doi.org/10.1007/s11160-020-09618-8

Black AN, Seears HA, Hollenbeck CM, Samollow PB (2017) Rapid Genetic and Morphologic Divergence Between Captive and Wild Populations of the Endangered Leon Springs Pupfish, *Cyprinodon bovinus*. Mol Ecol 26:2237–2256. https://doi.org/10.1111/mec.14028

Black AN, Snekser JL, Al-Shaer L, et al (2016) A Review of the Leon Springs Pupfish (*Cyprinodon bovinus*) Long-Term Conservation Strategy and Response to Habitat

Restoration. Aquatic Conservation: Marine and Freshwater Ecosystems 26:410–416. https://doi.org/10.1002/aqc.2608

Blažek R, Polačik M, Reichard M (2013) Rapid Growth, Early Maturation and Short Generation Time in African Annual Fishes. EvoDevo 4:24. https://doi.org/10.1186/2041-9139-4-24

Booy G, Hendriks RJJ, Smulders MJM, et al (2000) Genetic Diversity and the Survival of Populations. Plant Biol (Stuttg) 2:379–395. https://doi.org/10.1055/s-2000-5958

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

Debouzie D (1980) Estimate of Variance Effective Population Size in a Laboratory Ceratitis Population. Heredity 45:297–299

Do C, Waples RS, Peel D, et al (2014) NeEstimator v2: Re-Implementation of Software for the Estimation of Contemporary Effective Population Size (N_E) from Genetic Data. Mol Ecol Resour 14:209–214. https://doi.org/10.1111/1755-0998.12157

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) Late Cenozoic Drainage History of the Southwestern Great Basin and Lower Colorado River Region: Geologic and Biotic Perspectives. Geological Society of America, p 0

Echelle AA, Carson EW, Echelle AF, et al (2005) Historical Biogeography of the New-World Pupfish Genus *Cyprinodon* (Teleostei: Cyprinodontidae). cope 2005:320–339. https://doi.org/10.1643/CG-03-093R3

El Mousadik A, Petit RJ (1996) High Level of Genetic Differentiation for Allelic Richness Among Populations of the Argan Tree [*Argania spinosa* (L.) Skeels] Endemic to Morocco. Theoret Appl Genetics 92:832–839. https://doi.org/10.1007/BF00221895

Excoffier L, Lischer HEL (2010) Arlequin Suite Ver 3.5: A New Series of Programs to Perform Population Genetics Analyses Under Linux and Windows. Mol Ecol Resour 10:564–567. https://doi.org/10.1111/j.1755-0998.2010.02847.x

Fagan WF, Unmack PJ, Burgess C, Minckley WL (2002) Rarity, Fragmentation, and Extinction Risk in Desert Fishes. Ecology 83:3250–3256. https://doi.org/10.1890/0012-9658(2002)083[3250:RFAERI]2.0.CO;2

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. https://doi.org/10.1534/genetics.108.092221

Frankham R, Bradshaw CJA, Brook BW (2014a) Genetics in Conservation Management: Revised Recommendations for the 50/500 Rules, Red List Criteria and Population Viability Analyses. Biological Conservation 170:56–63. https://doi.org/10.1016/j.biocon.2013.12.036

Frankham R, Bradshaw CJA, Brook BW (2014b) 50/500 Rules Need Upward Revision to 100/1000 – Response to Franklin et al. Biological Conservation 176:286–286. https://doi.org/10.1016/j.biocon.2014.05.006

Goudet J (2005) Hierfstat, A Package for R to Compute and Test Hierarchical F-Statistics. Molecular Ecology Notes 5:184–186. https://doi.org/10.1111/j.1471-8286.2004.00828.x

Hothorn T, Bretz F, Westfall P (2008) Simultaneous Inference in General Parametric Models. Biometrical Journal 50:346–363. https://doi.org/10.1002/bimj.200810425

Hubbs C (1957) Distributional Patterns of Texas Fresh-Water Fishes. The Southwestern Naturalist 2:89. https://doi.org/10.2307/3669496

Jaeger KL, Olden JD, Pelland NA (2014) Climate Change Poised to Threaten Hydrologic Connectivity and Endemic Fishes in Dryland Streams. Proceedings of the National Academy of Sciences 111:13894–13899. https://doi.org/10.1073/pnas.1320890111

Jombart T (2008) Adegenet: A R Package for the Multivariate Analysis of Genetic Markers. Bioinformatics 24:1403–1405. https://doi.org/10.1093/bioinformatics/btn129

Jombart T, Devillard S, Balloux F (2010) Discriminant Analysis of Principal Components: A New Method for the Analysis of Genetically Structured Populations. BMC Genetics 11:94. https://doi.org/10.1186/1471-2156-11-94

Kennedy SE (1977) Life History of the Leon Springs Pupfish, *Cyprinodon bovinus*. Copeia 1977:93–103. https://doi.org/10.2307/1443509

Laub BG, Budy P (2015) Assessing the Likely Effectiveness of Multispecies Management for Imperilled Desert Fishes with Niche Overlap Analysis. Conservation Biology 29:1153–1163. https://doi.org/10.1111/cobi.12457

Leberg PL (2002) Estimating Allelic Richness: Effects of Sample Size and Bottlenecks. Molecular Ecology 11:2445–2449. https://doi.org/10.1046/j.1365-294X.2002.01612.x

Lewis RH, Allan NL, Stoops SB, et al (2013) Status of the Endangered Pecos Gambusia (*Gambusia nobilis*) and Comanche Springs Pupfish (*Cyprinodon elegans*) in Phantom Lake Spring, Texas. The Southwestern Naturalist 58:234–238

Liss SA, Lamer JT, Sass GG, Suski CD (2016) Physiological Consequences of Hybridization: Early Generation Backcrossing Decreases Berformance in Invasive Bigheaded Carps.

Journal of Freshwater Ecology 31:543–554. https://doi.org/10.1080/02705060.2016.1188426

Love Stowell SM, Pinzone CA, Martin AP (2017) Overcoming Barriers to Active Interventions for Genetic Diversity. Biodivers Conserv 26:1753–1765. https://doi.org/10.1007/s10531-017-1330-z

Maehr DS, Land ED, Shindle DB, et al (2002) Florida Panther Dispersal and Conservation. Biological Conservation 106:187–197. https://doi.org/10.1016/S0006-3207(01)00245-2

Meffe GK, Vrijenhoek RC (1988) Conservation Genetics in the Management of Desert Fishes. Conservation Biology 2:157–169. https://doi.org/10.1111/j.1523-1739.1988.tb00167.x

Meretsky VJ, Fischman RL, Karr JR, et al (2006) New Directions in Conservation for the National Wildlife Refuge System. BioScience 56:135–143. https://doi.org/10.1641/0006-3568(2006)056[0135:NDICFT]2.0.CO;2

Minckley WL, Deacon JE (2017) Battle Against Extinction: Native Fish Management in the American West. University of Arizona Press

Nei M (1973) Analysis of Gene Diversity in Subdivided Populations. Proc Natl Acad Sci U S A 70:3321–3323. https://doi.org/10.1073/pnas.70.12.3321

Novak BJ, Fraser D, Maloney TH (2020) Transforming Ocean Conservation: Applying the Genetic Rescue Toolkit. Genes 11:209. https://doi.org/10.3390/genes11020209

O'Leary SJ, Puritz JB, Willis SC, et al (2018) These Aren't the Loci You're Looking For: Principles of Effective SNP Filtering for Molecular Ecologists. Molecular Ecology 27:3193–3206. https://doi.org/10.1111/mec.14792

Pavlova A, Beheregaray LB, Coleman R, et al (2017) Severe Consequences of Habitat Fragmentation on Genetic Diversity of an Endangered Australian Freshwater Fish: A Call for Assisted Gene Flow. Evol Appl 10:531–550. https://doi.org/10.1111/eva.12484

Perry WL, Lodge DM, Feder JL (2002) Importance of Hybridization Between Indigenous and Nonindigenous Freshwater Species: An Overlooked Threat to North American Biodiversity. Systematic Biology 51:255–275. https://doi.org/10.1080/10635150252899761

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:e37135. https://doi.org/10.1371/journal.pone.0037135

Pollak E (1983) A New Method for Estimating the Effective Population Size from Allele Frequency Changes. Genetics 104:531–548. https://doi.org/10.1093/genetics/104.3.531

Puritz J, Gold J, Hollenbeck C dDocent: A RADseq, Variant-Calling Pipeline Designed for Population Genomics of Mon-Model Organisms. PeerJ. https://peerj.com/articles/431/

Rosenfield JA, Nolasco S, Lindauer S, et al (2004) The Role of Hybrid Vigor in the Replacement of Pecos Pupfish by It's Hybrids with Sheepshead Minnow. Conservation Biology 18:1589–1598. https://doi.org/10.1111/j.1523-1739.2004.00356.x

Schwartz M, Luikart G, Waples R (2007) Genetic Monitoring as a Promising Tool for Conservation and Management. Trends in Ecology & Evolution 22:25–33. https://doi.org/10.1016/j.tree.2006.08.009

Seutin G, White BN, Boag PT (1991) Preservation of Avian Blood and Tissue Samples for DNA Analyses. Can J Zool 69:82–90. https://doi.org/10.1139/z91-013

Shepta E (2022) Assessing the Live Bait Industry and the Ecological Status of Invasive Cyprinodontiformes (Sheepshead minnow: *Cyprinodon variegatus* and Gulf killifish: *Fundulus grandis*) in Texas Streams. Electronic Theses and Dissertations

Tech C (2006) Postzygotic Incompatibilities Between the Pupfishes, *Cyprinodon elegans* and *Cyprinodon variegatus*: Hybrid Male Sterility and Sex Ratio Bias. Journal of Evolutionary Biology 19:1830–1837. https://doi.org/10.1111/j.1420-9101.2006.01173.x

Todesco M, Pascual MA, Owens GL, et al (2016) Hybridization and Extinction. Evolutionary Applications 9:892–908. https://doi.org/10.1111/eva.12367

Wang J (2014) Marker-Based Estimates of Relatedness and Inbreeding Coefficients: An Assessment of Current Methods. Journal of Evolutionary Biology 27:518–530. https://doi.org/10.1111/jeb.12315

Waples RK, Larson WA, Waples RS (2016) Estimating Contemporary Effective Population Size in Non-Model Species Using Linkage Disequilibrium Across Thousands of Loci. Heredity 117:233–240. https://doi.org/10.1038/hdy.2016.60

Waples RS (1989) A Generalized Approach for Estimating Effective Population Size from Temporal Changes in Allele Frequency. Genetics 121:379–391. https://doi.org/10.1093/genetics/121.2.379

Ward RD, Zemlak TS, Innes BH, et al (2005) DNA Barcoding Australia's Fish Species. Philos Trans R Soc Lond B Biol Sci 360:1847–1857. https://doi.org/10.1098/rstb.2005.1716

Weir BS, Cockerham CC (1984) Estimating F-Statistic for the Analysis of Population Structure. Evolution 38:1358–1370. https://doi.org/10.1111/j.1558-5646.1984.tb05657.x

Whiteley AR, Fitzpatrick SW, Funk WC, Tallmon DA (2015) Genetic Rescue to the Rescue. Trends in Ecology & Evolution 30:42–49. https://doi.org/10.1016/j.tree.2014.10.009

Willis SC, Hollenbeck CM, Puritz JB, et al (2017) Haplotyping RAD Loci: An Efficient Method to Filter Paralogs and Account for Physical Linkage. Molecular Ecology Resources 17:955–965. https://doi.org/10.1111/1755-0998.12647

Figure 2.1. Photo of male *Cyprinodon bovinus* caught from Diamond Y Spring Preserve.

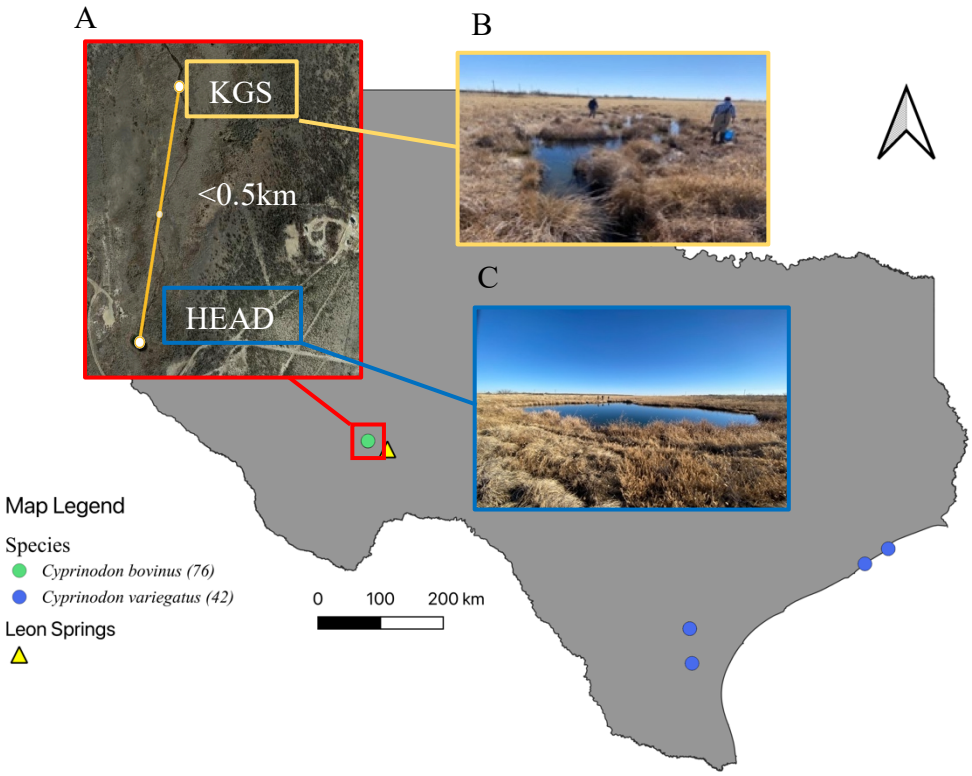

Figure 2.2. Map of Texas sampling locations of both species (*C. bovinus* and *C. variegatus*) with insets of (A) Diamond Y locations, (B) Karges Springs (KGS), (C) Headwater Spring (HEAD), and a triangle to denote the type locality of *C. bovinus*.

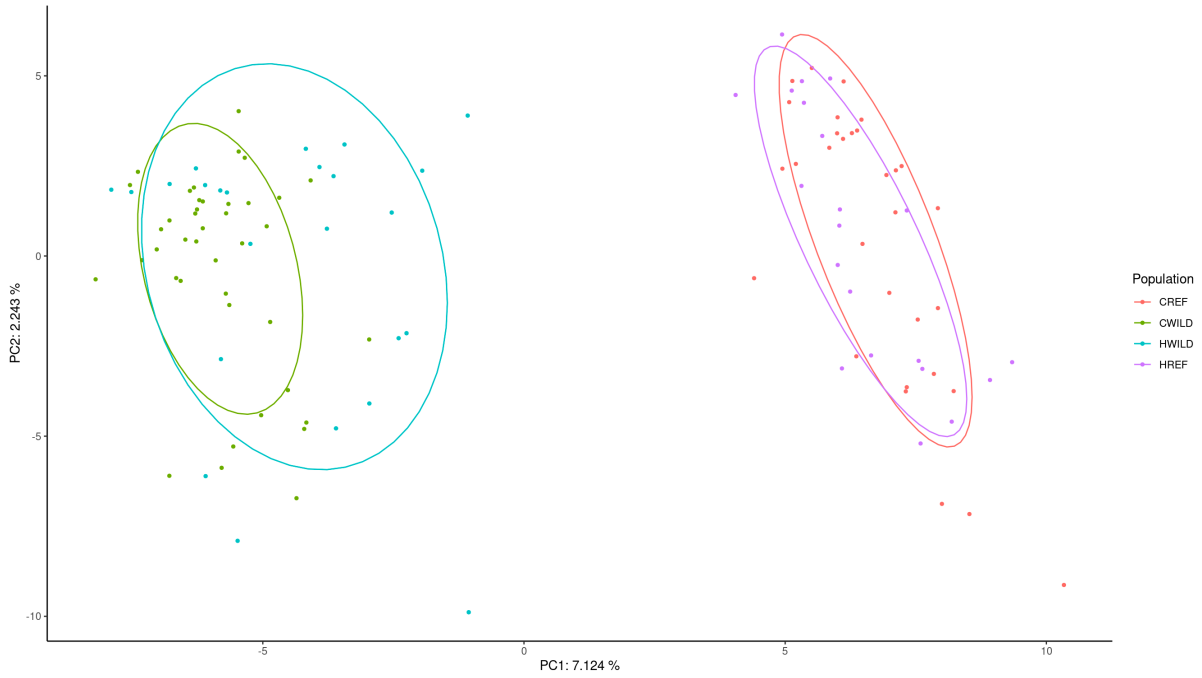

Figure 2.3. Principal component analysis of contemporary wild (CWILD) and refuge (CREF) and historical wild (HWILD) and refuge (HREF) samples of *C. bovinus*.

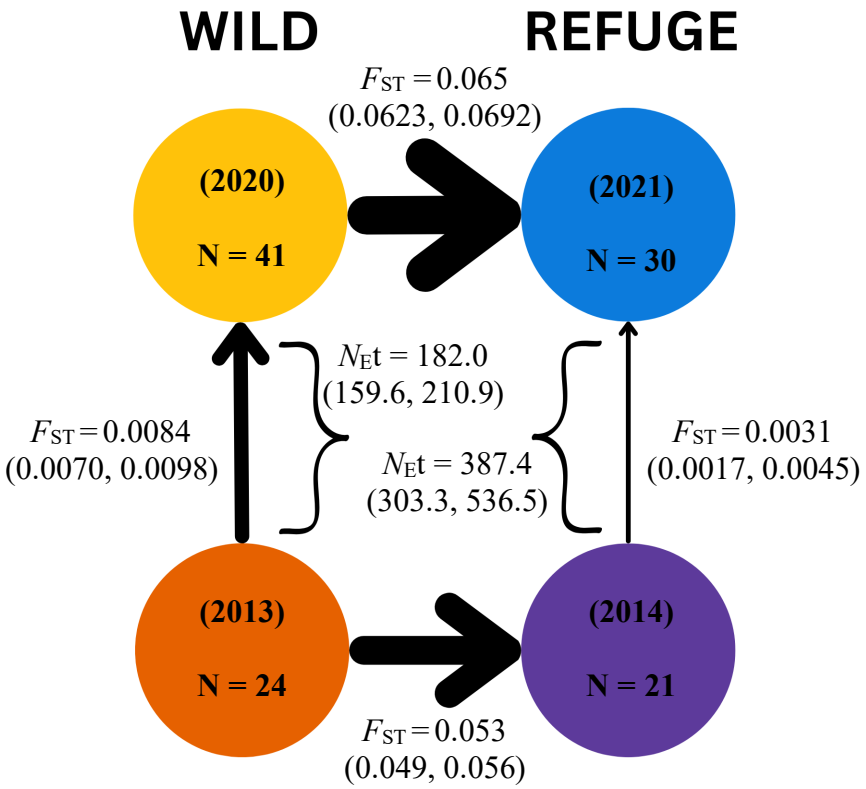

Figure 2.4. Sample sizes (N) of historical and contemporary C. bovinus, with arrows denoting the magnitude of population differentiation (FST) on spatial and temporal scales. Temporal estimates of effective population size (N_Et) are listed between temporal populations and confidence intervals for all estimated metrices are in parenthesis.

Table 2.1. AMOVA results (average over 4,568 loci) for three contemporary *C. bovinus* populations; showing sum of squares (SS), variance components (VC) and percentage of variance (%).

	SS	VC	%
Among locations	5340.949	41.917	5.49
Within locations	100389.276	722.225	94.51
Total	105730.225	764.142	

Table 2.2. F_{ST} estimates of contemporary *C. bovinus* populations with *p*-values above the diagonal, locations are Karges Springs (KGS) and Headwater Pool Spring (HEAD) and the refuge population (REF).

	KGS	HEAD	REF
KGS	-	0.0004	<0.0001
HEAD	0.00645 (0.00584-0.00835)	-	<0.0001
REF	0.0729 (0.0667-0.0732)	0.0658 (0.0602-0.0663)	-

Table 2.3. Effective population (N_E) estimates with lower and upper 95% confidence intervals done using a jackknife method, as well as point estimates for all three contemporary *C. bovinus* populations; locations are Karges Springs (KGS) and Headwater Pool Spring (HEAD) and the refuge population (REF).

	Lower	**Point**	**Upper**
KGS	160	252	559
HEAD	62	126	1701
REF	387	601	1325

Table 2.4. Mean expected heterozygosity (H_E) and mean allelic richness (A_R) by population, locations are Karges Springs (KGS) and Headwater Pool Spring (HEAD) and the refuge population (REF).

	$\boldsymbol{H_E}$	$\boldsymbol{A_R}$
KGS	0.32	2.05
HEAD	0.32	2.08
REF	0.35	2.21

Table 2.5. Post-hoc Wilcoxon test for H_e with test statistic (W) and p-value, locations are Karges Springs (KGS) and Headwater Pool Spring (HEAD) and the refuge population (REF).

	W	**P**
KGS x HEAD	-1.93	**<** 0.053
KGS x REF	-11.51	< 0.05
HEAD x REF	-11.28	< 0.05

Supplemental Information

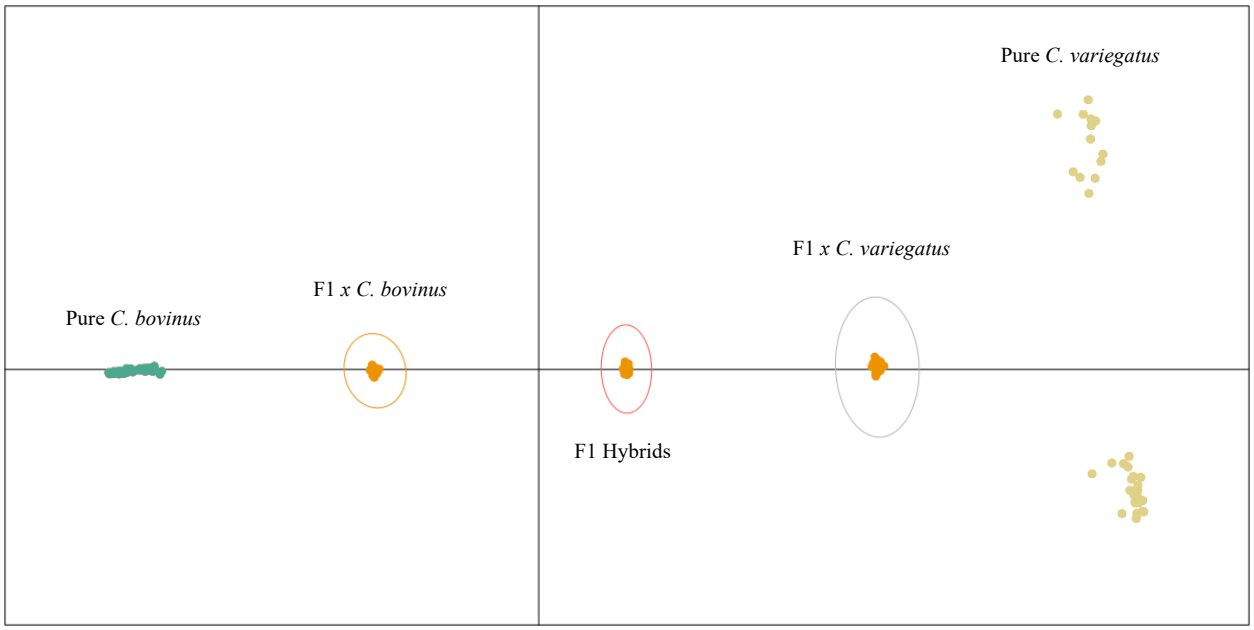

Figure S.2.1. Principal component analysis, including putative pure parent groups of C. bovinus and C. variegatus, 30 simulated F1 hybrids, 30 F1 x C. bovinus backcrosses, and 30 F1 x C. variegatus backcrosses.

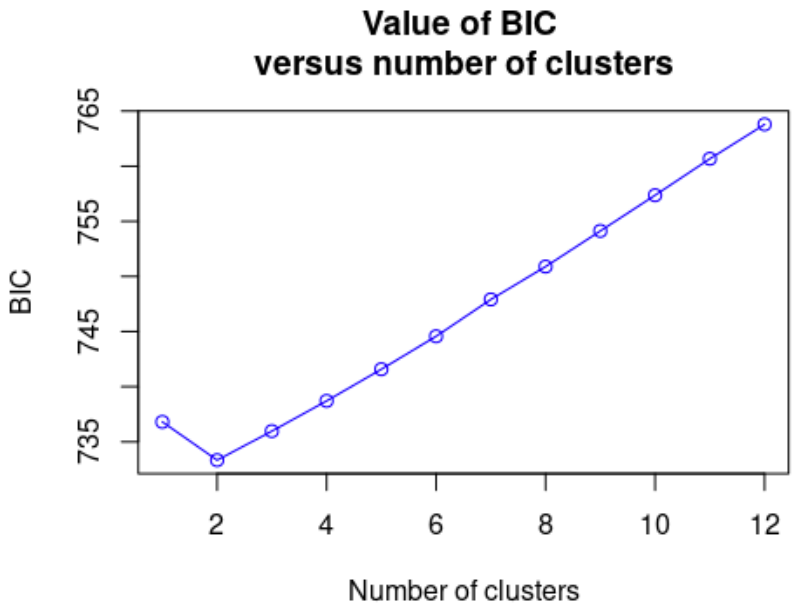

Figure S.2.2. Plot of Bayesian information criteria to inform *K*-means clustering; 2 is the optimized grouping for contemporary *C. bovinus*.

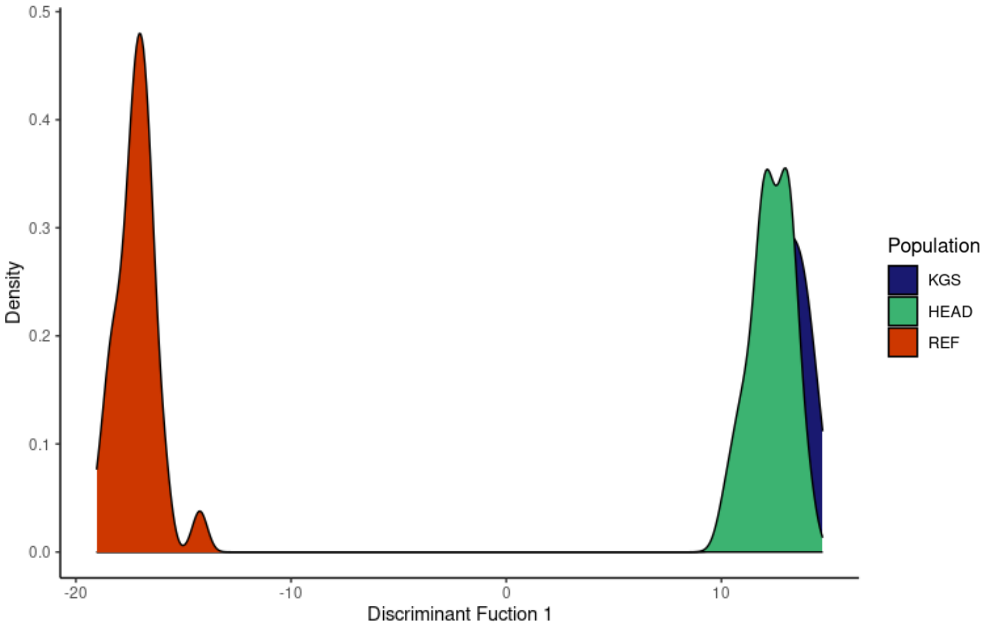

Figure S.2.3. Density plot for *K* =2 using contemporary *C. bovinus* dataset; locations are Karges Springs (KGS) and Headwater Pool Spring (HEAD) and the refuge population (REF).

Table S.2.1. Number of individuals from each species and sampling location randomized evenly across indices for pooling and NGS sequencing to decrease inflating library affects within a species or sampling location.

Species	Location	Total	I2	I4	I7	I10
Cyprinodon bovinus	Karges Spring	20	6	5	5	4
	Headwater Spring	23	5	6	6	6
	Refuge	33	8	8	8	9
Cyprinodon variegatus	Lake Corpus Christi	22	6	5	6	5
	Galveston Wall	8	2	2	2	2
	Tranquitas Creek	4	1	1	1	1
	San Louis Pass	8	2	2	2	2

CHAPTER III

CONSERVATION GENOMIC ASSESSMENT OF TWO GEOGRAPHICALY DISTICNT POPULATOINS OF PECOS GAMBUSIA (*GAMBUSIA NOBILIS*)

Abstract

The Pecos gambusia, *Gambusia nobilis,* is federally listed as endangered and is currently confined to fragmented habitat spread across the Pecos River Drainage in Texas and New Mexico due to several historical range contraction events. To inform conservation and management of the species, a comprehensive conservation population genomics study was performed in geographic localities inside the Diamond Y Spring Preserve, in Texas, and in habitat to the west (West Texas). Analysis including two invasive congener species, *G. affinis* and *G. geiseri*, to assess for evidence of hybridization and introgression. Evidence of hybrids and admixed individuals was found in one site inside Diamond Y, and the others were found in East Sandia and Phantom Lake Springs, respectively. Significant population structure was found within each geographic region (Diamond Y and West Texas) with all wild locations exhibiting significant heterogeneity. Estimates of contemporary genetic diversity was significant within each locality, and significantly different on a pairwise population level. Due to small sample sizes and high proportions of non-polymorphic loci, effective population sizes could not be estimated within the eastern localities, but effective population sizes were below the benchmark for long-term persistence in two of the three western localities.

Introduction

Freshwater fish biodiversity is a vital component of aquatic ecosystems, contributing to the health and balance of these environments (Leidy and Moyle 2021a). However, in arid regions like deserts, the decline of desert fish populations is a concerning trend. Desert fish have adapted

to survive in challenging conditions, often inhabiting isolated desert springs, streams, and rivers (Meffe and Vrijenhoek 1988). These specialized species face numerous threats, including habitat destruction due to water extraction, dams, and urban development, as well as water pollution and the introduction of non-native species (Meffe and Vrijenhoek 1988; Fagan et al. 2002). Because freshwater is a limiting resource in arid regions, freshwater fish generally exhibit small populations and fragmented distributions along their ranges and are susceptible to increased genetic drift (Pavlova et al. 2017b; Leidy and Moyle 2021b). Climate change exacerbates these challenges by altering temperature and flow patterns, resulting in further fragmentation and decreasing the ability of gene flow to counteract losses of genetic diversity caused by drift (Jaeger et al. 2014). The decline of desert fish populations not only threatens the existence of unique and often endemic species, but also disrupts the intricate web of life within desert ecosystems (Jaeger et al. 2014). Conservation efforts and sustainable water management are crucial for mitigating these declines and preserving the biodiversity of desert fishes, which have evolved to thrive in some of the world's harshest environments (Gumm et al. 2011). While management plans are put in place to mitigate these effects on desert fishes, long-term monitoring is crucial to maintain an updated snapshot of biodiversity along a temporal gradient. Conservation genomics can offer valuable insights into levels of genetic diversity within a population, and techniques can be implemented to measure several metrics of genetic diversity, as well as predict the likelihood of persistence of these species in the future (Minckley and Deacon 2017).

The Pecos gambusia, *Gambusia nobilis*, is a small freshwater fish species native to North America primarily found in the Pecos River Basin in New Mexico and Texas. *G. nobilis* is small, with an average length typically ranging from 1.5 to 2.5 inches (3.8 to 6.4 cm; Echelle et al.

1989; Lewis et al. 2013a). Their body is elongated and slender, with a dorsal fin situated close to the tail. Like other members of the genus *Gambusia*, *G. nobilis* exhibits sexual dimorphism with males exhibiting more vibrant coloration, often displaying shades of green, blue, or yellow, with distinct spots or speckles along their sides (Fig. 3.1; A: male *G. nobilis*, B: female *G. nobilis*). A male's anal fin is elongated into a specialized structure known as a gonopodium, which facilitates internal fertilization during reproduction. By contrast, females are generally more subdued in coloration, often appearing olive or brownish, are more rotund, and lack the gonopodium (Hopkins and Kodric-Brown 2015). Gonopodia are highly specialized structures, and within the genus *Gambusia* there is anatomical variation in the gonopodium that may act as a premating barrier between different species (Rodriguez 2017).

Gambusia nobilis is a highly adaptable species, capable of thriving in a variety of aquatic environments (Echelle et al. 1989). It predominantly inhabits slow-moving or stagnant waters such as ponds, pools, marshes, and backwaters, and have shown a remarkable ability to endure fluctuating water conditions and temperature ranges (Lewis et al. 2013b). Their natural range is concentrated in the Pecos River Basin, primarily in New Mexico and Texas, but they have also been introduced to various other locales in the United States such as a manmade Cienega in Balmorhea State Park (Hopkins and Kodric-Brown 2015). Gambusia are known for their beneficial ecological role as voracious consumers of mosquito larvae. They actively feed on these larvae, contributing to the control of mosquito populations, and consequently mitigating the risk of mosquito-borne diseases. *G. nobilis* are often observed foraging near the water's surface, making them an essential component of the ecosystems they inhabit (Hopkins and Kodric-Brown 2015).

Gambusia nobilis is a species of conservation concern due to habitat loss, competition with non-native species, and other environmental challenges (Echelle et al. 1989). In certain areas, it has been classified as a threatened species, warranting conservation efforts to safeguard its populations and restore its natural habitats. In 1987, *G. nobilis* was listed as endangered by USFW, and protected under the Endangered Species Act. Conservation initiatives are aimed at preserving this species not only for its unique ecological function but also as an emblematic representative of North America's freshwater biodiversity. Past studies focused on fine scale population distribution throughout the southwestern United States, but no studies used genomics to assess for the presence of hybridization and introgression in this species with invasive congeners, nor contemporary genetic diversity within and between discrete populations (Echelle et al. 1989; Hubbs et al. 2002; Rodriguez 2017). Thus, this study used discrete sampling of *G. nobilis* from across its known habitat in Texas to assess for contemporary patterns of genetic diversity within and among geographic locations. Because invasive congeners *G. affinis* and *G. geiseri* co-occur with *G. nobilis* throughout their range, both species were sampled inside and outside of *G. nobilis'* range to assess for contemporary hybridization and historical introgression.

Methods

Field Sampling

Fin clips were collected from 30 to 40 *G. nobilis* at Diamond Y from three discrete sampling locations identified in advance where *G. nobilis* were known to be present. A fourth site was included in planning, but upon arrival it had unexpectedly desiccated, leaving no habitat to sample. Thus, *G. nobilis* was sampled at DY from Headwater Spring (HEAD), Karges Spring (KGS), and Euphrasia Spring (EU), which are not directly connected, though HEAD and KGS are only separated by ~0.5 km and HEAD and EU are separated by ~4.5 km (Fig. 3.2).

Gambusia nobilis was also samples at Balmorhea State Park (CHS), Phantom Lake (PHL), and East Sandia (ES). Fin clips were also collected from *G. affinis* and *G. geiseri* at these locations when encountered. All other *G. affinis* and *G. geiseri* were sampled from locations outside of West Texas, with six locations for *G affinis*, (San Fernando Creek (SFR), Placedo Creek (PC), Perdido Creek (PEC), Lake Corpus Christi State Park (LCCSP), Aransas River (AR), and Galveston Seawall Pools (GSWPLS)), and one location for *G. geiseri* (San Marcos River (SMR)). All tissues were immersed in 20% salt-saturated DMSO buffer (Seutin et al. 1991) upon collection in the field and stored at room temperature until time of extraction. Tissues were collected from DY in March of 2020, while all other tissues were collected from September 2020 – September 2021. Fifty whole *G. nobilis* were collected from PHL in 2017 and provided to this project by Texas Parks and Wildlife (TPWD), as well as samples collected from PHL in 2023 after the water receded back into the cave. To capture individuals, a seine net was used when sites were large enough and dip nets at all other times. A sterile razor blade was used to remove the upper 25% of the caudal fin from each captured fish and all tools were cleaned using de-ionized water and 10% bleach solutions between fish. After handling, individuals were held in a bucket filled with water from their habitat, which was oxygenated using a bubbler, and placed in the shade to ensure survival after release. Three whole *G. nobilis* were euthanized during the collection in March 2020, preserved in 95% ethanol and are stored as voucher specimens in the Biodiversity, Research, and Teaching Collection at Texas A&M University in College Station. Whole individuals of *G. affinis and G. geiseri* were collected using dip nets, euthanized using clove oil, and preserved in 95% ethanol. Fin clips were later taken in the lab and immersed in 20% salt-saturated DMSO until time of extraction, while the rest of the tissue samples were stored at -20ºF in 95% ethanol at the Marine Genomics Lab at Texas A&M University-Corpus

Christi. The distribution of sampling sites is shown in Figure 3.2, including the total count of individuals sampled at each location that were included in library prep. Sampling for *G. nobilis*, *G. affinis*, and *G. geiseri* within Pecos County was conducted with TPWD and biologists from The Nature Conservancy under permit number TE814933. *G. affinis* and *G. geiseri* sampled outside of Pecos County were collected under permit number SPR-0614-111. All animal use and care followed IACUC animal protocol TAMU IACUC 2021-0001.

Genomic Library Preparation and Sequencing

Genomic DNA was extracted from fin clips using MagBind Blood and Tissue HDQ DNA extraction kits (Omega Bio-tek). Extractions were then electrophoresed through 1% agarose gel and quantified with AccuBlue High Sensitivity dsDNA Quantitation Kits (Biotium). Reduced representation libraries were assembled following a modified version of the double-digestion restriction-site associated DNA sequencing (ddRADseq) protocol of (Peterson et al. 2012). Briefly, libraries were assembled using ~500 ng of genomic DNA which was digested using restriction enzymes *Eco*RI and *Msp*I (New England Biolabs). After digestion, DNA was purified using Mag-Bind TotalPure NGS (Omega Bio-tek), quantified, and standardized to 100ng/μL. For each digestion one of 48 barcoded adapters was ligated to the *Eco*RI site and a common adaptor was ligated to the *Msp*I site. Ligation reactions were then PCR tested using a two-step PCR protocol (amplification for 18 & 32 cycles) and checked using gel electrophoresis. Successful ligation reactions were pooled into one of four indexed libraries (Table S.2.1), purified using PEG, and quantified again. Fragments were size selected between 313-437bp using Pippin Prep (Sage Science) and PCR amplified for 14 cycles to incorporate P2 adapters containing index sequences. Libraries were then purified one more time and quantified. To

ensure proper size selection, pooled fragments were run on a fragment analyzer. Libraries were paired end sequenced on a single lane of an Illumina HiSeq 4000 (2x150bp).

COI Sequencing

A subset of individuals from all three species were selected haphazardly for sequencing of the mitochondrially encoded (mtDNA) cytochrome oxidase subunit I (COI). DNA was extracted using MagBind Blood and Tissue HDQ DNA extraction kits (Omega Bio-tek) and a 698 bp fragment of CO1 was PCR amplified using universal F1/F2 & R1/R2 fish primers (Ward et al. 2005). PCR reactions (30 µL) contained 5 x GoTaq buffer, 1.5 mM magnesium chloride, 1% Tween, 2.5 mM dNTPs each, 0.25 mM of each primer, 0.03 units of *Taq* polymerase, 1 µL of template DNA, and water. Cycling was performed as follows: denaturation at 95 ºC for 2 minutes, followed by 35 cycles of denaturation at 95 ºC for 60 seconds, annealing at 52 ºC for 60 seconds, and extension at 72 ºC for 90 seconds, a final extension was done at 72 ºC for 10 minutes. Amplicons were purified using Mag-Bind TotalPure NGS. To prepare the samples for sequencing, amplicons were standardized to 10ng/µL in 10 µL of water and Sanger on an ABI 3730 XL DNA Analyzer sequenced at the Core Lab at Texas A&M University-Corpus Christi. Raw Sanger sequences were quality trimmed and edited by using TRIMMOMATIC v8.22 (Free Software Foundation, Inc.) and compared in bulk to the NCBI nucleotide collection using BLASTn (Altschul et al. 1990) to obtain individual species IDs.

Bioinformatics and Filtering

For downstream analysis, four datasets were produced. One dataset was created for hybrid analysis and contained all *G. affinis*, *G. geiseri*, and *G. nobilis*. The second dataset included only *G. nobilis*. The third data set contained all *G. nobilis* individuals collected from the western most region of Texas, including sampling sites ES, CHS, and PHL. Finally, the fourth

dataset included only *G. nobilis* individuals there were collected from DY (eastern Texas), including sampling sites KGS, HEAD, and EU. The former three data sets were used to assess patterns of contemporary genetic structure.

To create the four datasets, raw Illumina reads for *G. affinis*, *G. geiseri*, and *G. nobilis* were demultiplexed using a custom Perl script and processed using the DDOCENT v2.9.1 pipeline (Puritz et al.). For each dataset, appropriate demultiplexed sequences were quality trimmed and stacked into orthologous scaffolds and made into a de novo reference. The reference was then optimized for mapping using custom BASH scripts. Quality trimmed sequences were then mapped back onto the reference and genomic variants scored and compiled into a variant calling factor file (VCF). Variants were then quality filtered using a combination of VCFTOOLS (Danecek et al. 2011) and various BASH and Perl scripts. Filtering followed guidance of (O'Leary et al. 2018) to remove low quality or artefactual SNPs as well as potentially paralogs and low-quality individuals. Genotypes with quality < 15 and < 10 reads were coded as missing, retaining loci with quality > 15, genotype call rate > 50%, and mean depth 10. Additionally, loci were filtered for allele balance, mapping quality ratios, strand balance, paired status, depth/quality ratio, and heterozygosity. Individuals with > 20% missing data were removed. Using a custom Perl script, SNP variants found in the same contig were phased into microhaplotypes (hear after loci) following (Willis et al. 2017). Pair-wise relatedness (Wang 2014) was used to screen for duplicate pairs and identify potentially related individuals. To test for putative loci under selection the FDIST (method was implemented in ARLEQUIN v3.5.2.2 (Excoffier and Lischer 2010), using 20,000 coalescent simulations and an island model. A Bayesian approach for outlier detection was also implemented, using BAYESCAN (Foll and Gaggiotti 2008) with 30 pilot runs of 5,000 iterations, flowed by a burn in of 50,000 iterations

and 500,000 iterations sampled 5,000 times, , and a *q*-value of 0.05. While outlier loci provide evidence for localized adaptation, they may cause bias in neutral data sets (Fagan et al. 2002). Therefore, outliers were removed from downstream analyses because the focus of this study is understanding patterns of population structure and drift (Holderegger et al. 2006).

Hybrid Analysis

Using the hybrid dataset, diagnostics loci were identified for pure *G. affinis*, *G. geiseri*, and *G. nobilis*. NEWHYBRIDS v2.0 (Anderson and Thompson 2002) was then used to assign individuals back to the parent species or one of four hybrid classes (F1 *G. nobilis* x *G. affinis*, F1 x *G. nobilis* x *G. geiseri*, F1 *G. nobilis* x *G. affinis* backcrosses, and F1 x *G. nobilis* x *G. geiseri* backcrosses). The program uses a Bayesian clustering algorithm that maximizes the posterior probability of each individual's assignment, using an estimated proportion of hybrid individuals and the assumed genetic model. A secondary approach for hybrid detection was implemented in *Adegenet* v2.5.1 (Jombart 2008) in R. For this approach, 30 F1 *G. nobilis* x *G. affinis*, F1 x *G. nobilis* x *G. geiseri*, F1 *G. nobilis* x *G. affinis* backcrosses, and F1 x *G. nobilis* x *G. geiseri* backcrosses were simulated. Hybrid individuals and empirical individuals were then plotted using PCA to see how empirical samples grouped with simulated individuals. Finally, species identities based on COI were compared to species identities based on nuclear data to validate hybrid analyses and assesses for historical mtDNA introgression.

Population Genetic Analyses

For the full *G. nobilis* dataset, a hierarchical AMOVA was conducted in ARLEQUIN v3.5.2.2 using a locus by locus framework to account for uneven levels of missing data across loci (Weir and Cockerham 1984) and PCA was implemented. For the contemporary eastern and western *G. nobilis* datasets, a single-level analysis of molecular variance (AMOVA) was carried

out in ARLEQUIN, using the same locus by locus framework as the hierarchical AMOVA above. For each AMOVA, significance was determined at an α-level of 0.05 by permuting individuals among locations 10,000 times and 95% confidence intervals determined using 20,000 bootstrap replicates. Post-hoc pairwise F_{ST} was then estimated in ARLEQUIN with significance determined as above, and 95% confidence intervals calculated in *hierfstat* (Goudet 2005) in R. For all subsequent datasets, the following clustering methodology was used: to assess the number of genetic groups present in the dataset, discriminant analysis of principal components (DAPC; Jombart et al. 2010) was implemented in *adegenet* using *K*-means clustering (K = 1-6) with the number of clusters selected by comparing Bayesian information criterion values (BIC). The optimal number of PCs to retain was then determined using cross validation and membership probabilities for each individual to the inferred clusters were calculated. Effective population size (N_E) was estimated for each location using the linkage disequilibrium approach with an allele frequency cutoff of 0.1 and 95% confidence intervals determined using 1,000 jackknife replicates, as implemented in NEESTIMATOR v2.1 (Do et al. 2014). Mean expected heterozygosity (H_E; Nei 1973), and rarefied allelic richness (A_R; El Mousadik and Petit 1996), were estimating using *hierfstat*. Friedman's tests were conducted on both metrics to test for homogeneity among locations, and post-hoc Wilcoxon tests conducted to assess pairwise differences using the R package *coin* v.1.3.1 (Hothorn et al. 2008).

Results

Hybrid Detection

After filtering, the hybrid dataset contained 195 individuals genotyped at 4,132 loci, with an average of 5.2 alleles per locus. Six hybrids were detected using NEWHYBRIDS and the same six individuals grouped with simulated hybrids or backcrosses using PCA (Fig. S3.3). Of the six

putative hybrid samples, one individual was identified as a F1 hybrid and the other five were the result of backcrossing between an F1 hybrid and one of the parental species. A single individual, which resulted from a backcross between *G. affinis* and an F1 hybrid, was collected in Headwater Springs at Diamond Y, and two individuals which resulted from a backcross between *G. geiseri* and an F1 hybrid were collected at East Sandia. The remaining admixed individuals were sampled inside the cave at Phantom Lake Springs after the water outside the cave had receded back into the cave. All admixed individuals were removed from further population genetic analyses.

After filtering, the *G. nobilis* data contained 141 individuals genotyped at 3,384 loci, with an average of 2.6 alleles per locus. The samples were put into two groups, based on proximity, for AMOVA, with the eastern group containing all three samples collected at DY and a western group containing CHS, ES and PHL. Components of genetic variation attributable to differences among groups and samples within were highly significant (among groups: %V = 43.27, $P < 0.0001$; within groups: %V = 48.18, $P < 0.0001$; Table 3.1). A principal components analysis showed significant differentiation of the eastern and western groups, with the first principal component separating the samples between groups, and the second principal component separating the samples within groups (Fig. 3.4).

Population Genetic Analyses of Eastern Localities

After filtering, the contemporary eastern *G. nobilis* data contained 45 individuals genotyped across 2,813 loci, with an average of 1.47 alleles per locus. No relatedness values were above 0.9, leading to no removal of related individuals. The component of genetic variation attributable to differences among samples was highly significant, but more than an order of magnitude smaller than the amount of variation seen among groups in the previous analysis

(among samples: %V = 1.84, $P < 0.0001$, Table 3.1). Post-hoc estimates of pairwise F_{ST} were significant between all locations and smallest between the middle reach (KGS) and the lower reach (EU) (Table 3.2). The optimum BIC value was obtained for K=1 (Fig. S3.1), but cross-validation exhibited a 100% membership probability for K >= 3. Concordant with geographic isolation and significant differentiation, a K = 3 was used (Fig. S.3.2). Estimates of H_E were 0.216, 0.11, 0.218, and estimates of A_R were 1.91, 1.11, 1.98, for KGS, HEAD, and EU, respectively (Table 3.4). Friedman's tests for within population diversity were significant for both H_E and A_R (H_E: $X^2(2) = 22.525$, p-value = 1.29×10^{-5}, A_R: $X^2(2) = 22.976$, p-value = 1.03×10^{-5}). Post-hoc Wilcoxon tests were significant for both H_E or A_R for all pairwise comparisons between populations. Estimates of effective population size were not estimable with this dataset.

Population Genetic Analyses of Western Locality

After filtering, the contemporary western *G. nobilis* data contained 65 individuals genotyped across 5,416 loci, with an average of 2.28 alleles per locus. No pairs of individuals had a relatedness value above 0.9, leading to no removal of related individuals. The component of genetic variation attributable to differences among samples was highly significant, and larger than what was seen among eastern locations (among samples: %V = 18.7, $P < 0.0001$; Table 3.5). Post-hoc estimate of pairwise F_{ST} were significant between all samples, and smaller between CHS and ES relative to comparisons involving PHL (Table 3.6). The optimum BIC value was obtained for K=3 (Fig. S3.3) with 100% of individuals assigned back to their group of origin concordant with geographic isolation (Fig. S.3.4). Estimates of H_E were 0.23, 0.23, 0.18, and estimates of A_R were 1.65, 1.64, 1.51for CHS, ES, and PHL, respectively (Table 3.8). Friedman's tests for within population diversity were significant for both H_E and A_R (H_E: $X^2(2) = 22.976$, p-value = 1.03×10^{-5}, A_R: $X^2(2) = 768.4$, p-value = 2.2×10^{-16}). Post-hoc Wilcoxon tests

were significant for both H_E or A_R between all pairwise population comparisons. Estimated N_E for CHS, ES, and PHL were 470, 956, and 2.3 respectively (Table 3.7).

Discussion

Evidence of contemporary hybridization was detected in this study, and individual species identities inferred from COI and nuclear data corresponded in all but two cases. These results suggest that hybridization is not only feasible between the closely related *Gambusia* congeners and *G. nobilis*, but that there is evidence of historical mtDNA introgression of *G. affinis* mtDNA into *G. nobilis* (Stepien et al. 2019). We can also infer that there is evidence of direction introgression between *G. nobilis* and both invasive congeners due to all five admixed individuals grouping with the invasive parent species during the hybrid analysis. PHL had the largest number of admixed individuals (one F1 and two backcrosses with *G. nobilis*), but all of these individuals were collected after reduction in habitat which increased the likelihood of interspecific breeding (Hasselman et al. 2014). The remaining admixed individuals were found at low frequency in Diamond Y and East Sandia. It is important to note that the level of interactions between potentially hybridizing species is dynamic and continued monitoring will likely be necessary (Perry et al. 2002; Todesco et al. 2016). Furthermore, *G. affinis* and *G. geiseri* for this project were sampled at locations far from DY, CHS, ES, and PHL to ensure that pure *Gambusia* congener mtDNA was used to look for introgression. While the results of mtDNA analysis are clear, there could be undetected components of *G. geiseri* nuclear variation specific to populations in West Texas present in *G. nobilis*. However, this seems unlikely due to the robust sampling regime of this project (Swenton and Kodric-Brown 2012).

The western locality had higher estimated diversity measurements (expected heterozygosity) and significant differences in genetic diversity among populations. Of the three

populations, CHS and ES were the most diverse, suggesting greater potential for population persistence (Frankham et al. 2014a). PHL had the greatest number of admixed individuals and the lowest diversity of the western groups. The severely low expected heterozygosity of PHL in comparison to the other populations in the western locality confirms the reduction in habitat lead to a reduction in the effective and realized population size for *G. nobilis* (Balloux 2004). This population is likely to be extirpated in the near future if habitat outside the mouth of the cave is not restored. It would be beneficial to consider the remaining locations as separate populations when enacting management plans in the future as the habitats are very different and spread apart geographically. Without previous studies, or a temporal comparison, it is difficult to surmise how the populations of *G. nobilis* in West Texas has changed over time, but contemporary diversity and fine-scale population structure suggest that fluxes in habitat availability and quality has affected the contemporary state of this species. As management is updated for this species, it may be advisable to prioritize small-scale habitat restoration as time moves forward.

The eastern localities had lower estimated genetic diversity (expected heterozygosity) compared the western localities (except for PHL), but overall had a higher allelic richness and significant differences among populations. Of the three populations, the middle and lower reaches (KGS and EU respectively) were more diverse than HEAD. From the among population diversity, it can also be assumed that small realized population sizes, non-random mating, or a non-ideal operating sex-ratio might be contributing to low diversity and Inf effective population sizes, or furthermore, that low number of breeding individuals and recent cessation in gene flow might be the cause of the disparity between diversity metrics and infinite N_E estimates. Also, without knowledge of when gene flow halted between the upper, middle, and lower reaches of DY it is hard to interpret if low diversity and infinite effective populations sizes are due to a lack

in diversity or non-equilibrium processes. Nonetheless, the results of this study shows that the eastern groups in DY exhibit low estimated diversity and low sample sizes evident of populations that are at risk of collapse.

To summarize, the results suggest low levels of gene flow between *G. nobilis* and *G. geiseri* and *G. affinis* in the wild, and that continued monitoring may be necessary to ensure that *G. nobilis* does not suffer from a decrease in local adaptation due to high levels of hybridization with invasive species, otherwise known as genetic swamping, in the future (Perry et al. 2002). Population genetic analyses revealed deep divergence between Diamond Y and West Texas, with smaller levels of divergence between samples within the groups. This indicates a lack of gene flow between eastern and western groups over long periods of time, but more recent isolation between sites within the groups. Confidence estimates of N_E at CHS and ES were between 500-1000, suggesting maintenance of adaptive variation may be more of a long term than short term problem (Holderegger et al. 2006). PHL on the other hand had an estimate of N_E of two; however, the spring-fed water body has receded entirely into the mouth of the cave and what few *Gambusia* (of any species) remained will likely be extirpated in the near future. Estimates of N_E for the eastern sites were infinite, but the numbers of individuals sampled at these sites (particularly HEAD) were low relative to the western sites as was genetic diversity (expected heterozygosity), so it is not possible to tell if the results indicate larger numbers of breeding individuals in DY.

References

Altschul SF, Gish W, Miller W, et al (1990) Basic Local Alignment Search Tool. Journal of Molecular Biology 215:403–410. https://doi.org/10.1016/S0022-2836(05)80360-2

Anderson EC, Thompson EA (2002) A Model-Based Method for Identifying Species Hybrids Using Multilocus Genetic Data. Genetics 160:1217–1229

Balloux F (2004) Heterozygote Excess in Small Populations and the Heterozygote-Excess Effective Population Size. Evolution 58:1891–1900. https://doi.org/10.1111/j.0014-3820.2004.tb00477.x

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

Do C, Waples RS, Peel D, et al (2014) NeEstimator v2: Re-Implementation of Software for the Estimation of Contemporary Effective Population Size (N_E) from Genetic Data. Mol Ecol Resour 14:209–214. https://doi.org/10.1111/1755-0998.12157

Echelle AF, Echelle AA, Edds DR (1989) Conservation Genetics of a Spring-Dwelling Desert Fish, the Pecos Gambusia (*Gambusia nobilis*, Poeciliidae). Conservation Biology 3:159–169

El Mousadik A, Petit RJ (1996) High Level of Genetic Differentiation for Allelic Richness Among Populations of the Argan Tree [*Argania spinosa* (L.) Skeels] Endemic to Morocco. Theoret Appl Genetics 92:832–839. https://doi.org/10.1007/BF00221895

Excoffier L, Lischer HEL (2010) Arlequin Suite Ver 3.5: A New Series of Programs to Perform Population Genetics Analyses Under Linux and Windows. Mol Ecol Resour 10:564–567. https://doi.org/10.1111/j.1755-0998.2010.02847.x

Fagan WF, Unmack PJ, Burgess C, Minckley WL (2002) Rarity, Fragmentation, and Extinction Risk in Desert Fishes. Ecology 83:3250–3256. https://doi.org/10.1890/0012-9658(2002)083[3250:RFAERI]2.0.CO;2

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. https://doi.org/10.1534/genetics.108.092221

Frankham R, Bradshaw CJA, Brook BW (2014) Genetics in Conservation Management: Revised Recommendations for the 50/500 Rules, Red List Criteria and Population Viability Analyses. Biological Conservation 170:56–63. https://doi.org/10.1016/j.biocon.2013.12.036

Goudet J (2005) Hierfstat, A Package for R to Compute and Test Hierarchical F-Statistics. Molecular Ecology Notes 5:184–186. https://doi.org/10.1111/j.1471-8286.2004.00828.x

Gumm JM, Snekser JL, Leese JM, et al (2011) Management of Interactions Between Endangered Species Using Habitat Restoration. Biological Conservation 144:2171–2176. https://doi.org/10.1016/j.biocon.2011.05.006

Hasselman DJ, Argo EE, McBride MC, et al (2014) Human Disturbance Causes the Formation of a Hybrid Swarm Between Two Naturally Sympatric Fish Species. Molecular Ecology 23:1137–1152. https://doi.org/10.1111/mec.12674

Holderegger R, Kamm U, Gugerli F (2006) Adaptive vs. Neutral Genetic Diversity: Implications for Landscape Genetics. Landscape Ecol 21:797–807. https://doi.org/10.1007/s10980-005-5245-9

Hopkins A, Kodric-Brown A (2015) Life History of *Gambusia nobilis* (Pecos Gambusia) from Bitter Lake National Wildlife Refuge. Environ Biol Fish 98:1833–1844. https://doi.org/10.1007/s10641-015-0401-9

Hothorn T, Bretz F, Westfall P (2008) Simultaneous Inference in General Parametric Models. Biometrical Journal 50:346–363. https://doi.org/10.1002/bimj.200810425

Hubbs C, Edwards RJ, Garrett GP (2002) Threatened Fishes of the World: *Gambusia nobilis* Baird & Girard, 1853 (Poeciliidae). Environmental Biology of Fishes 64:428–428. https://doi.org/10.1023/A:1016154429254

Jaeger KL, Olden JD, Pelland NA (2014) Climate Change Poised to Threaten Hydrologic Connectivity and Endemic Fishes in Dryland Streams. Proceedings of the National Academy of Sciences 111:13894–13899. https://doi.org/10.1073/pnas.1320890111

Jombart T (2008) Adegenet: A R Package for the Multivariate Analysis of Genetic Markers. Bioinformatics 24:1403–1405. https://doi.org/10.1093/bioinformatics/btn129

Jombart T, Devillard S, Balloux F (2010) Discriminant Analysis of Principal Components: A New Method for the Analysis of Genetically Structured Populations. BMC Genetics 11:94. https://doi.org/10.1186/1471-2156-11-94

Leidy RA, Moyle PB (2021a) Keeping Up with the Status of Freshwater Fishes: A California (USA) Perspective. Conservation Science and Practice 3:e474. https://doi.org/10.1111/csp2.474

Leidy RA, Moyle PB (2021B) Keeping Up with the Status of Freshwater Fishes: A California (USA) Perspective. Conservation Science and Practice 3:e474. https://doi.org/10.1111/csp2.474

Lewis RH, Allan NL, Stoops SB, et al (2013) Status of the Endangered Pecos Gambusia (*Gambusia nobilis*) and Comanche Springs Pupfish (*Cyprinodon elegans*) in Phantom Lake Spring, Texas. The Southwestern Naturalist 58:234–238

Lewis RH, Allan NL, Stoops SB, et al (2013b) Status of the Endangered Pecos Gambusia (*Gambusia nobilis*) and Comanche Springs Pupfish (*Cyprinodon elegans*) in Phantom Lake Spring, Texas. swna 58:234–238. https://doi.org/10.1894/0038-4909-58.2.234

Meffe GK, Vrijenhoek RC (1988) Conservation Genetics in the Management of Desert Fishes. Conservation Biology 2:157–169. https://doi.org/10.1111/j.1523-1739.1988.tb00167.x

Minckley WL, Deacon JE (2017) Battle Against Extinction: Native Fish Management in the American West. University of Arizona Press

Nei M (1973) Analysis of Gene Diversity in Subdivided Populations. Proc Natl Acad Sci U S A 70:3321–3323. https://doi.org/10.1073/pnas.70.12.3321

O'Leary SJ, Puritz JB, Willis SC, et al (2018) These Aren't the Loci You're Looking For: Principles of Effective SNP Filtering for Molecular Ecologists. Molecular Ecology 27:3193–3206. https://doi.org/10.1111/mec.14792

Pavlova A, Beheregaray LB, Coleman R, et al (2017) Severe Consequences of Habitat Fragmentation on Genetic Diversity of an Endangered Australian Freshwater Fish: A Call for Assisted Gene Flow. Evol Appl 10:531–550. https://doi.org/10.1111/eva.12484

Perry WL, Lodge DM, Feder JL (2002) Importance of Hybridization Between Indigenous and Nonindigenous Freshwater Species: An Overlooked Threat to North American Biodiversity. Systematic Biology 51:255–275. https://doi.org/10.1080/10635150252899761

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:e37135. https://doi.org/10.1371/journal.pone.0037135

Puritz J, Gold J, Hollenbeck C dDocent: A RADseq, Variant-Calling Pipeline Designed for Population Genomics of Mon-Model Organisms. PeerJ. https://peerj.com/articles/431/

Rodriguez V (2017) The Hybridization Between the Endangered *Gambusia nobilis* and Introduced *Gambusia geiseri* in Texas. Thesis

Seutin G, White BN, Boag PT (1991) Preservation of Avian Blood and Tissue Samples for DNA Analyses. Can J Zool 69:82–90. https://doi.org/10.1139/z91-013

Stepien CA, Snyder MR, Elz AE (2019) Invasion Genetics of the Silver Carp *Hypophthalmichthys molitrix* Across North America: Differentiation of Fronts, Introgression, and eDNA Metabarcode Detection. PLOS ONE 14:e0203012. https://doi.org/10.1371/journal.pone.0203012

Swenton DM, Kodric-Brown A (2012) Habitat and Life History Differences Between Two Species of Gambusia. Environ Biol Fish 94:669–680. https://doi.org/10.1007/s10641-011-9973-1

Todesco M, Pascual MA, Owens GL, et al (2016) Hybridization and Extinction. Evolutionary Applications 9:892–908. https://doi.org/10.1111/eva.12367

Wang J (2014) Marker-Based Estimates of Relatedness and Inbreeding Coefficients: An Assessment of Current Methods. Journal of Evolutionary Biology 27:518–530. https://doi.org/10.1111/jeb.12315

Ward RD, Zemlak TS, Innes BH, et al (2005) DNA Barcoding Australia's Fish Species. Philos Trans R Soc Lond B Biol Sci 360:1847–1857. https://doi.org/10.1098/rstb.2005.1716

Weir BS, Cockerham CC (1984) Estimating F-Statistic for the Analysis of Population Structure. Evolution 38:1358–1370. https://doi.org/10.1111/j.1558-5646.1984.tb05657.x

Willis SC, Hollenbeck CM, Puritz JB, et al (2017) Haplotyping RAD Loci: An Efficient Method to Filter Paralogs and Account for Physical Linkage. Molecular Ecology Resources 17:955–965. https://doi.org/10.1111/1755-0998.12647

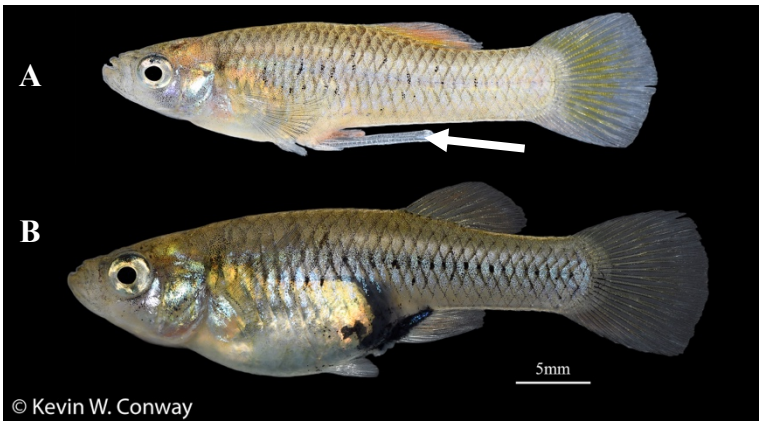

Figure 3.1. Photos of (A) male *Gambusia nobilis* and (B) female *Gambusia nobilis* (with an arrow pointing to the male gonopodium) collected from Diamond Y Spring Preserve.

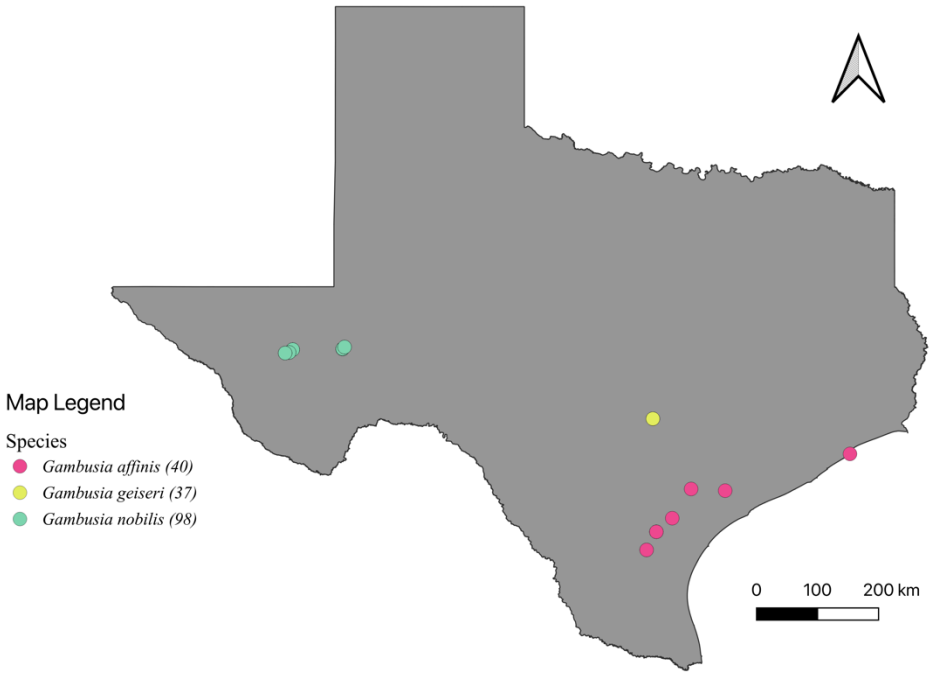

Figure 3.2. Map of Texas sampling locations of three species (*G. affinis, G. geiseri,* and *G. nobilis*).

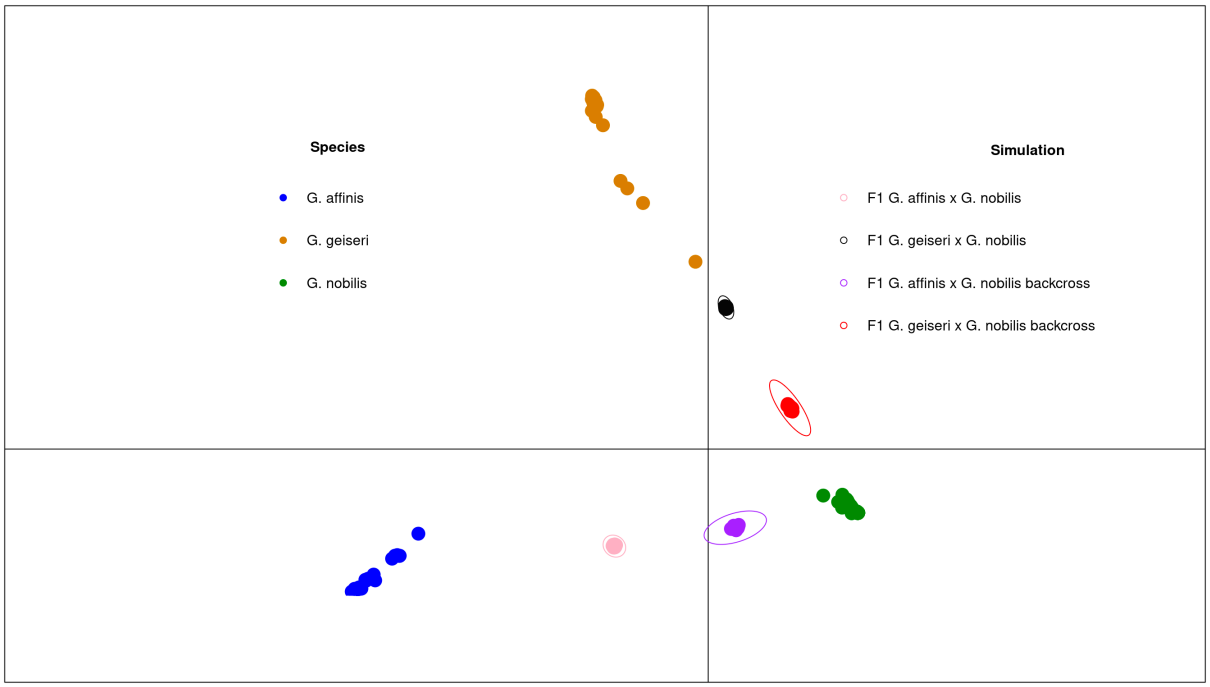

Figure 3.3. Principal components analysis, including putative pure parent groups of *G. affinis, G. geiseri, and G. nobilis*, 30 F1 *G. nobilis* x *G. affinis*, F1 x *G. nobilis* x *G. geiseri*, F1 *G. nobilis* x *G. affinis* backcrosses, and F1 x *G. nobilis* x *G. geiseri* backcrosses.

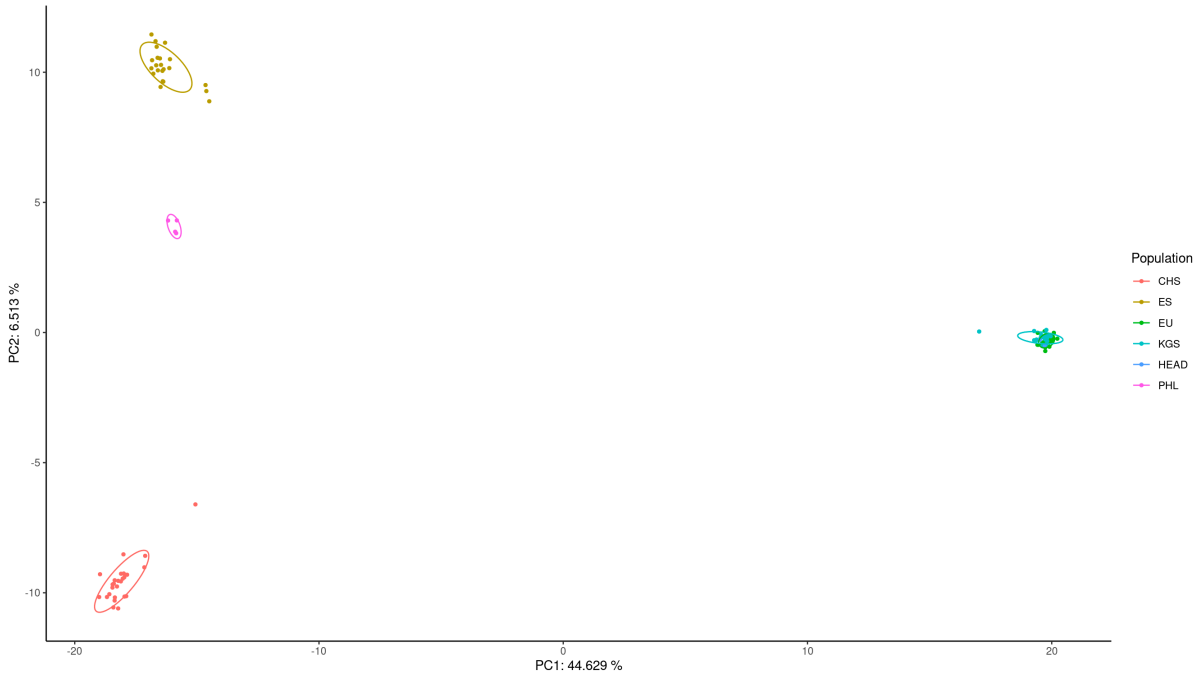

Figure 3.4. Principal components analysis, including all populations of *G. nobilis*: locations include Clark Hubbs Cienega (CHS), East Sandia (ES), Euphrasia Spring (EU), Karges Spring (KGS), Headwater Spring (HEAD), and Phantom Lake Spring (PHL).

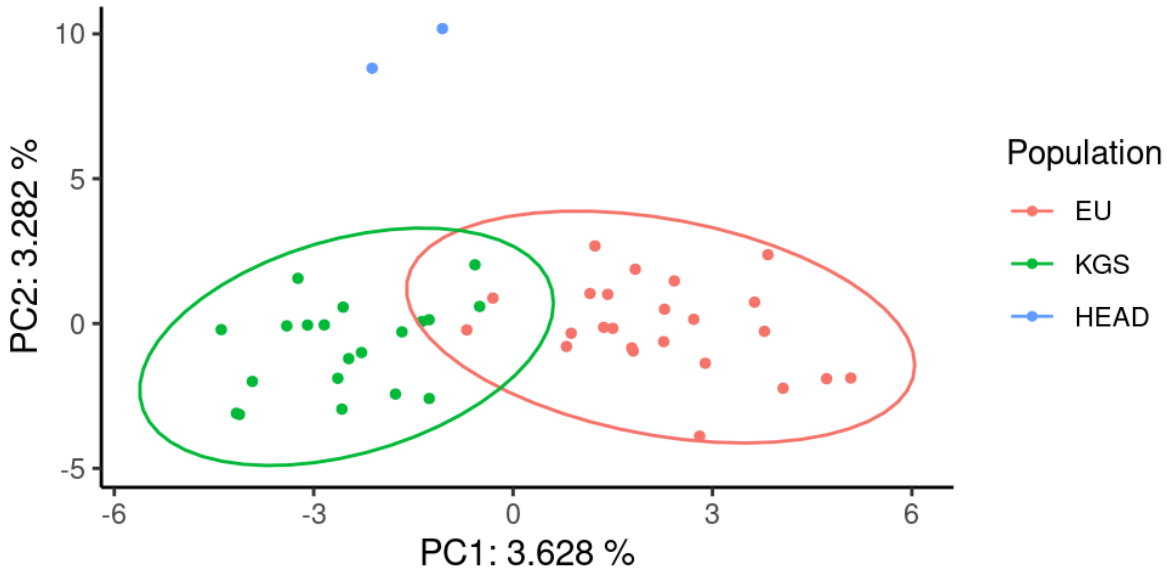

Figure 3.5. Principal components analysis of contemporary samples of *G. nobilis* from the eastern locality; locations are Euphrasia Spring (EU), Karges Spring (KGS), and Headwater Spring (HEAD).

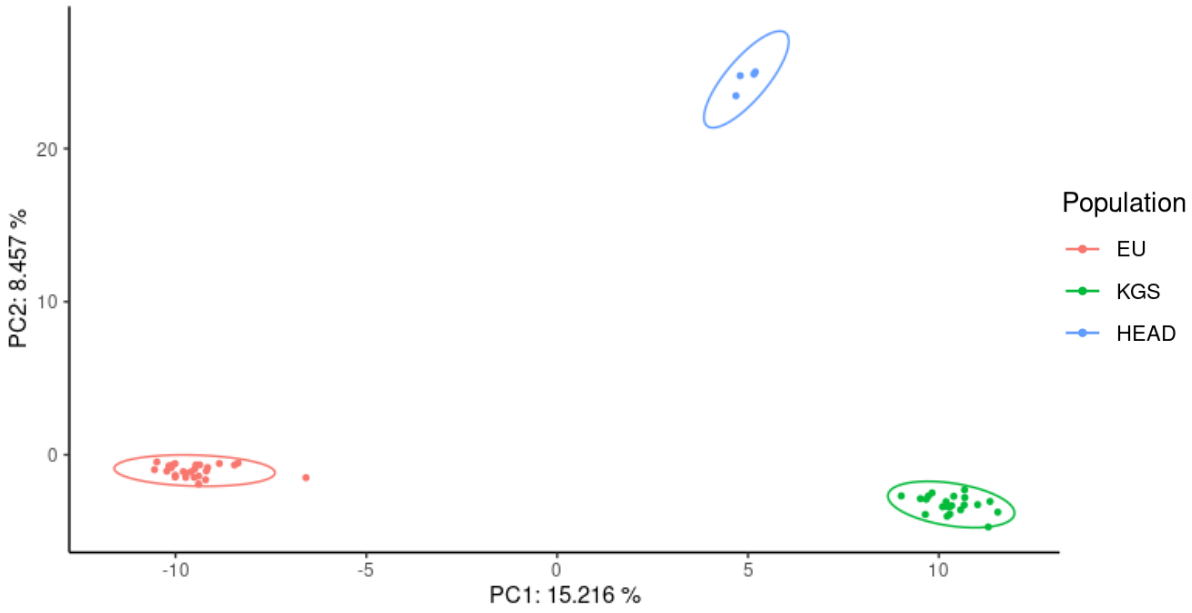

Figure 3.6. Principal components analysis of contemporary samples of G. nobilis from the western locality; locations are Clark Hubbs Cienega (CHS), East Sandia (ES), and Phantom Lake Spring (PHL).

Table 3.1. Hierarchical AMOVA results for contemporary *G.nobilis*; showing sum of squares (SS), variance components (VC) and percentage of variance (%).

	SS	VC	%
Among groups	32853.107	305.87471	43.27
Among locations within groups	7898.403	60.42781	8.55
Within groups	64718.679	340.62463	48.18
Total	17709.772	196.268	

Table 3.2. AMOVA results for three contemporary G. nobilis in the eastern locality; showing sum of squares (SS), variance components (VC) and percentage of variance (%).

	SS	VC	%
Among locations	564.043	3.619	1.84
Within locations	17145.729	193.649	98.16
Total	17709.772	196.268	

Table 3.3. FST estimates of contemporary G. nobilis populations with p-values above the diagonal, locations are Karges Springs (KGS), Euphrasia Spring (EU), and Headwater Pool Spring (HEAD).

	KGS	**EU**	**HEAD**
KGS	-	<0.0001	0.00376
EU	0.01179 (0.00584-0.00835)	-	0.00436
HEAD	0.05251 (0.0667-0.0732)	0.05374 (0.0602-0.0663)	-

Table 3.4. Mean expected heterozygosity (H_E) and mean allelic richness (A_R) by population, locations are Karges Springs (KGS) and Headwater Pool Spring (HEAD) and Euphrasia Spring (EU).

	H_E	A_R
KGS	0.216	1.91
HEAD	0.11	1.11
EU	0.218	1.98

Table 3.5. AMOVA results for three contemporary *G. nobilis* in the western locality; showing sum of squares (SS), variance components (VC) and percentage of variance (%).

	SS	**VC**	**%**
Among locations	6768.077	100.317	18.70
Within locations	44061.644	436.254	81.30
Total	50829.721	536.571	

Table 3.6. F_{ST} estimates of contemporary *G. nobilis* populations with *p*-values above the diagonal, locations are locations are East Sandia (ES), Clark Hubbs Cienega (CHS), and Phantom Lake Spring (PHL).

	ES	**CHS**	**PHL**
ES	-	<0.0001	<0.0001
CHS	0.153 (0.145-0.158)	-	<0.0001
PHL	0.254 (0.241-0.265)	0.281 (0.266-0.290)	-

Table 3.7. Effective population (N_E) estimates with lower and upper 95% confidence intervals done using a jackknife method, as well as point estimates for all three contemporary *G. nobilis* populations in the western locality; locations are East Sandia (ES), Clark Hubbs Cienega (CHS), and Phantom Lake Spring (PHL).

	Lower	**Point**	**Upper**
ES	444.7	470	501
CHS	903.1	956	970
PHL	1.4	2.2	4.5

Table 3.8. Mean expected heterozygosity (H_E) and mean allelic richness (A_R) by population, locations are East Sandia (ES), Clark Hubbs Cienega (CHS), and Phantom Lake Spring (PHL).

	H_E	A_R
ES	0.23	1.64
CHS	0.23	1.65
PHL	0.18	1.51

Supplemental Information

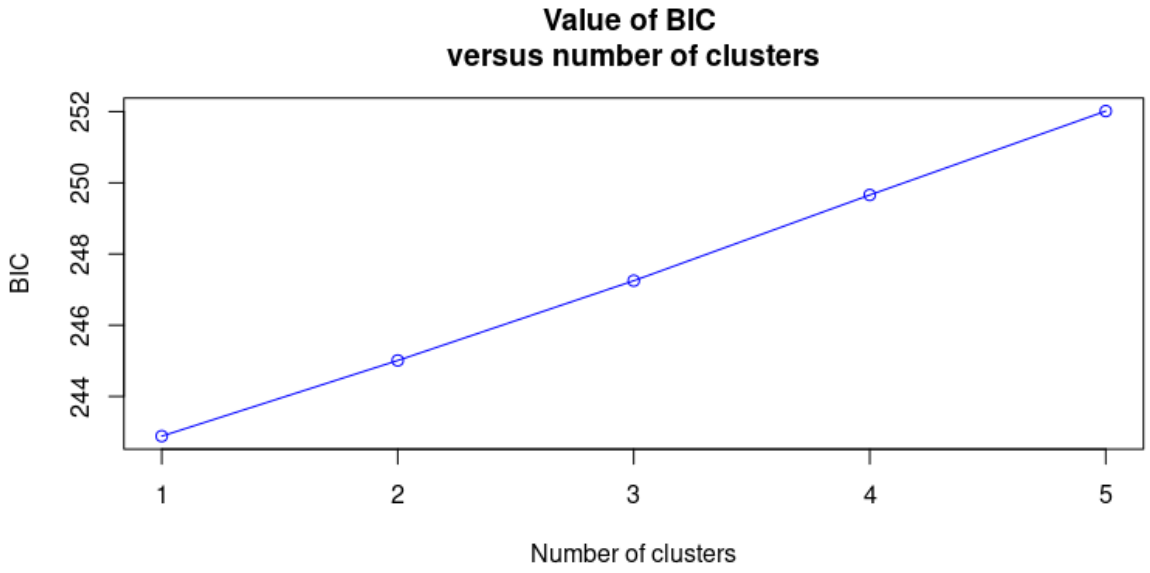

Figure S.3.1. Plot of Bayesian information criteria to inform *K*-means clustering; 1 is the optimized grouping for contemporary *G. nobilis* in the eastern locality.

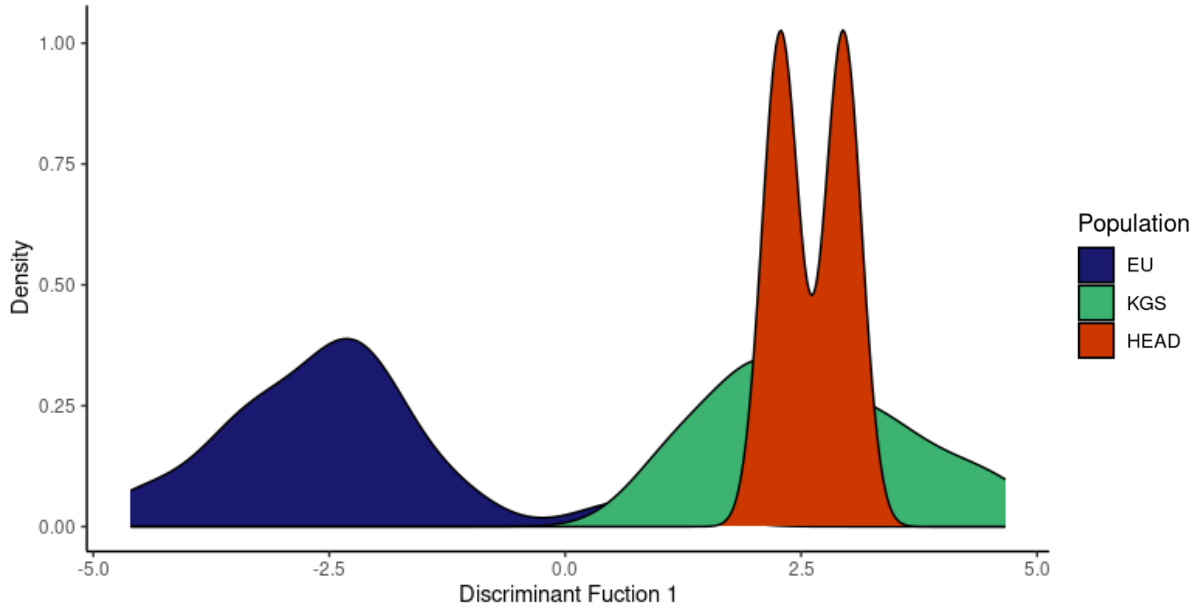

Figure S.3.2. Density plot for *K* =2 using contemporary *G. nobilis* dataset from the western locality; locations are Euphrasia Spring (EU), Karges Spring (KGS), Headwater Spring (HEAD), and Euphrasia Spring (EU).

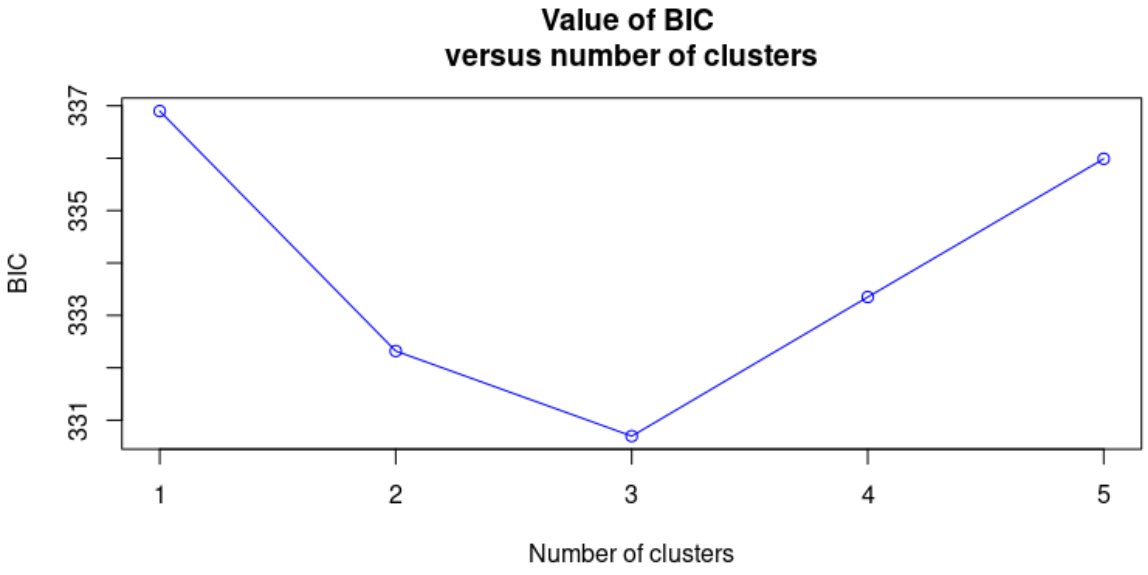

Figure S.3.3. Plot of Bayesian information criteria to inform K-means clustering; 3 is the optimized grouping for contemporary G. nobilis in the western locality.

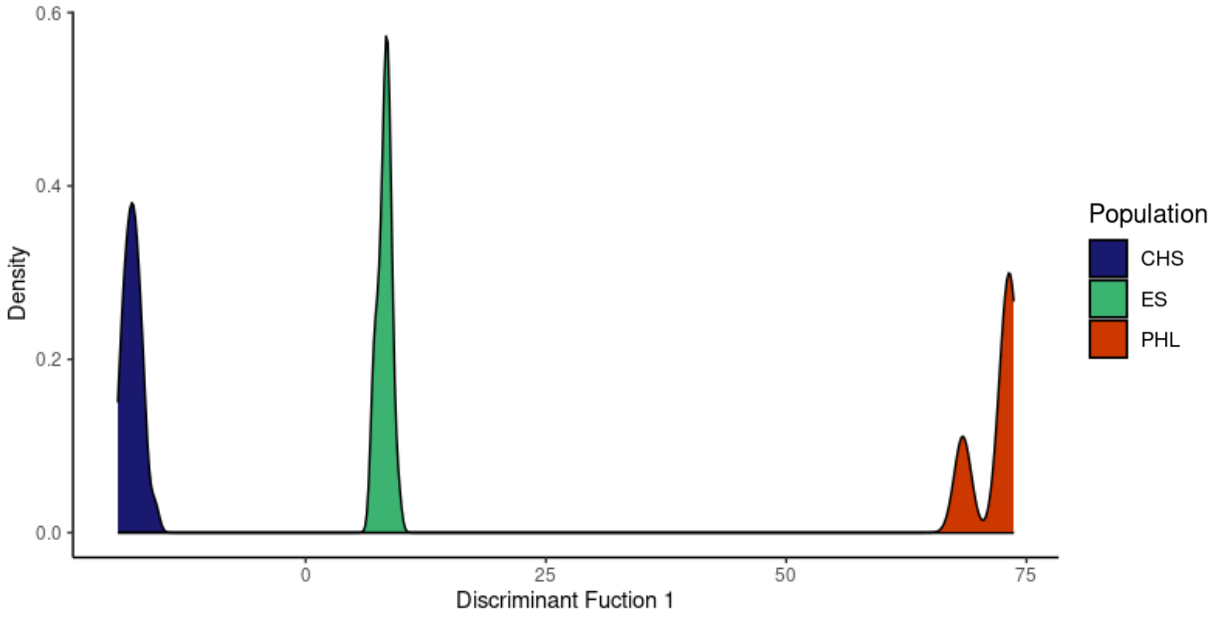

Figure S.3.4. Density plot for *K* =3 using contemporary *G. nobilis* dataset from the western locality; locations are Euphrasia Spring (EU), Karges Spring (KGS), Headwater Spring (HEAD), and Phantom Lake Spring (PHL).

CHAPTER IV

CONCLUSIONS

A Pupfish Imperiled

Over the past 40 to 50 years, the focus of conservation efforts on *C. bovinus* has been the detection and removal of hybrids between *C. bovinus* and *C. variegatus*. The present study suggests that removal and mitigation efforts have been successful to this point; however, continued vigilance is necessary. However, structuring at small spatial scales, low level of contemporary diversity, and small contemporary effective size estimates indicate that long term, and potentially short-term persistence, of the species in the wild is of concern. Despite the recent introduction of thousands of animals from the reserve population, *C. bovinus* in the wild seems to display the hallmark signs of a species that is rapidly approaching or has already entered the extinction vortex (Blomqvist et al. 2010; Minckley and Deacon 2017) and steps may need to be taken to increase the number of breeding animals and genetic diversity in the wild. One method that could introduce genetic diversity into the wild is a bi-directional assisted migration of wild and captive *C. bovinus*. Bi-directional gene flow could help the wild populations, while increasing the amount of genetic diversity present in the refuge. This will help combat the homogenizing effect that transplanting refuge individuals has on the wild population gene pools (Balloux 2004). Assisted migration between wild populations also could help increase the effective population sizes in the wild to increase short term persistence (Frankham et al. 2014a, b). Based on the natural history of *C. bovinus* in DY, prioritizing breeding habitat availability could help with increasing the number of breeding events and increase the effective and realized population sizes over time (Al-Shaer et al. 2016). In this study alone, two locations in Diamond Y known very recently to harbor *C. bovinus* had none, and reestablishing more wild populations

should be considered. As management plans are put into place, monitoring to track small-scale and large-scale genetic changes and to make sure that refuge gene pool remains appropriate for augmenting the wild population will be necessary as well (Meretsky et al. 2006).

Pecos Gambusia Management

This study suggests that fine-scale population structure exists for both the eastern and western localities of *G. nobilis*. While between group differentiation was significant and large, localities within groups exhibited divergence from each other as well. Structure at fine-scales and low levels of contemporary diversity indicate that long term, and potentially short-term persistence, of the species in the wild is of concern (Attard et al. 2022). Currently no refuge population exists, and with continued habitat changes occurring and water usage increasing, decreases in habitat availability might exacerbate the risks of extirpation for these small and fragmented populations (Frankham et al. 2014a). As management plans are put into place, monitoring to track small-scale and large-scale genetic changes is necessary to provide the necessary information needed to propose a refuge population be implemented in the event these wild populations collapse (Meretsky et al. 2006). If a refuge is proposed, special considerations will have to be made to consider how many separate refuges should be necessary to account for deep genomic divergence between the eastern and western groups. Due to long-term disconnection between the eastern and western populations, it could be detrimental to homogenize those populations into one refuge (Booy et al. 2000). Also, further studies need to be conducted, including populations of *G. nobilis* from New Mexico, where other fragmented habitat is located. An assessment including those localities is necessary for identifying if they need to be included in the western group, or if a third separate refuge populations should be implemented for them (Echelle and Echelle 1986; Echelle et al. 1989; Gill 2021).

Future Directions

This project served as a steppingstone in providing information for the conservation management of two imperiled freshwater fishes native to West Texas. Fine-scale sampling regimes and comprehensive genomic studies need to be conducted on a regular basis to continually update management plans for desert fishes that currently exhibit low genomic diversity and small and fragmented populations. Robust sampling regimes across these fine scales might help better inform levels of contemporary diversity and effective population sizes in the future. While habitat use and availability will continually be an issue in arid regions, prioritizing small scale habitat restoration at sites where populations sizes are the smallest might be the first step in helping increase and stabilize populations. Finally, the techniques used in this study are not specific to these groups of fishes. Comprehensive genomic studies can be implemented to update conservation management plans of any imperiled vertebrate where strict and discrete sampling can be conducted.

References

Al-Shaer L, Bloch A, Paciorek T, et al (2016) Renovated Breeding Habitat Use in Wild & Captive-bred Populations of an Endangered Desert Pupfish. Journal of Biodiversity & Endangered Species 4:. https://doi.org/10.4172/2332-2543.1000156

Attard CRM, Sandoval-Castillo J, Brauer CJ, et al (2022) Fish out of Water: Genomic Insights Into Persistence of Rainbowfish Populations in the Desert. Evolution 76:171–183. https://doi.org/10.1111/evo.14399

Balloux F (2004) Heterozygote Excess in Small Populations and the Heterozygote-Excess Effective Population Size. Evolution 58:1891–1900. https://doi.org/10.1111/j.0014-3820.2004.tb00477.x

Blomqvist D, Pauliny A, Larsson M, Flodin L-Å (2010) Trapped in the Extinction Vortex? Strong Genetic Effects in a Declining Vertebrate Population. BMC Evolutionary Biology 10:33. https://doi.org/10.1186/1471-2148-10-33

Booy G, Hendriks RJJ, Smulders MJM, et al (2000) Genetic Diversity and the Survival of Populations. Plant Biol (Stuttg) 2:379–395. https://doi.org/10.1055/s-2000-5958

Echelle AF, Echelle AA (1986) Geographic Variation in Morphology of a Spring-Dwelling Desert Fish, *Gambusia nobilis* (Poeciliidae). The Southwestern Naturalist 31:459–468. https://doi.org/10.2307/3671700

Echelle AF, Echelle AA, Edds DR (1989) Conservation Genetics of a Spring-Dwelling Desert Fish, the Pecos Gambusia (*Gambusia nobilis*, Poeciliidae). Conservation Biology 3:159–169

Frankham R, Bradshaw CJA, Brook BW (2014) Genetics in Conservation Management: Revised Recommendations for the 50/500 Rules, Red List Criteria and Population Viability Analyses. Biological Conservation 170:56–63. https://doi.org/10.1016/j.biocon.2013.12.036

Gill GS (2021) Morphological Signatures of Hybridization Between the Endemic Endangered Pecos Gambusia (*Gambusia nobilis*) and Invasive *Gambusia spp*. at Bitter Lake National Wildlife Refuge, New Mexico.

Meretsky VJ, Fischman RL, Karr JR, et al (2006) New Directions in Conservation for the National Wildlife Refuge System. BioScience 56:135–143. https://doi.org/10.1641/0006-3568(2006)056[0135:NDICFT]2.0.CO;2

Minckley WL, Deacon JE (2017) Battle Against Extinction: Native Fish Management in the American West. University of Arizona Press