UNIVERSITY OF CALIFORNIA

Los Angeles

Conservation genetics in the age of high-throughput sequencing:

new tools for old dilemmas

A dissertation submitted in partial satisfaction of the

requirements for the degree Doctor of Philosophy

in Biology

by

Rachel Turba de Paula

2022

ABSTRACT OF THE DISSERTATION

Conservation genetics in the age of high-throughput sequencing:

new tools for old dilemmas

by

Rachel Turba de Paula
Doctor of Philosophy in Biology
University of California, Los Angeles, 2022
Professor David K. Jacobs, Chair

A combination of complex geological and climatological history in California have created a biodiversity "hotspot" with numerous endemic species. Most of the wetlands in the region, such as rivers and coastal lagoons, were shaped by uplift, sea-level change and climatic fluctuations creating a physical template on which genetic diversification of the aquatic biota could unfold. Coastal lagoons, for example, provide important ecological services as well as habitat and nursery for endemic and endangered animals, such as the steelhead (*Oncorhynchus mykiss*), red-legged frog (*Rana aurora draytonii*), and the tidewater goby (*Eucyclogobius newberryi*). However, anthropogenic impacts on the landscape, such as urbanization and water management for human use and/or flood control, have had major impacts on the landscape threatening much of the fauna associated with these wetlands. Such is the case of the rare and endangered unarmored threespine stickleback, *Gasterosteus aculeatus williamsoni*, found in only a few drainages in southern California. Most of the populations are now extirpated and those that persist have suffered isolation

and bottlenecks, as well as inadvertent mixture of distinct lineages and introduction of non-native predators. Many conservation and restoration efforts led by numerous governmental units and NGOs are underway in California. These require decisions informed by more comprehensive understanding of current and extirpated populations, as such information will lead to more successful and effective planning and management. Advancements in molecular techniques, particularly the development of high-throughput sequencing, has made it possible to not only deepen our knowledge on evolution but also to use it as a tool in conservation biology. This revolution allows us to unravel genomic signals of demography and selection, investigate historic data from museum material, and even look at environmental samples collected from soil and water. All the work presented here explores the use of high-throughput sequencing and bioinformatic tools to target relevant conservation issues related to endangered species and habitats in southern California. We expect that results will have an impact on conservation efforts, and techniques can be applied to other systems and fields of interest.

The dissertation of Rachel Turba de Paula is approved.

Kirk E. Lohmueller

Matteo Pellegrini

Beth Shapiro

Ryan Kelly

David K. Jacobs, Committee Chair

University of California, Los Angeles

2022

Table of Contents

List of Tables

List of Supplementary Tables

List of Figures

List of Supplementary Figures

Acknowledgements

Funding was provided by the National Council for Scientific and Technological Development of Brazil under Grant No. 209261/2014-5. All the work listed here used computational and storage services associated with the Hoffman2 Shared Cluster provided by UCLA Institute for Digital Research and Education's Research Technology Group (IDRE), and access was sponsored by the Pellegrini lab.

Chapter 1 is a version of Turba, R., Richmond, J.Q., Fitz-Gibbon, S., Morselli, M., Fisher, R.N., Swift, C.C., Ruiz-Campos, G., Backlin, A.R., Dellith, C. and Jacobs, D.K. (2022), Genetic structure and historic demography of endangered unarmored threespine stickleback at southern latitudes signals a potential new management approach. Molecular Ecology. https://doi.org/10.1111/mec.16722. Rachel Turba, Jonathan Q. Richmond, David K. Jacobs and Robert N. Fisher designed the research. Jonathan Q. Richmond, Camm C. Swift, Gorgonio Ruiz-Campos, Adam R. Backlin, Chris Dellith and David K. Jacobs collected samples. Rachel Turba and Jonathan Q. Richmond performed microsatellite and genomic analysis. Marco Morselli contributed reagents and wet laboratory expertise. Sorel Fitz-Gibbon contributed to bioinformatic analysis and expertise. Rachel Turba and Jonathan Q. Richmond analyzed the data. Rachel Turba, Jonathan Q. Richmond and David K. Jacobs wrote the manuscript. All authors reviewed the manuscript. Low coverage whole-genome sequencing was provided by Matteo Pellegrini (UCLA/MCDB) through the Broad Stem Cell Research Center (BSCRC) at UCLA. Gratitude goes to Annabel Beichman for aid with the MSMC analyses.

Chapter 2 is still being prepared for publication. Rachel Turba and David K. Jacobs designed the research. Marco Morselli and Beth Shapiro contributed reagents and laboratory expertise. Rachel Turba analyzed the data. Rachel Turba and David Jacobs wrote the manuscript.

For **Chapters 1 and 2**, funding was provided by the U.S. Fish and Wildlife Service Recovery Challenge Grant No. F19AC00543, and the U.S. Geological Survey's Ecosystems Mission area. Steve Howard, Tim Hovey, Eric Morrissette, Elizabeth Gallegos, Russell Barabe, Morgan Ball and John LaBonte assisted with sample collection. Catherine Darst and Eric Morrissette of the U.S. Fish and Wildlife Service and John O'Brien and Jennifer Pareti of the California Department of Fish and Wildlife facilitated logistics, permitting, and overall support.

Chapter 3 is a version of Turba, R., Thai, G., Jacobs, K. D. (2022). To freeze or to scoop? Dealing with the turbid waters of California's coastal lagoons. *BioRxiv*. https://doi.org/10.1101/2022.06.17.495388. This chapter has been submitted to Peer Community in Ecology and is under revision. Rachel Turba and David K. Jacobs designed the research. Rachel Turba and Glory Thai collected samples and processed them in the laboratory. Rachel Turba analyzed the data. Rachel Turba, Glory Thai and David K. Jacobs wrote the manuscript. Funding was provided by NOAA Sea Grant 120651698:1, and the UCLA La Kretz Center/Stunt Ranch Reserve. Funding for the CALeDNA sample processing, infrastructure, and personnel was provided by the University of California Research Initiatives (UCRI) Catalyst grant CA-16-376437 and Howard Hughes Medical Institute (HHMI) Professors Grant GT10483. We are very grateful for all the help provided by the CALeDNA team, but would like to give special thanks to Teia Schweizer, who personally trained us in the wet laboratory skills. Huge thanks to Ryan Kelly, for helping streamline the design and analysis of this paper and for always being so responsive and available.

VITA

Rachel Turba de Paula

EDUCATION

House of Oswaldo Cruz (COC/Fiocruz)
2015 *Specialization Course in Popularization of Science, Technology and Health*
Federal University of Rio de Janeiro/National Museum (MNRJ)
2013 *M.Sc. Degree in Zoology*
Federal University of Rio de Janeiro (UFRJ)
2011 B.Sc. in Biology (emphasis in Zoology)

PROFESSIONAL EXPERIENCE

UCLA Department of Ecology and Evolutionary Biology
2015 – present *Graduate Researcher*

- Analyze genomics of endangered populations of freshwater species in southern California
- Explore techniques to extract DNA from museum samples to generate high-quality genomic analyses
- Utilize environmental DNA and metabarcoding approaches to assess biological communities and habitat vulnerability

Caiman Ecological Refuge
2015 *Bilingual Tour Guide*

- Facilitated conversations between 100+ eco-tourists from across the globe and local field guides
- Led tours to educate eco-tourists about local fauna/flora, and local conservation actions, and shared insights about science, biology and conservation in general
- Developed star gazing activity drawing upon Native Americans' knowledge of astronomy to educate eco-tourists on star systems visible in Brazil

International Virtual Institute of Global Change (IVIG/COPPE-UFRJ)
2012 – 13 *Intern*

- Supported a team of +7 researchers with field work, collecting data on synanthropic fauna (e.g. pigeons, rats, cockroaches) at Rio de Janeiro' seaport, as part of a management compliance program for all Brazilian ports
- Developed manuals of 'Seaport Good Practices' for Brazil's 22 seaports including management actions and plans
- Developed strategies to build virtual training resources in 'Good Practices' for seaport workers

TEACHING EXPERIENCE

UCLA Division of Life Sciences
2021 – 22 LS192 - LA Program - Collaborative Learning Theory & Practice: Introduction, Methods & Applications
UCLA Department of Ecology and Evolutionary Biology
2018 (W) Life Sciences 7B – Genetics, Evolution and Ecology
2016 (Sp/Su) Life Sciences 1 – Evolution, Ecology, and Biodiversity
2017 (Sp) Field Marine Biology Quarter

PRESENTATIONS & MEETINGS

2020	EEB Research Symposium, UCLA, USA *Genomics for Recovery of the Unarmored Threespine Stickleback in Southern California: Three Independent Origins?* **(Poster)**
2020	Student Conference on Conservation Science, New York, USA *Use Of Museum Material To Reconstruct The Extirpated Fauna Of The Los Angeles Basin* **(Oral)**
2020	North American Congress for Conservation Biology, Colorado, USA *Through colorful lenses: addressing eDNA methods variability in a threatened habitat in southern California* **(Oral)**
2019	Western Society of Naturalists, Ensenada, UMS *Use of museum material to reconstruct the extirpated fauna of the Los Angeles basin* **(Oral)**
2019	Coastal Wetland Workshop, UCLA, USA *Use of environmental DNA in a dynamic habitat: The case of coastal lagoons in southern California* **(Oral)**
2018	University of California Conservation Genomics Consortium Workshop, Asilomar, USA *Bias and strengths of eDNA methods in a dynamic habitat – the case of coastal lagoons in southern California* **(Oral)**
2013	World Congress of Malacologists, Ponta Delgada, PT *Taxonomic review of Turbonilla (Gastropoda: Pyramidellidae) collected through the MD55 expedition, off SE coast of Brazil* **(Oral)**

PUBLICATIONS

2022	**Turba, R.**, Richmond, J.Q., Fitz-Gibbon, S., Morselli, M., Fisher, R.N., Swift, C.C., Ruiz-Campos, G., Backlin, A.R., Dellith, C. and Jacobs, D.K. (2022), Genetic structure and historic demography of endangered unarmored threespine stickleback at southern latitudes signals a potential new management approach. *Molecular Ecology*. https://doi.org/10.1111/mec.16722
2022	Blanchette, A., Spies, B., Eminhizer, S. **Turba, R.**, *et al.* Effects of onshore development and damselfish (*Stegastes nigricans*) on coral richness in Opunohu Bay and Cook's Bay in Moorea, French Polynesia. *Coral Reefs* **41,** 987–999 (2022). https://doi.org/10.1007/s00338-022-02282-3
2021	Lin, M., **Turba, R.** et al. Landscape analyses using eDNA metabarcoding and Earth observation predict community biodiversity in California. *Ecological* Applications. https://doi.org/10.1002/eap.2379
2021	Sura, S.A., Bell, A., Kunes, K.L., **Turba, R.**, Songer R., Fong, P., 2021. Responses of two common coral reef macroalgae to nutrient addition, sediment addition, and mechanical damage. *Journal of Experimental Marine Biology and Ecology*, 536. https://doi.org/10.1016/j.jembe.2021.151512
2019	Sura, S.A., Delgadillo, A., Franco, N., Gu, K., **Turba, R.** and Fong, P., 2019. Macroalgae and nutrients promote algal turf growth in the absence of herbivores. *Coral Reefs*, 38(3), pp. 425-429. https://doi.org/10.1007/s00338-019-01793-w
2019	Blanchette, A., Ely, T., Zeko, A., Sura, S.A., **Turba, R.** and Fong, P., 2019. Damselfish Stegastes nigricans increase algal growth within their territories on shallow coral reefs via enhanced nutrient supplies. *Journal of Experimental Marine Biology and Ecology*, 513, pp.21-26. https://doi.org/10.1016/j.jembe.2019.02.001
Pre-print	Gomes. D. G. E., …, **Turba, R.** et al. Why don't we share data and code? Perceived barriers and benefits to public archiving practices. *MetaArXiv*. Doi: 10.31222/osf.io/gaj43
Pre-print	**Turba, R.**, Thai, G., Jacobs, K. D. To freeze or to scoop? Dealing with the turbid waters of California's coastal lagoons. *BioRxiv*. Doi: 10.1101/2022.06.17.495388

Chapter 1 . Genetic structure and historic demography of endangered unarmored threespine stickleback at southern latitudes signals a potential new management approach

Abstract

Habitat loss, flood control infrastructure, and drought have left most of southern California and northern Baja California's native freshwater fish near extinction, including the endangered unarmored threespine stickleback (*Gasterosteus aculeatus williamsoni*). This subspecies, an unusual morph lacking the typical lateral bony plates of the *G. aculeatus* complex, occurs at arid southern latitudes in the eastern Pacific Ocean and survives in only three inland locations. Managers have lacked molecular data to answer basic questions about the ancestry and genetic distinctiveness of unarmored populations. These data could be used to prioritize conservation efforts. We sampled *G. aculeatus* from 36 localities and used microsatellites and whole genome data to place unarmored populations within the broader evolutionary context of *G. aculeatus* across southern California/northern Baja California. We identified three genetic groups, with none consisting solely of unarmored populations. Unlike *G. aculeatus* at northern latitudes, where Pleistocene glaciation has produced similar historical demographic profiles across populations, we found markedly different demographies depending on sampling location, with inland unarmored populations showing steeper population declines and lower heterozygosity compared to low armored populations in coastal lagoons. One exception involved the only high elevation population in the region, where the demography and alleles of unarmored fish were similar to low armored populations near the coast, exposing one of several cases of artificial translocation. Our results suggest that the current 'management-by-phenotype' approach, based on lateral plates, is

incidentally protecting the most imperiled populations; however, redirecting efforts toward evolutionary units, regardless of phenotype, may more effectively preserve adaptive potential.

Introduction

Some of the most highly modified habitats in southwestern North America include rivers and streams (Benke, 1990; Minckley & Marsh, 2009; Pringle et al., 2000; Propst et al., 2021; Ricciardi & Rasmussen, 1999), particularly in the southern California-northern Baja California coast ecoregion, where lowland reaches of major rivers are now channelized in concrete and flows are regulated via dam releases from artificial lakes and reservoirs (Gumprecht, 1999; Orsi, 2004). This infrastructure was developed to protect the rapidly growing human population of the mid-20th century from flood damage and to store and divert water in a landscape that was largely devoid of perennial water (Stephenson & Calcarone, 1999; Van Wormer, 1991). However, the effects on the native freshwater fish fauna have been dramatic, reducing the amounts of suitable habitat (Faber et al., 1989; Stephenson & Calcarone, 1999; Swift et al., 1993a), eliminating the capacity for gene exchange within and between drainage basins (Benjamin et al., 2016; Richmond et al., 2018), and altering the dynamics of an ecosystem with historically ephemeral and intermittent surface water (Levick et al., 2008; Richmond et al., 2018). Climatic stressors that promote oscillating demography (e.g., strong seasonal and multi-year variation in amounts of precipitation and duration of drought: Levick et al., 2008; Mount, 1995; Stephenson & Calcarone, 1999) and other factors may be accelerating losses of local genetic diversity and pushing these species closer to the brink of extinction (Moyle et al., 2015). Because most of the region's freshwater fish species lack genetic data to guide management (Leidy & Moyle, 2021), more studies are needed to characterize their ancestry, population structure, and historical demography to provide better snapshots of population dynamics prior to 20th century urbanization.

One of the most iconic taxa of this fauna is the unarmored threespine stickleback, which lacks lateral bony plates, or 'body armor', that typically line the flanks of the *G. aculeatus* species complex. The complex comprises marine, anadromous, and freshwater populations, and is distributed across the mid to high latitude in the Northern Hemisphere along the margins of the Pacific, Arctic and Atlantic oceans (Bell & Foster, 1994). Morphology varies widely across this distribution, with the unarmored morph being the rarest of four recognized plate morphs (i.e., fully armored, partial, low, and unarmored: Bell, 1976; Miller & Hubbs, 1969). Fully armored morphs have a continuous row of 30+ plates per side and predominantly occur in marine environments; partially armored morphs have a discontinuous row of 11+ plates per side and occur in brackish water; and low and unarmored morphs occur exclusively in freshwater and have fewer than 11 plates per side or no plates at all (Bell & Foster, 1994). Diversification of the plate phenotypes is often adaptive (reviewed in Bell, 2001; Colosimo et al., 2005; Hagen & Gilbertson, 1973; Reimchen, 1994, 2000). Plates are associated with greater piscivorous predation, while their absence is hypothesized to confer advantages in mobility, buoyancy, calcium availability and faster growth rates (possibly in avoidance of insect predators) (Barrett et al., 2008; Bell et al., 1993; Bell & Foster, 1994; Bergstrom, 2002).

Unarmored populations were historically more widespread in riverine settings of coastal southern California, but most have become extirpated since the 1940s (Bell, 1978; Miller, 1961; Miller & Hubbs, 1969; Swift et al., 1993a). Currently, native populations of unarmored stickleback remain in three upstream areas: San Antonio Creek on Vandenberg Space Force Base (VSFB, Santa Barbara County); the upper Santa Clara River (Los Angeles County); and the upper Santa Ana River and Baldwin Lake drainage basin in the San Bernardino mountains (San Bernardino County) (USFWS, 2021; Figure 1.1). *Gasterosteus aculeatus williamsoni* (Girard, 1854) was

described on the basis of specimens collected in the upper Santa Clara River in Soledad Canyon (Figure 1.1: #25-26, type-locality). It was listed as endangered by the U.S. federal government in 1970 (35 Federal Register 16047) and the state of California (https://wildlife.ca.gov/Conservation/CESA) in 1971. Limited habitat and continued population decline are described in a recovery plan for *G. a. williamsoni*, with status updates conducted every five years to describe current research, management efforts, and progress towards recovery of this endangered species.

Most conservation management efforts in southern California have focused on the unarmored stickleback population occurring near the type locality in the upper Santa Clara River (USFWS, 2021). However, such emphasis on the phenotype may be misguided, as *G. aculeatus* is well-known for parallel evolution of different plate morphs. This phenomenon was poorly understood at the time the taxonomy was developed, but genetic data have since shown that transitions from fully armored marine morphs (with over 30 plates) to partially or low armored freshwater morphs (typically between 30 and one plate) have occurred repeatedly and independently following deglaciation at northern latitudes over the last ~15,000 years (Deagle et al., 2013a; Hohenlohe & Magalhaes, 2020; F. C. Jones, Grabherr, Chan, Russell, Mauceli, Johnson, Swofford, Pirun, Zody, & White, 2012; Reimchen, 1994; Schluter, 2000). Parallel evolution is widely viewed as a phylogenetic signature of adaptive trait evolution (Schluter & Nagel, 1995), as standing genetic variation in closely related populations increases the probability that traits will evolve in the same way in similar environments (Hendry, 2013; Morris et al., 2018; Peichel & Marques, 2017).

In this study, we used microsatellite markers and low coverage whole-genome sequencing (lcWGS) to characterize the genetic structure and historical demography of *G. aculeatus* in

southern California, with emphasis on a geographic area known as the Southern California Bight (SCB; Figure 1.1). The SCB is formed by a 685 km arc of coastline along the west coast of the United States and Mexico, from Point Conception in California south to Punta Colonet in Baja California. Many upstream and now isolated populations of *G. aculeatus* occur within the SBC, including extant unarmored populations in the upper Santa Clara (Figure 1.1: #20-22, 25-26) and in the upper Santa Ana drainage in the San Bernardino Mountains (Figure 1.1: #28), while the third native unarmored population occurs immediately to the north of the SCB in San Antonio Creek (Figure 1.1: # 8) at VSFB. However, questions about the relationship between these unarmored populations and the low armored populations in the region remain unanswered. In addition, unarmored fish have been transplanted from the type locality to other areas to protect against ecological instability and safeguard the genetic integrity of this population. Numerous inadvertent transplants have also occurred during trout stocking and many are not well documented. Given this complex history, comprehensive genetic data can help focus conservation efforts where they are needed most.

The broader goal of this work is to use modern genetic approaches to suggest a management strategy that more accurately reflects evolutionary units. Genomic data in this study are part of a larger, ongoing whole genome sequencing project on *G. aculeatus* throughout the region that will include more samples with higher sequencing coverage and address additional questions related to adaptive evolution. However, these current data are sufficient to address the following key questions: (1) What is the genetic structure of *G. aculeatus* at multiple geographic scales in southern California?; (2) Do unarmored populations of *G. aculeatus* form a cohesive group when compared to other populations of *G. aculeatus* across the region?; (3) Do temporal

trends in historical demography vary between up- and downstream reaches, and are the same trends mirrored across drainages?

Materials and Methods

Sampling

We obtained tissue samples from monitoring surveys conducted by the U.S. Geological Survey (U.S. Federal Recovery Permits TE-045994, TE 793644-6 &7; California State Scientific Collecting Permits SCP-2679, SCP-90); Mexican federal collecting permit (Permiso de Pesca de Fomento) DGOPA 14253.101005.6950, and its extension DGOPA 06435.210606.2640 issued by the Comisión Nacional de Acuacultura y Pesca of the Secretaría de Agricultura, Ganadería, Desarrollo Rural, Pesca y Alimentación (SAGARPA); professional colleagues (federal recovery and state scientific collecting permits available on request); and museum collections. Fish were captured using minnow traps, seines, and or dip nets and muscle tissue samples were stored in 95% ethanol. The dataset included low and unarmored *G. aculeatus* from 34 localities across southern California to what used to be the southernmost extent of the species range at Bocana El Rosario in northern Baja California, Mexico. We also included representatives of two fully plated populations collected in the San Francisco Bay area (Figure 1.1, Table 1.1).

We performed DNA extractions using a QIAGEN DNeasy Blood & Tissue Kit on muscle tissue and genotyped 470 individuals for 12 microsatellite loci known to be polymorphic in populations of *G. aculeatus* in southern California following the approach of Richmond et al. (2015). We then performed low coverage whole genome sequencing (lcWGS) on a subset of these samples from six locations (two individuals per location; $N = 12$). Three of the six locations involved each of the upstream, unarmored populations, and the remaining three were from low

armored lagoonal populations (Figure 1.1: yellow highlighted dots). For two drainages, San Antonio Creek-VSFB and the Santa Clara River, we could directly compare upstream unarmored fish to low armored lagoonal fish at the mouth. However, because unarmored stickleback have been extirpated from the lower Santa Ana River, we compared the third unarmored cohort from the upper part of this drainage at Sugarloaf Meadow (elevation 2055 m) to all lagoonal fish, including those from El Rosario in Baja California. Sugarloaf Meadow is one of three isolated ponds in the San Bernardino Mountains at ~2000 m elevation occupied by unarmored morphs that are collectively referred to as 'Shay Creek stickleback' (Haglund & Buth, 1988; Malcolm, 1992; Moyle, 2002).

Microsatellites

We amplified microsatellites in sets of 10 µl multiplex reactions using a QIAGEN Multiplex PCR kit and 50-100 ng of DNA. We used negative controls in each PCR and ran a subset of the samples as repeats to verify the genotyping calls (i.e., positive controls). We performed genotyping runs on an ABI 3100S Automated Sequencer (Applied Biosystems, Foster City, CA, USA) at the San Diego State University Microbiology Core Facility using a LIZ600 size standard (Applied Biosystems) and scored alleles using GeneMarker v1.85 (Softgenetics LLC, State College, PA). We obtained a total of 261 alleles with 0.39% missing data.

Population structure

We used STRUCTURE v2.3.1 (Pritchard et al., 2000) on the microsatellite data to assign individuals to natural groups, test for admixture across populations, and identify populations that were established through artificial transplantation. We performed assignment tests using the uncorrelated allele frequencies setting for $K = 2$ through $K = 10$, running 25 iterations at each K

for 500,000 replicates after a burn-in of 500,000 iterations. We used the Δ*K* method (Evanno et al., 2005) to select the optimal *K*, but report the assignments at other *K* values in the Supplementary Information section. To summarize the results, we generated alignments of the assignment coefficient matrices at each *K* using CLUMPAK (Kopelman et al., 2015).

We also used principal components analysis (PCA) as an exploratory tool to complement the results of the STRUCTURE analyses. We used the function dudi.pca in the adegenet package (Jombart & Ahmed, 2011) and ggplot2 in R v.3.6.2 (R Core Team, 2018a) using RStudio v.2022.02.3 (RStudio Team, 2020a) to generate the final PCA plot. We replaced missing data using the "mean" method and retained all axes.

Genome sequencing

We used the sparQ DNA Frag & Library Prep Kit (Quanta Bio) to generate the libraries from the purified, extracted DNAs. From these libraries we generated single-end 100 base-pair sequences on an Illumina Hiseq4000 at the Broad Stem Cell Research Center (BSCRC) at UCLA. We aligned the sequences to the Bear Paw Lake threespine stickleback reference genome (F. C. Jones, Grabherr, Chan, Russell, Mauceli, Johnson, Swofford, Pirun, Zody, & White, 2012) using BWA-MEM v. 0.7.12 (H. Li, 2013a), and performed variant discovery using HaplotypeCaller and GenotypeGVCFs from GATK (McKenna et al., 2010) using default parameters (Table S 1.1 contains the summary quality report of the BAM files). The final VCF file was filtered using the GATK standard thresholds (for SNPs: QD < 2.0, MQ < 40.0, FS > 60.0, MQRankSum < -12.5, ReadPosRankSum < -8.0; for INDELs: QD < 2.0, FS > 200.0, ReadPosRankSum < -20.0, InbreedingCoeff < -0.8).

Inferring historical demography

We used the multiple sequentially Markovian coalescent method (MSMC: Schiffels & Durbin, 2014; an extension of PSMC: Li & Durbin, 2011) to reconstruct the historical demography for a subset of populations in the dataset. This method tracks coalescence across genomic regions and estimates how many coalescent events occur within certain time frames across the genome. The inverse of the rate of these events within each time frame, the inverse instantaneous coalescence rate (IICR: Mazet, Rodríguez, Grusea, Boitard, & Chikhi, 2016) serves as a proxy for N_e, provided that certain conditions are met. If not, other factors influencing the IICR (e.g., non-random mating, inbreeding, admixture) need to be accounted for when interpreting demography through time.

We performed our analyses using the unphased sequences of single individuals. Simulations have shown that MSMC on single genomes (also referred to as PSMC') performs equally well or better than multiple genomes under certain conditions and is preferable for lower coverage sequencing (Beichman et al., 2017).

We filtered VCF files for indels, missing data, multi-allelic loci and read depth (DP) before running MSMC using default parameters (number of iterations = 20; recombination rate/mutation rate = 0.25; pattern of fixed time segments = 10*1+15*2; fixed recombination rate = NO). Filtering for read depth improves the ability to characterize demographic change when mean coverage is ≤10 (Nadachowska-Brzyska et al., 2016). Therefore, we filtered out any SNPs with coverage below DP≤8 (GATK: -SelectVariants -select "DP≥8") based on the average coverage of our samples and the percentage of missing data after filtering (see Table S 1.3 for number of SNPs retained and percentage of missing data for different filtering thresholds and Figure S 1.4 for effects on MSMC curves). To remove INDELs, multi-allelic loci and missing data from our VCF

files, we used VCFtools (–remove-indels –max–allele 2 –min-allele 2 –max-missing 1; Danecek et al., 2011).

We used bootstrap replicates (n = 100) to estimate the variance around the estimates of IICR. We also explored results using two different mutation rates, $3.7x10^{-8}$ and $6.6x10^{-8}$ (Liu et al., 2016), and generation times of 1 and 2 years (Liu et al., 2016; Rollins, 2017), as neither parameter is known with certainty for populations in southern California or whether they differ between inland and coastal localities due to environmental heterogeneity (e.g., higher summer temperatures and greater UV exposure at inland sites). These parameters convert the values calculated by the model to estimates of time and population size. Shorter generation time and higher mutation rates yield younger dates and lower population size estimates (i.e., curve shifted down and to the left), while the reverse is true for longer generation times and lower mutation rates (i.e., curve shifted up and to the right). We plotted the results using the R packages ggplot2, scales and ggpubr.

As a verification step for our MSMC approach, we re-analyzed two individuals from our sample from Santa Clara River (lagoon/upstream) and two from Alaska (river/lake comparison) from the work of Liu et al. (2016). Because their genomes had higher coverage than ours (26.7X and 30.2X), we filtered out SNPs with coverage below 8 and higher than 20, and reran the analysis using 20 autosomal chromosomes following Liu et al. (2016).

Lastly, we generated PCA plots using the R package SNPrelate (Zheng et al., 2012) to detect population structure. We filtered the dataset to include only biallelic SNPs (snpgdsVCF2GDS, method = “biallelic.only”) and reduced linkage among sites (snpgdsLDpruning, method = “corr”, ld.threshold = sqrt(0.1)).

Comparisons of genetic diversity

For the microsatellite dataset, we calculated expected (H_s) and observed (H_o) heterozygosity for each population using GenoDive v3.06 (Meirmans, 2020) (except those that had <8 samples: Arroyo Grande, Shuman Lagoon, Sweetwater River) and report values in the Supplementary Information section (Table S 1.4).

For the lcWGS data, we compared heterozygosity between pairs of samples from different sites, with each pair involving one of the three unarmored populations and a corresponding low armored, lagoonal population near the coast (Figure 1.1). To perform this analysis, we used BCFtools v1.10.2 (Danecek et al., 2021) to calculate numbers of heterozygous and homozygous sites based on SNP data only, and then calculated the heterozygosity ratio by dividing nHets (RA) / nHom (AA) (R = reference allele; A = alternative/non-reference allele). We used this index because it is less sensitive to density of genotyping (Samuels et al., 2016). The ratio vary between 0-2, with 2 representing the highest level of genetic diversity based on Hardy-Weinberg equilibrium (Guo et al., 2014). We also calculated the number of singletons (i.e., unique variants) for each individual (Table 1.2).

Results

Microsatellites

Unarmored stickleback populations do not form an exclusive group at any *K*, and plots of the estimated ln (log-normal) probability of the data (D) at different *K* values (lnPD|*K*) and Δ*K* identified three clusters that roughly align with geography (Figure 1.1A-B). We report the

assignments for K = 2–7 in the Supplementary Information section (Figure S 1.1), as informative clustering patterns emerged at $K > 3$.

At K = 2 (north to south), one cluster corresponds to fish from Bodega Bay south along the central coast to Salsispuedes Creek (a tributary to the lower Santa Ynez River in Santa Barbara County), the upper Santa Ana River in the San Bernardino Mountains, San Juan Creek (Trabuco Creek and Bell Canyon tributaries), and Bocana El Rosario in Baja California. Except for the upper Santa Ana River, these coastal populations form what we refer to hereafter as the Outer Coast (OC) group. The second cluster consists exclusively of populations within the SCB, and includes the Ventura, Santa Clara, and San Jacinto rivers, Big Rock Creek, and three drainages in San Diego County (San Felipe Creek, Pine Valley Creek, and the Sweetwater River).

Further distinction of the upper Santa Clara River, San Jacinto River, Big Rock Creek and the three San Diego County drainages emerged at K = 3. The population in Bouquet Creek, a tributary to the upper Santa Clara River, is admixed with alleles from both the upper and lower sections of the Santa Clara mainstem due to contamination during trout stocking (see Discussion). At K = 3, we also detected that admixture in Mission Creek (Figure 1.1: #10) and Apple Valley (Figure 1.1: #20) reflects shared ancestry between the Santa Clara River (upper and lower reaches) and the OC group.

Successive increases in K continued to show strong association between low armored *G. aculeatus* in the lower Santa Clara and Ventura rivers; an affinity between the upstream Santa Ana River and San Juan Creek tributaries appeared at $K \geq 4$; and distinction of the unarmored population in San Antonio Creek-VSFB from all others became evident at $K \geq 6$ (Figure S 1.1).

Results of the PCA were consistent with the output from STRUCTURE (Figure 1.1B), with the percent variance explained by the first 10 principal components (PC) ranging from 14.64 to 2.15 (Figure 1.2). The OC group (plus upstream Santa Ana River) forms its own cluster, but with some distinction for San Antonio Creek-VSFB. Individuals in the SCB group are distributed along PC1, with the Ventura and lower Santa Clara river samples clustered on the lower left, and upstream Santa Clara samples clustered in the upper right. Admixed fish from Mission Creek and Apple Valley were intermediate between the OC and SCB clusters, and individuals from the upper part of the Bouquet Creek tributary were associated with fish from the lower Santa Clara River (i.e., Figure 1.2: #23).

Genomes

Individuals had an average of ~8X coverage, ranging from 6.2-10X (Table S 1.2). We recovered a total of 8,340,853 SNPs; missing data varied from 3.2-7.0% per individual. We reduced the dataset to 10,308 biallelic SNPs after removing linked loci, with the percent of variance explained ranging from 13.14 to 5.07 for the first 11 PCs. While the sampling was limited for lcWGS dataset, we still detected clear separation of the three unarmored stickleback populations, and a general affinity between up- and downstream fish in the same drainage regardless of plate morphology (Figure 1.3). Fish collected at the same locations tended to cluster in similar PCA space, with partially armored fish from the lagoon at San Antonio Creek-VSFB showing more divergence than all other pairs. The only drainage that did not involve a direct up- versus downstream comparison was the upper Santa Ana River and Bocana El Rosario (Baja California). Paired samples from these two locations were broadly separated along PC2, which contrasts with the PCA using microsatellites, where both were clustered in the OC group.

Historical demography under the MSMC

We present results for two combinations of mutation rate and generation time: $\mu = 3.7 \times 10^{-8}$; age = two years (Figure 1.4) and $\mu = 6.6 \times 10^{-8}$; age = one year (Figure S 1.2) and show that the shape of the IICR curve remains unchanged regardless of the parameter values used in the analysis. For clarity, populations referred to as 'upstream' are unarmored, whereas those from 'lagoons' are low armored.

Plots show a general congruence in curve shape between fish from the same sampling location, except for the upper Santa Clara River, which we attribute to differences in the amount of coverage (a more stringent filtering threshold produced curves with less overlap between the samples due to a higher percentage of missing data for one of the individuals; Figure S 1.3). Bootstrapped data showed little variation from the observed data, although there was slightly more uncertainty in the IICR curves at the deeper time intervals for upstream fish.

Initial population sizes appear larger at upstream sites compared to the lagoons, although estimates of both time and population size for the upstream Santa Ana River are more similar to lagoonal populations than to either of the upstream populations further north.

All fish reveal marked demographic declines from initially large population sizes but differed in the general shape of the IICR curves depending on whether the fish were from upstream areas or lagoons. Plots from lagoonal fish have more of a sigmoid shape, with an inflection in the middle part of the curve that reflects a transition from early, rapid decline to a period of mild "growth" (Figure 1.4). The timing and duration of these "growth" periods is roughly congruent across lagoons. In contrast, none of the upstream fish show evidence of a "growth" phase, although the Santa Ana River fish reveal a long period of stable demography, expressed as a plateau in the

IICR plot, that leads to a sigmoid-like curve shape. The chronology of this plateau also overlaps with the "growth" phase in lagoonal fish. In contrast, plots of the other upstream fish from San Antonio Creek-VSFB and the Santa Clara River reveal steeper and more continuous declines across their histories, with the upper Santa Clara River fish showing the steepest trajectory and no inflection in the curves at all.

Our re-analysis of data from *G. aculeatus* sampled in Alaska recovered the same overlap in IICR curves shown in Liu et al. (2016, Figure S 1.5). We also note that they look distinctively flatter, a pattern we attribute to lower depth of sequencing coverage (which is known to flatten IICR curve shapes: Nadachowska-Brzyska et al., 2016) and possibly the use of an updated SMC algorithm (Schiffels & Durbin, 2014). These results suggest that the non-overlapping curves between up- and downstream fish in this study is not an artifact of how the data was processed and analyzed, but rather a result of distinct and divergent demographic histories between populations from the same drainage.

Genetic diversity

Measurements of heterozygosity based on SNP data show that upstream fish have lower gene diversity (H_o) than lagoonal fish in the cases where within-drainage comparisons are made (e.g., San Antonio Creek-VSFB and in the Santa Clara River). In the comparison involving the upper Santa Ana River, H_o was comparable to lagoonal fish from Bocana El Rosario and the lower Santa Clara River, and considerably higher than either of the other unarmored stickleback populations (Table 1.2). Individuals with the highest heterozygosity were from the San Antonio Creek-VSFB lagoon and the lowest from upper Santa Clara River. The same pattern is observed for the number of singletons (Table 1.2) and is also reflected in the microsatellite data (Table S 1.3).

Discussion

Rapid decline of the freshwater fish fauna of southern California (Moyle et al., 2015) and northern Baja California (Ruiz-Campos et al., 2014) comes as no surprise given the extensive loss of habitat and permanent alteration of the region's natural hydrology over the past century (Swift et al., 1993a). Yet, results from this study indicate that demographic decline is more the norm than the exception for *G. aculeatus* in the region, even at the deepest time scales. This may reflect the realities of survival in a climate that has become increasingly more arid over the Holocene, combined with warmer conditions at subtropical latitudes that reduces the availability of perennial water. We explore these factors below by describing population structure and characterizing demographic histories across different parts of the study area. We then discuss the implications of the current data for management and suggest that the emphasis on plate phenotype has both aided and detracted from conserving the important genetic variation across this part of the species range.

Population structure across southern California

We detected three regional clusters within *G. aculeatus* across southern California and show that unarmored populations assign to separate groups. The clearest distinction is between populations within the SCB, specifically those from Ventura and Santa Clara river drainages, and those from the OC group. High admixture between Ventura River and lower Santa Clara River populations has been reported previously and is likely tied to the proximity of the river mouths, which are separated by ≤8 km of coastline and well within the dispersal distance of *G. aculeatus* in the ocean (Richmond et al., 2015). In contrast, the OC group has maintained its cohesion across a much larger expanse of coastline, consistent with long-distance movement in marine environments (Fang et al., 2018; D. H. Jones & John, 1978; Mäkinen et al., 2006).

An unexpected finding was the clustering of the upstream Santa Ana River population within the OC group (at K = 2–5, Figure 1.1B) given its inland and isolated location in the San Bernardino Mountains (elev. = 2055 m: Figure 1.1A). It is the only population that occurs at high elevation, more than ~1200 m higher than any other native *G. aculeatus* in the region. This retention of ancestral polymorphism and connection with the OC group, especially San Juan Creek (Trabuco Creek and Bell Canyon tributaries; $K \geq 4$, Figure S 1.1), suggests that the current isolation of the source population (i.e., Shay Creek) may be recent. However, the absence of any known hydrologic connection between Shay Creek and the coast (Flint et al., 2012; French & Busby, 1974) presents a challenge to this explanation (see next section for further details). Whole genome data now being collected will help shed more light on this issue.

Our results also reveal cases where naturally occurring populations of *G. aculeatus* have been mixed with non-local fish, usually through accidental introduction. One example is Bouquet Creek (Figure 1.1: #23–24;), where a resident population of unarmored *G. aculeatus* was mixed with low armored individuals from the lower Santa Clara River as a by-product of trout stocking (Richmond et al., 2015). A second example is Mission Creek (Figure 1.1: #10), which is situated along the Santa Barbara coastline between the OC sites and the mouth of the Santa Clara and Ventura rivers. In this case, admixture could be due to either accidental introduction with trout stocking or natural dispersal from the nearby Ventura or Santa Clara rivers (Figure 1.1).

Another set of populations are likely the result of pure incidental transport during trout stocking, with no evidence that *G. aculeatus* existed at the site previously. These include San Jacinto Creek (Figure 1.1: #30) and the Sweetwater River (Figure 1.1: #33), which also support populations of hatchery-sourced trout (Abadía-Cardoso et al., 2016). The population at Apple Valley (Figure 1.1: #29) (i.e., Mojave River) also fits this category; however, unarmored *G.*

aculeatus were also purposely introduced to this drainage prior to 1940, but then were eliminated by subsequent incidental transport of hatchery fish sometime thereafter (Buth et al., 1984; Miller & Hubbs, 1969; Swift et al., 1993a).

Still, a third set of populations involve cases of purposeful translocations aimed at safeguarding the gene pool at the type locality for *G. a. williamsoni* in Soledad Canyon (USFWS, 2021). The most well-documented example in this category is the San Felipe Creek population at the edge of the Anza Borrego Desert in San Diego County, where translocations occurred in 1972, 1973, and 1981 (Swift et al., 1993a). The population in Pine Valley Creek is also a suspected, intentional transplant based on archival memos and hearsay, and cluster assignments confirm that *G. aculeatus* in this drainage share the same genetic background as those from the upper Santa Clara River.

Historical demography varies with geography

The history of glaciation and transitions from wetter to drier climates in southern California makes it challenging to specify points in time that unequivocally coincide with transitions in the IICR curves, especially when mutation rates and generation times are estimated from *G. aculeatus* in other regions. However, qualitative comparison of the demographic trajectories can provide useful information about the approximate timing and location of population expansions and contractions, and some indication of how their contemporary status compares to the backdrop of their deeper history. Moreover, our results show similarities to those in Liu et al. (2016) for *G. aculeatus* from Vancouver Island, a possible glacial refuge near the edge of the unglaciated region at the Last Glacial Maximum (18-20 ky). These fish may represent a set of interacting populations that

extended from this area through the large unglaciated region to the south, including southern California and northern Baja California (Bell, 1976; Glover et al., 2021; L. Heusser, 1998).

A key piece of information provided by the MSMC analysis is that all populations of *G. aculeatus* have declined dramatically from early peaks, and that the patterns of decline show some consistency with respect to location (Figure 1.4). Lagoonal populations also show a later transient period of mild "growth" that varies slightly in chronology and duration among sites, whereas upstream unarmored stickleback populations do not. One explanation for these 'growth phases' is that they represent admixture with non-local migrants along the coast, which could offset declines in the IICR to produce inflections in the estimated trajectories. Other work documents the genetic connectivity of estuarine populations on the Pacific coast as a result of habitat expansion as sea level rose (Dolby et al., 2016, 2018, 2020; Stiller et al., 2021). Sea level rise would also lead to the formation and broad distribution of lagoons following maturation of the coast (D. Jacobs et al., 2011; Masters, 2006).

Lagoons provide freshwater habitat for most of the year but can open to the ocean during heavy rain events (D. Jacobs et al., 2011). The complex coastal geomorphology and hydrology of lagoons likely created dispersal opportunities that help explain the cohesion of the OC group, a pattern consistent with *G. aculeatus* worldwide where marine fish tend to show genetic uniformity across large geographic distances (Mäkinen et al., 2006). It may also explain why the lagoonal populations tend to have higher heterozygosity.

That the initial population sizes in the early IICR curves appear larger in upstream areas compared to lagoons could reflect a wetter landscape and cooler climate during the time when *G. aculeatus* first expanded into the region (Glover et al., 2021; L. Heusser, 1998). However, we interpret these relative population size estimates in the earliest part of the IICR curves with

skepticism (see *Caveats to the MSMC* below). Wetter and cooler conditions within the SCB were prevalent from ~65 ka up to ~14 ka (Faribanks & Carey, 1910; Glover et al., 2017, 2020; L. E. Heusser et al., 2015; Kirby et al., 2013; Owen et al., 2003; Sharp et al., 1959) and likely provided more expansive freshwater habitat that could have sustained larger and more interconnected populations of *G. aculeatus* in inland areas. In fact, habitat supporting unarmored *G. aculeatus* extended well into the Los Angeles Basin as recently as the mid-20th century (Culver & Hubbs, 1917; Mendenhall, 1908; Swift et al., 1993a).

In contrast to lagoonal fish, none of the upstream fish show evidence of a reversal in demographic decline, although the IICR for upper Santa Ana River samples reveals an interval of stable size that produces a sigmoid-shaped curve that more closely resembles OC fish. The remaining unarmored fish in San Antonio Creek-VSFB and the Santa Clara River instead show a steeper, more continuous decline across their full histories, with the latter showing no inflection in the IICR at all. We interpret these steeper, more linear trajectories as evidence of isolation with limited or no gene flow, particularly the upper Santa Clara River, where several large dry gaps separate the population from lower reaches of the main channel (Richmond et al., 2015). Multiple molecular data sources confirm that, while fish occasionally get flushed downstream across these gaps, there is no evidence of genes moving upstream (Buth et al., 1984; Richmond et al., 2015). Susceptibility to recurrent bottlenecks due to fluctuating hydrology and El Niño-related climate events (reviewed in Richmond et al., 2015), and more recently increased fire frequency (Flint et al., 2019), may also be reducing the IICR of these fish at a contemporary time scale. It is likely that this history of isolation between upstream and lagoonal fish and sequential bottlenecking from fluctuating climatic events led to distinct and divergent demographic histories, as captured by the lack of overlap in their IICR curves.

The greater similarity between unarmored *G. aculeatus* in the upper Santa Ana River and low armored fish in the OC group, in terms of the shape of the IICR plot (more sigmoid than linear), age and size of the founder population, chronology of the transition in the IICR, level of heterozygosity, and clustering affinity, is peculiar given the isolation and high elevation of the source population at Shay Creek in the San Bernardino Mountains. Two hypotheses explain this conundrum. The first is that Shay Creek supports a naturally occurring, relictual population that became isolated during uplift of the San Bernardino Mountains, and that suitable freshwater habitat has potentially persisted in the upland plateau as far back as the Last Glacial Maximum (18-20 ka; Owen et al., 2003; Sharp et al., 1959). However, Shay Creek drains northward into Baldwin Lake instead of toward the coast via the Santa Ana River drainage basin, challenging this view of natural genetic connectivity with outer coast fish.

A second, more plausible hypothesis, is that the Shay Creek population is introduced. This better explains the presence of outer coast alleles in these small, isolated ponds, and why it is the only population to occur at high elevation. Sticklebacks have also been introduced to the Big Bear Reservoir just 3 km to the west of Baldwin Lake and there has been at least one accidental and several known introductions of Shay Creek fish to other high elevation ponds in the area (Swift et al., 1993a). These introductions further attest to the ease and regularity that this species has been moved artificially, although a definitive outer coast source(s) for the Shay Creek population has yet to be identified.

Caveats to the MSMC

There are limits as to how far back in time the MSMC can reliably infer demographic parameters, as alleles with deep coalescence become increasingly rare at older time scales (Beichman et al.,

2017; Mazet et al., 2016; Takahata & Nei, 1985). This explains the greater noise in the bootstrapped data, and perhaps the differences in IICR between upstream and lagoonal populations at the oldest time intervals. For this reason, IICR estimates for early population history are often viewed with skepticism (Beichman et al., 2017). Natural selection for specific mutations and any linked neutral variants can also distort the IICR, particularly if the type of selection reduces polymorphism (Ewing & Jensen, 2016; Schrider et al., 2016), and there is good reason to assume that selection is a factor in this system. Phenomena that contribute to population structure in mating/migration also act as confounding factors, and may be expected in this case, given the known demographic fluctuation of landlocked populations in upstream areas (Moyle, 2002). For example, inbreeding (as a nonrandom mating process) can affect interpretation of the IICR curve because it increases the rate of coalescence and leads to a reduction in N_e. Sequences from the larger ongoing WGS study will allow us to screen for long runs of homozygous sequence, in which case we can remove such runs and repeat the analyses to test the sensitivity of our results (Freedman et al., 2014; Mather et al., 2020).

Rethinking management

Our results are consistent with separate origins for unarmored populations in southern California, although phylogenetic analyses are needed to polarize the relationships. Nonetheless, the current data provide evidence that the morphological entity now classified as *G. a. williamsoni* consists of distinctive genetic units, and that managing according to these units may be prudent.

A number of partners have focused management on unarmored fish from the type locality in Soledad Canyon (upper Santa Clara River) (USFWS, 2021), and until now there were questions as to whether this made sense from a genetics perspective. The answer is yes, given that the

population has low genetic diversity, is geographically isolated, and the habitat is ecologically unstable and subject to persistent human disturbance (e.g., fires, garbage dumps, recreational vehicle abandonment, illegal water diversions, homeless encampments, off-highway vehicle activity, etc.). Climate change may also be impacting the population as perennial water no longer occurs at the type locality of *G. a. williamsoni* (Girard, 1854). Whether low genetic diversity is cause for alarm requires further study, as the population may have survived for much of its history with low diversity, and deleterious alleles that were potentially hidden from selection in the heterozygous state could have been purged over time due to drift (Robinson et al., 2018). If it can be shown that specific polymorphism is responsible for parallel loss of body armor, managing the three unarmored populations as separate units may provide greater assurance that those alleles will be preserved.

Increased knowledge about the demographic history of *G. aculeatus* in southern California also lays the groundwork for understanding how these fish have adaptively responded to selection, as the evolutionary potential of populations may rely on their pre-existing genetic variation (Barrett et al., 2008; F. C. Jones, Grabherr, Chan, Russell, Mauceli, Johnson, Swofford, Pirun, Zody, & White, 2012; Lai et al., 2019). Pre-existing variation may be key to understanding the complete loss of body armor, a rare condition in *G. aculeatus* worldwide that could be related to the absence of predators, as it has been extensively documented in the transition from armored to low armored morphs (e.g. Eriksson et al., 2021; Paccard et al., 2018; Reimchen, 1994; Wasserman et al., 2020, and references therein). However, whether the unarmored condition is driven by predation or lack thereof remains unresolved (Reimchen, 1994).

Considerable knowledge exists on the genetic underpinnings of plate reduction in fully armored fish (typically marine), where parallel evolution of partially armored fish in freshwater

habitat is largely the product of selection for an allele of the ectodysplasin gene (EDA), a signaling protein that is important for the development of the skeleton, skin and other tissues (Colosimo et al., 2005; Cresko et al., 2004; O'Brown et al., 2015b). The allele is recessive and rare in the marine environment and presents a striking example of how standing variation can lead to rapid shifts in phenotype depending on the selective environment (Barrett et al., 2008). Earlier work found no association between EDA polymorphism and unarmoredness, suggesting that other genes or regulatory regions may be involved with the complete loss of plates (Richmond et al., 2015). This subject will be explored in further detail in our ongoing WGS project.

Results of this work also raise the question of whether low armored stickleback populations might also be the focus of conservation efforts, given the genetic distinctiveness and geographic isolation of some populations in the SCB. For example, low armored *G. aculeatus* in San Juan Creek (Figure 1.1: #34-35) share close ancestry with the unarmored stickleback population now in the San Bernardino Mountains and are the only remaining representatives of the OC group in the Los Angeles Basin. The low armored stickleback population at Bocana El Rosario is also unique in representing the southernmost tip of the species' distribution in the eastern Pacific Ocean. However, it is currently considered extirpated (Ruiz-Campos & González-Acosta, in press). The only other extant population occurring in Mexico is in El Descanso lagoon, Rosarito (Ruiz-Campos et al., 2014). This edge population may still harbor critical genetic diversity that is adaptive in transitional environments that define the range edge itself.

These findings highlight the importance of using population genetic and genomic data to re-examine management strategies that were initially developed according to phenotype-based taxonomies. In this case, efforts to manage populations with a rare phenotype have likely helped to preserve important adaptive polymorphism, but emphasis on the unarmored phenotype alone

may exclude other populations in the region that have geographically unique variation that was once widespread across the coastal lagoons and drainages of southern California and northern Baja California.

Tables

Table 1.1: Location and plate phenotype data. Map ID = map number in Figure 1.1A. Numbers with asterisk (*) represent the 12 individuals selected for the lcWGS analysis. Genetic Groups (*K*=3): Outer Coast (OC); Southern California Bight (SCB); Lower Santa Clara/Ventura (L); Upper Santa Clara (U); Admixed (Ad).

Map ID	Population	Drainage	County	Genetic Group	Armor phenotype	Lat	Long
1	Bodega Bay	Bodega Bay	Sonoma	OC	fully armored	38.327	-123.055
2	Bothin Marsh	Mill Valley	Marin	OC	low armored	37.888	-122.524
3	Toro Creek	Toro Creek	San Luis Obispo	OC	low armored	35.413	-120.872
4	San Luis Obispo Creek	San Luis Obispo Creek	San Luis Obispo	OC	low armored	35.209	-120.696
5	Arroyo Grande Creek	Arroyo Grande Creek	San Luis Obispo	OC	low armored	35.099	-120.628
6	Shuman Lagoon	Shuman Canyon	Santa Barbara	OC	low armored	34.845	-120.597
7*	San Antonio Creek Lagoon	San Antonio Creek	Santa Barbara	OC	low armored	34.795	-120.621
8*	San Antonio Creek-VAFB	San Antonio Creek	Santa Barbara	OC	unarmored	34.782	-120.530
9	Salsipuedes Creek	Santa Ynez River	Santa Barbara	OC	low armored	34.621	-120.423
10	Mission Creek	Mission Creek	Santa Barbara	OC-Ad	low armored	34.466	-119.710
11	Matilija Creek	Ventura River	Ventura	SCB-L	low armored	34.501	-119.344
12	San Antonio Creek	Ventura River	Ventura	SCB-L	low armored	34.433	-119.251
13	Seaside Park	Ventura River main stem	Ventura	SCB-L	low armored	34.282	-119.309
14	Foster Park	Ventura River main stem	Ventura	SCB-L	low armored	34.352	-119.307
15*	McGrath	Santa Clara River Lagoon	Ventura	SCB-L	low armored	34.230	-119.261
16	Oxnard	Santa Clara River main stem	Ventura	SCB-L	low armored	34.241	-119.192
17	Upper Sespe Creek	Santa Clara River	Ventura	SCB-L	low armored	34.558	-119.253
18	Lower Sespe Creek	Santa Clara River	Ventura	SCB-L	low armored	34.406	-118.932
19	Piru Creek	Santa Clara River	Ventura	SCB-L	low armored	34.417	-118.790
20	Newhall Ranch	Santa Clara River main stem	Los Angeles	SCB-U	low armored	34.435	-118.603

Map ID	Population	Drainage	County	Genetic Group	Armor phenotype	Lat	Long
21	Valencia	Santa Clara River main stem	Los Angeles	SCB-U	low armored	34.427	-118.577
22	San Francisquito Creek	Santa Clara River	Los Angeles	SCB-U	unarmored	34.546	-118.516
23	Upper Bouquet Creek	Santa Clara River	Los Angeles	SCB-L-Ad	low armored	34.554	-118.416
24	Lower Bouquet Creek	Santa Clara River	Los Angeles	SCB-L-Ad	low armored	34.510	-118.451
25*	Soledad Canyon, Robin's Nest	Santa Clara River main stem	Los Angeles	SCB-U	unarmored	34.438	-118.277
26	Soledad Canyon, Thousand Trails	Santa Clara River main stem	Los Angeles	SCB-U	unarmored	34.442	-118.211
27	Big Rock Creek	Mojave River	Los Angeles	SCB-U	low armored	34.452	-117.856
28*	Santa Ana River, Sugarloaf Meadow	Santa Ana River	San Bernardino	OC	unarmored	34.178	-116.830
29	Apple Valley	Mojave River	San Bernardino	OC-Ad	low armored	34.524	-117.277
30	San Jacinto River	Lake Elsinore	Riverside	SCB-U	low armored	33.736	-116.819
31	San Felipe Creek	Salton Sea	San Diego	SCB-U	unarmored	33.098	-116.473
32	Pine Valley Creek	Otay River	San Diego	SCB-U	unarmored	32.830	-116.552
33	Sweetwater River	Sweetwater River	San Diego	SCB-U	low armored	32.885	-116.600
34	Trabuco Creek	San Juan River	Orange	OC	low armored	33.563	-117.651
35	Bell Creek	San Juan Creek	Orange	OC	low armored	33.629	-117.555
36*	Bocana El Rosario	El Rosario River	San Quintin, Baja California	OC	low armored	30.041	-115.788

Table 1.2: Heterozygosity ratio and proportion of singletons for the 12 lcWGS. Column shows values for both individuals analyzed at each location (lower/higher value).

Map ID	Locality	PCA ID	HetRatio	Numbers of singletons
7	San Antonio Ck (lagoon)	SAC_L	0.69/0.83	224,330/505,449
8	San Antonio Ck VSFB (upstream)	SAC_U	0.19/0.29	47,5461/56,900
15	Santa Clara River (lagoon)	SCR_L	0.51/0.55	115,463/157,823
25	Santa Clara River (upstream)	SCR_U	0.12/0.15	34,392/40,937
28	Santa Ana River (upstream)	SAR_U	0.43/0.51	255,408/323,017
36	Bocana El Rosario (lagoon)	ERO_L	0.40/0.44	133,055/121,347

Figures

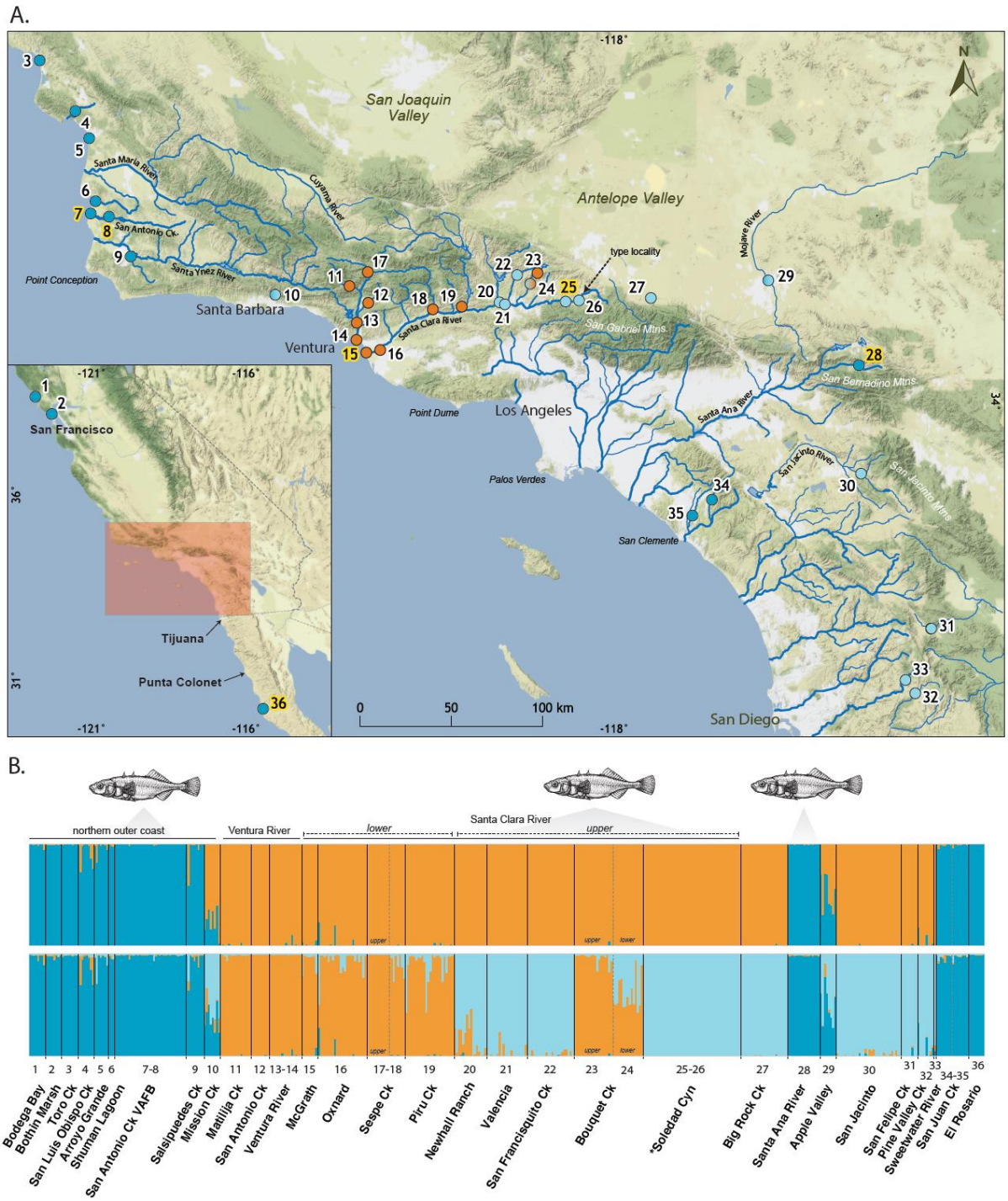


Figure 1.1: Map showing (A) sampling localities and (B) estimates of individual assignments (K = 2-3). (A) Red square on the map inset highlights the Southern California Bight (SCB). Sample locations (circles) are color coded according to the assignments at K = 3, and numbers highlighted in yellow indicate the sites used for lcWGS. Locations of each sampling point are provided in Table 1. (B) (lnPD|K) and ΔK identified

three as the best number of clusters, which roughly align with geography: Light blue = upstream Santa Clara River; orange = downstream Santa Clara and Ventura Rivers; dark blue = outer coast (OC) group, including the upper Santa Ana River. Cartoon fish lacking body armor denote the three extant populations of unarmored *G. aculeatus*, which does not form a single cohesive group at any *K* (Figure S 1.1). Lighter dashed lines in the assignment plots indicate upper and lower tributary reaches. Map was made on QGIS 3.16.2 and edited in Adobe Illustrator. River dataset from the National Weather Service (NOAA).

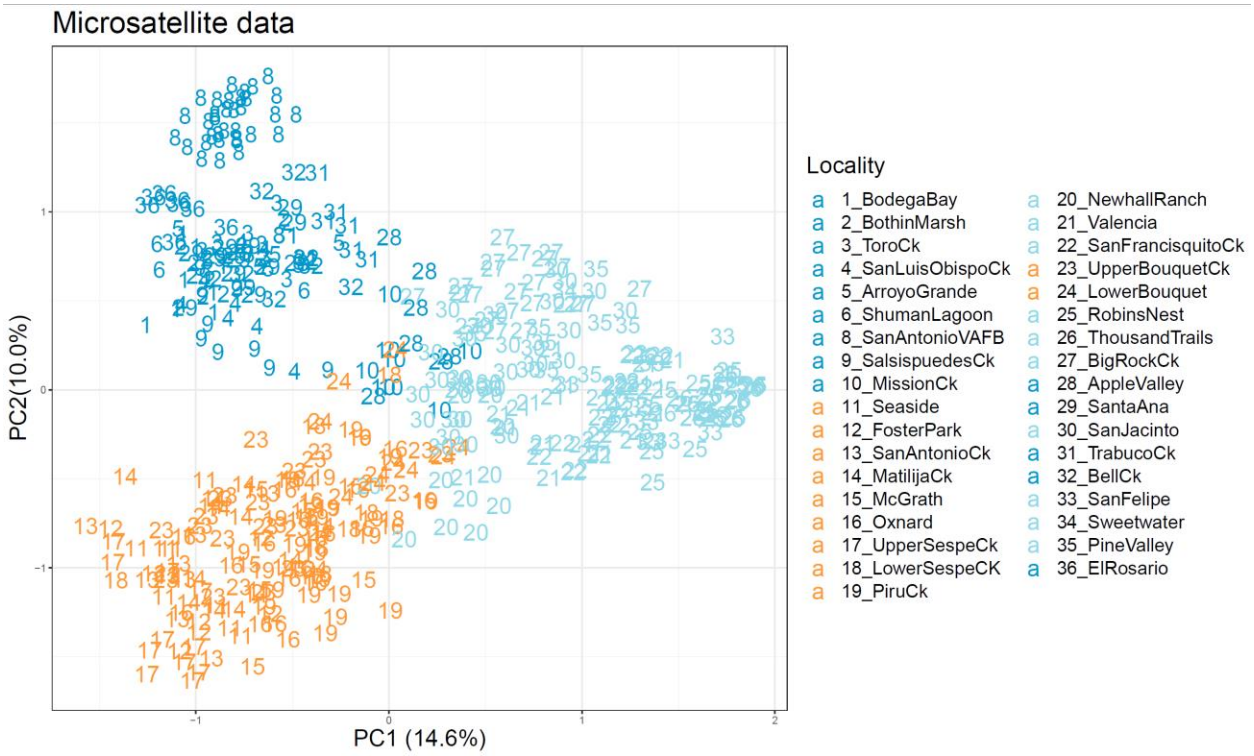


Figure 1.2: PCA plot of 470 individuals in the microsatellite dataset. PC1 explains 14.6% of the variance; PC2 explains 10.0% of the variance. The coloring scheme follows the assignment plots in Fig 1B for K = 3; unarmored populations = 8, 20-22, 25-26, and 29. The clustering pattern confirms the STRUCTURE results (Figure 1.1B), with three separate groups roughly matching the populations' geographic distribution. Admixed fish from Mission Creek (10) and Apple Valley (28) are clustered together in the middle of the plot.

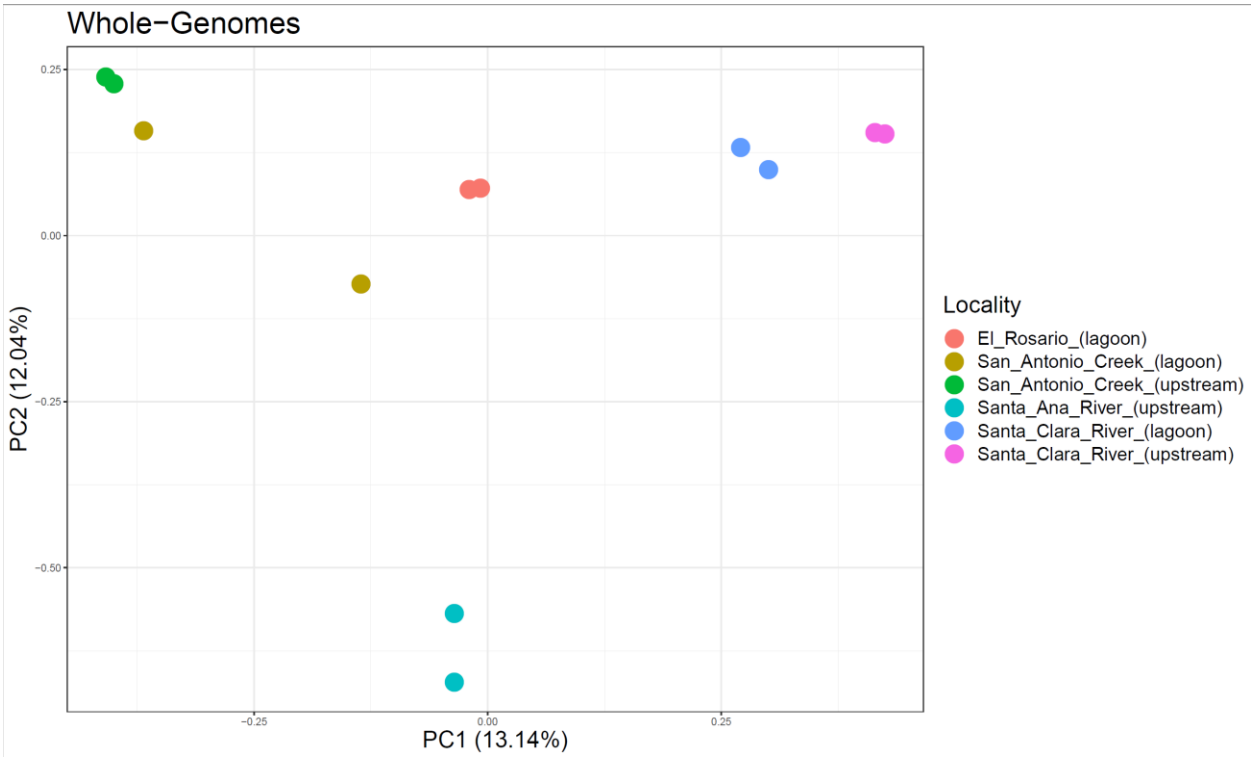


Figure 1.3: PCA plot of the 12 individuals from the lcWGS analysis. PC1 explains 12.62% of variance; PC2 explains 11.65% of the variance. Although sampling was limited for the lcWGS, we still detected the separation of the three unarmored populations, and drainage-level affinity between lagoonal and upstream fish. Fish from Bocana El Rosario and upstream Santa Ana River are separated along PC2, which differs from the microsatellite PCA, where both were clustered in the OC group.

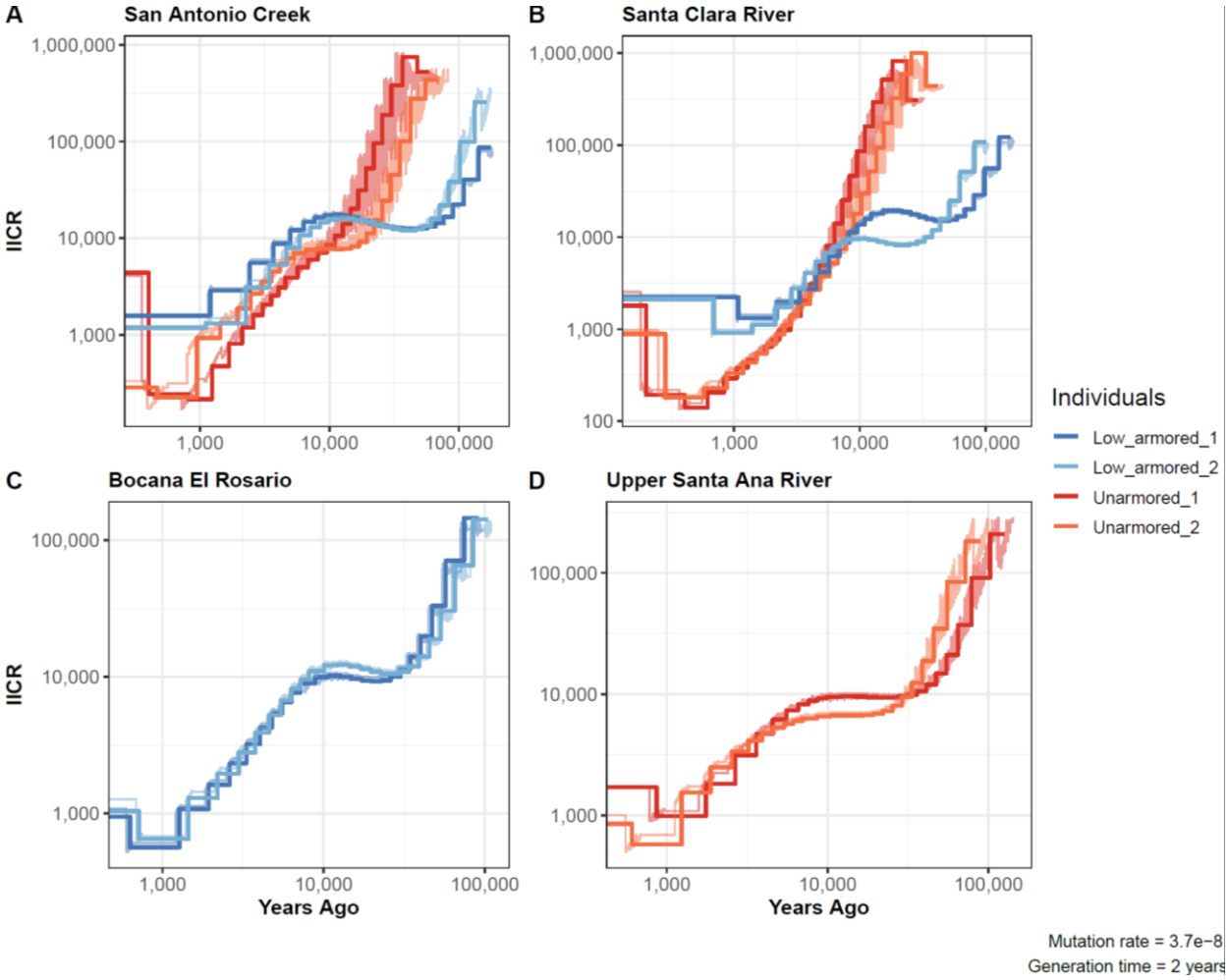


Figure 1.4: MSMC plots of upstream, unarmored (red) and lagoonal, low armored (blue) individuals. Thick lines are the point estimates, and faint lines represent the bootstrap replicates. Age and IICR estimates reflect a generation time of 2 years and mutation rate of 3.7 x 10^{-8} respectively. Plots for Bocana El Rosario and upper Santa Ana River are shown separately because they occur in different drainage basins (see text for additional details). There is lack of congruence between the curves of fish from the same drainage (lagoonal/upstream), which is the result of divergent demographic history.

Supplementary Information

Library Preparation

Extracted DNA was purified with a Zymo-DCC-5 column (Zymo Research) to remove impurities that could affect library preparation. Approximately 80 ng of purified DNA was subject to library preparation using the sparQ DNA Frag & Library Prep Kit (Quanta Bio) according to the manufacturer's recommendations. Final libraries were quantified using the Qubit dsDNA BR assay (Life Technologies) and quality control was performed with the D1000 Assay on an Agilent TapeStation 2200 (Agilent Technologies). The final libraries were diluted to normalize the concentration and sequenced as 100bp single end reads on a HiSeq4000.

Variant Calling and Filtering

The 12 samples were sequenced on an Illumina HiSeq 2000 at the Broad Stem Cell Research Center (BSCRC) at UCLA, producing an average of 3.75 Gb of 100bp single end reads for each sample. We aligned the reads to the Bear Paw Lake threespine stickleback reference genome (https://www.broadinstitute.org/stickleback/stickleback-genome-project) (F. C. Jones, Grabherr, Chan, Russell, Mauceli, Johnson, Swofford, Pirun, Zody, White, et al., 2012) using BWA-MEM v. 0.7.12-r1039 (H. Li, 2013b). Variant discovery was done using HaplotypeCaller and GenotypeGVCFs from GATK v3.6-0-g89b7209 (McKenna et al., 2010), using default parameters. We applied the following filters using GATK's VariantFiltration:

- for SNPs: QD < 2.0, MQ < 40.0, FS > 60.0, MQRankSum < -12.5, ReadPosRankSum < -8.0;
- for INDELs: QD < 2.0, FS > 200.0, ReadPosRankSum < -20.0, InbreedingCoeff < -0.8.

We then used BCFtools v1.10.2 (Danecek et al., 2021) and the function stats on the filtered VCF file to calculate summary statistics. Individuals had an average of ~8X coverage, ranging from 6.2-10X (Table S 1.2). We recovered a total of 8,340,853 SNPs; missing data varied from 3.2-7.0% per individual.

Supplementary Tables

(See attached supplemental material)

Table S 1.1: Summary quality report of BAM files for each individual generated using Qualimap. Each individual was pooled in two separate lanes. Numbers 7 and 8 after each sample name refer to the sequencing lane.

Table S 1.2: Summary statistics of the filtered VCF file generated using bcftools (stats command). Each column reports the number of variants, except Avg. depth.

Table S 1.3: Individual summary of missing data percentage averaged across all chromosomes for each level of read depth (DP) filtration.

Table S 1.4: Estimates of heterozygosity for sites with >8 samples. Pop ID = population identity (see Table S 1.1 and Figure 1.1); *N* = no. of samples; No. *A* = no. of observed alleles; Ho = observed heterozygosity; Hs = heterozygosity within populations, or gene diversity. Note that for all but one population, H_s was equal to or greater than Ho.

Supplementary Figures

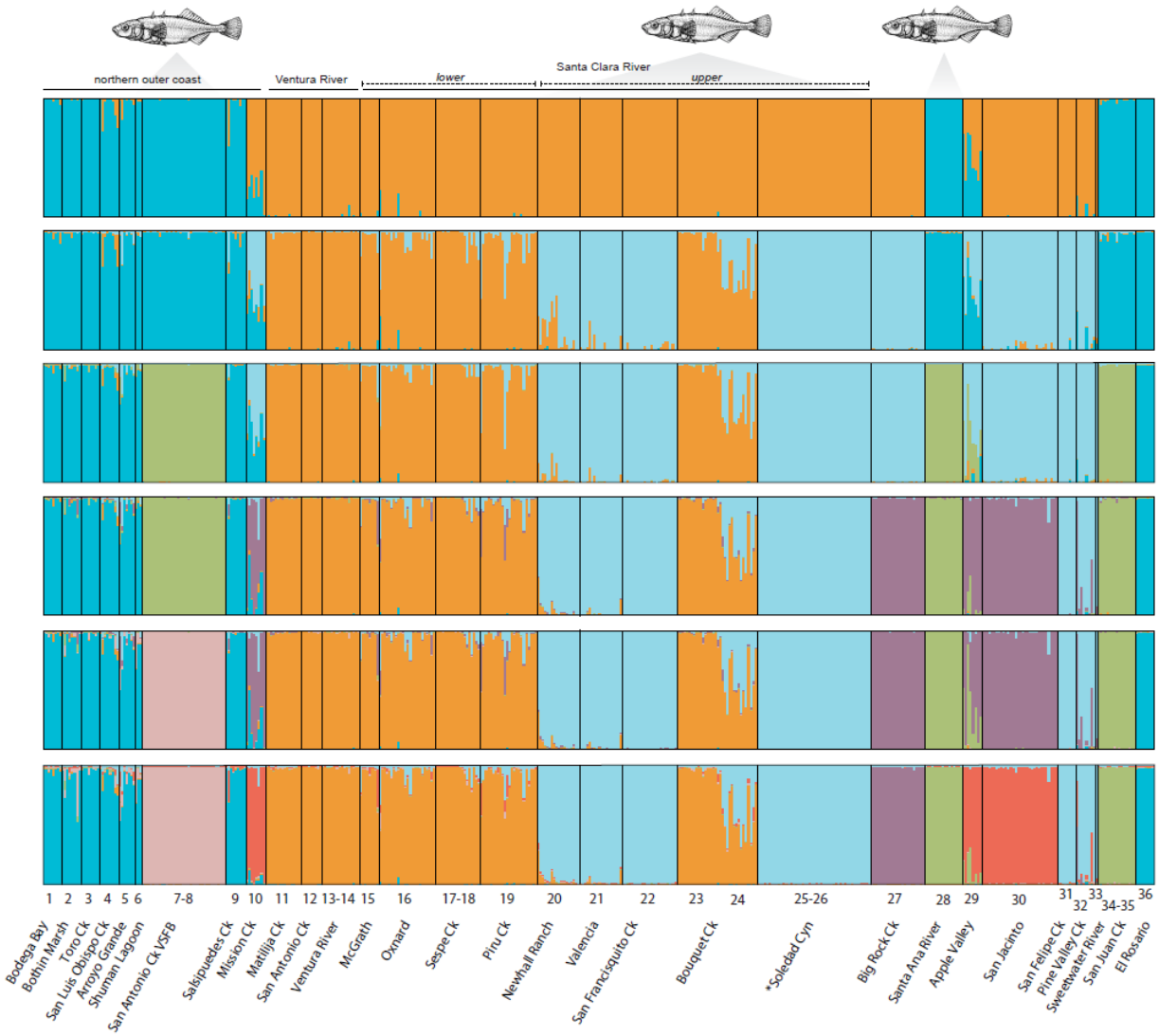


Figure S 1.1: Assignment plots using STRUCTURE from *K* = 2-7. Cartoon fish denote the three extant populations of *G. a. williamsoni*. *K* = 2-3 is presented in more detail in the main paper (Figure 1.1B). At *K* = 4-5, upstream fish from Santa Ana River show affinity with San Juan Ck individuals and San Antonio Ck. However, at *K* = 6-7 the latter gets distinctive from all the rest. At higher *K* (>4), southern outer coast fish start getting more differentiated from northern OC except for El Rosario.

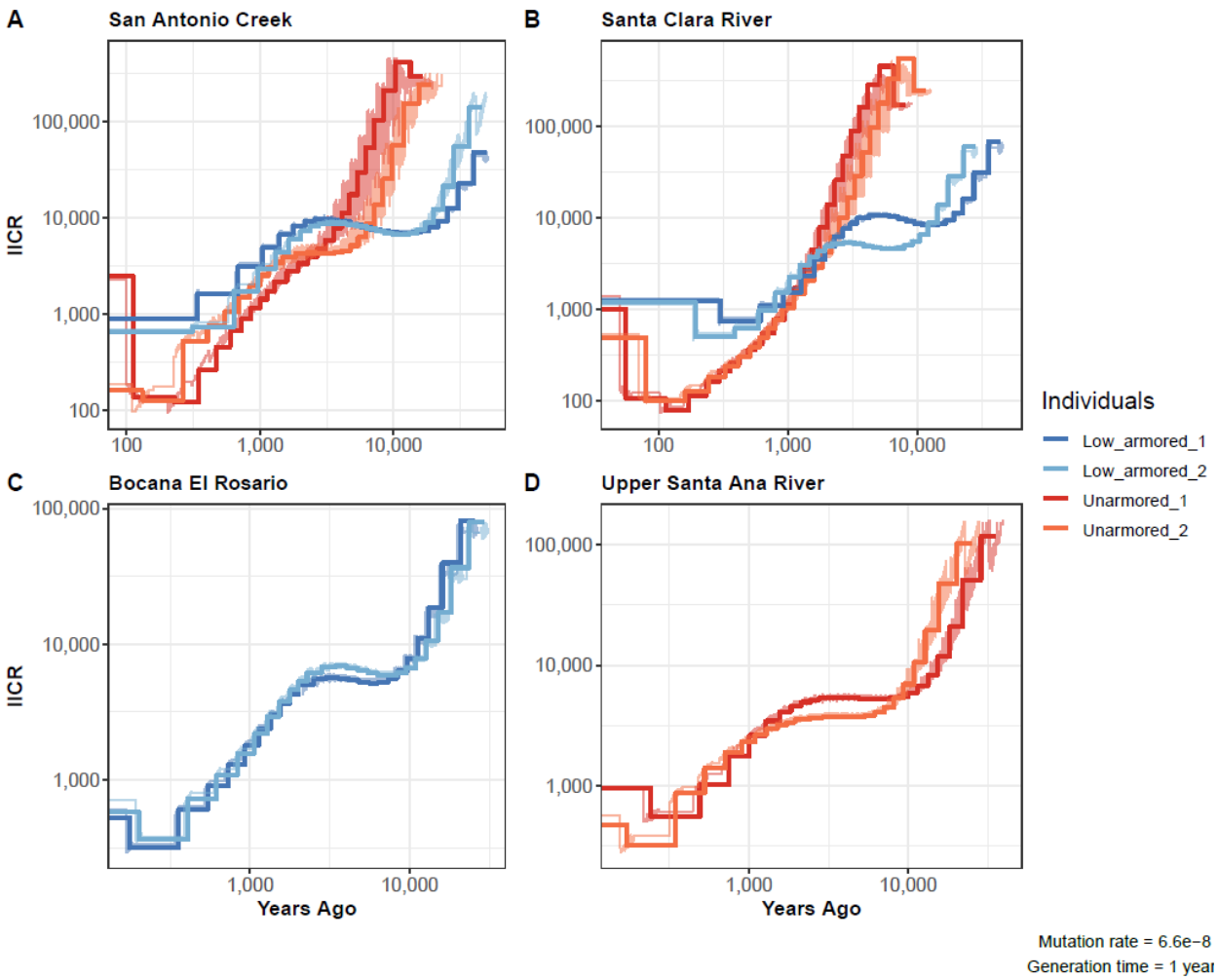


Figure S 1.2: MSMC plots of upstream, unarmored (red) and lagoonal, low armored (blue) individuals generated using default parameters and per-site filtering >= 8. The thick line is the point estimate, and faint lines represent the 100 bootstraps. Scaling for real time (in years) and IICR was done with generation time of 1 year and mutation rate of 6.6x10-8 respectively. Plots show upstream (unarmored) and lagoonal (low armored) fish superimposed for (A) San Antonio Creek and (B) Santa Clara River. Curves are the same as in Figure 1.4, but with differences in the scale of both axes. Shorter generation time decreases the TMRCA, and higher mutation rate increases IICR.

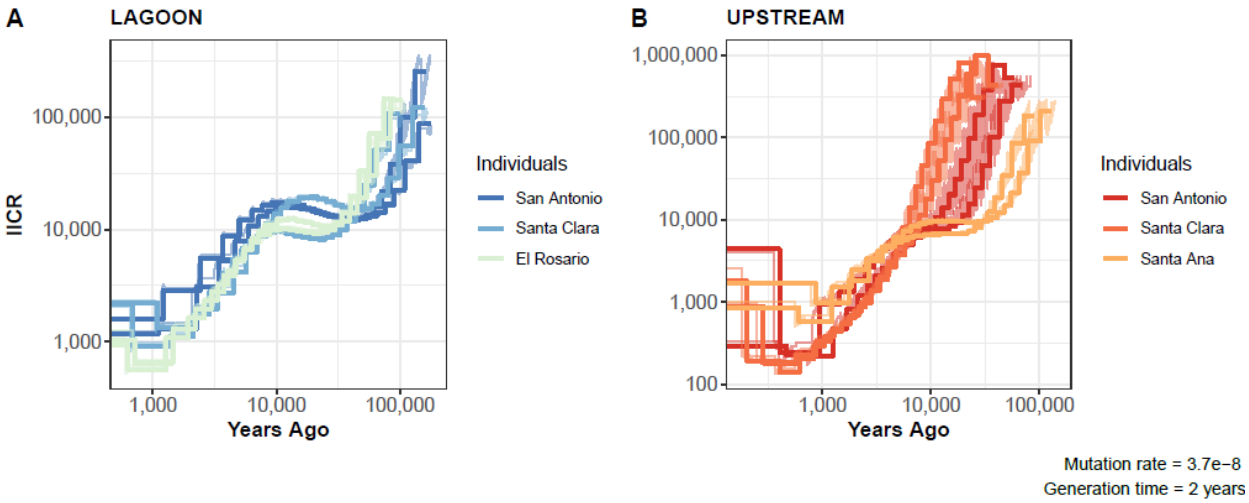


Figure S 1.3: MSMC plots of combined (A) lagoonal, low armored individuals and (B) upstream, unarmored individuals. Scaling is the same as in Figure 1.4 (main text): μ = 3.7 x 10-8; age = 2 years.

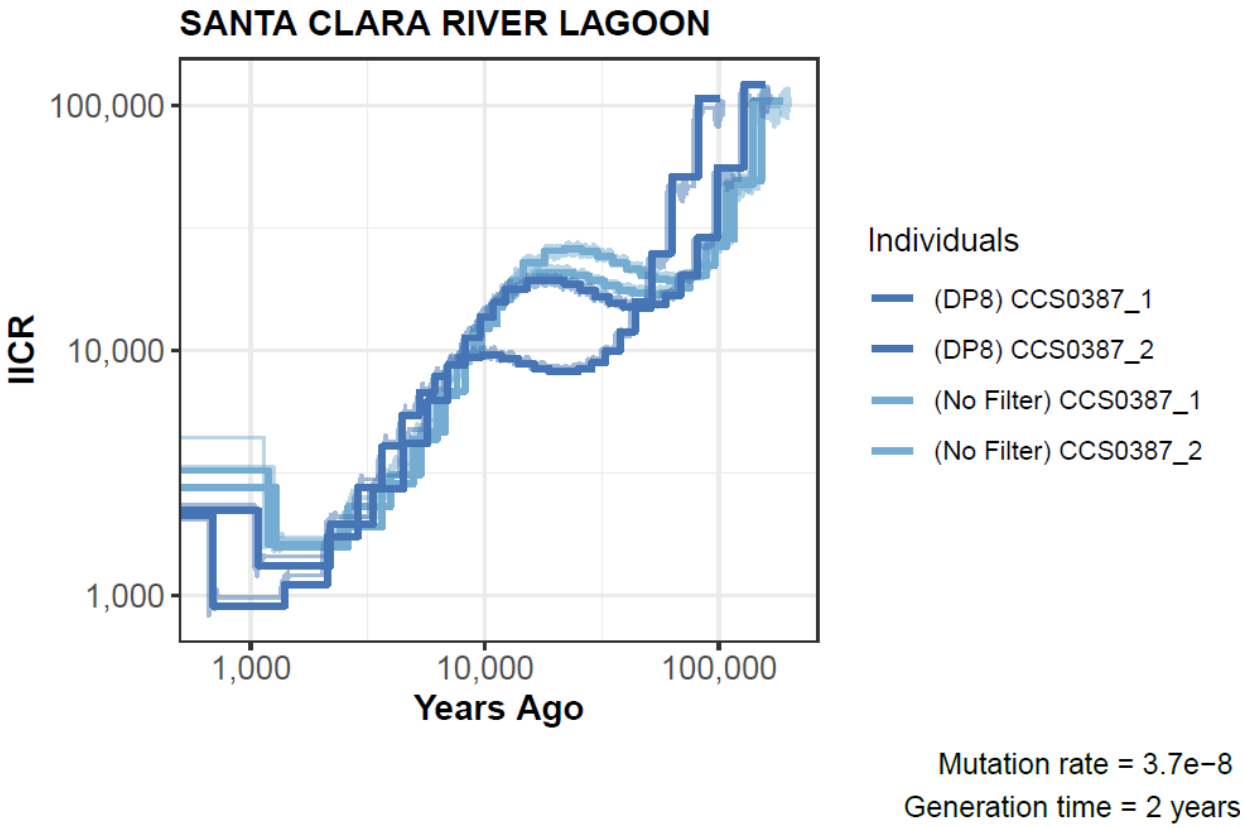


Figure S 1.4: MSMC plots of lagoonal (low armored) individuals from the Santa Clara River generated using default parameters. The thick line is the point estimate, and faint lines represent the 100 bootstrap replicates. Scaling for real time (in years) and IICR was done with generation time of 2 years and mutation rate of $3.7x10^{-8}$, respectively. Darker blue represents individuals after per-site filtering >= 8. Light blue represents the same individuals without filtering. Individual CCS0387_2 shows a stronger deviation in its curve compared to individual CCS0387_1 due to a higher percentage of missing data after filtration (41% versus 18%, respectively).

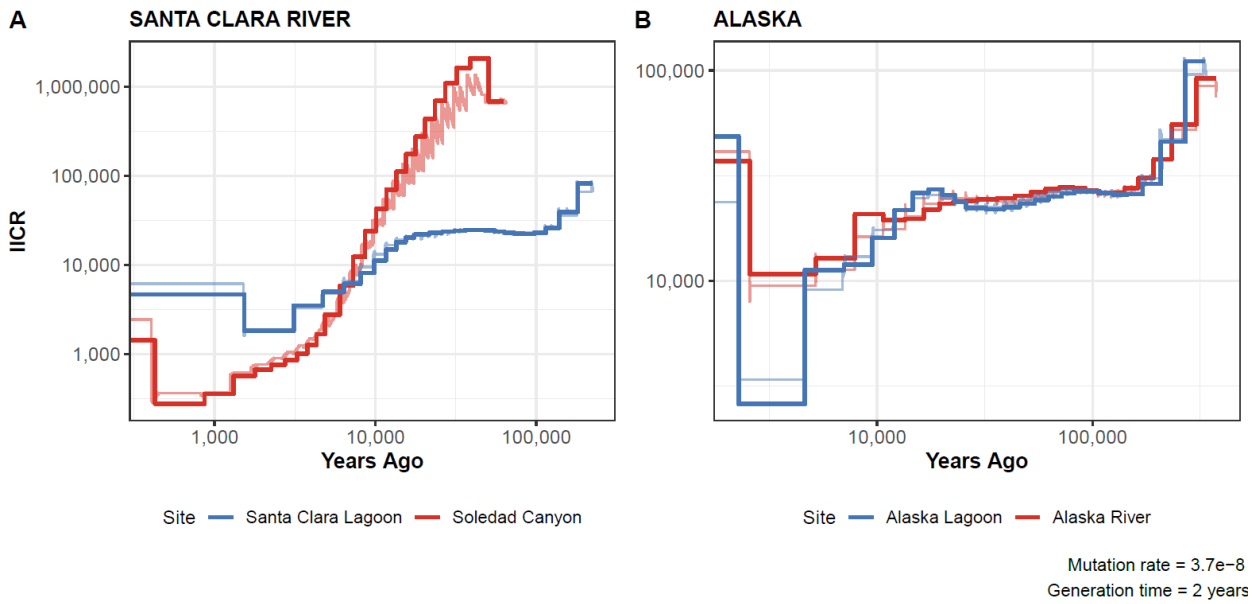


Figure S 1.5: MSMC plots from two individuals from our dataset (A: Santa Clara River) and two individuals from Liu et al. (2016) work (B: Alaska). Genomes from Alaskan individuals had their site coverage filtered to fall within our samples' site coverage (8<DP>20). As it can be seen, the curves from the Alaskan individuals are flatter than in their original publication, which could be due to artifacts from the lower coverage employed here, as it has been shown in Nadachowska-Brzyska et al. (2016) and discussed in the main text. It could also be related to the different algorithm employed in this work. Here, we used the MSMC method, which uses the SMC' algorithm, while Liu et al. (2016) used the PSMC method. Nevertheless, the lack of overlap between our samples does not seem to be an issue with our analysis, since we were able to recover it for the Alaskan samples. This lack of overlap is likely due to the demographics of these populations in southern CA, reflecting ancient isolation of upstream populations, and sequential admixture events between lagoonal and potentially marine/anadromous fish.

Bibliography

Alberdi, A., Aizpurua, O., Gilbert, M. T. P., & Bohmann, K. (2018). Scrutinizing key steps for reliable metabarcoding of environmental samples. *Methods in Ecology and Evolution*, *9*(1), 134–147. https://doi.org/10.1111/2041-210X.12849

Ballard, J., Pezda, J., Spencer, D., & Plantinga, A. (n.d.). *An Economic Valuation of Southern California Coastal Wetlands*. http://scwrp.org/wp-content/uploads/2017/06/SoCalWetlands_FinalReport.pdf

Barrett, R. D., Rogers, S. M., & Schluter, D. (2008). Natural selection on a major armor gene in threespine stickleback. *Science*, *322*(5899), 255–257.

Baumgartner, J. V. (1986). The genetics of differentiation in a stream population of the threespine stickleback, Gasterosteus aculeatus. *Heredity*, *57*(2), 199–208.

Beichman, A. C., Phung, T. N., & Lohmueller, K. E. (2017). Comparison of Single Genome and Allele Frequency Data Reveals Discordant Demographic Histories. *G3: Genes, Genomes, Genetics*, *7*(11), 3605–3620. https://doi.org/10.1534/g3.117.300259

Bell, M. A. (1976). Evolution of Phenotypic Diversity in Gasterosteus Aculeatus Superspecies on the Pacific Coast of North America. *Systematic Biology*, *25*(3), 211–227. https://doi.org/10.2307/2412489

Bell, M. A. (1978). Fishes of the Santa Clara River system, southern California. *Natural History Museum of Los Angeles County Contributions in Science*, *295*, 1–20.

Bell, M. A. (2001). Lateral plate evolution in the threespine stickleback: Getting nowhere fast. In A. P. Hendry & M. T. Kinnison (Eds.), *Microevolution Rate, Pattern, Process* (Vol. 8, pp. 445–461). Springer.

Bell, M. A., & Foster, S. A. (1994). *The Evolutionary Biology of the Threespine Stickleback.pdf*. Oxford University Press.

Bell, M. A., Orti, G., Walker, J. A., & Koenings, J. P. (1993). Evolution of Pelvic Reduction in Threespine Stickleback Fish: A Test of Competing Hypotheses. *Evolution*, *47*(3), 906. https://doi.org/10.2307/2410193

Bell, M. A., & Richkind, K. E. (1981). Clinal Variation of Lateral Plates in Threespine Stickleback Fish. *The American Naturalist*, *117*(2), 113–132. https://doi.org/10.1086/283693

Benjamin, A., May, B., O'Brien, J., & Finger, A. J. (2016). Conservation Genetics of an Urban Desert Fish, the Arroyo Chub. *Transactions of the American Fisheries Society*, *145*(2), 277–286. https://doi.org/10.1080/00028487.2015.1121925

Benke, A. C. (1990). A Perspective on America's Vanishing Streams. *Journal of the North American Benthological Society*, *9*(1), 77–88. https://doi.org/10.2307/1467936

Bergstrom, C. A. (2002). Fast-start swimming performance and reduction in lateral plate number in threespine stickleback. *Canadian Journal of Zoology*, *80*(2), 207–213. https://doi.org/10.1139/z01-226

Bohmann, K., Chua, P., Holman, L. E., & Lynggaard, C. (2021). DNAqua-Net conference unites participants from around the world with the quest to standardize and implement DNA-based aquatic biomonitoring. *Environmental DNA*, *3*(5), 884–888. https://doi.org/10.1002/edn3.207

Buth, D. G., Crabtree, C. B., Orton, R. D., & Rainboth, W. J. (1984). Genetic differentiation between the freshwater subspecies of Gasterosteus aculeatus in Southern California.

Biochemical Systematics and Ecology, *12*(4), 423–432. https://doi.org/10.1016/0305-1978(84)90075-9

Buxton, A. S., Groombridge, J. J., & Griffiths, R. A. (2017). Is the detection of aquatic environmental DNA influenced by substrate type? *PLOS ONE*, *12*(8), e0183371. https://doi.org/10.1371/journal.pone.0183371

Callahan, B. J., McMurdie, P. J., Rosen, M. J., Han, A. W., Johnson, A. J. A., & Holmes, S. P. (2016). DADA2: High-resolution sample inference from Illumina amplicon data. *Nature Methods*, *13*(7), 581–583. https://doi.org/10.1038/nmeth.3869

Campos, P. F., & Gilbert, T. M. P. (2012). DNA Extraction from Formalin-Fixed Material. In B. Shapiro & M. Hofreiter (Eds.), *Ancient DNA: Methods and Protocols* (pp. 81–85). Humana Press. https://doi.org/10.1007/978-1-61779-516-9_11

Caroe, C., & Bohmann, K. (2020). Tagsteady: A metabarcoding library preparation protocol to avoid false assignment of sequences to samples. *BioRxiv*.

Chen, H. (2018). *VennDiagram: Generate High-Resolution Venn and Euler Plots* (1.6.20). https://CRAN.R-project.org/package=VennDiagram

Colosimo, P. F., Hosemann, K. E., Balabhadra, S., Villarreal, G., Dickson, M., Grimwood, J., Schmutz, J., Myers, R. M., Schluter, D., & Kingsley, D. M. (2005). Widespread parallel evolution in sticklebacks by repeated fixation of ectodysplasin alleles. *Science*, *307*(5717), 1928–1933.

Colosimo, P. F., Peichel, C. L., Nereng, K., Blackman, B. K., Shapiro, M. D., Schluter, D., & Kingsley, D. M. (2004). The genetic architecture of parallel armor plate reduction in threespine sticklebacks. *PLoS Biol*, *2*(5), e109.

Conte, G. L., Arnegard, M. E., Best, J., Chan, Y. F., Jones, F. C., Kingsley, D. M., Schluter, D., & Peichel, C. L. (2015). Extent of QTL Reuse During Repeated Phenotypic Divergence of Sympatric Threespine Stickleback. *Genetics*, *201*(3), 1189–1200. https://doi.org/10.1534/genetics.115.182550

Cresko, W. A., Amores, A., Wilson, C., Murphy, J., Currey, M., Phillips, P., Bell, M. A., Kimmel, C. B., & Postlethwait, J. H. (2004). Parallel genetic basis for repeated evolution of armor loss in Alaskan threespine stickleback populations. *Proceedings of the National Academy of Sciences*, *101*(16), 6050–6055. https://doi.org/10.1073/pnas.0308479101

Cruickshank, T. E., & Hahn, M. W. (2014). Reanalysis suggests that genomic islands of speciation are due to reduced diversity, not reduced gene flow. *Molecular Ecology*, *23*(13), 3133–3157. https://doi.org/10.1111/mec.12796

Culver, G. B., & Hubbs, C. L. (1917). The fishes of the Santa Ana system of streams in Southern California. *Lorguinia*, *1*, 82–83.

Curd, E., Gomer, J., Kandlikar, G., Gold, Z., Ogden, M., & Shi, B. (2018). *The Anacapa Toolkit*. https://github.com/limey-bean/Anacapa

Curd, E., Kandlikar, G., & Gomer, J. (2018). *CRUX: Creating Reference libraries Using eXisting tools*. https://github.com/limey-bean/CRUX_Creating-Reference-libraries-Using-eXisting-tools

Cuvier, G., & Valenciennes, M. (1829). *Histoire naturelle des poissons*. Chez F. G. Levrault; [etc., etc.]. https://www.biodiversitylibrary.org/item/54024

Danecek, P., Auton, A., Abecasis, G., Albers, C. A., Banks, E., DePristo, M. A., Handsaker, R. E., Lunter, G., Marth, G. T., Sherry, S. T., McVean, G., & Durbin, R. (2011). The variant

call format and VCFtools. *Bioinformatics*, *27*(15), 2156–2158. https://doi.org/10.1093/bioinformatics/btr330

Danecek, P., Bonfield, J. K., Liddle, J., Marshall, J., Ohan, V., Pollard, M. O., Whitwham, A., Keane, T., McCarthy, S. A., Davies, R. M., & Li, H. (2021). Twelve years of SAMtools and BCFtools. *GigaScience*, *10*(2), giab008. https://doi.org/10.1093/gigascience/giab008

Deagle, B. E., Jones, F. C., Absher, D. M., Kingsley, D. M., & Reimchen, T. E. (2013a). Phylogeography and adaptation genetics of stickleback from the H aida G waii archipelago revealed using genome-wide single nucleotide polymorphism genotyping. *Molecular Ecology*, *22*(7), 1917–1932.

Deagle, B. E., Jones, F. C., Absher, D. M., Kingsley, D. M., & Reimchen, T. E. (2013b). Phylogeography and adaptation genetics of stickleback from the Haida Gwaii archipelago revealed using genome-wide single nucleotide polymorphism genotyping. *Molecular Ecology*, *22*(7), 1917–1932. https://doi.org/10.1111/mec.12215

Deiner, K., Bik, H. M., Mächler, E., Seymour, M., Lacoursière-Roussel, A., Altermatt, F., Creer, S., Bista, I., Lodge, D. M., & De Vere, N. (2017). Environmental DNA metabarcoding: Transforming how we survey animal and plant communities. *Molecular Ecology*, *26*(21), 5872–5895.

Deiner, K., Walser, J.-C., Mächler, E., & Altermatt, F. (2015). Choice of capture and extraction methods affect detection of freshwater biodiversity from environmental DNA. *Biological Conservation*, *183*, 53–63.

Dejean, T., Valentini, A., Miquel, C., Taberlet, P., Bellemain, E., & Miaud, C. (2012). Improved detection of an alien invasive species through environmental DNA barcoding: The example of the American bullfrog *Lithobates catesbeianus*: *Alien invasive species*

detection using eDNA. Journal of Applied Ecology, *49*(4), 953–959. https://doi.org/10.1111/j.1365-2664.2012.02171.x

Dell'Anno, A., & Corinaldesi, C. (2004). Degradation and Turnover of Extracellular DNA in Marine Sediments: Ecological and Methodological Considerations Degradation and Turnover of Extracellular DNA in Marine Sediments: Ecological and Methodological Considerations. *Applied and Environmental Microbiology*, *70*(7), 4384–4386. https://doi.org/10.1128/AEM.70.7.4384

Des Roches, S., Bell, M. A., & Palkovacs, E. P. (2020). Climate-driven habitat change causes evolution in Threespine Stickleback. *Global Change Biology*, *26*(2), 597–606.

Doi, H., Uchii, K., Matsuhashi, S., Takahara, T., Yamanaka, H., & Minamoto, T. (2017). Isopropanol precipitation method for collecting fish environmental DNA. *Limnology and Oceanography: Methods*, *15*(2), 212–218. https://doi.org/10.1002/lom3.10161

Dolby, G. A., Bedolla, A. M., Bennett, S. E. K., & Jacobs, D. K. (2020). Global physical controls on estuarine habitat distribution during sea level change: Consequences for genetic diversification through time. *Global and Planetary Change*, *187*, 103128. https://doi.org/10.1016/j.gloplacha.2020.103128

Dolby, G. A., Ellingson, R. A., Findley, L. T., & Jacobs, D. K. (2018). How sea level change mediates genetic divergence in coastal species across regions with varying tectonic and sediment processes. *Molecular Ecology*, *27*(4), 994–1011. https://doi.org/10.1111/mec.14487

Dolby, G. A., Hechinger, R., Ellingson, R. A., Findley, L. T., Lorda, J., & Jacobs, D. K. (2016). Sea-level driven glacial-age refugia and post-glacial mixing on subtropical coasts, a

palaeohabitat and genetic study. *Proceedings of the Royal Society B: Biological Sciences*, *283*(1843), 20161571.

Earl, D. A., Louie, K. D., Bardeleben, C., Swift, C. C., & Jacobs, D. K. (2010). Rangewide microsatellite phylogeography of the endangered tidewater goby, a genetically subdivided coastal fish with limited marine dispersal. *Conservation Genetics*, *11*, 103–104. https://doi.org/10.1007/s10592-009-0008-9

Enciso-Romero, J., Pardo-Díaz, C., Martin, S. H., Arias, C. F., Linares, M., McMillan, W. O., Jiggins, C. D., & Salazar, C. (2017). Evolution of novel mimicry rings facilitated by adaptive introgression in tropical butterflies. *Molecular Ecology*, *26*(19), 5160–5172. https://doi.org/10.1111/mec.14277

Eriksson, B. K., Yanos, C., Bourlat, S. J., Donadi, S., Fontaine, M. C., Hansen, J. P., Jakubavičiūtė, E., Kiragosyan, K., Maan, M. E., Merilä, J., Austin, Å. N., Olsson, J., Reiss, K., Sundblad, G., Bergström, U., & Eklöf, J. S. (2021). Habitat segregation of plate phenotypes in a rapidly expanding population of three-spined stickleback. *Ecosphere*, *12*(6), 1–18. https://doi.org/10.1002/ecs2.3561

Esling, P., Lejzerowicz, F., & Pawlowski, J. (2015). Accurate multiplexing and filtering for high-throughput amplicon-sequencing. *Nucleic Acids Research*, *43*(5), 2513–2524. https://doi.org/10.1093/nar/gkv107

Evanno, G., Regnaut, S., & Goudet, J. (2005). Detecting the number of clusters of individuals using the software structure: A simulation study. *Molecular Ecology*, *14*(8), 2611–2620. https://doi.org/10.1111/j.1365-294X.2005.02553.x

Ewing, G. B., & Jensen, J. D. (2016). The consequences of not accounting for background selection in demographic inference. *Molecular Ecology*, *25*(1), 135–141.

Faber, P. M., Keller, E., Sands, A., & Massey, B. M. (1989). *The Ecology of Riparian Habitats of the Southern California Coastal Region: A Community Profile: Vol. 7.27*. U.S. Department of the Interior, Fish and Wildlife Service, Research and Development, National Wetlands Research Center.

Fang, B., Kemppainen, P., Momigliano, P., Feng, X., & Merilä, J. (2020). On the causes of geographically heterogeneous parallel evolution in sticklebacks. *Nature Ecology & Evolution*, *4*(8), 1105–1115.

Fang, B., Merilä, J., Ribeiro, F., Alexandre, C. M., & Momigliano, P. (2018). Worldwide phylogeny of three-spined sticklebacks. *Molecular Phylogenetics and Evolution*, *127*, 613–625.

Faribanks, H. W., & Carey, E. P. (1910). Glaciation in the San Bernardino Range, California. *Science*, *31*(784), 32–33.

FASTX-Toolkit. (n.d.). Retrieved January 11, 2018, from http://hannonlab.cshl.edu/fastx_toolkit/

Ficetola, G. F., Miaud, C., Pompanon, F., & Taberlet, P. (2008). Species detection using environmental DNA from water samples. *Biology Letters*, *4*(4), 423–425. https://doi.org/10.1098/rsbl.2008.0118

Fitak, R. R. (2021). OptM: Estimating the optimal number of migration edges on population trees using Treemix. *Biology Methods and Protocols*, *6*(1), bpab017. https://doi.org/10.1093/biomethods/bpab017

Flint, L. E., Brandt, J., Christensen, A. H., Flint, A. L., Hevesi, J. A., Jachens, R., Kulongoski, J. T., Martin, P., & Sneed, M. (2012). Geohydrology of Big Bear Valley, California: Phase 1--geologic framework, recharge, and preliminary assessment of the source and age of groundwater. In *Geohydrology of Big Bear Valley, California: Phase 1—Geologic*

framework, recharge, and preliminary assessment of the source and age of groundwater (USGS Numbered Series No. 2012–5100; Scientific Investigations Report, Vols. 2012–5100, p. 130). U.S. Geological Survey. https://doi.org/10.3133/sir20125100

Flint, L. E., Underwood, E. C., Flint, A. L., & Hollander, A. D. (2019). Characterizing the Influence of Fire on Hydrology in Southern California. *Natural Areas Journal*, *39*(1), 108–121.

Freedman, A. H., Gronau, I., Schweizer, R. M., Vecchyo, D. O.-D., Han, E., Silva, P. M., Galaverni, M., Fan, Z., Marx, P., Lorente-Galdos, B., Beale, H., Ramirez, O., Hormozdiari, F., Alkan, C., Vilà, C., Squire, K., Geffen, E., Kusak, J., Boyko, A. R., … Novembre, J. (2014). Genome Sequencing Highlights the Dynamic Early History of Dogs. *PLOS Genetics*, *10*(1), e1004016. https://doi.org/10.1371/journal.pgen.1004016

French, J. J., & Busby, M. W. (1974). Flood-hazard study, 100-year stage for Baldwin Lake, San Bernardino County, California. In *Flood-hazard study, 100-year stage for Baldwin Lake, San Bernardino County, California* (USGS Numbered Series No. 74–26; Water-Resources Investigations Report, Vols. 74–26). U.S. Geological Survey. https://doi.org/10.3133/wri7426

Gao, X., Lin, H., Revanna, K., & Dong, Q. (2017). A Bayesian taxonomic classification method for 16S rRNA gene sequences with improved species-level accuracy. *BMC Bioinformatics*, *18*(1), 247. https://doi.org/10.1186/s12859-017-1670-4

Gbif.Org. (2022). *Occurrence Download* (p. 170487) [Darwin Core Archive]. The Global Biodiversity Information Facility. https://doi.org/10.15468/DL.HTJ3HT

Girard, C. F. (1854). *Descriptions of new fishes, collected by Dr. AL Heermann, naturalist attached to the survey of the Pacific railroad route, under Lieut. RS Williamson, USA. 7*, 129–140.

Glover, K. C., Chaney, A., Kirby, M. E., Patterson, W. P., & MacDonald, G. M. (2020). Southern California vegetation, wildfire, and erosion had nonlinear responses to climatic forcing during marine isotope stages 5–2 (120–15 ka). *Paleoceanography and Paleoclimatology*, *35*(2), 1–15.

Glover, K. C., George, J., Heusser, L., & MacDonald, G. M. (2021). West Coast vegetation shifts as a response to climate change over the past 130,000 years: Geographic patterns and process from pollen data. *Physical Geography*, 1–19.

Glover, K. C., MacDonald, G. M., Kirby, M. E., Rhodes, E. J., Stevens, L., Silveira, E., Whitaker, A., & Lydon, S. (2017). Evidence for orbital and North Atlantic climate forcing in alpine Southern California between 125 and 10 ka from multi-proxy analyses of Baldwin Lake. *Quaternary Science Reviews*, *167*, 47–62.

Goldberg, C. S., Turner, C. R., Deiner, K., Klymus, K. E., Thomsen, P. F., Murphy, M. A., Spear, S. F., McKee, A., Oyler-McCance, S. J., & Cornman, R. S. (2016). Critical considerations for the application of environmental DNA methods to detect aquatic species. *Methods in Ecology and Evolution*, *7*(11), 1299–1307.

Gumprecht, B. (1999). *The Los Angeles River: Its Life, Death, and Possible Rebirth* (2001st ed.). Johns Hopkins University Press.

Guo, Y., Ye, F., Sheng, Q., Clark, T., & Samuels, D. C. (2014). Three-stage quality control strategies for DNA re-sequencing data. *Briefings in Bioinformatics*, *15*(6), 879–889. https://doi.org/10.1093/bib/bbt069

Hagen, D. W., & Gilbertson, L. G. (1973). Selective predation and the intensity of selection acting upon the lateral plates of threespine sticklebacks. *Heredity, 30*(3), 273–287.

Haglund, T., & Buth, D. (1988). Allozymes of the unarmored threespine stickleback (Gasterosteus aculeatus williamsoni) and identification of the Shay Creek population. *Isozyme Bull, 21*, 196.

Han, F., Lamichhaney, S., Grant, B. R., Grant, P. R., Andersson, L., & Webster, M. T. (2017). Gene flow, ancient polymorphism, and ecological adaptation shape the genomic landscape of divergence among Darwin's finches. *Genome Research, 27*(6), 1004–1015. https://doi.org/10.1101/gr.212522.116

Harper, L. R., Buxton, A. S., Rees, H. C., Bruce, K., Brys, R., Halfmaerten, D., Read, D. S., Watson, H. V., Sayer, C. D., & Jones, E. P. (2019). Prospects and challenges of environmental DNA (eDNA) monitoring in freshwater ponds. *Hydrobiologia, 826*(1), 25–41.

Hendry, A. P. (2013). Key questions in the genetics and genomics of eco-evolutionary dynamics. *Heredity, 111*(6), 456–466. https://doi.org/10.1038/hdy.2013.75

Hendy, I. (2009). A fresh perspective on the Cordilleran Ice Sheet. *The Geological Society of America, 19*(4–5), 60. https://doi.org/10.1130/0091

Heusser, L. (1998). Direct correlation of millennial-scale changes in western North American vegetation and climate with changes in the California Current System over the past ~60 kyr. *Paleoceanography, 13*(3), 252–262. https://doi.org/10.1029/98PA00670

Heusser, L. E., Kirby, M. E., & Nichols, J. E. (2015). Pollen-based evidence of extreme drought during the last Glacial (32.6–9.0 ka) in coastal southern California. *Quaternary Science Reviews, 126*, 242–253.

Higuchi, M., Sakai, H., & Goto, A. (2014). A new threespine stickleback, Gasterosteus nipponicus sp. nov.(Teleostei: Gasterosteidae), from the Japan Sea region. *Ichthyological Research*, *61*(4), 341–351.

Hinlo, R., Gleeson, D., Lintermans, M., & Furlan, E. (2017). Methods to maximise recovery of environmental DNA from water samples. *PloS One*, *12*(6), e0179251.

Hohenlohe, P. A., & Magalhaes, I. S. (2020). The Population Genomics of Parallel Adaptation: Lessons from Threespine Stickleback. In M. F. Oleksiak & O. P. Rajora (Eds.), *Population Genomics: Marine Organisms* (pp. 249–276). Springer International Publishing. https://doi.org/10.1007/13836_2019_67

Indjeian, V. B., Kingman, G. A., Jones, F. C., Guenther, C. A., Grimwood, J., Schmutz, J., Myers, R. M., & Kingsley, D. M. (2016). Evolving new skeletal traits by cis-regulatory changes in bone morphogenetic proteins. *Cell*, *164*(1–2), 45–56.

Jacobs, D. K., Haney, T. A., & Louie, K. D. (2004). Genes, diversity, and geologic process on the Pacific coast. *Annual Review of Earth and Planetary Sciences*, *32*(1), 601–652. https://doi.org/10.1146/annurev.earth.32.092203.122436

Jacobs, D. K., Stein, E. D., & Longcore, T. (2011a). Classification of California Estuaries Based on Natural Closure Patterns: Templates for Restoration and Management Management. *Technical Report, August*, 1–72.

Jacobs, D. K., Stein, E. D., & Longcore, T. (2011b). Classification of California Estuaries Based on Natural Closure Patterns: Templates for Restoration and Management Management. *Technical Report, August*, 1–72.

Jacobs, D., Stein, E. D., & Longcore, T. (2011). *Classification of California Estuaries Based on Natural Closure Patterns: Templates for Restoration and Management* (619.a; p. 72).

Jombart, T., & Ahmed, I. (2011). adegenet 1.3-1: New tools for the analysis of genome-wide SNP data. *Bioinformatics*, *27*(21), 3070–3071. https://doi.org/10.1093/bioinformatics/btr521

Jones, D. H., & John, A. W. G. (1978). A three-spined stickleback, Gasterosteus aculeatus L. from the North Atlantic. *Journal of Fish Biology*, *13*(2), 231–236. https://doi.org/10.1111/j.1095-8649.1978.tb03430.x

Jones, F. C., Grabherr, M. G., Chan, Y. F., Russell, P., Mauceli, E., Johnson, J., Swofford, R., Pirun, M., Zody, M. C., & White, S. (2012). The genomic basis of adaptive evolution in threespine sticklebacks. *Nature*, *484*(7392), 55–61.

Jones, F. C., Grabherr, M. G., Chan, Y. F., Russell, P., Mauceli, E., Johnson, J., Swofford, R., Pirun, M., Zody, M. C., White, S., Birney, E., Searle, S., Schmutz, J., Grimwood, J., Dickson, M. C., Myers, R. M., Miller, C. T., Summers, B. R., Knecht, A. K., … Kingsley, D. M. (2012). The genomic basis of adaptive evolution in threespine sticklebacks. *Nature*, *484*(7392), 55–61. https://doi.org/10.1038/nature10944

Jónsson, H., Ginolhac, A., Schubert, M., Johnson, P. L. F., & Orlando, L. (2013). MapDamage2.0: Fast approximate Bayesian estimates of ancient DNA damage parameters. *Bioinformatics*, *29*(13), 1682–1684. https://doi.org/10.1093/bioinformatics/btt193

Kandlikar, G. (2020). *ranacapa: Utility Functions and "shiny" App for Simple Environmental DNA Visualizations and Analyses* (0.1.0). https://github.com/gauravsk/ranacapa

Kelly, R. P., Shelton, A. O., & Gallego, R. (2019). Understanding PCR Processes to Draw Meaningful Conclusions from Environmental DNA Studies. *Scientific Reports*, *9*(1), 12133. https://doi.org/10.1038/s41598-019-48546-x

Kingman, G. A. R., Vyas, D. N., Jones, F. C., Brady, S. D., Chen, H. I., Reid, K., Milhaven, M., Bertino, T. S., Aguirre, W. E., & Heins, D. C. (2021). Predicting future from past: The genomic basis of recurrent and rapid stickleback evolution. *Science Advances*, *7*(25), eabg5285.

Kirby, M. E., Feakins, S. J., Bonuso, N., Fantozzi, J. M., & Hiner, C. A. (2013). Latest Pleistocene to Holocene hydroclimates from Lake Elsinore, California. *Quaternary Science Reviews*, *76*, 1–15. https://doi.org/10.1016/j.quascirev.2013.05.023

Kircher, M., Sawyer, S., & Meyer, M. (2012). Double indexing overcomes inaccuracies in multiplex sequencing on the Illumina platform. *Nucleic Acids Research*, *40*(1), e3–e3. https://doi.org/10.1093/nar/gkr771

Konijnendijk, N., Shikano, T., Daneels, D., Volckaert, F. A., & Raeymaekers, J. A. (2015). Signatures of selection in the three-spined stickleback along a small-scale brackish water–freshwater transition zone. *Ecology and Evolution*, *5*(18), 4174–4186.

Kopelman, N. M., Mayzel, J., Jakobsson, M., Rosenberg, N. A., & Mayrose, I. (2015). Clumpak: A program for identifying clustering modes and packaging population structure inferences across K. *Molecular Ecology Resources*, *15*(5), 1179–1191. https://doi.org/10.1111/1755-0998.12387

Kumar, G., Farrell, E., Reaume, A. M., Eble, J. A., & Gaither, M. R. (2022). One size does not fit all: Tuning eDNA protocols for high- and low-turbidity water sampling. *Environmental DNA*, *4*(1), 167–180. https://doi.org/10.1002/edn3.235

Kwambana, B. A., Mohammed, N. I., Jeffries, D., Barer, M., Adegbola, R. A., & Antonio, M. (2011). Differential effects of frozen storage on the molecular detection of bacterial taxa

that inhabit the nasopharynx. *BMC Clinical Pathology*, *11*(1), 2. https://doi.org/10.1186/1472-6890-11-2

Lai, Y.-T., Yeung, C. K., Omland, K. E., Pang, E.-L., Hao, Y., Liao, B.-Y., Cao, H.-F., Zhang, B.-W., Yeh, C.-F., & Hung, C.-M. (2019). Standing genetic variation as the predominant source for adaptation of a songbird. *Proceedings of the National Academy of Sciences*, *116*(6), 2152–2157.

Langmead, B., & Salzberg, S. L. (2012). Fast gapped-read alignment with Bowtie 2. *Nature Methods*, *9*(4), 357–359. https://doi.org/10.1038/nmeth.1923

Laramie, M. B., Pilliod, D. S., Goldberg, C. S., & Strickler, K. M. (2015). Environmental DNA sampling protocol—Filtering water to capture DNA from aquatic organisms. *U.S Geological Survey Techniques and Methods*, *Book 2*(Chapter A13), 15 p. https://doi.org/10.3133/TM2A13

Larsson, A. J. M., Stanley, G., Sinha, R., Weissman, I. L., & Sandberg, R. (2018). Computational correction of index switching in multiplexed sequencing libraries. *Nature Methods*, *15*(5), 305–307. https://doi.org/10.1038/nmeth.4666

Leidy, R. A., & Moyle, P. B. (2021). Keeping up with the status of freshwater fishes: A California (USA) perspective. *Conservation Science and Practice*, *3*(8). https://doi.org/10.1111/csp2.474

Levick, L. R., Goodrich, D. C., Hernandez, M., Fonseca, J., Semmens, D. J., Stromberg, J. C., Tluczek, M., Leidy, R. A., Scianni, M., & Guertin, D. P. (2008). *The ecological and hydrological significance of ephemeral and intermittent streams in the arid and semi-arid American Southwest*. Environmental Protection Agency, Office of Research and Development.

Levy-Booth, D. J., Campbell, R. G., Gulden, R. H., Hart, M. M., Powell, J. R., Klironomos, J. N., Pauls, K. P., Swanton, C. J., Trevors, J. T., & Dunfield, K. E. (2007). Cycling of extracellular DNA in the soil environment. *Soil Biology and Biochemistry*, *39*(12), 2977–2991. https://doi.org/10.1016/j.soilbio.2007.06.020

Li, H. (2013a). *Aligning sequence reads, clone sequences and assembly contigs with BWA-MEM*. *00*(00), 1–3. https://doi.org/10.48550/arXiv.1303.3997

Li, H. (2013b). *Aligning sequence reads, clone sequences and assembly contigs with BWA-MEM*. *00*(00), 1–3. https://doi.org/arXiv:1303.3997 [q-bio.GN]

Li, H., & Durbin, R. (2011). Inference of human population history from individual whole-genome sequences. *Nature*, *475*(7357), 493–496. https://doi.org/10.1038/nature10231

Li, J., Lawson Handley, L.-J., Read, D. S., & Hänfling, B. (2018). The effect of filtration method on the efficiency of environmental DNA capture and quantification via metabarcoding. *Molecular Ecology Resources*, *18*(5), 1102–1114.

Liang, Z., & Keeley, A. (2013). Filtration Recovery of Extracellular DNA from Environmental Water Samples. *Environmental Science & Technology*, *47*(16), 9324–9331. https://doi.org/10.1021/es401342b

Lipson, M. (2020). Applying f_4-statistics and admixture graphs: Theory and examples. *Molecular Ecology Resources*, *20*(6), 1658–1667. https://doi.org/10.1111/1755-0998.13230

Liu, S., Hansen, M. M., & Jacobsen, M. W. (2016). Region-wide and ecotype-specific differences in demographic histories of threespine stickleback populations, estimated from whole genome sequences. *Molecular Ecology*, *25*(20), 5187–5202. https://doi.org/10.1111/mec.13827

Love, M. I., Huber, W., & Anders, S. (2014). Moderated estimation of fold change and dispersion for RNA-seq data with DESeq2. *Genome Biology*, *15*(12), 550. https://doi.org/10.1186/s13059-014-0550-8

Maier, R., Flegontov, P., Flegontova, O., Changmai, P., & Reich, D. (2022). *On the limits of fitting complex models of population history to genetic data* (p. 2022.05.08.491072). bioRxiv. https://doi.org/10.1101/2022.05.08.491072

Majaneva, M., Diserud, O. H., Eagle, S. H., Boström, E., Hajibabaei, M., & Ekrem, T. (2018). Environmental DNA filtration techniques affect recovered biodiversity. *Scientific Reports*, *8*(1), 1–11.

Mäkinen, H. S., Cano, J. M., & Merilä, J. (2006). Genetic relationships among marine and freshwater populations of the European three-spined stickleback (Gasterosteus aculeatus) revealed by microsatellites. *Molecular Ecology*, *15*(6), 1519–1534.

Malcolm, J. R. (1992). *Supporting information for a petition to list as endangered or threatened: Shay Creek stickleback, Gasterosteus sp* (P. B. Moyle & R. M. Yoshiyama, Eds.; pp. 213–222).

Marques, D. A., Jones, F. C., Di Palma, F., Kingsley, D. M., & Reimchen, T. E. (2018). Experimental evidence for rapid genomic adaptation to a new niche in an adaptive radiation. *Nature Ecology & Evolution*, *2*(7), 1128–1138. https://doi.org/10.1038/s41559-018-0581-8

Martin, M. (2013). Cutadapt removes adapter sequences from high-throughput sequencing reads. *EMBnet.Journal*, *17*(1), 10–12.

Martin, S. H., Davey, J. W., & Jiggins, C. D. (2015). Evaluating the Use of ABBA–BABA Statistics to Locate Introgressed Loci. *Molecular Biology and Evolution*, *32*(1), 244–257. https://doi.org/10.1093/molbev/msu269

Masters, P. M. (2006). Holocene sand beaches of southern California: ENSO forcing and coastal processes on millennial scales. *Palaeogeography, Palaeoclimatology, Palaeoecology*, *232*(1), 73–95.

Mather, N., Traves, S. M., & Ho, S. Y. W. (2020). A practical introduction to sequentially Markovian coalescent methods for estimating demographic history from genomic data. *Ecology and Evolution*, *10*(1), 579–589. https://doi.org/10.1002/ece3.5888

Mattern, M. Y. (2004). Molecular phylogeny of the Gasterosteidae: The importance of using multiple genes. *Molecular Phylogenetics and Evolution*, *30*(2), 366–377.

Mazet, O., Rodríguez, W., Grusea, S., Boitard, S., & Chikhi, L. (2016). On the importance of being structured: Instantaneous coalescence rates and human evolution—Lessons for ancestral population size inference? *Heredity*, *116*(4), 362–371. https://doi.org/10.1038/hdy.2015.104

Mazzarella, A. B., Boessenkool, S., Østbye, K., Vøllestad, L. A., & Trucchi, E. (2016). Genomic signatures of the plateless phenotype in the threespine stickleback. *Ecology and Evolution*, *6*(10), 3161–3173. https://doi.org/10.1002/ece3.2072

McKenna, A., Hanna, M., Banks, E., Sivachenko, A., Cibulskis, K., Kernytsky, A., Garimella, K., Altshuler, D., Gabriel, S., Daly, M., & DePristo, M. A. (2010). The Genome Analysis Toolkit: A MapReduce framework for analyzing next-generation DNA sequencing data. *Genome Research*, *20*(9), 1297–1303. https://doi.org/10.1101/gr.107524.110

McMurdie, P. J., & Holmes, S. (2013). phyloseq: An R Package for Reproducible Interactive Analysis and Graphics of Microbiome Census Data. *PLoS ONE*, *8*(4), e61217. https://doi.org/10.1371/journal.pone.0061217

McMurdie, P. J., & Holmes, S. (2014). Waste Not, Want Not: Why Rarefying Microbiome Data Is Inadmissible. *PLoS Computational Biology*, *10*(4), e1003531. https://doi.org/10.1371/journal.pcbi.1003531

Meier, J. I., Marques, D. A., Wagner, C. E., Excoffier, L., & Seehausen, O. (2018). Genomics of Parallel Ecological Speciation in Lake Victoria Cichlids. *Molecular Biology and Evolution*, *35*(6), 1489–1506. https://doi.org/10.1093/molbev/msy051

Meirmans, P. G. (2020). GenoDive version 3.0: Easy-to-use software for the analysis of genetic data of diploids and polyploids. *Molecular Ecology Resources*, *20*(4), 1126–1131. https://doi.org/10.1111/1755-0998.13145

Mendenhall, W. C. (1908). Ground waters and irrigation enterprises in the foothill belt, southern California. In *Water Supply Paper* (No. 219; Water Supply Paper, p. 180). https://doi.org/10.3133/wsp219

Miller, R. R. (1961). *Man and the changing fish fauna of the American southwest*. *46*, 365–404.

Miller, R. R., & Hubbs, C. L. (1969). Systematics of Gasterosteus aculeatus, with Particular Reference to Intergradation and Introgression along the Pacific Coast of North America: A Commentary on a Recent Contribution. *Copeia*, *1*(1), 52–69.

Minckley, W. L., & Marsh, P. C. (2009). *Inland Fishes of the Greater Southwest: Chronicle of a Vanishing Biota*. University of Arizona Press.

Morris, M. R., Bowles, E., Allen, B. E., Jamniczky, H. A., & Rogers, S. M. (2018). Contemporary ancestor? Adaptive divergence from standing genetic variation in Pacific marine threespine stickleback. *BMC Evolutionary Biology*, *18*(1), 1–21.

Mount, J. F. (1995). *California Rivers and Streams: The Conflict Between Fluvial Process and Land Use*. University of California Press.

Moyle, P. B. (2002). *Inland Fishes of California: Revised and Expanded*. University of California Press.

Moyle, P. B., Quiñones, R. M., Katz, J. V., & Weaver, J. (2015). *Fish species of special concern in California* (p. 842). California Department of Fish and Wildlife. https://wildlife.ca.gov/Conservation/SSC/Fishes

Nadachowska-Brzyska, K., Burri, R., Smeds, L., & Ellegren, H. (2016). PSMC analysis of effective population sizes in molecular ecology and its application to black-and-white Ficedula flycatchers. *Molecular Ecology*, *25*(5), 1058–1072. https://doi.org/10.1111/mec.13540

Nath, S., Shaw, D. E., & White, M. A. (2020). Improved contiguity of the threespine stickleback genome using long-read sequencing. *BioRxiv*, 2020.06.30.170787. https://doi.org/10.1101/2020.06.30.170787

Nico, L., Fuller, P., & Neilson, M. (2022). *Acanthogobius flavimanus (Temminck and Schlegel, 1845)*. U.S. Geological Survey, Nonindigenous Aquatic Species Database. https://nas.er.usgs.gov/queries/FactSheet.aspx?speciesID=707

Noor, M. a. F., & Bennett, S. M. (2009). Islands of speciation or mirages in the desert? Examining the role of restricted recombination in maintaining species. *Heredity*, *103*(6), Article 6. https://doi.org/10.1038/hdy.2009.151

O'Brown, N. M., Summers, B. R., Jones, F. C., Brady, S. D., & Kingsley, D. M. (2015a). A recurrent regulatory change underlying altered expression and Wnt response of the stickleback armor plates gene EDA. *Elife, 4*, e05290.

O'Brown, N. M., Summers, B. R., Jones, F. C., Brady, S. D., & Kingsley, D. M. (2015b). A recurrent regulatory change underlying altered expression and Wnt response of the stickleback armor plates gene EDA. *ELife*, *4*(4), 1–17. https://doi.org/10.7554/eLife.05290

Oksanen, J., Blanchet, F. G., Friendly, M., Kindt, R., Legendre, P., McGlinn, D., Minchin, P. R., O'Hara, R. B., Simpson, G. L., Solymos, P., Stevens, M. H. H., Szoecs, E., & Wagner, H. (2019). *vegan: Community Ecology Package* (2.5-6). https://CRAN.R-project.org/package=vegan

Orsi, J. P. (2004). *Hazardous metropolis: Flooding and urban ecology in Los Angeles* [Ph.D., The University of Wisconsin - Madison]. https://www.proquest.com/docview/304537402/abstract/68D98B12099C4AE4PQ/1

Owen, L. A., Finkel, R. C., Minnich, R. A., & Perez, A. E. (2003). Extreme southwestern margin of late Quaternary glaciation in North America: Timing and controls. *Geology*, *31*(8), 729–732.

Paccard, A., Wasserman, B. A., Hanson, D., Astorg, L., Durston, D., Kurland, S., Apgar, T. M., El-Sabaawi, R. W., Palkovacs, E. P., & Hendry, A. P. (2018). Adaptation in temporally variable environments: Stickleback armor in periodically breaching bar-built estuaries. *Journal of Evolutionary Biology*, *31*(5), 735–752.

Patterson, N., Moorjani, P., Luo, Y., Mallick, S., Rohland, N., Zhan, Y., Genschoreck, T., Webster, T., & Reich, D. (2012). Ancient Admixture in Human History. *Genetics, 192*(3), 1065–1093. https://doi.org/10.1534/genetics.112.145037

Pawlowski, J., Bruce, K., Panksep, K., Aguirre, F. I., Amalfitano, S., Apothéloz-Perret-Gentil, L., Baussant, T., Bouchez, A., Carugati, L., Cermakova, K., Cordier, T., Corinaldesi, C., Costa, F. O., Danovaro, R., Dell'Anno, A., Duarte, S., Eisendle, U., Ferrari, B. J. D., Frontalini, F., … Fazi, S. (2022). Environmental DNA metabarcoding for benthic monitoring: A review of sediment sampling and DNA extraction methods. *Science of The Total Environment*, *818*, 151783. https://doi.org/10.1016/j.scitotenv.2021.151783

Peichel, C. L., & Marques, D. A. (2017). The genetic and molecular architecture of phenotypic diversity in sticklebacks. *Philosophical Transactions of the Royal Society B: Biological Sciences, 372*(1713), 20150486. https://doi.org/10.1098/rstb.2015.0486

Perkins, T. L., Clements, K., Baas, J. H., Jago, C. F., Jones, D. L., Malham, S. K., & McDonald, J. E. (2014). Sediment Composition Influences Spatial Variation in the Abundance of Human Pathogen Indicator Bacteria within an Estuarine Environment. *PLoS ONE, 9*(11), e112951. https://doi.org/10.1371/journal.pone.0112951

Pickrell, J., & Pritchard, J. (2012). Inference of population splits and mixtures from genome-wide allele frequency data. *Nature Precedings*, 1–1.

Pietramellara, G., Ascher, J., Borgogni, F., Ceccherini, M. T., Guerri, G., & Nannipieri, P. (2009). Extracellular DNA in soil and sediment: Fate and ecological relevance. *Biology and Fertility of Soils, 45*(3), 219–235. https://doi.org/10.1007/s00374-008-0345-8

Pilliod, D. S., Goldberg, C. S., Arkle, R. S., Waits, L. P., & Richardson, J. (2013). Estimating occupancy and abundance of stream amphibians using environmental DNA from filtered

water samples. *Canadian Journal of Fisheries and Aquatic Sciences*, *70*(8), 1123–1130. https://doi.org/10.1139/cjfas-2013-0047

Port, J. A., O'Donnell, J. L., Romero-Maraccini, O. C., Leary, P. R., Litvin, S. Y., Nickols, K. J., Yamahara, K. M., & Kelly, R. P. (2016). Assessing vertebrate biodiversity in a kelp forest ecosystem using environmental DNA. *Molecular Ecology*, *25*(2), 527–541. https://doi.org/10.1111/mec.13481

Pringle, C. M., Freeman, M. C., & Freeman, B. J. (2000). Regional Effects of Hydrologic Alterations on Riverine Macrobiota in the New World: Tropical-Temperate Comparisons: The massive scope of large dams and other hydrologic modifications in the temperate New World has resulted in distinct regional trends of biotic impoverishment. While neotropical rivers have fewer dams and limited data upon which to make regional generalizations, they are ecologically vulnerable to increasing hydropower development and biotic patterns are emerging. *BioScience*, *50*(9), 807–823. https://doi.org/10.1641/0006-3568(2000)050[0807:REOHAO]2.0.CO;2

Pritchard, J. K., Stephens, M., & Donnelly, P. (2000). Inference of Population Structure Using Multilocus Genotype Data. *Genetics*, *155*(2), 945 LP – 959.

Propst, D. L., Williams, J. E., Bestgen, K. R., & Hoagstrom, C. W. (2021). *Standing between Life and Extinction: Ethics and Ecology of Conserving Aquatic Species in North American Deserts*. University of Chicago Press.

R Core Team. (2018a). *R: A language and environment for statistical computing*. R Foundation for Statistical Computing. https://www.R-project.org/

R Core Team. (2018b). *R: A language and environment for statistical computing*. R Foundation for Statistical Computing. https://www.R-project.org/

Raj, A., Stephens, M., & Pritchard, J. K. (2014). FastSTRUCTURE: Variational Inference of Population Structure in Large SNP Data Sets. *Genetics*, *197*(2), 573–589. https://doi.org/10.1534/genetics.114.164350

Ravinet, M., Faria, R., Butlin, R. K., Galindo, J., Bierne, N., Rafajlović, M., Noor, M. a. F., Mehlig, B., & Westram, A. M. (2017). Interpreting the genomic landscape of speciation: A road map for finding barriers to gene flow. *Journal of Evolutionary Biology*, *30*(8), 1450–1477. https://doi.org/10.1111/jeb.13047

Ravinet, M., Prodöhl, P. A., & Harrod, C. (2013). Parallel and nonparallel ecological, morphological and genetic divergence in lake–stream stickleback from a single catchment. *Journal of Evolutionary Biology*, *26*(1), 186–204.

Rees, H. C., Maddison, B. C., Middleditch, D. J., Patmore, J. R. M., & Gough, K. C. (2014). REVIEW: The detection of aquatic animal species using environmental DNA – a review of eDNA as a survey tool in ecology. *Journal of Applied Ecology*, *51*(5), 1450–1459. https://doi.org/10.1111/1365-2664.12306

Reid, K., Bell, M. A., & Veeramah, K. R. (2021). Threespine Stickleback: A Model System For Evolutionary Genomics. *Annual Review of Genomics and Human Genetics*, *22*(1), 357–383. https://doi.org/10.1146/annurev-genom-111720-081402

Reimchen, T. E. (1994). Predators and evolution in threespine stickleback. In M. A. Bell & S. A. Foster (Eds.), *Evolution of the threespine stickleback* (pp. 240–273). Oxford University Press.

Reimchen, T. E. (2000). Predator handling failures of lateral plate morphs in Gasterosteus aculeatus: Functional implications for the ancestral plate condition. *Behaviour*, *137*(7–8), 1081–1096. https://doi.org/10.1163/156853900502448

Ricciardi, A., & Rasmussen, J. B. (1999). Extinction Rates of North American Freshwater Fauna. *Conservation Biology*, *13*(5), 1220–1222. https://doi.org/10.1046/j.1523-1739.1999.98380.x

Richmond, J. Q., Backlin, A. R., Galst-Cavalcante, C., O'Brien, J. W., & Fisher, R. N. (2018). Loss of dendritic connectivity in southern California's urban riverscape facilitates decline of an endemic freshwater fish. *Molecular Ecology*, *27*(2), 369–386. https://doi.org/10.1111/mec.14445

Richmond, J. Q., Jacobs, D. K., Backlin, A. R., Swift, C. C., Dellith, C., & Fisher, R. N. (2015). Ephemeral stream reaches preserve the evolutionary and distributional history of threespine stickleback in the Santa Clara and Ventura River watersheds of southern California. *Conservation Genetics*, *16*(1), 85–101. https://doi.org/10.1007/s10592-014-0643-7

Robinson, J. A., Brown, C., Kim, B. Y., Lohmueller, K. E., & Wayne, R. K. (2018). Purging of Strongly Deleterious Mutations Explains Long-Term Persistence and Absence of Inbreeding Depression in Island Foxes. *Current Biology*, *28*(21), 3487-3494.e4. https://doi.org/10.1016/j.cub.2018.08.066

Robson, H. L. A., Noble, T. H., Saunders, R. J., Robson, S. K. A., Burrows, D. W., & Jerry, D. R. (2016). Fine-tuning for the tropics: Application of eDNA technology for invasive fish detection in tropical freshwater ecosystems. *Molecular Ecology Resources*, *16*(4), 922–932. https://doi.org/10.1111/1755-0998.12505

Rollins, J. L. (2017). Body-size and growth-rate divergence among populations of threespine stickleback (*Gasterosteus aculeatus*) in Cook Inlet, Alaska, USA. *Canadian Journal of Zoology*, *95*(11), 877–884. https://doi.org/10.1139/cjz-2017-0092

RStudio Team. (2020a). *RStudio: Integrated Development for R*. RStudio, PBC. http://www.rstudio.com/

RStudio Team. (2020b). *RStudio: Integrated Development for R*. RStudio, PBC. http://www.rstudio.com/

Ruiz-Campos, G., Camarena-Rosales, F., González-Acosta, A. F., Maeda-Martinez, A. M., García de León, F. J., Varela-Romero, A., & Andreu-Soler, A. (2014). Estatus actual de conservación de seis especies de peces dulceacuícolas de la península de Baja California, México. *Revista Mexicana de Biodiversidad*, *85*(4), 1235–1248. https://doi.org/10.7550/rmb.43747

Ruiz-Campos, G., & González-Acosta, A. F. (in press). *Peces de las aguas continentales de Baja California, México*. Universidad Autónoma de Baja California.

Sales, N. G., Wangensteen, O. S., Carvalho, D. C., & Mariani, S. (2019). Influence of preservation methods, sample medium and sampling time on eDNA recovery in a neotropical river. *Environmental DNA*, *1*(2), edn3.14. https://doi.org/10.1002/edn3.14

Samuels, D. C., Wang, J., Ye, F., He, J., Levinson, R. T., Sheng, Q., Zhao, S., Capra, J. A., Shyr, Y., Zheng, W., & Guo, Y. (2016). Heterozygosity Ratio, a Robust Global Genomic Measure of Autozygosity and Its Association with Height and Disease Risk. *Genetics*, *204*(3), 893–904. https://doi.org/10.1534/genetics.116.189936

Santoro, A. E., Dupont, C. L., Richter, R. A., Craig, M. T., Carini, P., McIlvin, M. R., Yang, Y., Orsi, W. D., Moran, D. M., & Saito, M. A. (2015). Genomic and proteomic characterization of "Candidatus Nitrosopelagicus brevis": An ammonia-oxidizing archaeon from the open ocean. *Proceedings of the National Academy of Sciences*, *112*(4), 1173–1178. https://doi.org/10.1073/pnas.1416223112

Sard, N. M., Herbst, S. J., Nathan, L., Uhrig, G., Kanefsky, J., Robinson, J. D., & Scribner, K. T. (2019). Comparison of fish detections, community diversity, and relative abundance using environmental DNA metabarcoding and traditional gears. *Environmental DNA*, *1*(4), 368–384. https://doi.org/10.1002/edn3.38

Schaarschmidt, F., & Gerhard, D. (2019). *PairwiseCI: Confidence Intervals for Two Sample Comparisons* (0.1-27). https://CRAN.R-project.org/package=pairwiseCI

Schiffels, S., & Durbin, R. (2014). Inferring human population size and separation history from multiple genome sequences. *Nature Genetics*, *46*(8), 919–925. https://doi.org/10.1038/ng.3015

Schluter, D. (2000). *The Ecology of Adaptive Radiation*. OUP Oxford.

Schluter, D., & Nagel, L. M. (1995). Parallel Speciation by Natural Selection. *The American Naturalist*, *146*(2), 292–301. https://doi.org/10.1086/285799

Schnell, I. B., Bohmann, K., & Gilbert, M. T. P. (2015). Tag jumps illuminated—Reducing sequence-to-sample misidentifications in metabarcoding studies. *Molecular Ecology Resources*, *15*(6), 1289–1303. https://doi.org/10.1111/1755-0998.12402

Schrider, D. R., Shanku, A. G., & Kern, A. D. (2016). Effects of linked selective sweeps on demographic inference and model selection. *Genetics*, *204*(3), 1207–1223.

SCWRP. (2018). *Wetlands on the Edge: The Future of Southern California's Wetlands: Regional Strategy 2018* (p. 142). California State Coastal Conservancy. scwrp.databasin.org

Sekar, R., Kaczmarsky, L. T., & Richardson, L. L. (2009). Effect of Freezing on PCR Amplification of 16S rRNA Genes from Microbes Associated with Black Band Disease

of Corals. *Applied and Environmental Microbiology*, *75*(8), 2581–2584. https://doi.org/10.1128/AEM.01500-08

Shaffer, H. B., Fellers, G. M., Randal Voss, S., Oliver, J. C., & Pauly, G. B. (2004). Species boundaries, phylogeography and conservation genetics of the red-legged frog (Rana aurora/draytonii) complex. *Molecular Ecology*, *13*(9), 2667–2677. https://doi.org/10.1111/j.1365-294X.2004.02285.x

Shanfelter, A. F., Archambeault, S. L., & White, M. A. (2019). Divergent fine-scale recombination landscapes between a freshwater and marine population of threespine stickleback fish. *Genome Biology and Evolution*, *11*(6), 1573.

Sharp, R. P., Allen, C. R., & Meier, M. F. (1959). Pleistocene glaciers on southern California mountains. *American Journal of Science*, *257*(2), 81–94.

Shirazi, S., Meyer, R. S., & Shapiro, B. (2021). Revisiting the effect of PCR replication and sequencing depth on biodiversity metrics in environmental DNA metabarcoding. *Ecology and Evolution*, *11*(22), 15766–15779. https://doi.org/10.1002/ece3.8239

Smart, A. S., Weeks, A. R., Rooyen, A. R., Moore, A., McCarthy, M. A., & Tingley, R. (2016). Assessing the cost-efficiency of environmental DNA sampling. *Methods in Ecology and Evolution*, *7*(11), 1291–1298. https://doi.org/10.1111/2041-210X.12598

Stein, E. D., Cayce, K., Salomon, M., Bram, D. L., De Mello, D., Grossinger, R., & Dark, S. (2014). *Wetlands of the Southern California Coast: Historical Extent and Change Over Time* (SFEI Report 720; SCCWRP Technical Report 826; p. 58). Southern California Coastal Water Research Project and San Francisco Estuary Institute. https://www.caltsheets.org/socal/download.html

Stephenson, J. R., & Calcarone, G. M. (1999). *Southern California Mountains and Foothills Assessment: Habitat and Species Conservation Issues*. U.S. Department of Agriculture, Forest Service, Pacific Southwest Research Station.

Stiller, J., da Fonseca, R. R., Alfaro, M. E., Faircloth, B. C., Wilson, N. G., & Rouse, G. W. (2021). Using ultraconserved elements to track the influence of sea-level change on leafy seadragon populations. *Molecular Ecology*, *30*(6), 1364–1380.

Suomalainen, L.-R., Reunanen, H., Ijäs, R., Valtonen, E. T., & Tiirola, M. (2006). Freezing Induces Biased Results in the Molecular Detection of *Flavobacterium columnare*. *Applied and Environmental Microbiology*, *72*(2), 1702–1704. https://doi.org/10.1128/AEM.72.2.1702-1704.2006

Swift, C. C., Haglund, T. R., Ruiz, M., & Fisher, R. N. (1993a). The Status and Distribution of the Freshwater Fishes of Southern California. *Bulletin of the Southern California Academy of Sciences*, *92*(3), 101–167.

Swift, C. C., Haglund, T. R., Ruiz, M., & Fisher, R. N. (1993b). The Status and Distribution of the Freshwater Fishes of Southern California. *Bulletin of the Southern California Academy of Sciences*, *92*(3), 101–167.

Swift, C. C., Spies, B., Ellingson, R. A., & Jacobs, D. K. (2016). A New Species of the Bay Goby Genus Eucyclogobius, Endemic to Southern California: Evolution, Conservation, and Decline. *PloS One*, *11*(7), e0158543. https://doi.org/10.1371/journal.pone.0158543

Taberlet, P., Coissac, E., Pompanon, F., Brochmann, C., & Willerslev, E. (2012). Towards next-generation biodiversity assessment using DNA metabarcoding: NEXT-GENERATION DNA METABARCODING. *Molecular Ecology*, *21*(8), 2045–2050. https://doi.org/10.1111/j.1365-294X.2012.05470.x

Taberlet, P., Prud'Homme, S. M., Campione, E., Roy, J., Miquel, C., Shehzad, W., Gielly, L., Rioux, D., Choler, P., Clément, J.-C., Melodelima, C., Pompanon, F., & Coissac, E. (2012). Soil sampling and isolation of extracellular DNA from large amount of starting material suitable for metabarcoding studies: EXTRACTION OF EXTRACELLULAR DNA FROM SOIL. *Molecular Ecology*, *21*(8), 1816–1820. https://doi.org/10.1111/j.1365-294X.2011.05317.x

Takahara, T., Minamoto, T., & Doi, H. (2015). Effects of sample processing on the detection rate of environmental DNA from the Common Carp (Cyprinus carpio). *Biological Conservation*, *183*, 64–69. https://doi.org/10.1016/j.biocon.2014.11.014

Takahata, N., & Nei, M. (1985). Gene Genealogy and Variance of Interpopulational Nucleotide Differences. *Genetics*, *110*(2), 325–344.

Thomsen, P. F., & Willerslev, E. (2015). Environmental DNA – An emerging tool in conservation for monitoring past and present biodiversity. *Biological Conservation*, *183*, 4–18. https://doi.org/10.1016/j.biocon.2014.11.019

Torti, A., Lever, M. A., & Jørgensen, B. B. (2015). Origin, dynamics, and implications of extracellular DNA pools in marine sediments. *Marine Genomics*, *24*, 185–196. https://doi.org/10.1016/j.margen.2015.08.007

Tsuji, S., Takahara, T., Doi, H., Shibata, N., & Yamanaka, H. (2019). The detection of aquatic macroorganisms using environmental DNA analysis—A review of methods for collection, extraction, and detection. *Environmental DNA*, *1*(2), 99–108. https://doi.org/10.1002/edn3.21

Turba, R., Richmond, J. Q., Fitz-Gibbon, S., Morselli, M., Fisher, R. N., Swift, C. C., Ruiz-Campos, G., Backlin, A. R., Dellith, C., & Jacobs, D. K. (2022). Genetic structure and

historic demography of endangered unarmoured threespine stickleback at southern latitudes signals a potential new management approach. *Molecular Ecology*, mec.16722. https://doi.org/10.1111/mec.16722

Turner, C. R., Barnes, M. A., Xu, C. C. Y., Jones, S. E., Jerde, C. L., & Lodge, D. M. (2014). Particle size distribution and optimal capture of aqueous macrobial eDNA. *Methods in Ecology and Evolution*, *5*(7), 676–684. https://doi.org/10.1111/2041-210X.12206

Turner, C. R., Miller, D. J., Coyne, K. J., & Corush, J. (2014). Improved methods for capture, extraction, and quantitative assay of environmental DNA from Asian bigheaded carp (hypophthalmichthys spp.). *PLoS ONE*, *9*(12), 1–20. https://doi.org/10.1371/journal.pone.0114329

Turner, C. R., Uy, K. L., & Everhart, R. C. (2015). Fish environmental DNA is more concentrated in aquatic sediments than surface water. *Biological Conservation*, *183*, 93–102. https://doi.org/10.1016/j.biocon.2014.11.017

USFWS. (2021). *Unarmored threespine stickleback (Gasterosteus aculeatus williamsoni) 5-Year Review: Summary and Evaluation*. Ventura Fish and Wildlife Office. https://ecos.fws.gov/ecp/species/7002

van der Loos, L. M., & Nijland, R. (2021). Biases in bulk: DNA metabarcoding of marine communities and the methodology involved. *Molecular Ecology*, *30*(13), 3270–3288. https://doi.org/10.1111/mec.15592

Van Wormer, S. R. (1991). A History of Flood Control in the Los Angeles County Drainage Area. *Southern California Quarterly*, *73*(1), 55–94. https://doi.org/10.2307/41171560

Wasserman, B. A., Paccard, A., Apgar, T. M., Des Roches, S., Barrett, R. D. H., Hendry, A. P., & Palkovacs, E. P. (2020). Ecosystem size shapes antipredator trait evolution in estuarine threespine stickleback. *Oikos*, *129*(12), 1795–1806. https://doi.org/10.1111/oik.07482

WeatherSpark.com. (n.d.). *Historical Weather Summer 2018 at Point Mugu Naval Air Warfare Center*. WeatherSpark.Com. Retrieved June 8, 2022, from https://weatherspark.com/h/s/145310/2018/1/Historical-Weather-Summer-2018-at-Point-Mugu-Naval-Air-Warfare-Center;-California;-United-States#Figures-Rainfall

Weir, B. S., & Cockerham, C. C. (1984). Estimating F-Statistics for the Analysis of Population Structure. *Evolution*, *38*(6), 1358–1370. https://doi.org/10.2307/2408641

Wickham, H. (2016). *ggplot2: Elegant Graphics for Data Analysis*.

Williams, K. E., Huyvaert, K. P., & Piaggio, A. J. (2017). Clearing muddied waters: Capture of environmental DNA from turbid waters. *PLOS ONE*, *12*(7), e0179282. https://doi.org/10.1371/journal.pone.0179282

Wolf, J. B. W., & Ellegren, H. (2017). Making sense of genomic islands of differentiation in light of speciation. *Nature Reviews Genetics*, *18*(2), Article 2. https://doi.org/10.1038/nrg.2016.133

Zheng, X., Levine, D., Shen, J., Gogarten, S. M., Laurie, C., & Weir, B. S. (2012). A high-performance computing toolset for relatedness and principal component analysis of SNP data. *Bioinformatics*, *28*(24), 3326–3328.

Zinger, L., Lionnet, C., Benoiston, A.-S., Donald, J., Mercier, C., & Boyer, F. (2020). *metabaR: An R package for the evaluation and improvement of DNA metabarcoding data quality* [Preprint]. Bioinformatics. https://doi.org/10.1101/2020.08.28.271817

Chapter 2 . Parallel evolution of the unarmored condition in southern California threespine stickleback

Abstract

Threespine stickleback have been the target of many studies throughout the years in a range of biological fields, including evolution and ecological adaptation. However, much of this work has been focused on populations distributed at higher latitudes in the northern hemisphere. In those cases, the genetic basis of post-glacial transition from fully armored, marine morphs to low armored, freshwater morphs in rivers and lakes has been a topic of considerable interest. Nevertheless, the endangered unarmored populations of southern California remain poorly studied and suffer a history of habitat loss that is intensifying in the form of fires and drought. Here, we investigate the history of introgression of populations in the southernmost distribution of eastern Pacific threespine stickleback, and identify genomic regions putatively associated with the uncommon unarmored condition. We used whole-genomes from 86 individuals distributed across 36 sites in and around southern California, including a museum specimen from the Los Angeles River collected in 1925. We used a combination of approaches to investigate introgression and genome scans to look for evidence of parallel adaptation of the unarmored trait in the region. Our results reinforce previous findings related to the subdivision of populations in two broad groups, defined as Outer Coast and Southern California Bight and independent evolution of the unarmored form in each of these. Divergent unarmored forms appear independently derived, with one population related to the Outer Coast group, in San Antonio Creek, while the unarmored morph, *Gasterosteus aculeatus williamsoni*, near the type locality, derives from the Bight group in the upper Santa Clara River. Thus, unarmored populations in southern California appear to represent another case of multiple parallel adaptation in threespine stickleback, although further analyses are

needed to unambiguously identify regions under selection that are associated with the loss of plates. Additionally, there is evidence supporting multiple historic transplants of fish and human mediated introgression of populations. Of particular interest is the genetic similarity of the museum sample from the LA River to populations from the San Bernardino Mts, suggesting sourcing of these mountain fish naturally, or more likely, artificially from the river systems of the LA Basin, which was historically known to have had unarmored populations. This suite of information provides fundamental insights to inform conservation priorities and actions regarding recovery of populations of special interest in the region.

Introduction

Threespine stickleback (*Gasterosteus acuelatus*) is one of the best-studied systems regarding radiation of form, physiology and behavior and provide an example of ecological adaptation in vertebrates (Hohenlohe & Magalhaes, 2020; Peichel & Marques, 2017; Reid et al., 2021). This species complex comprises marine, anadromous and freshwater populations, and is widely distributed across the Northern Hemisphere, with a Pacific origin in the Late Pleistocene and subsequent colonization of the Atlantic through the Arctic Ocean ca. 44.6 Kya (Fang et al., 2018).

This complex comprises marine, anadromous, and freshwater populations and morphology varies widely across its broad geographical distribution (Bell & Foster, 1994). Early on, bony armor was used to diagnose species (Cuvier & Valenciennes, 1829) and even to this day is by far the best studied morphological change in sticklebacks, with four recognized plate morphs: (i) fully armored, with +30 plates per side and predominantly found in marine and anadromous populations; (ii) partial, with a discontinuous row of +11 plates per side, and occurring in brackish waters; (iii) low, with fewer than 11 plates, and (iv) unarmored, with no plates at all. The last two

are predominantly found in freshwater environments (Bell, 1976; Bell & Foster, 1994; Miller & Hubbs, 1969).

Diversification of morphology, but also behavior, life history and physiology traits, have occurred after colonization of freshwater habitats by marine and anadromous stickleback. Modification of the landscape facilitated by glacial and interglacial periods has favored colonization and isolation of populations in freshwater systems (Hohenlohe & Magalhaes, 2020; Reid et al., 2021), and many freshwater populations at higher latitude formed after the end of the Last Glacial Maxima (Colosimo et al., 2005; Deagle et al., 2013b; Liu et al., 2016). However, the southern region of western North America remained largely unglaciated during the Last Glacial Maximum (Hendy, 2009), and the complex hydro and geomorphological history of southern California have shaped much of the intricate history of threespine stickleback populations in the region.

Southern California is also special in that it harbors the rare unarmored morph of threespine stickleback. The unarmored morph is recognized as a subspecies, *G. a. williamsoni*, and listed as endangered by the state of California (https://wildlife.ca.gov/Conservation/CESA) and the U.S. federal government (35 Federal Register 16047). The original listing relates to individuals from the type locality in upper Santa Clara River, Soledad Canyon (Girard, 1854), but this morphology has been documented upstream of at least other four different drainages in southern California, including broadly the rivers of the Los Angeles Basin (Swift et al., 1993a). Nowadays, it is present in only three: San Antonio Creek on Vandenberg Space Force Base (VSFB, Santa Barbara County); the upper Santa Clara River (Los Angeles County); and the upper Santa Ana River and associated Baldwin Lake drainage basin in the San Bernardino Mountains (San Bernardino County) (USFWS, 2021). Previous work (Turba et al., 2022) has shown that the unarmored

populations represent distinct genetic units, with different demographic histories, and signals of isolation and bottleneck, which would indicate another instance of parallel change in plate morphology for this system.

Recurrent and widespread instances of parallel change in plate number suggests a strong selective pressure on this phenotype, which is known to be driven by predation and habitat change (Bergstrom, 2002; Des Roches et al., 2020; Eriksson et al., 2021; Hagen & Gilbertson, 1973; Morris et al., 2018; Reimchen, 1994, 2000; Wasserman et al., 2020). However, most investigations have been restricted into looking at transitions from fully armored (marine) to low armored (freshwater) populations, especially at higher latitudes (e.g., (Colosimo et al., 2005; Cresko et al., 2004; Indjeian et al., 2016; O'Brown et al., 2015a). These studies have shown that change in plate number is largely (ca. 70%, (Colosimo et al., 2004) the result of selective pressures on a single large-effect QTL on chromosome four associated with the Ectodysplasin A gene (*Eda*) (Colosimo et al., 2004). This gene is highly pleiotropic and is important for the development of other tissues, and even behavior (Peichel & Marques, 2017). Alleles have a largely Mendelian inheritance, with fully armored stickleback typically homozygous for the fully armored allele and low armored, homozygous for the low allele (Barrett et al., 2008; Colosimo et al., 2004).

Nevertheless, other regions have also been associated with the development of lateral plates (e.g., Colosimo et al., 2004; Conte et al., 2015; Cresko et al., 2004; Peichel & Marques, 2017). A study investigating plateless in Atlantic populations (Mazzarella et al., 2016) found at least 18 genomic regions potentially involved with within-morph plate number variation, suggesting a polygenic effect. However, there is understanding that standing genetic variation between the Pacific and Atlantic populations differs, with a substantially reduced pool in the latter after colonization and subsequent isolation of the Atlantic by ancestral Pacific populations (Fang et al.,

2020; Kingman et al., 2021). Moreover, variation is expected to largely occur in regulatory elements, especially when related to large-effect QTLs that have important pleiotropic effects. For example, mutations associated with differential expression of another gene of intermediate effect size in threespine stickleback, Growth/Differentiation Factor 6 (*GDF6*), has also been shown to relate to armor-plate size in freshwater fish (Indjeian et al., 2016).

Much remains to be understood of the genetic basis for the complete loss of armor, as also the evolutionary mechanisms governing this change (e.g., independent evolution of alleles or dispersion over the landscape). Latitudinal distribution of the *Eda* alleles in the eastern Pacific shows a nearly complete fixation of the low armored allele at lower latitudes (Des Roches et al., 2020; Morris et al., 2018), where bar-built lagoons are the most prevalent coastal wetland (D. K. Jacobs et al., 2011a). Therefore, we hypothesize that other, potentially multiple, genomic regions will be involved in the unarmored phenotype, as has been shown in previous genome-wide works (Conte et al., 2015; Kingman et al., 2021; Marques et al., 2018; Mazzarella et al., 2016).

Here, we further develop from previous work on the population structure of *G. aculeatus* in southern California (Turba et al., 2022) by using a larger whole-genome dataset with deeper sequencing. Moreover, we aim to identify genomic regions associated with the unarmored condition and potential parallelism of divergence in two unarmored populations, one in the upper Santa Clara River, the type locality of the unarmored morph, and the other in the San Bernardino Mountains. We used a sliding window approach, which has been extensively used in studies of adaptation in a myriad of systems (Enciso-Romero et al., 2017; Han et al., 2017; Meier et al., 2018), to calculate levels of relative (F_{st}) and absolute (D_{xy}) divergence, as well as windows of shared variation (f_d), based on ABBA-BABA test. Results from this work will help inform

management and conservation of the unarmored morph in southern California, by helping narrow down relevant genomic regions of the protected morph and safeguarding them for the future.

Material and Methods

Sites

Samples include 86 individuals of full, low and unarmored *G. aculeatus* from 36 localities, most of those distributed across southern California (Figure 2.1, Table 2.1). We included individuals from Bocana El Rosario in northern Baja California, Mexico, representing the southernmost population of *G. aculeatus* in the eastern Pacific, recently extirpated (Ruiz-Campos & González-Acosta, in press); and two fully plated populations collected in the San Francisco Bay area (Bodega Bay and San Luis Obispo). Museum samples from the Los Angeles River collected in 1925 were provided by Christine Thacker and Rick Feeney of Ichthyology at the Natural History Museum of Los Angeles County (LACMNH), and one individual provided enough endogenous material for further high-throughput sequencing. Two outgroups were also included in the dataset: *G. nipponicus* from the Japan Sea (Higuchi et al., 2014) and *G. wheatlandi* from the northwestern Atlantic (Mattern, 2004).

DNA extractions

Details of how tissue samples were obtained and processed are described in Richmond et al. (2015) and Turba et al. (2022). DNA extraction was performed using a QIAGEN DNeasy Blood & Tissue Kit. DNA extraction for the museum samples followed the protocol from (Campos & Gilbert, 2012) at a separate, clean facility where only degraded material (e.g., ancient and museum samples) is processed. Care was taken to avoid contamination by wearing protective gear, i.e.

plastic gown, shoe covers, hairnet, surgical mask and double gloves; and also by sterilizing surfaces and utensils with 20% bleach, 70% ethanol, DNA and RNA removal solution and UV light.

Library preparation and Sequencing

DNA extracts were purified using a Zymo cleaning kit (ZR-96 Genomic DNA Clean and Concentrator-5 kit, Cat: D4066). Concentrations were measured using a Qubit 2.0. Samples followed the SparQ DNA library prep protocol aiming at a final concentration of 80 ng with a few modifications: (i) half reaction volumes (20 µl); (ii) 1 µl of undiluted adapters; (iii) half PCR volume (13.5 µl); (iv) 8 PCR cycles; (v) 0.9X post amplification clean-up. Samples were quantified and pooled to reach 10 nM concentration. The first round of sequencing, with 72 individuals, was done in a NovaSeq S4, 2x150 bp at the Technology Center for Genomics & Bioinformatics (TCGB) at UC Los Angeles.

DNA extracts from museum samples were sent to the Paleogenomics lab at UC Santa Cruz for single-stranded library preparation. After quality check, one successful library was chosen to be pooled with a new batch of modern individuals, adding to the previous run, in a NovaSeq SP 2x150 bp at TCGB (UCLA).

Alignment, Variant Calling and Filtering

We used the revised assembly of the Bear Paw Lake threespine stickleback reference genome (Nath et al., 2020) for the alignment. Alignment was done using BWA-MEM v. 0.7.12 (H. Li, 2013a), and variant discovery, using HaplotypeCaller and GenotypeGVCFs from GATK (McKenna et al., 2010) using default parameters to generate the VCF file of modern samples.

Whole-genomes for the outgroup species were downloaded from NCBI: *G. nipponicus*, a sympatric species to *G. aculateaus* in the Japan Sea (SRA DRR147065) (hereafter referred to as JAP); and *G. wheatlandi*, which is a species that occurs in the northwestern Atlantic (DRR013347) (hereafter referred to as ATL). Alignment followed the same pipeline described above.

The museum sample followed a modified protocol (for details see Supplementary Information, 'Scripts and code availability'), where alignment was done using the BWA-ALN option, which is more suitable for short reads. After damage analysis using mapDamage2 (Jónsson et al., 2013, Figure S 1.1), quality scores of BAM files were corrected using the --rescale option prior to alignment to the reference genome.

SNPs and INDELs were filtered using standard metrics from GATK. For SNPs: QD < 2.0, MQ < 40.0, FS > 60.0, SOR > 3.0, MQRankSum < -12.5, ReadPosRankSum < -8.0. For INDELs: QD < 2.0, ReadPosRankSum < -20.0, InbreedingCoeff < -0.8, FS > 200.0, SOR > 10.0. The final dataset was then divided into different sets and converted to plink (details in Supplementary Information):

- Set 1a: modern samples with both outgroups and museum individual;
- Set 1b: modern samples with both outgroups;
- Set 2: modern samples with museum individual;
- Set 3: only modern samples.

PCA and FastSTRUCTURE

All analysis done in R in this work used version v.3.6.2 (R Core Team, 2018a) and RStudio v.2022.02.3 (RStudio Team, 2020a). For the Principal Component Analysis (PCA), we used Set 2. The package SNPrelate (Zheng et al., 2012) and ggplot (Wickham, 2016) were used to generate

the PCA plots in R. Variants were filtered to retain only biallelic SNPs (snpgdsVCF2GDS, method = "biallelic.only"), and pruned for linkage (snpgdsLDpruning, ld.threshold = sqrt(0.1), method = "corr"). For further details on population structure, we looked at admixture proportions with the fastSTRUCTURE software (Raj et al., 2014), see Supplementary Information for details).

Heterozygosity

Bcftools v1.10.2 (Danecek et al., 2021) was used to calculate numbers of heterozygous and homozygous sites. For this calculation, the software only considers SNP data and not INDELs. The heterozygosity ratio (nHets (RA) / nHom (AA) (Samuels et al., 2016) was calculated for each individual and plotted in R. This metric was chosen for its stability regarding genotyping density and genomic location, and based on Hardy-Weinberg equilibrium (Guo et al., 2014) the expected ratio should be 2 (meaning twice the amount of heterozygous SNPs compared to non reference homozygous SNPs).

Fst

We used a custom R script (for details see Supplementary Information, *Scripts and code availability*) to calculate pairwise FST for each location using the SNPrelate package and method = 'W&C84' (Weir & Cockerham, 1984). In the script developer's analysis, the mean F_{st} has a systematic downward bias compared to the other packages, so we used the weighted F_{st} on Set 2 filtered to contain only biallelic SNPs and no missing data (missing.rate = 0).

TreeMix

We used TreeMix to infer patterns of population splitting and mixing from allele frequency data. This involves two steps: (i) generating a maximum likelihood tree for the set of populations; then

(ii) inferring a number of admixture events that improve the fit of the modeled tree (Pickrell & Pritchard, 2012). One limitation of this approach is that it models migration events as occurring at a single, instantaneous point; and assumptions break down when the number of admixed populations is higher than the number of non admixed ones. All of these factors can become a challenge to our dataset, considering the extensive history of movement of *G. aculeatus* in the landscape.

The plink binary fileset for Set 1a (modern + outgroup + museum specimens) was cleaned of any missing data (plink --geno 0), converted to a cluster-stratified allele frequency report (plink --freq --within options), then converted to a TreeMix input file using an already-made script (plink2treemix.py). TreeMix was run in 10 iterations for each round of increasing migration edges (from 0 to 12), with bootstrap replicates by resampling blocks of 10 SNPs (-bootstrap -k 10), -global option and -root on the ATL species. We plotted the residuals for each number of migration edges, and also calculated the optimal number of migration edges using the optM package (Fitak, 2021) in R using the Evanno and SiZer methods.

Admixtools

f-statistics was computed using the R package ADMIXTOOLS2 (Maier et al., 2022). This statistic measures allele frequency correlations among populations, with f_3 providing a formal test of admixture and f_4 also providing directionality of gene flow. Both f_3- and f_4-statistics are based on f_2, which measures the amount of genetic drift separating two populations (Patterson et al., 2012). f_2-statistics was computed for Set 1a without excluding any SNPs (maxmiss = 1), since depth of coverage for the Los Angeles River (LAR) individual was low (~2.5X, Table S1-S2). This approach is less conservative, but it is considered reasonably safe when missing data is generated

at random (e.g., when samples were sequenced and genotyped following the same protocols), and work well when using more exploratory tools, such as the graph functions (Maier et al., 2022).

In total, 41,460,503 SNPs remained after filtering, of which 632,955 are monomorphic (i.e., a single form or allele is identified). We used the function find_graphs() to generate admixture graphs compatible with the calculated *f*-statistics fixing the number of admixing events on two (based on the TreeMix results), and the outgroup on the Atlantic species (*G. wheatlandi*, ATL).

We used outgroup f_3- and f_4-statistics to formally test admixture between the LAR individual to other populations in the region. The outgroup f_3-statistics had the JAP as the outgroup (Pop A), LAR as Pop B, and the remaining populations as Pop C (removing the ATL outgroup). We also tested individuals from San Bernardino Mountains (SBM) (Shay Creek: SHC; and Juniper Creek: JUC) as P2 compared to all other populations.

For the f_4-statistics, we ran two tests, both of them had JAP as the outgroup. The first test was looking into admixture between the LAR individual to populations from SBM, admixed individuals from Apple Valley, populations in San Juan Creek (Trabuco Canyon and Bell Creek) and the Outer Coast (OC) group (San Luis Obispo, Arroyo Grande and San Antonio Creek). The second test was focused on admixture between the SBM populations and admixed Apple Valley individuals to individuals from San Juan Creek and the OC. In this second test we excluded the Sugarloaf Meadow population from analysis since it is a known transplant from Shay Creek. Therefore we only included the latter for simplicity.

Sliding windows of differentiation

We used custom scripts developed by Martin (for details see Supplementary Information) to perform divergence analysis in sliding windows across the genomes. We converted our VCF file

to GENO files using the parseVCF.py script. We defined populations for the pairwise comparisons based on results from previous analyses: (1) low armored Outer Coast (San Luis Obispo, Arroyo Grande and Shuman Lagoon) versus unarmored San Bernardino Mountains (Shay Creek, Sugarloaf Meadow and Juniper Creek); and (2) low-armored Santa Clara Lagoon (McGrath and Oxnard) versus unarmored upstream Santa Clara River (Robins Nest, Thousand Trails and San Felipe Creek). The outgroup was defined as the *G. nipponicus* species (JAP). Then, we ran the popgenWindows.py script to compute *pi*, F_{st} and D_{xy} in non-overlapping 10kb windows for both groups, with the minimal number of sites (-m) set to 100. Unlike relative measures of divergence such as F_{st}, absolute measures of divergence (D_{xy}) are less sensitive to processes affecting within-population variation, such as background selection (Cruickshank & Hahn, 2014; Ravinet et al., 2013). We calculated the mean distribution of both F_{st} and D_{xy} across the whole genome, and extracted windows at the 90th percentile distribution of each metric in each pairwise comparison, as well as windows shared by both, and calculated the mean distribution in each case. We also looked at overlapping windows within the top 10% F_{st} and D_{xy} values and calculated the mean distribution of each metric.

For a validation of this method, we also computed *pi*, F_{st} and D_{xy} for the pairwise comparison between fully armored individuals from Bodega Bay and low armored individuals from Bocana El Rosario. We expect to see a divergence peak at the *Eda* locus on chromosome 4 (Colosimo et al., 2004; Peichel & Marques, 2017), where alleles for full and low armored are expected to be nearly fixed, respectively.

We also computed *D* and f_d statistics using the same window parameters as above but defining - m 10. Both are based on the ABBA-BABA test (S. H. Martin et al., 2015) and are used to measure allele sharing between two populations by comparing allele frequencies among four

populations, but f_d is more robust to low numbers of SNPs (Meier et al., 2018). To estimate ABBA-BABA patterns three ingroup and one outgroup population are required following the relationship (((P1,P2),P3),O). The outgroup must carry the ancestral allele (A) while P3 the derived allele (B), with P1 and P2 carrying either (A or B). Under a scenario with no hybridization or selection pressure, both patterns should be equally frequent (ABBA-BABA). Excess sharing between P2 and P3 is a signal of either gene flow or parallel selection on shared ancestral polymorphisms (Enciso-Romero et al., 2017; Meier et al., 2018). For this analysis we defined the outgroup as JAP, P1 as the OC group, P2 as the unarmored, upstream Santa Clara River populations and P3 as the unarmored SB populations. For negative values of D, f_d values were converted to 0 (S. H. Martin et al., 2015). We calculated the mean f_d distribution across the whole genome and checked if windows at the 90th percentile distribution matched the overlapping windows within the 90th percentile F_{st} and D_{xy} values.

Recombination rates will have an influence on the degree of divergence between groups as regions of low recombination will show reduced nucleotide diversity. Therefore, we downloaded the dataset on recombination rates generated by (Shanfelter et al., 2019) for the 20 autosomal chromosomes, and looked at the correlation between rates of recombination and metrics of divergence (F_{st} and D_{xy}). The dataset was averaged in 10 kb windows before analysis. Rates were calculated using an older version of the threespine stickleback genome (F. C. Jones, Grabherr, Chan, Russell, Mauceli, Johnson, Swofford, Pirun, Zody, & White, 2012) than the one used in this work, therefore we manually edited the windows to match the chromosome positions as best as possible by trimming the start and ends of scaffolds (Table S 1.4). Because recombination rates were calculated on a different reference genome, we did not look at the specific correlations of the overlapping windows within the 90th percentile of F_{st} and D_{xy} values and the recombination

windows, since chances of mismatch at this scale would be high. Plots were done in R (see Supplementary Information for details).

Results

Alignment, Variant Calling and Filtration

After filtration, individuals had an average of ~15X coverage, ranging from 6.9-23.4X (with ca. 2-4% missing data, Table S 2.1-Table S 2.2). One individual with 0.6X mean coverage from the San Antonio Creek tributary of the Ventura River (JQR808) was removed from all downstream analyses, resulting in 87 individuals representing 38 populations, including the outgroup species (Figure 2.1, Table 2.1). The museum individual (LAR) had 2.7X coverage with ~29% missing data; *G. nipponicus* (JAP), 64.4X coverage with ~4% missing data; and *G. wheatlandi* (ATL), 50.8X coverage with 18% missing data. Final filtered dataset including modern, museum samples and both outgroup species had a total of 37,248,702 variants (SNPs and INDELs).

PCA and FastStructure

The reduced dataset (after filtering to retain only biallelic SNPs and pruning for linkage disequilibrium) had 83,570 variants (Set 2). The percent of variance retained by the 32 PCs ranged from 3.26-1.16%. Along PC1 (Figure 2.2A) there is separation between populations in the Southern California Bight (SCB), i.e., Santa Clara and Ventura Rivers, and populations from the OC and SBM. PC2, on the other hand, shows separation between the SCB populations following a gradient from Ventura to lower and upstream Santa Clara River, where the unarmored populations are situated, at Soledad Canyon (Robins Nest and Thousand Trails) (Figure 2.2B). On

PC3, the unarmored population of San Antonio Creek is distinct from the remainder of the OC populations, and this axis effectively separates all three unarmed populations from each other.

Individuals from the upper part of the Bouquet Creek tributary are more associated with fish from the lower reaches of the Santa Clara River, and is a well-documented case of introgression within the Santa Clara River (Richmond et al., 2015). Other individuals at intermediate positions in the PCA are Mission Creek, San Jacinto, Valyermo, Apple Valley and the unarmored museum individual (LAR), and many of these may relate to known or suspected cases of introgression. While three San Antonio Creek individuals show associations with populations from OC (JQR893, RNF8218 and RNF8220), two others (RNF8267 and RNF8268) show instead closer association with those from Valyermo, casting doubts to its origins and assignment to the San Antonio Creek.

Results from the fastSTRUCTURE analysis mostly corroborate the PCA results, showing overall clustering between OC and SCB populations, with a third one showing affinities between the SBM, Apple Valley and LAR populations (Figure S 2.2, Supplementary Information). Assignments of the LAR individual should be taken with caution, however, considering the level of degradation of the material and low sequence coverage.

Heterozygosity

The northernmost outer coast samples had the highest levels of genetic diversity compared to other populations, with heterozygosity ratios above one (i.e., the proportion of heterozygosity compared to non-reference homozygous sites is higher than 1:1). Coastal and stream sites south of Point Conception show reduced levels of genetic diversity, with heterozygosity ratios falling below one

(Table S 2.2; Figure S 2.3). Exceptions are individuals from Apple Valley and one individual from Mission Creek, both localities with suspected introgression.

Upstream populations are expected to show lower heterozygosity. However, the unarmored populations, all of which are upstream, show different amounts of reduction in heterozygosity ratio. The lowest heterozygosities were observed in populations upstream the Santa Clara River, including individuals transplanted from there, the type locality (Soledad Canyon) to San Felipe Creek and presumptively to Sweetwater River. Unarmored fish from SBM show relatively high heterozygosity ratios, in about the same range as other fish from more coastal sites in the SCB, with the Juniper Creek population showing the highest level among the three sites. While two (out of three) individuals from San Antonio Creek (SAC) show low levels of heterozygosity, similar to those seen in the upper Santa Clara River, one individual had levels comparable to the SBM and other coastal populations from SCB.

The unarmored LAR museum sample and one individual of the Ventura River tributary, San Antonio Creek (SAV, Ventura Co), also show very reduced levels of heterozygosity, but in the case of the San Antonio fish, it is not a consistent pattern, with the other two individuals showing heterozygosity levels similar to other lower SCB populations.

F_{st}

The pairwise F_{st} table (Table S 2.3) supports the overall trend observed in the PCA analysis. The OC, SCB, and SBM populations form groups with lower F_{st} within and higher F_{st} between groups. Low F_{st} (arbitrarily defined as below 0.1) are evident between geographically close and hydrologically connected streams, such as sites near the type locality, Thousand Trails and Robin's Nest in Soledad Canyon, and sites derived from them (e.g., San Felipe Creek); Newhall Ranch and

Valencia; lower sites in the Santa Clara and Ventura Rivers, up to Piru Creek; upper and lower Bouquet Canyon; adjacent tributaries of San Juan Creek (Trabuco and Bell Canyons).

The SBM populations (Sugarloaf Meadow, Shay and Juniper Creeks) are similarly minimally differentiated (F_{st}<0.1), and they are likely the result of management transplantation. Sugarloaf Meadow is a founder population from Shay Creek and is the most differentiated amongst these sites.

The northernmost OC populations (Bodega Bay, San Luis Obispo Creek, Arroyo Grande and Shuman Lagoons) show relatively lower F_{st} amongst them, with individuals from Arroyo Grange and Shuman Lagoon falling below F_{st}<0.1. Shuman is an ephemeral site, therefore the low F_{st} is likely due to local extirpation and recolonization from Arroyo Grande Lagoon, which is situated up the coast. There is increasing F_{st} within the OC group as sites are found at lower latitudes and populations transition from more connected to the ocean, such as Bodega Bay, to more closed lagoonal systems and up streams. For example, individuals from the San Antonio Creek system were collected upstream and showed higher F_{st} compared to other OC populations nearby.

Upstream populations far from the coast, such as the unarmored populations from the SBM and Santa Clara River, show consistently high F_{st} when compared to other populations and with each other. F_{st} values are lower when SBM or OC populations are compared to APV, suggesting an introgressive origin of this population (as is also suggested in the PCA, Figure 2.2). Mission Creek individuals showed the lowest F_{st} when compared to individuals from Salsipuede and San Jacinto, also suggesting a shared history of introgression amongst these samples.

TreeMix

The maximum likelihood tree generated by TreeMix shows the southern California threespine stickleback populations as sister to the Japanese species, *G. nipponicus* (Figure 2.3). Within the southern California clade, the SBM populations form a group with the LAR individual. The OC group is paraphyletic, but a group is formed with individuals from San Luis Obispo, Arroyo Grande Lagoon, Shuman Creek and San Antonio Creek. Trabuco and Bell Canyon also show as sisters to one another.

Results from the optM (Figure S 2.5) and residuals (Figure S 2.6) for the optimal number of migration edges did not converge. Best deltaM was two, but there was a second peak on m = 6. The SiZer map also shows that derivatives of the smoothed curve plateau around m = 6 as well. However, the residual map improves only up to the number of migration edges at five. After that, the fit gets worse, and it starts to over/underestimate the observed covariances. Therefore, we discuss results for the two migration edges. Results for the remainder migration numbers can be found in the Supplementary Information section.

This lack of convergence in the results for optimal number of edges could be driven by the proximity of the populations and their complicated history of admixture. Increased complexity can lead to different graphs with identical covariance matrices, and large numbers of admixture can be accommodated by allowing slight underestimation between populations (Pickrell & Pritchard, 2012).

Nevertheless, migration edges reflect patterns of admixture that were also captured on the PCA and fastSTRUCTURE results. One edge is connecting populations from the SBM to individuals from Apple Valley, and another one is connecting individuals from Valyermo to San

Jacinto. Both populations (Apple Valley and San Jacinto) are the result of transplants from trout stocking. In the case of Apple Valley, this site used to host a hatchery, while San Jacinto had no previous evidence of native *G. aculeatus* inhabiting the system.

Admixtools

The best admixture graph (Figure S 2.7) generated based on our calculated f_2-statistics and fixed on two migration events show admixture events that were not recovered by other methods (TreeMix, fastSTRUCTURE, PCA), and are also unrelated to known and suspected histories of translocation. It is evident that the drift branch lengths are zero for most cases, and therefore this method might not have had enough power to perform admixture tests (Lipson, 2020). This limitation was also likely driven by the inclusion of the museum individual, which had a high proportion of missing data and prevented us to use the same SNPs for every f_4-statistic. It is possible that the use of a less stringent approach (details in the Material and Methods) impacted results, even though graph analysis is supposed to be less sensitive (https://uqrmaie1.github.io/admixtools/articles/fstats.html).

$f3$-statistics show that the SBM populations are the ones most closely related to the LA River individual (Figure 2.4-Figure 2.5), with f_4-statistics (Table 2.2) showing significant shared alleles between LAR and Sugarloaf Meadow, Shay and Juniper Creeks. Admixture to Apple Valley was only significant when Pop C was one of the populations in San Juan Creek (Trabuco and Bell Canyons). Results were similar when testing specifically populations from the SBM (Sugarloaf Meadow, Shay and Juniper Creeks) (Figure 2.5), showing significant shared alleles between those populations and individuals from Apple Valley and LAR. There is also evidence of

shared alleles between the SBM and Apple Valley populations with those from San Juan Creek (Table 2.3).

Sliding windows of differentiation

Windows within the 90th percentile of relative (F_{st}) and absolute (D_{xy}) divergence for each pairwise comparison were about twice as high as the average F_{st} and D_{xy} calculated across the whole-genome, and values increased about 0.2-0.3 more when the top F_{st} and D_{xy} windows overlapped (Table 2.4-Table 2.5). The number of F_{st} and D_{xy} windows in the 90th percentile between the SBM and OC populations was 4,386 and 4,384, respectively, with 1,164 windows overlapping between these metrics. For the lower versus upper Santa Clara populations, there were 4,391 and 4,392 windows, respectively, with 701 overlapping ones (Table 2.4). There were 55 windows shared amongst the 90th percentile windows of F_{st} and D_{xy} between the two comparison groups, distributed across chromosomes 1-2, 4-8, 10-12, 17-20 and U (undefined) (Table 2.5, Figure 2.6A).

Distribution of F_{st} across the genomes showed low correlation between the pairwise comparisons (r^2 = 0.004, Figure S 2.9A). Levels of D_{xy} were more correlated between the population pairs (Figure S 2.9B), but the SBM versus OC pairwise comparison was on average higher than the levels of divergence of the Santa Clara Lagoon (SCL) versus upstream Santa Clara (SCU) comparison (0.08 vs 0.04, respectively, Table 2.4, Figure 2.6B). Correlations between F_{st} and D_{xy} for each pairwise comparison were low (Figure S 2.10). There was a slight improvement in fit for the SBM vs OC comparison for the 55 shared windows, but not for the SCL vs SCU pair, which in fact decreased (Figure S 2.11). *Pi* and D_{xy} were highly correlated for each pairwise comparison ($r^2 > 7$, Figure S 2.12).

f_d windows showed a heterogeneous pattern across chromosomes, with some having a more even distribution of values throughout its length, while others showed more variation and distinct peaks of shared alleles between both upstream populations, Santa Clara River and SBM (Figure 2.6C). f_d showed little correlation between F_{st} and D_{xy} windows (Figure S 2.13), although there was a slight improvement when looking only at the shared 55 windows (Figure S 2.14). Of these, 35 windows were in regions with f_d values higher than the genome-wide mean of 0.02 (Table S 2.5).

Genomic windows of differentiation, including the top shared F_{st} values, also had low correlation to recombination rates (Figure S 2.15). The average recombination rate across the genome is ~5.4, meanwhile the average rate for the shared top 90th percentile F_{st} windows was higher, at ~5.9. The unique top 90th percentile F_{st} windows for each pairwise comparison was ~4.1 for the SBM vs OC comparison, and ~5.6 for the SCL vs SCU comparison.

Our validation test was able to capture the expected divergence peak for plate transition on chromosome 4, around the region where the *Eda* gene is located (chrIV:12,783,708-12,793,934), for the Bodega Bay and Bocana El Rosario comparison (Figure S 2.16-Figure S 2.17). In addition, there is also a major peak on chromosome 21, where there is a known inversion with a supergene cluster also associated with marine-freshwater divergence in *G. aculeatus*. In total, there were 1,291 windows that overlapped between the top 90th percentile of F_{st} and D_{xy}.

Discussion

In this study, we have generated the first comprehensive genomic dataset of the southernmost populations of *G. aculeatus* in the eastern Pacific. This dataset has great potential for the exploration of evolutionary and ecological changes in populations that have been going through

environmental pressures on a progressively drier and warmer landscape throughout geologic times that precede the Last Glacial Maximum (18-20 ka; Owen et al., 2003; Sharp et al., 1959). Our work generally supports previous results, albeit adding more nuance, to those based on microsatellite data and limited low coverage whole-genomes (Turba et al., 2022), where populations are structured in two broad geographic groups, the OC and SCB, with the unarmored condition showing independently in each group. Below, we explore these findings in more detail, coupled with a sliding window of differentiation approach, that reinforces the idea of independent evolution of unarmored populations. Lastly, we address the current limitations of this analysis to find genomic regions associated with the unarmored condition and suggest future work to be done to explore putative signals of selection related to this phenotype.

Population Demographics

Similar to previous work (Turba et al., 2022), we have found that populations in southern California form two major clusters: OC and SCB. The PCA also captured a distinction between the SBM populations and individuals located along the coast (Figure 2.2), a pattern also observed in the fastSTRUCTURE results at $K = 3$ (Figure S 2.2), which included the LA River individual recovered from museum material.

In the PCA and fastSRUCTURE results, the OC group shows a relatively large-scale cohesion, grouping individuals from north of Point Conception, south to Baja California (Bocana El Rosario). As argued in earlier (Turba et al., 2022), this cohesion could be driven by change in sea level and widespread formation of lagoons (D. K. Jacobs et al., 2011a; Masters, 2006) facilitating gene flow between populations. Interestingly, based on the calculated pairwise F_{st} values (Table S 1.3) and the heterozygosity ratio levels (Figure S 2.3, Table S 2.2), the OC group

does not appear as cohesive as suggested. Individuals from the northernmost sites are less differentiated amongst themselves than from OC fish at lower latitudes. Excluding the San Antonio Creek site, all other northern populations are located north of Point Conception, a geographical barrier that delimits the change to warmer temperature waters in the California Bight (D. K. Jacobs et al., 2004). However, this segregation may be more a product of geographic distance and upstream isolation. Individuals from upstream San Antonio Creek, for example, show relatively high differentiation compared with other OC populations. The environment also responds to a latitudinal change that, as it shifts to a more arid, Mediterranean climate, it facilitates the formation of bar-built lagoons (Des Roches et al., 2020) that remain closed to the ocean for more prolonged periods of time, and therefore increase isolation between sites.

The gradient of admixture between fish from the SCB reported earlier (Richmond et al., 2015; Turba et al., 2022) has also been recovered in the current study (Figure 2.2-Figure S 2.2). In the PCA (Figure 2.2), it is visible the separation between populations from the upper Santa Clara River from the lower reaches, which in turn show closer association to individuals from the Ventura River. The mouths of both rivers are indeed geographically very close to one another, which facilitates gene flow between sites (<8 km apart, Richmond et al., 2015).

Populations at intermediate positions in the PCA suggest cases of introgression, as is also shown in the fastSTRUCTURE results (K = 2) and TreeMix analysis (Figure 2.3). Cases within the Santa Clara River have already been reported and discussed in detail in previous work (Richmond et al., 2015), demonstrating upstream introgression of lower Santa Clara River fish in Bouquet Canyon. Turba et al. (2022) documents other examples of introgression between major groups, based primarily on microsatellite work, many of which we confirm here. These comprise cases of known intentional and suspected admixture from regional sources of *G. aculeatus* from

the SCB, such as the case of the San Jacinto Creek and Sweetwater River. The Apple Valley site hosts a state fish hatchery with transplants from the upper Santa Clara River (Buth et al., 1984; Miller & Hubbs, 1969; Swift et al., 1993a), and individuals from Valyermo in the Mojave drainage may have been affected by transplants from this site. However, it also seems that individuals from Apple Valley might have been introgressed by fish that are more related to the SBM population and LA River fish (Figure S 2.2). The case of the Mission Creek population still remains elusive, as the site is located in the coast, between the northern OC sites and the SCB populations, and could be a case of natural introgression or introgression associated with transplanted trout.

Individuals from Matilija Creek represent a population that have been isolated from the remainder of the drainage since after the dam construction, in 1947 (https://matilijadam.org/background/). It has been noted that individuals at this site show increased genetic divergence from the remainder of the Ventura River populations, and increased plate count numbers (Bell & Richkind, 1981; Richmond et al., 2015), which was then attributed to selective pressures from the presence of predatory trout. Nevertheless, introgression via translocations cannot be completely ruled out.

The outgroup f_3- and f_4-statistics tests (Figure 2.4-Figure 2.5, Table 2.2-Table 2.3) also identify the SBM populations as the ones most closely related to the museum specimen from the LA River. Although data generated for the LA River individual is less robust, it is known that the river was historically physically and genetically connected to the Santa Ana and San Gabriel Rivers (Richmond et al., 2018). Besides, the LA River individual was collected in 1925, before the hatchery in Apple Valley was established in the 1940's, and any potential introgression happened. The Apple valley population also shows association with the SBM and LA River individuals and is likely introgressed.

The data, therefore, seems to suggest that individuals from the SBM, LA River and Apple Valley may retain the genetic remnants of an interconnected population that used to inhabit the LA Basin region (Culver & Hubbs, 1917; Swift et al., 1993a). Even though the SBM individuals may not represent natural populations that have been uplifted with the mountains, they could still be the result of transplants from the lower reaches of the Santa Ana River. This cluster formed by the SBM, LA River and Apple Valley populations also seems to be more closely related to individuals from San Juan Creek than the northernmost OC populations. This association is expected considering the proximity of the Santa Ana River and San Juan Creek mouths. This would have allowed for increased gene flow between these sites following historic floods, as is evident in our data for the Ventura and Santa Clara Rivers, and is evident in historic processes across the rivers of the LA Basin (Moyle, 2002; Richmond et al., 2018).

These results reinforce our previous finding that the unarmored populations in southern California are separate genetic units. But most importantly, it suggests that the unarmored individuals in the SBM potentially retain some ancestral, genetic diversity relative to this larger extirpated group from the LA Basin.

Sliding windows of differentiation

Further evidence that unarmored populations are following independent evolutionary processes comes from the genome-wide analysis of differentiation of the unarmored individuals from the upper Santa Clara River (including the San Felipe transplant), and the SBM. The majority of F_{st} windows within the 90th percentile were unique to each pairwise comparison, with very little correlation between them (r^2 = 0.004, Figure S 2.9A). As mentioned earlier, F_{st} is a measure of

relative divergence, i.e., it compares differentiation between groups to within groups, and is more readily affected by recent processes that increase or decrease within group allelic variation.

In contrast, D_{xy} is a measure of absolute divergence that is not affected by forces that decrease within-population variation (Cruickshank & Hahn, 2014; Noor & Bennett, 2009; Ravinet et al., 2017), and for recently diverged taxa it is expected that D_{xy} will relate to mutation rates and ancestral polymorphism (Cruickshank & Hahn, 2014; Meier et al., 2018). Indeed, D_{xy} had high correlation to nucleotide diversity (*pi*) for each pairwise comparison (r^2 = 0.707, SBM vs OC; r^2 = 0.744, SCL vs SCU, Figure S 2.12), and also had relatively high correlation between both groups (r^2 = 0.393, Figure S 2.9B).

Levels of mean D_{xy} across the genome were also different between groups, and were almost twice as high in the SBM vs OC pairwise comparison (0.078) than in the SCL vs SCU comparison (0.045). This could be the result of a history of longer isolation between the first group. Within the Santa Clara River, although there is limited gene flow between upstream fish with fish from the lower reaches, there is molecular evidence of genes occasionally flowing downstream during high flood events (Buth et al., 1984; Richmond et al., 2015).

The number of shared windows between groups within the 90th percentile of D_{xy} was five times greater than for the F_{st} windows (438 vs 2336). Considering the long and persistent history of bottleneck of these unarmored populations (Turba et al., 2022), drift will play an important role in the creation of genome-wide regions of differentiation that has nothing to do with selection of adaptive (or maladaptive) alleles. Therefore, by looking at convergence of both metrics that are more or less affected by within-group variation, and comparing populations that are diverging in parallel with little gene flow, we can narrow down regions of interest that could potentially be related to the selection of our trait of interest.

Convergence between F_{st} and D_{xy}, identified 55 windows distributed across the genome. Of these 35 were in regions that also showed higher than average f_d (i.e., shared alleles between both unarmored populations). This effectively excludes regions found in chromosome 6, 10 and 18, but still retains windows in other chromosomes with known QTLs of large effect, such as the one found in chromosome 4 and related to the *Eda* locus (Colosimo et al., 2004). Chromosomes 2 and 7 also have regions associated with plate count numbers (Conte et al., 2015). This clustering of genes in chromosomes is a recognized process in threespine stickleback genomics (Peichel & Marques, 2017; Reid et al., 2021). Through linkage and recombination cold/hot spots, these allow for the quick assembly and reassembly of haploblocks associated with important traits for ecological adaptation to new systems.

While there is evidence for clustering, there are many other regions that are scattered throughout the genome as well (Figure 2.6). These could represent regions unrelated to the unarmored trait, and instead be diverging through linkage. On the other hand, they could be actually related to other phenotypes that are important and adaptive for these upstream populations (Baumgartner, 1986; Konijnendijk et al., 2015), but that we have not taken into consideration in this work. Nevertheless, we cannot rule out the possibility that these may represent polygenic regions of small-effect associated with the unarmored condition, as those are expected to occur at higher frequency than regions of large-effect (Reid et al., 2021). The resolution of the recombination map available at this time is suboptimal to determine where specifically these 55 windows fall in relation to recombination rates (there was no overall correlation to F_{st} and D_{xy}, Figure S 2.11), and future work on developing a recombination map for our genomes will be helpful in determining where these regions of high divergence fall in relation to recombination rates.

Although the power and reproducibility of these metrics in finding regions under selection is limited (Wolf & Ellegren, 2017), the ability to compare pairs of two independently evolving lineages of unarmored populations and find regions of convergence between them increase our confidence in our findings. Our validation test and ability to recover expected regions of divergence between fully and low armored phenotypes also adds to this confidence. These should, nevertheless, be tested in conjunction with other systematic approaches to find evidence for selection, which can include the development of demographic models for the null distributions of outlier regions genome-wide, and finding regions of selective sweeps using extended haplotype length approach (e.g., Marques et al., 2018; Meier et al., 2018).

Tables

Table 2.1: List of individual fish sequenced for this study and relevant metadata. The museum individual from the LA River did not have Lat/Long information. Map ID = map number in Figure 2.1. Regional groups based on previous work (Turba et al., 2022): Outer Coast (OC); Southern California Bight (SCB); Lower Santa Clara/Ventura (L); Upper Santa Clara (U); San Bernardino Mountains (SBM); Admixed (Ad).

Map ID	Pop Name	Pop Code	Area	Regional Group	Collection ID	Lat	Long	Ecotype	Morphotype	Sex
3	Bodega Bay	BDG	Sonoma	OC	JQR928	38.327	-123.055	Marine	Fully-plated	Female
3	Bodega Bay	BDG	Sonoma	OC	JQR929	38.327	-123.055	Marine	Fully-plated	Male
4	San Luis Obispo Creek	SLO	San Luis Obispo	OC	RNF8761	35.196	-120.697	Freshwater	Fully-plated	Female
4	San Luis Obispo Creek	SLO	San Luis Obispo	OC	RNF8762	35.196	-120.697	Freshwater	Fully-plated	Female
5	Arroyo Grande Lagoon	AGL	San Luis Obispo	OC	JQR835	35.099	-120.629	Freshwater	Low-plate	Male
5	Arroyo Grande Lagoon	AGL	San Luis Obispo	OC	JQR836	35.099	-120.629	Freshwater	Low-plate	Female
6	Shuman Lagoon	SHU	Santa Barbara	OC	JQR832	34.504	-120.354	Freshwater	Low-plate	Female
6	Shuman Lagoon	SHU	Santa Barbara	OC	JQR833	34.504	-120.354	Freshwater	Low-plate	Female
7	San Antonio Creek	SAC	Santa Barbara	OC	JQR893	34.783	-120.53	Freshwater	Unarmored	Female
7	San Antonio Creek	SAC	Santa Barbara	OC	RNF8218	34.785	-120.481	Freshwater	Unarmored	Male
7	San Antonio Creek	SAC	Santa Barbara	OC	RNF8220	34.785	-120.481	Freshwater	Unarmored	Male

Map ID	Pop Name	Pop Code	Area	Regional Group	Collection ID	Lat	Long	Ecotype	Morphotype	Sex
8	Salsispuede	SSP	Santa Barbara	OC	JQR921	34.621	-120.423	Freshwater	Low-plate	Male
8	Salsispuede	SSP	Santa Barbara	OC	JQR922	34.621	-120.423	Freshwater	Low-plate	Male
9	Mission	MIS	Santa Barbara	OC-Ad	RNF8785	34.414	-119.685	Freshwater	Low-plate	Female
9	Mission	MIS	Santa Barbara	OC-Ad	RNF8792	34.414	-119.685	Freshwater	Low-plate	Female
10	Seaside	SEA	Ventura	SCB-L	RNF8797	34.282	-119.309	Freshwater	Low-plate	Male
10	Seaside	SEA	Ventura	SCB-L	RNF8798	34.282	-119.309	Freshwater	Low-plate	Male
11	Foster Park	FOP	Ventura	SCB-L	RNF8780	34.352	-119.307	Freshwater	Low-plate	Female
11	Foster Park	FOP	Ventura	SCB-L	RNF8781	34.352	-119.307	Freshwater	Low-plate	Female
12	San Antonio	SAV	Ventura	SCB-L	JQR808	34.433	-119.251	Freshwater	Low-plate	Female
12	San Antonio	SAV	Ventura	SCB-L	JQR809	34.433	-119.251	Freshwater	Low-plate	Female
12	San Antonio	SAV	Ventura	SCB-L	JQR811	34.433	-119.251	Freshwater	Low-plate	Female
13	Matilija	MTJ	Ventura	SCB-L	JQR812	34.501	-119.251	Freshwater	Low-plate	Female
13	Matilija	MTJ	Ventura	SCB-L	JQR813	34.501	-119.251	Freshwater	Low-plate	Male
14	McGrath	MCG	Santa Clara	SCB-L	JQR869	34.230	-119.261	Freshwater	Low-plate	Female
14	McGrath	MCG	Santa Clara	SCB-L	JQR870	34.230	-119.261	Freshwater	Low-plate	Female
15	Oxnard	OXN	Santa Clara	SCB-L	RNF8696	34.241	-119.192	Freshwater	Low-plate	Male
15	Oxnard	OXN	Santa Clara	SCB-L	RNF8697	34.241	-119.192	Freshwater	Low-plate	Female
16	Sespe Lower	SSL	Santa Clara	SCB-L	JQR1039	34.407	-118.932	Freshwater	Low-plate	Male
16	Sespe Lower	SSL	Santa Clara	SCB-L	JQR1040	34.407	-118.932	Freshwater	Low-plate	Female

Map ID	Pop Name	Pop Code	Area	Regional Group	Collection ID	Lat	Long	Ecotype	Morphotype	Sex
17	Sespe Upper	SSU	Santa Clara	SCB-L	RNF8728	34.558	-119.253	Freshwater	Low-plate	Male
17	Sespe Upper	SSU	Santa Clara	SCB-L	RNF8729	34.558	-119.253	Freshwater	Low-plate	Male
18	Piru	PIR	Santa Clara	SCB-L	RNF7987	34.418	-118.79	Freshwater	Low-plate	Male
18	Piru	PIR	Santa Clara	SCB-L	RNF7991	34.418	-118.79	Freshwater	Low-plate	Female
19	Newhall Ranch	NHR	Santa Clara	SCB-U	RNF7905	34.427	-118.577	Freshwater	Low-plate	Female
19	Newhall Ranch	NHR	Santa Clara	SCB-U	RNF7906	34.427	-118.577	Freshwater	Low-plate	Female
19	Newhall Ranch	NHR	Santa Clara	SCB-U	RNF7907	34.427	-118.577	Freshwater	Low-plate	Female
20	Valencia	VAL	Santa Clara	SCB-U	JQR1087	34.435	-118.603	Freshwater	Low-plate	Female
20	Valencia	VAL	Santa Clara	SCB-U	JQR1088	34.435	-118.603	Freshwater	Low-plate	Male
20	Valencia	VAL	Santa Clara	SCB-U	JQR1095	34.435	-118.603	Freshwater	Low-plate	Male
20	Valencia	VAL	Santa Clara	SCB-U	JQR1096	34.435	-118.603	Freshwater	Low-plate	Female
20	Valencia	VAL	Santa Clara	SCB-U	JQR1097	34.435	-118.603	Freshwater	Low-plate	Female
20	Valencia	VAL	Santa Clara	SCB-U	JQR1098	34.435	-118.603	Freshwater	Low-plate	Female
21	San Francisquito	SFQ	Santa Clara	SCB-U	RNF8333	34.546	-118.516	Freshwater	Low-plate	Male
21	San Francisquito	SFQ	Santa Clara	SCB-U	RNF8334	34.546	-118.516	Freshwater	Low-plate	Female
22	Bouquet Lower	BOL	Santa Clara	SCB-L-Ad	JQR1076	34.510	-118.451	Freshwater	Low-plate	Female
22	Bouquet Lower	BOL	Santa Clara	SCB-L-Ad	JQR1077	34.510	-118.451	Freshwater	Low-plate	Female

Map ID	Pop Name	Pop Code	Area	Regional Group	Collection ID	Lat	Long	Ecotype	Morphotype	Sex
22	Bouquet Lower	BOL	Santa Clara	SCB-L-Ad	JQR1078	34.510	-118.451	Freshwater	Low-plate	Female
23	Bouquet Upper	BOU	Santa Clara	SCB-L-Ad	RNF8679	34.554	-118.416	Freshwater	Low-plate	Female
23	Bouquet Upper	BOU	Santa Clara	SCB-L-Ad	RNF8680	34.554	-118.416	Freshwater	Low-plate	Male
23	Bouquet Upper	BOU	Santa Clara	SCB-L-Ad	RNF8681	34.554	-118.416	Freshwater	Low-plate	Female
24	Robins Nest	RBN	Santa Clara	SCB-U	RNF8118	34.439	-118.317	Freshwater	Unarmored	Male
24	Robins Nest	RBN	Santa Clara	SCB-U	RNF8119	34.439	-118.317	Freshwater	Unarmored	Female
25	Thousand Trails	THT	Santa Clara	SCB-U	RNF8223	34.442	-118.211	Freshwater	Unarmored	Female
25	Thousand Trails	THT	Santa Clara	SCB-U	RNF8224	34.442	-118.211	Freshwater	Unarmored	Female
26	Valyermo	VLM	Mojave Desert	SCB-U-Ad	RNF8347	34.452	-117.856	Freshwater	Low-plate	Male
26	Valyermo	VLM	Mojave Desert	SCB-U-Ad	RNF8348	34.452	-117.856	Freshwater	Low-plate	Male
26	Valyermo	VLM	Mojave Desert	SCB-U-Ad	RNF8351	34.452	-117.856	Freshwater	Low-plate	Male
27	Los Angeles River	LAR	Los Angeles	SBM-Ad	NHMLA982_1	N/A	N/A	Freshwater	Unarmored	Male
28	Apple Valley	APV	San Bernardino	SBM-Ad	RNF8745	34.528	-117.282	Freshwater	Low-plate	Female
28	Apple Valley	APV	San Bernardino	SBM-Ad	RNF8746	34.528	-117.282	Freshwater	Low-plate	Female
29	Sugarloaf	SLM	San Bernardino	SBM	RNF2867 #1	34.178	-116.826	Freshwater	Unarmored	Male

Map ID	Pop Name	Pop Code	Area	Regional Group	Collection ID	Lat	Long	Ecotype	Morphotype	Sex
	Meadow									
29	Sugarloaf Meadow	SLM	San Bernardino	SBM	RNF2867 #2	34.178	-116.826	Freshwater	Unarmored	Male
30	Shay Creek	SHC	San Bernardino	SBM	JQR1226	34.254	-116.808	Freshwater	Unarmored	Male
30	Shay Creek	SHC	San Bernardino	SBM	JQR1227	34.254	-116.808	Freshwater	Unarmored	Male
31	Juniper Creek	JUC	San Bernardino	SBM	JQR1312	34.220	-116.719	Freshwater	Unarmored	Female
31	Juniper Creek	JUC	San Bernardino	SBM	JQR1313	34.220	-116.719	Freshwater	Unarmored	Male
32	Trabuco Canyon	TRC	Orange	OC	RNF8753	33.563	-117.651	Freshwater	Low-plate	Female
32	Trabuco Canyon	TRC	Orange	OC	RNF8754	33.563	-117.651	Freshwater	Low-plate	Male
33	Bell Canyon	BEC	Orange	OC	JQR843	33.629	-117.555	Freshwater	Low-plate	Male
33	Bell Canyon	BEC	Orange	OC	JQR844	33.629	-117.555	Freshwater	Low-plate	Female
34	San Jacinto	SJT	Riverside	SCB-U-Ad	RNF8162	33.739	-116.835	Freshwater	Low-plate	Male
34	San Jacinto	SJT	Riverside	SCB-U-Ad	RNF8163	33.739	-116.835	Freshwater	Low-plate	Female
35	San Felipe Creek	SFC	San Diego	SCB-U	JQR850	33.098	-116.473	Freshwater	Unarmored	Male
35	San Felipe Creek	SFC	San Diego	SCB-U	JQR851	33.098	-116.473	Freshwater	Unarmored	Female
35	San Felipe Creek	SFC	San Diego	SCB-U	JQR852	33.098	-116.473	Freshwater	Unarmored	Male
36	Pine Valley	PIV	San Diego	SCB-U	RNF8738	32.836	-116.543	Freshwater	Unarmored	Female
36	Pine Valley	PIV	San Diego	SCB-U	RNF8739	32.836	-116.543	Freshwater	Unarmored	Female

Map ID	Pop Name	Pop Code	Area	Regional Group	Collection ID	Lat	Long	Ecotype	Morphotype	Sex
36	Pine Valley	PIV	San Diego	SCB-U	RNF8741	32.836	-116.543	Freshwater	Unarmored	Male
37	Sweetwater	SWE	San Diego	SCB-U	RNF3618 #2	32.896	-116.595	Freshwater	Low-plate	Male
37	Sweetwater	SWE	San Diego	SCB-U	RNF3618 #3	32.896	-116.595	Freshwater	Low-plate	Male
38	El Rosario	BRO	Baja California	OC	RNF8770	30.041	-115.788	Freshwater	Low-plate	Female
38	El Rosario	BRO	Baja California	OC	RNF8771	30.041	-115.788	Freshwater	Low-plate	Female

Table 2.2: f_4-statistics with outgroup (Pop A) as JAP and Pop B as LAR. Pop C includes populations from SBM and Apple Valley, and Pop D, representatives of the OC group (except Bocana El Rosario). Based on the z-score (ordered from lowest to highest), gene flow occurred between Pop B and D if positive, or Pop B and C if negative. The LA River individual shows strong evidence of shared alleles with populations from SBM compared to the OC group.

Pop A	Pop B	Pop C	Pop D	est	se	z	p-value
JAP	LAR	JUC	SAC	-0.00696	0.000288	-24.1312	1.2E-128
JAP	LAR	JUC	AGL	-0.00786	0.000331	-23.7565	9.4E-125
JAP	LAR	SHC	SAC	-0.00703	0.000308	-22.8153	3.2E-115
JAP	LAR	SHC	AGL	-0.00793	0.000349	-22.7461	1.6E-114
JAP	LAR	SLM	AGL	-0.00904	0.00041	-22.061	7.5E-108
JAP	LAR	SLM	SAC	-0.00814	0.000369	-22.0258	1.6E-107
JAP	LAR	JUC	TRC	-0.00562	0.000276	-20.3339	6.5E-92
JAP	LAR	JUC	BEC	-0.00559	0.000293	-19.0567	5.8E-81
JAP	LAR	SLM	TRC	-0.0068	0.000358	-19.0058	1.5E-80
JAP	LAR	SHC	TRC	-0.00569	0.000301	-18.9082	9.8E-80
JAP	LAR	SLM	BEC	-0.00677	0.000374	-18.1328	1.8E-73
JAP	LAR	SHC	BEC	-0.00567	0.000316	-17.9515	4.7E-72
JAP	LAR	JUC	SLO	-0.00797	0.000486	-16.3908	2.2E-60
JAP	LAR	SLM	SLO	-0.00915	0.000564	-16.2219	3.5E-59
JAP	LAR	SHC	SLO	-0.00804	0.000523	-15.3731	2.5E-53
JAP	LAR	APV	SLO	-0.00305	0.000209	-14.5998	2.8E-48
JAP	LAR	APV	AGL	-0.00294	0.000435	-6.75939	1.4E-11
JAP	LAR	APV	SAC	-0.00204	0.000431	-4.74253	2.1E-06
JAP	LAR	APV	TRC	-0.0007	0.00041	-1.71708	8.6E-02
JAP	LAR	APV	BEC	-0.00068	0.00042	-1.61875	1.1E-01

Table 2.3: f_4-statistics with outgroup (Pop A) as JAP and Pop B as SBM populations or Apple Valley. Pop C includes populations from San Juan Creek (Trabuco and Bell Canyons), and Pop D, representatives of the northernmost OC group (except Bodega Bay). Based on the z-score (ordered from lowest to highest), gene flow occurred between Pop B and D if positive, or Pop B and C if negative. There is strong evidence of shared alleles between SBM and Apple Valley populations and those at San Juan Creek (part of the southern OC group), when compared to the northernmost OC populations.

Pop A	Pop B	Pop C	Pop D	est	se	z	p-value
JAP	SHC	TRC	SAC	-0.00204	0.000154	-13.3124	1.96E-40
JAP	JUC	TRC	SAC	-0.00197	0.000148	-13.2721	3.36E-40
JAP	JUC	TRC	AGL	-0.00228	0.000174	-13.1233	2.42E-39
JAP	SHC	TRC	AGL	-0.00235	0.000181	-13.0174	9.75E-39
JAP	SLM	TRC	AGL	-0.00252	0.000195	-12.9425	2.59E-38
JAP	SLM	TRC	SAC	-0.00211	0.000173	-12.1953	3.29E-34
JAP	JUC	BEC	AGL	-0.00219	0.000179	-12.195	3.31E-34
JAP	SHC	BEC	AGL	-0.00226	0.000187	-12.1063	9.79E-34
JAP	SLM	BEC	AGL	-0.00237	0.000198	-11.9495	6.53E-33
JAP	SHC	BEC	SAC	-0.00195	0.00017	-11.4298	2.97E-30
JAP	JUC	BEC	SAC	-0.00187	0.000167	-11.2404	2.58E-29
JAP	JUC	TRC	SLO	-0.00244	0.000218	-11.183	4.94E-29
JAP	SLM	TRC	SLO	-0.00271	0.000245	-11.0844	1.49E-28
JAP	SHC	TRC	SLO	-0.00258	0.000236	-10.9532	6.42E-28
JAP	SLM	BEC	SAC	-0.00196	0.000187	-10.4495	1.47E-25
JAP	JUC	BEC	SLO	-0.00235	0.000228	-10.2985	7.16E-25
JAP	SHC	BEC	SLO	-0.00248	0.000246	-10.1063	5.18E-24
JAP	SLM	BEC	SLO	-0.00256	0.000254	-10.0843	6.48E-24
JAP	APV	TRC	SLO	-0.00134	0.000143	-9.37091	7.19E-21
JAP	APV	BEC	SLO	-0.00134	0.000149	-8.9989	2.28E-19
JAP	APV	TRC	SAC	-0.00099	0.000122	-8.10706	5.19E-16
JAP	APV	TRC	AGL	-0.00121	0.000162	-7.43071	1.08E-13
JAP	APV	BEC	SAC	-0.00099	0.00014	-7.06081	1.66E-12
JAP	APV	BEC	AGL	-0.00121	0.000173	-6.97009	3.17E-12

Table 2.4: Summary statistics of windows for each pairwise comparison. Mean is the average of F_{st} and D_{xy} for the whole-genome windows and windows within the 90th percentile of value distribution. Top F_{st} vs Top D_{xy} refers to windows within the 90th percentile for each metric that overlaps (55 windows).

	SBM vs OC		**SCL vs SCU**		**Shared**
	Mean	Number of windows	Mean	Number of windows	Number of windows
F_{st} (whole-genome)	0.21		0.22		
F_{st} (90th percentile)	0.44	4386	0.56	4391	438
D_{xy} (whole-genome)	0.08		0.04		
D_{xy} (90th percentile)	0.16	4384	0.12	4392	2336
Top F_{st} vs Top D_{xy}		1164		701	55

Table 2.5: Average F_{st} and D_{xy} values of the 55 overlapping windows.

Shared 55 windows	**SBM vs OC**	**SCL vs SCU**
F_{st}	0.47	0.57
D_{xy}	0.17	0.14

Figures

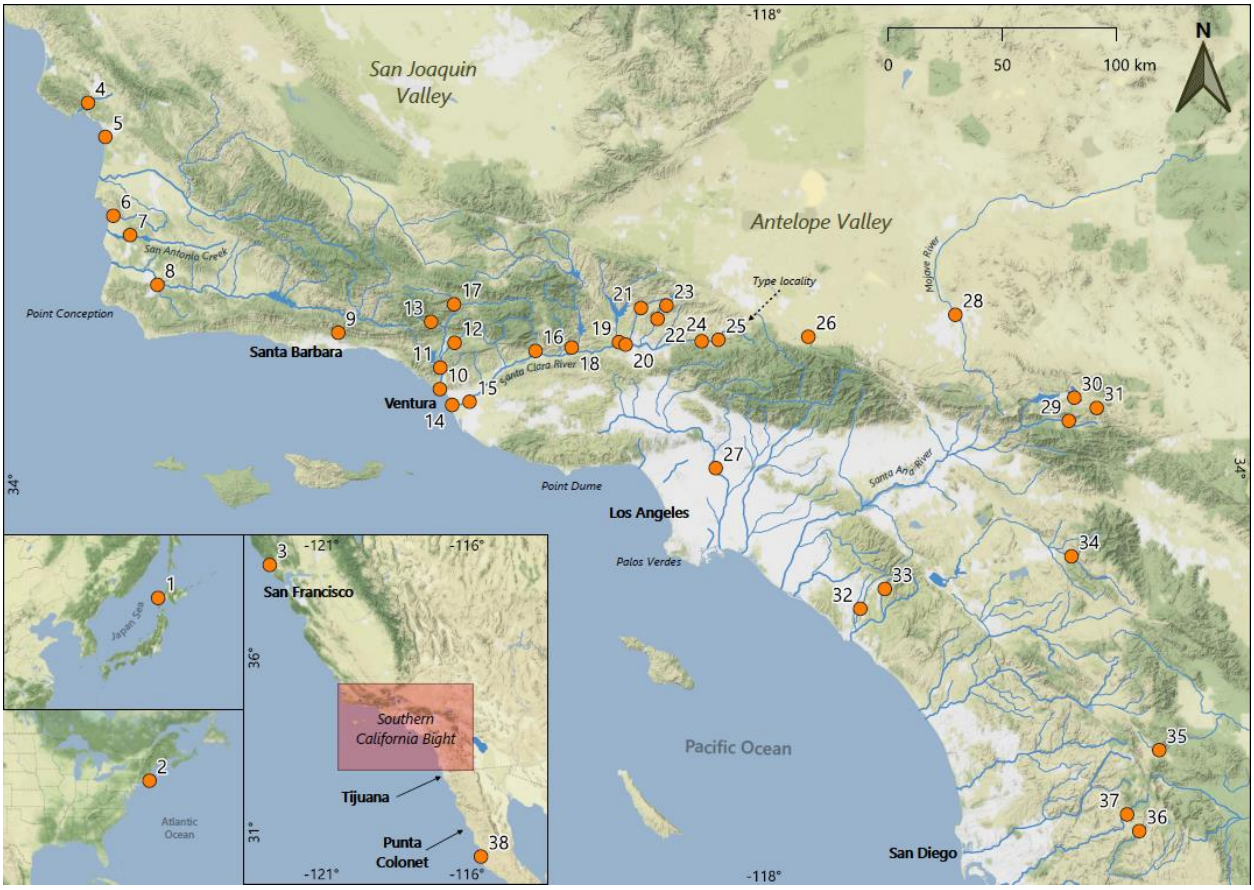


Figure 2.1: Map of sampling localities as listed on Table 2.1. Insets on bottom left depicts locality for samples outside the Southern California Bight (SCB) region (red square). 1: *Gasterosteus nipponicus* (Higuchi et al., 2014); 2: *G. wheatlandii* (Mattern, 2004). Lat/long information for sites 1 and 2 were estimated based on original publications. This map was made on QGIS 3.16.2 and edited on Adobe Illustrator. River dataset from the National Weather Service (NOAA).

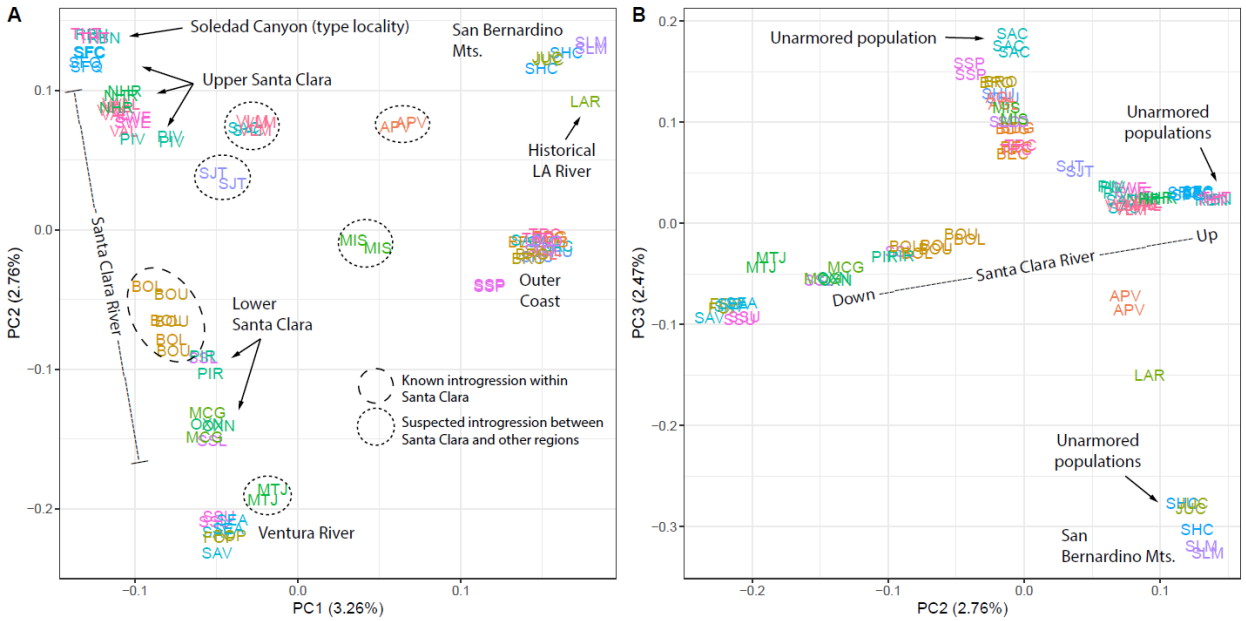


Figure 2.2: PCA plots of dataset 2 (modern samples plus LA River museum fish) containing 83,570 biallelic SNPs. A) PC1 vs PC2; B) PC2 vs PC3. The analysis largely recapitulates previous groupings (Turba et al., 2022): Outer Coast (OC) and Southern California Bight (SCB). Additionally, we see the segregation of the San Bernardino Mountain (SBM) populations from the OC group, and a gradient from the Ventura and lower Santa Clara Rivers towards sites at the upper Santa Clara. Unarmored populations are clustered independently within their separate groups. Two ambiguous San Antonio Creek (SAC) individuals clustering with the Valyermo samples were removed from downstream admixture analysis (TreeMix and Admixtools).

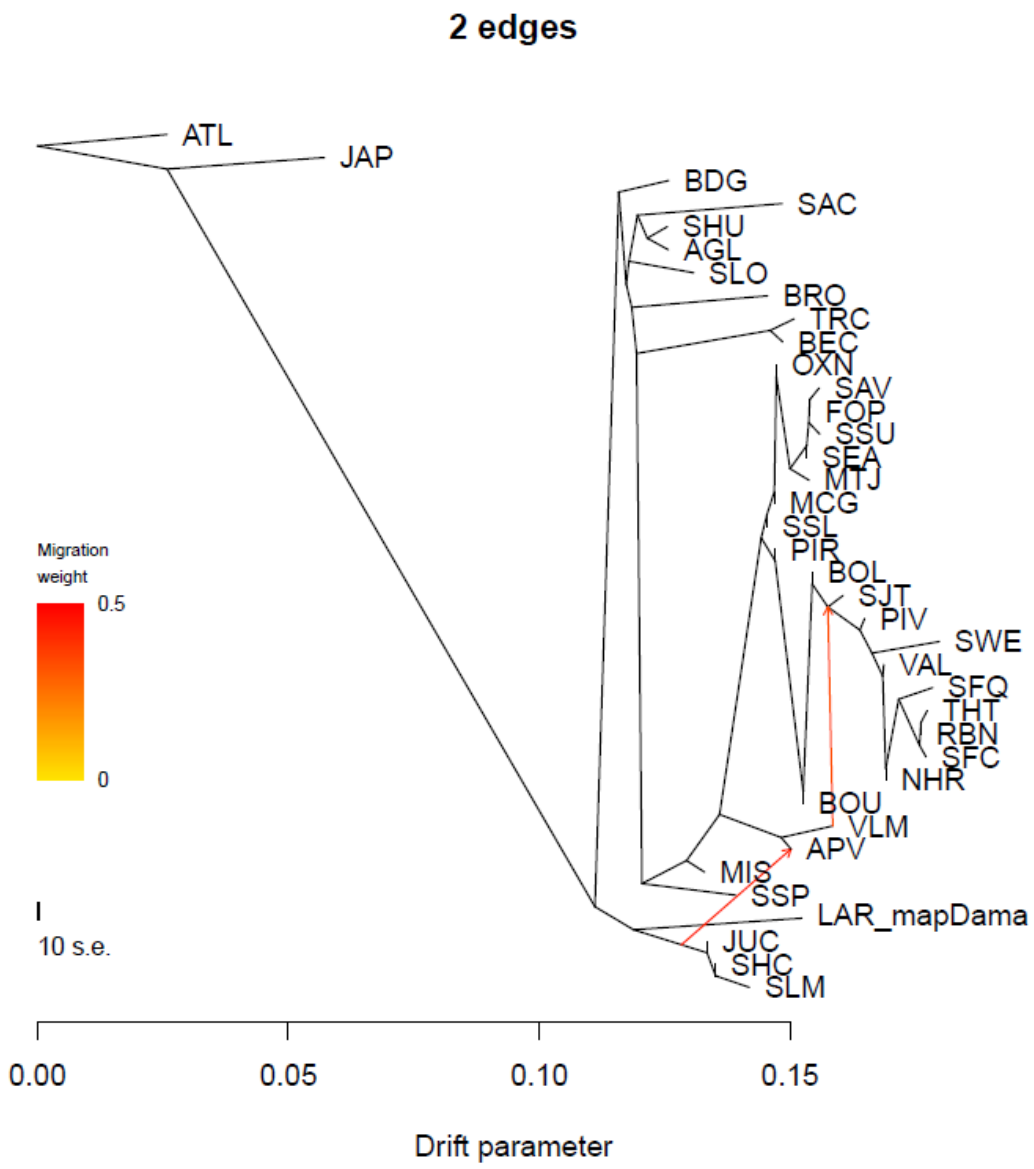


Figure 2.3: Maximum-likelihood TreeMix tree with 2 migration edges (as defined by the Evanno method). Horizontal branches are proportional to the amount of genetic drift. Migration arrows are colored according to weight (see legend for scale). Admixture is shown from the SBM population branch to Apple Valley; and from Valyermo to San Jacinto population. Residual fit is shown in Figure S 2.6.

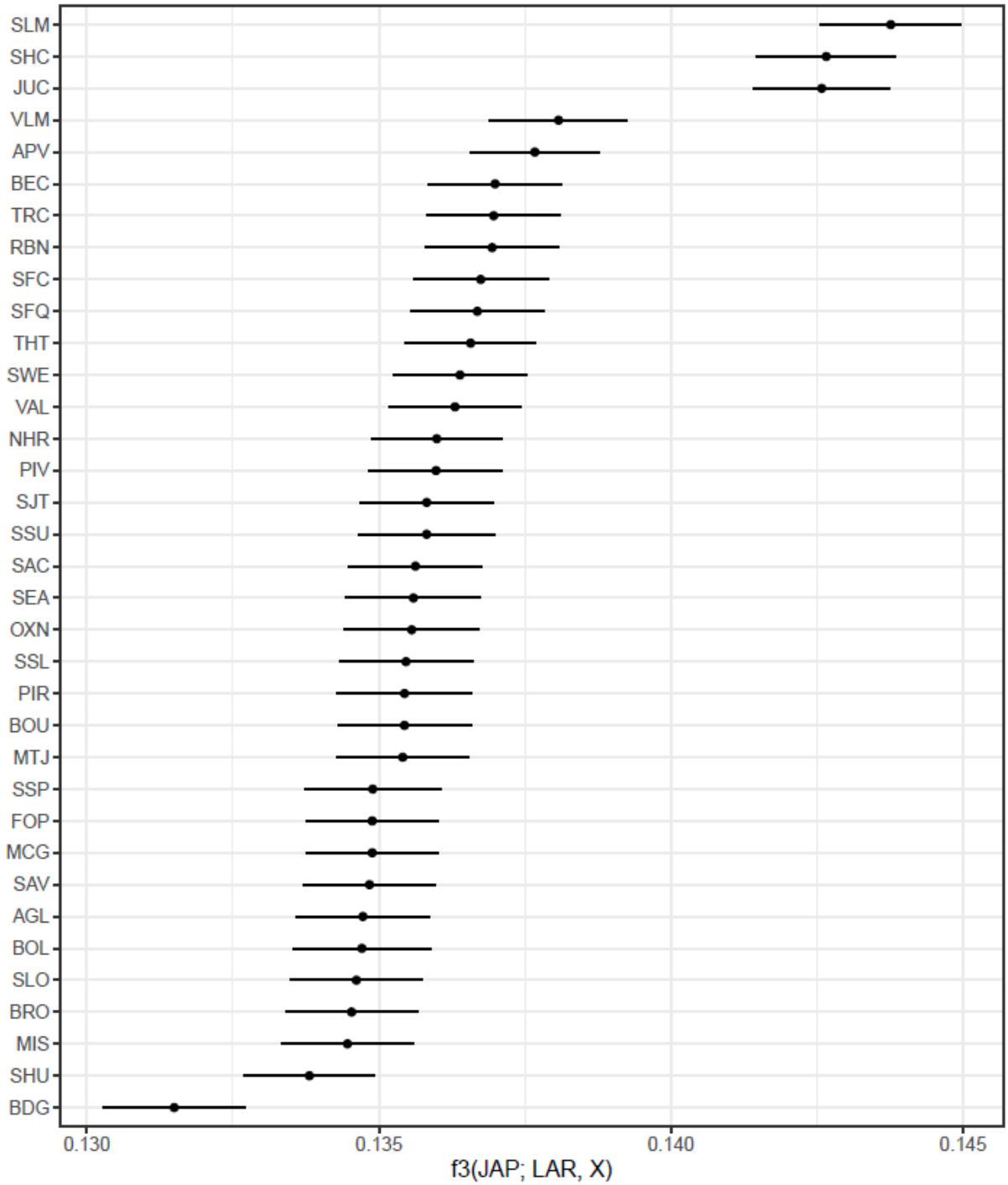


Figure 2.4: Outgroup f_3-statistics comparing the LA River individual to other populations (JAP; LAR, X). Populations from the SBM are the closest to the LA River individual.

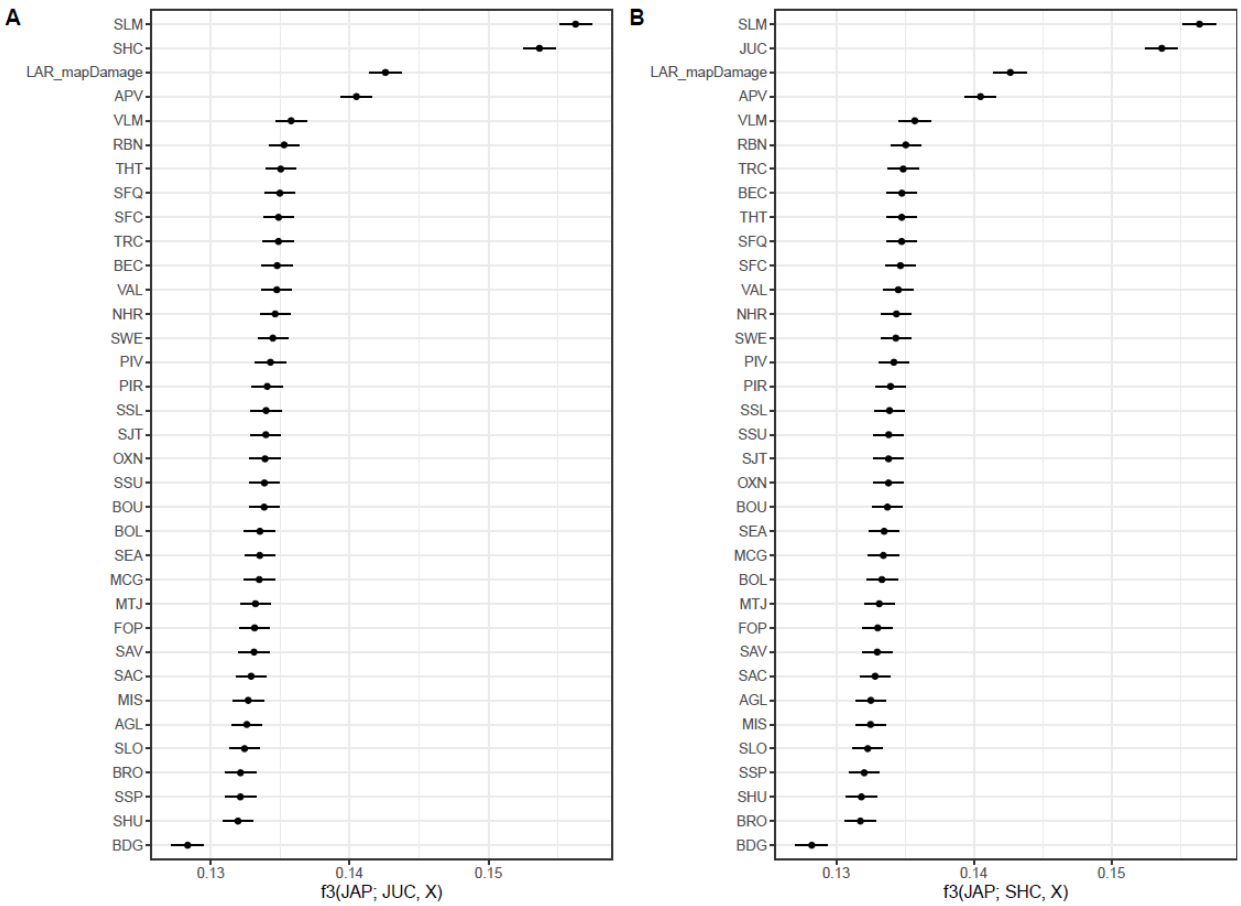


Figure 2.5: Outgroup f_3-statistics comparing the SBM populations to other populations (A) JUC: Juniper Creek; (B) SHC: Shay Creek (JAP; SHC/JUC, X). Populations from the SBM are more closely related to individuals from the LA River and Apple Valley than to the remainder of the populations.

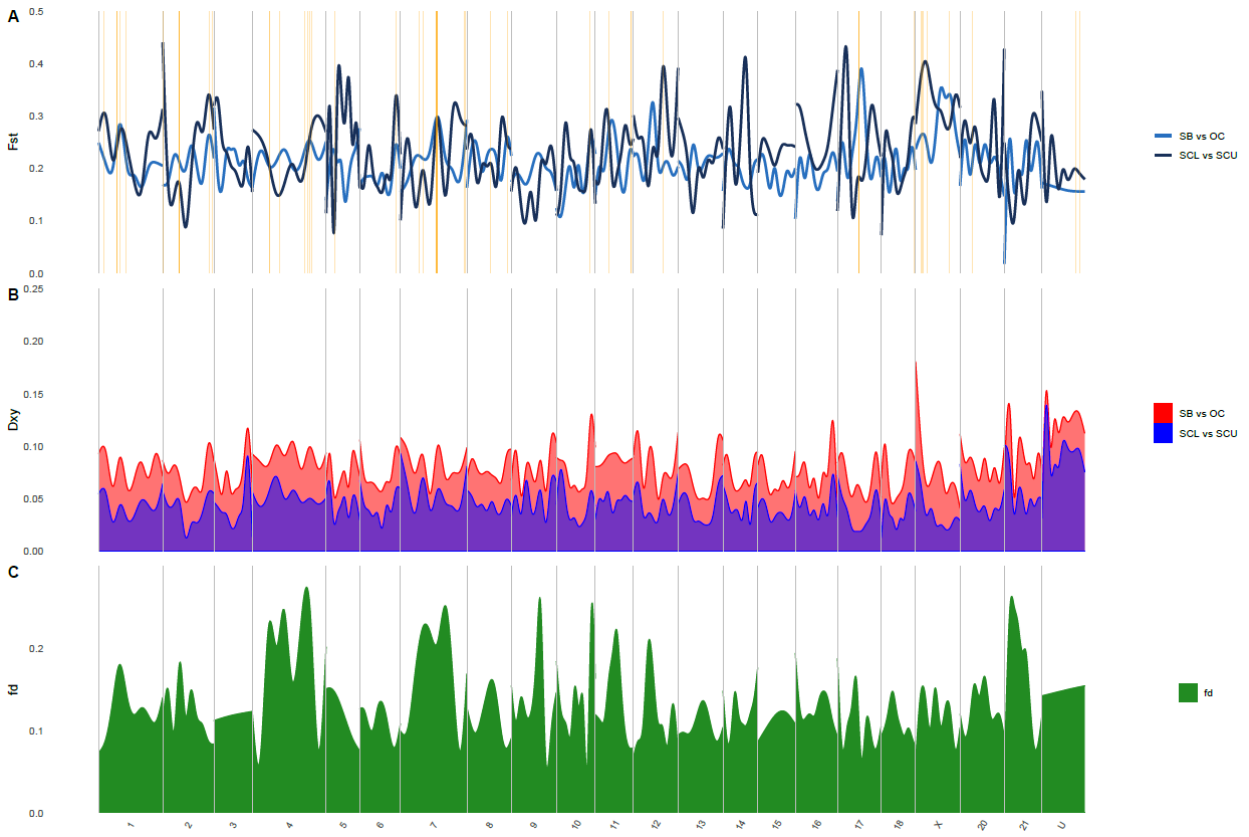


Figure 2.6: Genome-wide sliding windows approach of (A) relative (F_{st}) and (B) absolute (D_{xy}) divergence. (C) Windows of shared alleles (f_d) between unarmored populations from Santa Clara River and San Bernardino Mountains. Yellow bars on top plot (A) denote the 55 windows where the 90th percentile of F_{st} and D_{xy} values overlap for both pairwise comparisons (Table S 2.5).

VCF to plink

The VCF file was edited to replace chromosome code from roman numerals to arabic, then transformed into plink (v.1.9) binary fileset (following instructions from https://apol1.blogspot.com/2014/11/best-practice-for-converting-vcf-files.html) to contain only bi-allelic SNPs and unambiguous INDELs. Then, the FIDs and sex information were updated in the .fam file (--update-ids and –update-sex, respectively). For the fastSTRUCTURE and TreeMix analysis, we also pruned the files for linkage disequilibrium (--indep-pairwise <window size>['50 kb'] <step size (10 bp)> <r^2 threshold (0.1)>).

FastSTRUCTURE

The plink file was used as input for the analysis of admixture proportions using fastSTRUCTURE (Raj et al., 2014). Inference was calculated using the simple prior for a range of *K*, from 1-10. Model complexity was calculated for each prior and then plotted. The appropriate K number that maximizes marginal likelihood using the simple prior was $K = 2$, but $K = 3$ is best at explaining structure in data (Figure S 2.2).

At $K = 2$ patterns correspond to two general clusters. The first, involving fish from Bodega Bay south to Salsipuedes Creek, the LAR, populations in the SBM, San Juan Creek (Trabuco Creek and Bell Canyon tributaries), and Bocana El Rosario in Baja California. The second, consisting of populations within the SCB and includes individuals from the Ventura, Santa Clara, and San Jacinto Rivers, Valyermo, and three drainages in San Diego County (San Felipe Creek, Pine Valley Creek, and the Sweetwater River). Apple Valley and Mission Creek show about even levels of admixture between these two clusters. Three individuals from San Antonio Creek show

allele frequency patterns more consistent with the first group, while two individuals show a pattern more similar to the second, mirroring results from the PCA.

At K = 3 a distinct admixture pattern emerges, with shared allele frequencies occurring between individuals from the SBM, Apple Valley and the museum individual from the LA River. This individual shows equal admixture proportions between SBM and OC populations.

Scripts and code availability

1. Pairwise Fst analysis (Gosselin, 2020): https://thierrygosselin.github.io/assigner/articles/web_only/fst_comparisons.html#snprelate
2. Paleogenomics pipeline for museum sample (Ottoni, 2021): https://physalia-paleogenomics-2020.readthedocs.io/_/downloads/en/latest/pdf/
3. Sliding window approach (Martin, n.a.): https://github.com/simonhmartin/genomics_general
4. Scripts for all analysis can be found on the author's (RT) GitLab: https://gitlab.com/rturba/uts-genomics

Supplementary Tables

(See attached supplemental material)

Table S 2.1: Summary quality report of BAM files for each individual generated using Qualimap.

Table S 2.2: Summary statistics of the filtered VCF file generated using bcftools (stats command).

Table S 2.3: Pairwise weighted F_{st} on dataset 2 (modern samples with museum individual) containing only biallelic SNPs and no missing data.

Table S 2.4: Sliding window summary statistics of 10kb windows for both pairwise comparisons (OC vs. SBM; SCL vs. SCU). Mid = base-pair position at the middle of the window for each chromosome. Mid2 = same as previous but in a continuous count (not separated by chromosome). Start = base-pair position of start of window. End = base-pair position of end of window.

Table S 2.5: Detailed statistics of the 55 windows within the 90^{th} percentile of F_{st} and D_{xy}, including results from the ABBA-BABA statistics.

Supplementary Figures

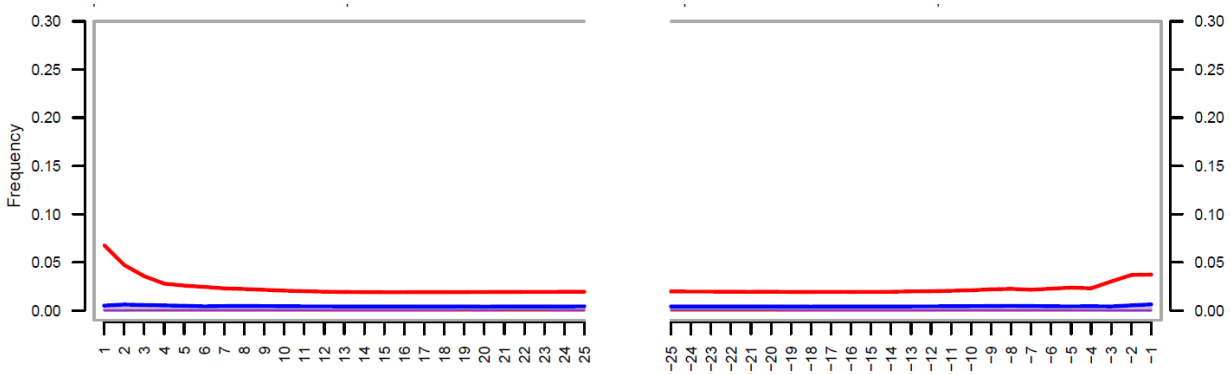


Figure S 2.1: MapDamage profile of the museum Los Angeles River (LAR) individual. The pattern of cytosine deamination (C>T) at the end of reads provides a valuable authentication tool that reads were from a degraded sample and not the result of contamination.

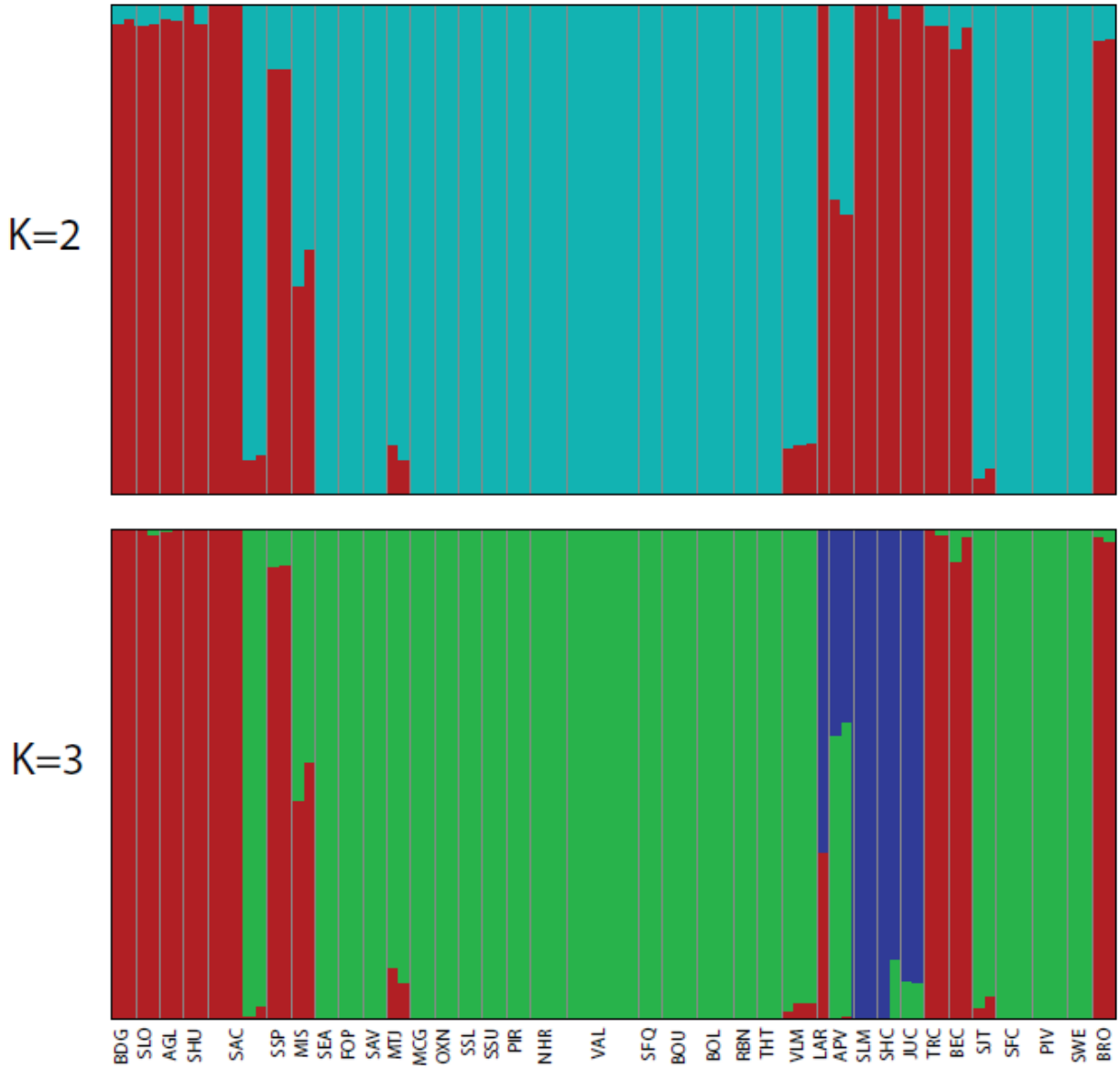


Figure S 2.2: fastSTRUCTURE assignment plots using simple prior. $K = 2$ maximizes marginal likelihood and $K = 3$ best explains structure in the data. These results largely replicate findings in previous work (Turba et al., 2022), with two major groups assigned to Outer Coast (OC) and Southern California Bight (SCB) at $K = 2$. At $K = 3$, a separate assignment arises between individuals from the San Bernardino Mountains (SBM), Apple Valley and the museum individual from Los Angeles River (LAR). Apple Valley individuals show admixture with SCB, and LAR with OC populations. The two ambiguous San Antonio Creek (SAC) individuals that also showed associations with the Valyermo samples in the PCA (Figure 2.2) show similar admixture proportions here, and therefore were removed from downstream admixture analysis (TreeMix and Admixtools).

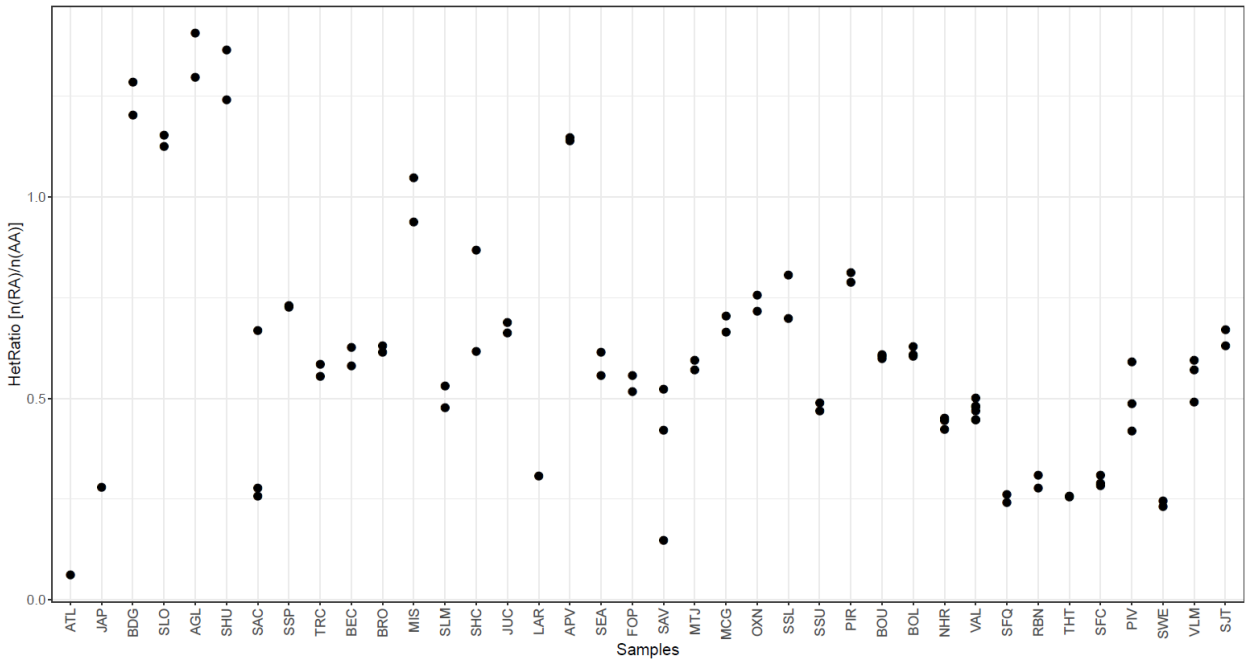


Figure S 2.3: Plot of heterozygosity ratios for each population (values can be found in Table S 2.2). Dots represent individuals. Ordering follows the regional groupings from Table 2.1. Northernmost OC populations show the highest levels of heterozygosity. Most of the remaining populations fall below this threshold, with unarmored individuals from the upper Santa Clara River falling consistently in the lowest range.

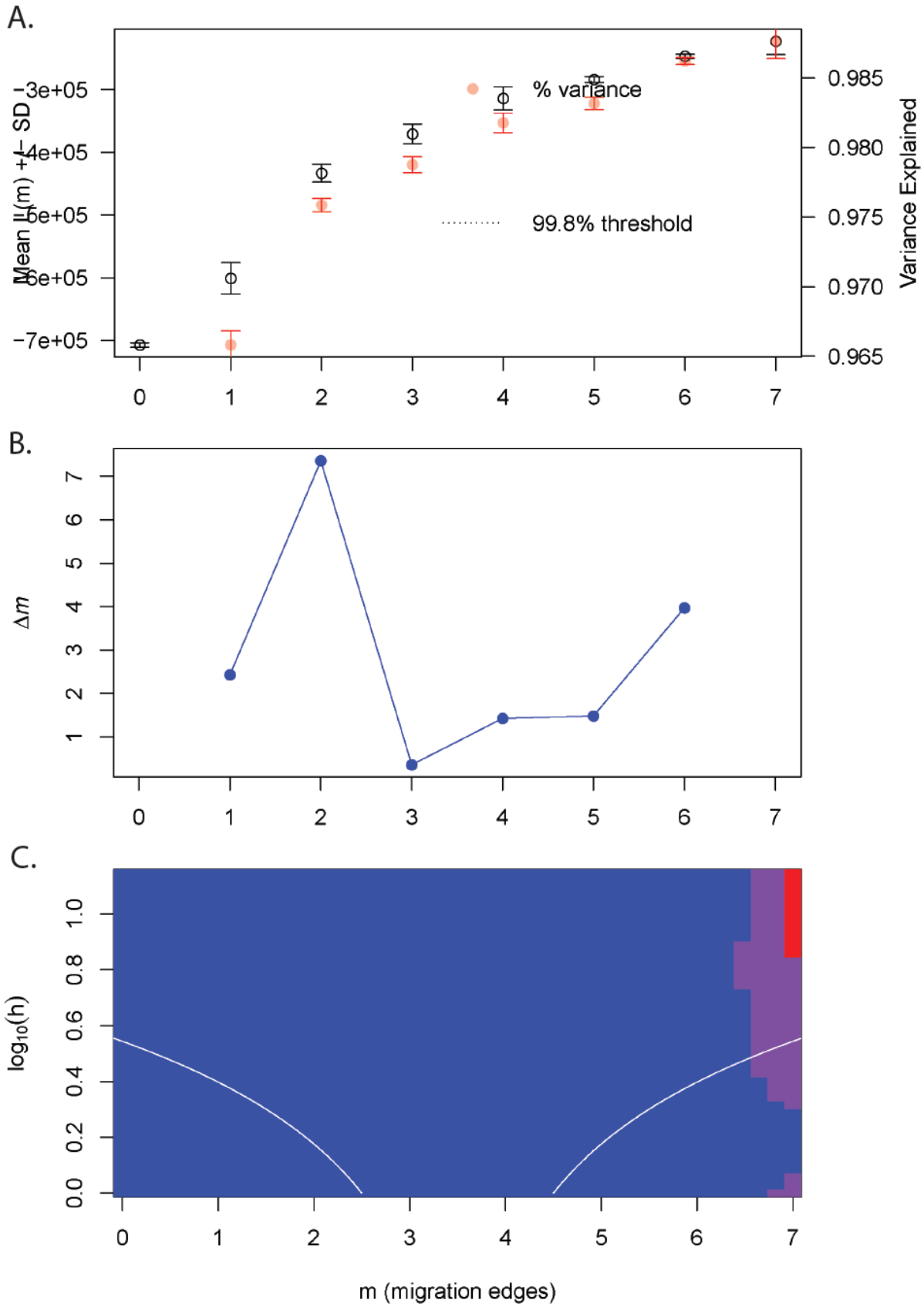

A.
Mean L(m) +/- SD
−3e+05
−4e+05
−5e+05
−6e+05
−7e+05
% variance
99.8% threshold
0.985
0.980
0.975
0.970
0.965
Variance Explained
0
1
2
3
4
5
6
7
B.
Δm
1
2
3
4
5
6
7
C.
log10(h)
0.0
0.2
0.4
0.6
0.8
1.0
m (migration edges)

Figure S 2.4: Calculated optimal number of migration edges of TreeMix using the (A) Evanno and (B) SiZer methods. The best number of migration edges based on deltaM was 2. Based on the SiZer map, derivatives of the smoothed curve reach a plateau around m = 6.

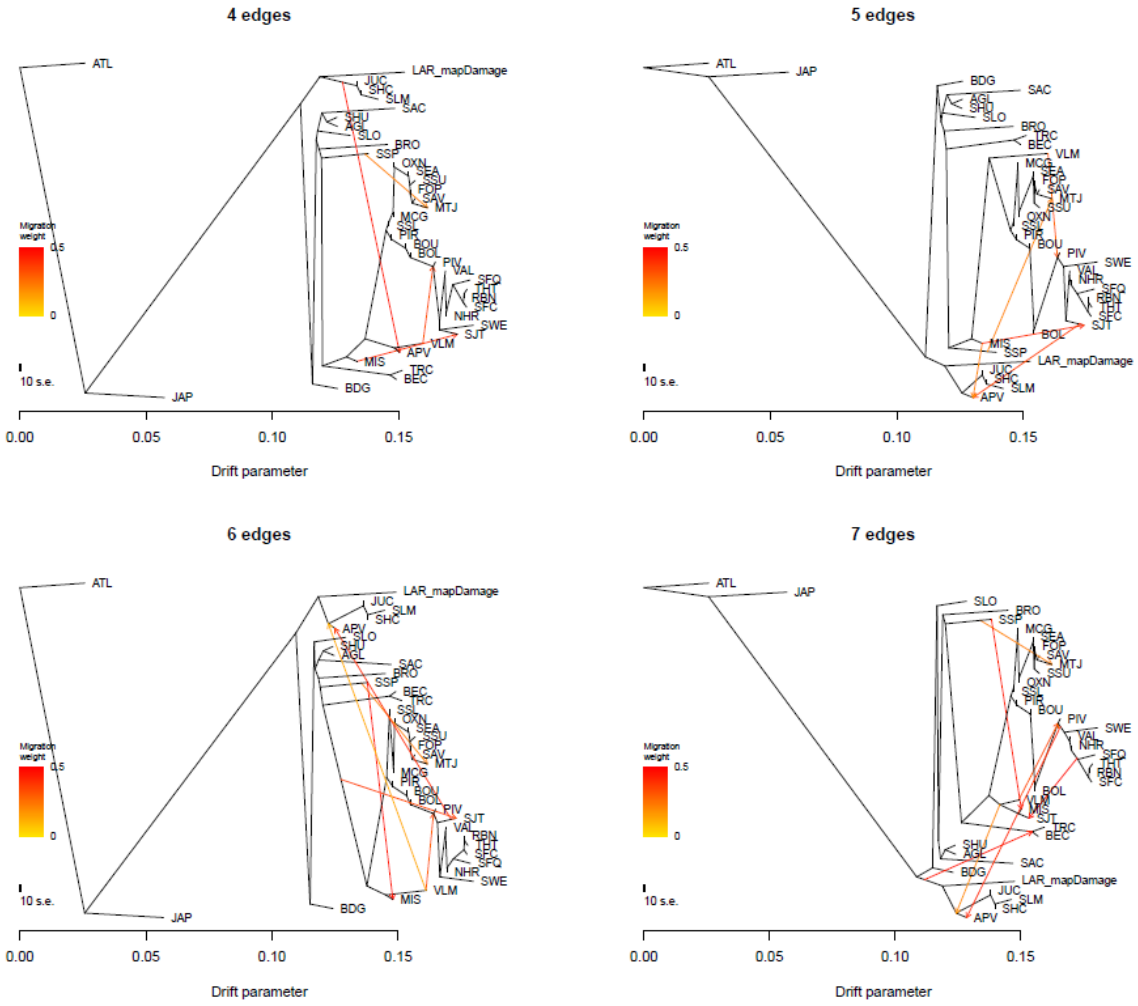


Figure S 2.5: Maximum likelihood TreeMix results for trees with 0-7 number of edges. Horizontal branches are proportional to the amount of genetic drift. Migration arrows are colored according to weight (see legend for scale).

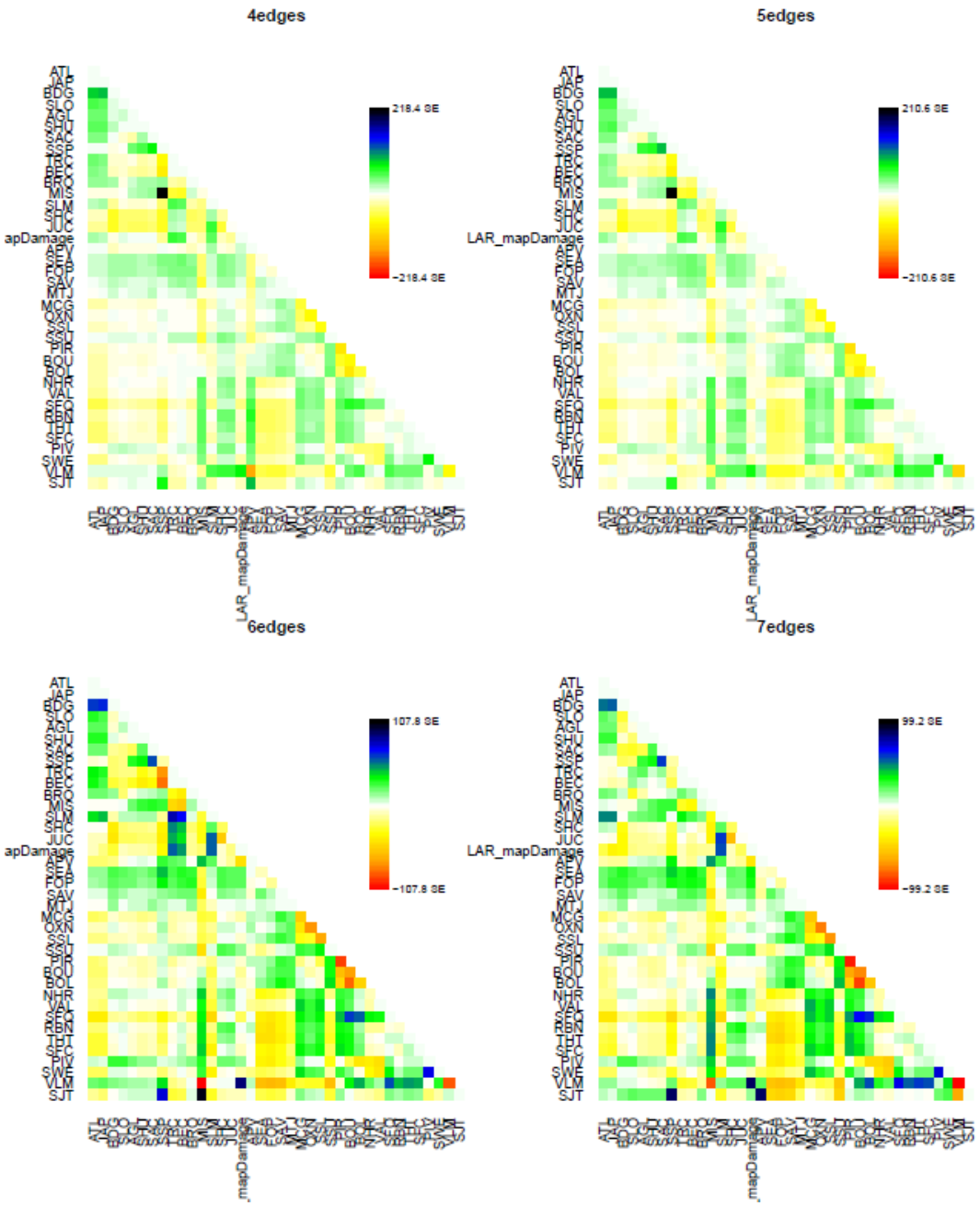


Figure S 2.6: Heatmap of TreeMix residual fits for 0-7 number of edges. Cool colors indicate positive residuals (underestimates of observed covariance) and pairs of populations that are more closely related to each other. In this case fit can be improved by adding edges. Warm colors indicate negative residuals and is a necessary outcome of the positive residuals, but can also mean pairs of populations that were forced too close due to unmodeled migration elsewhere in the graph (Pickrell & Pritchard, 2012).

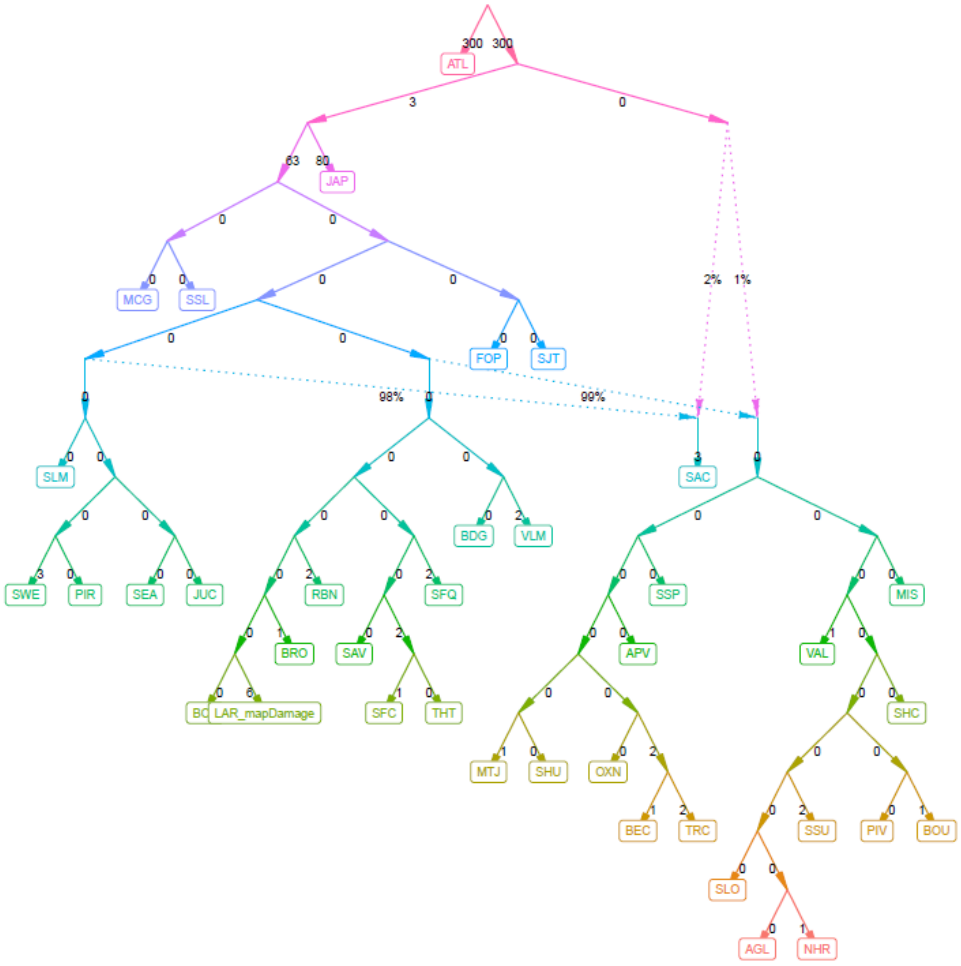


Figure S 2.7: Best admixture graph generated on calculated *f*-statistics by constraining number of admixture events as two (based on TreeMix), and fixing root on the Atlantic (ATL) outgroup species. Solid lines are drift edges and dotted lines are admixture proportions. Results do not recover events from other methods (TreeMix, fastSTRUCTURE, PCA), or known/suspected histories of translocation. Issues related to proportion of missing data when including the museum individual might have limited the accuracy of this method.

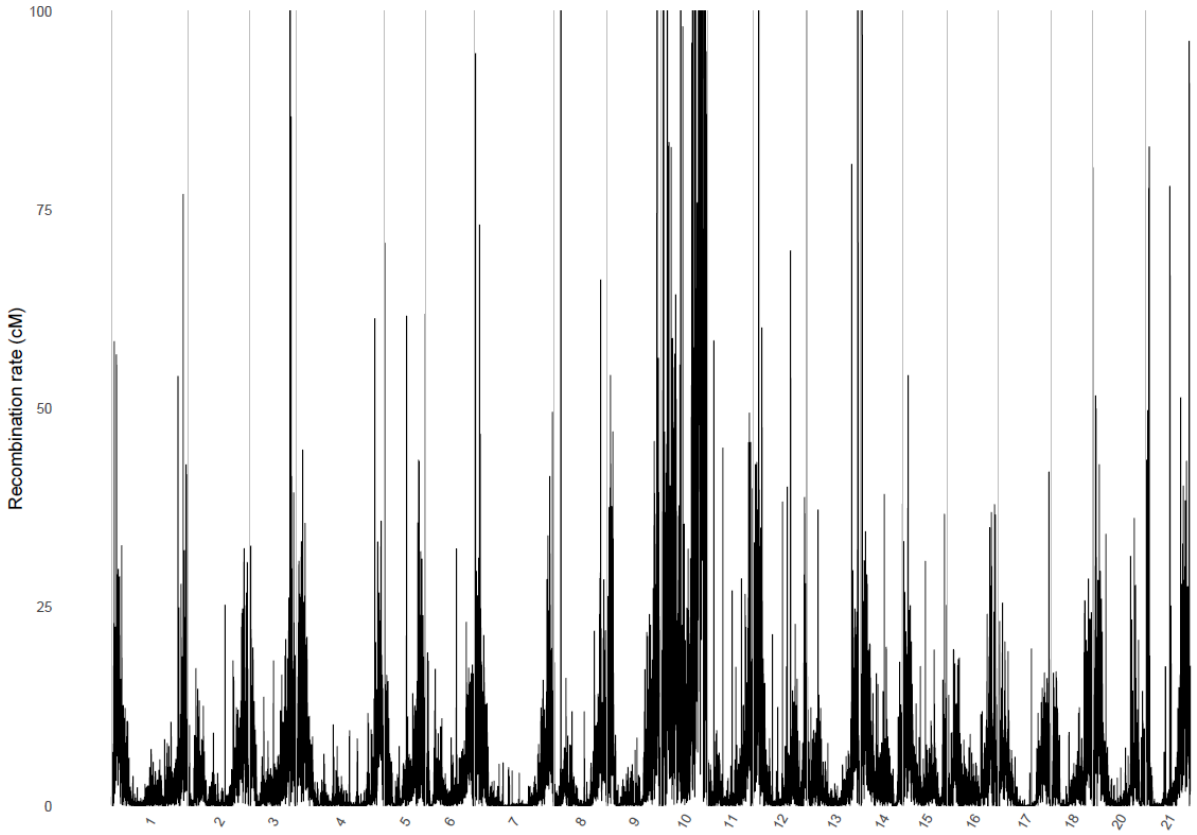


Figure S 2.8: Recombination rates generated by Shanfelter, Archambeault and White (2019) on the 20 autosomal chromosomes of *G. aculeatus*. Data was generated based on the reference genome published by Jones et al. (2012). Rates (cM) have been averaged across 10kb windows.

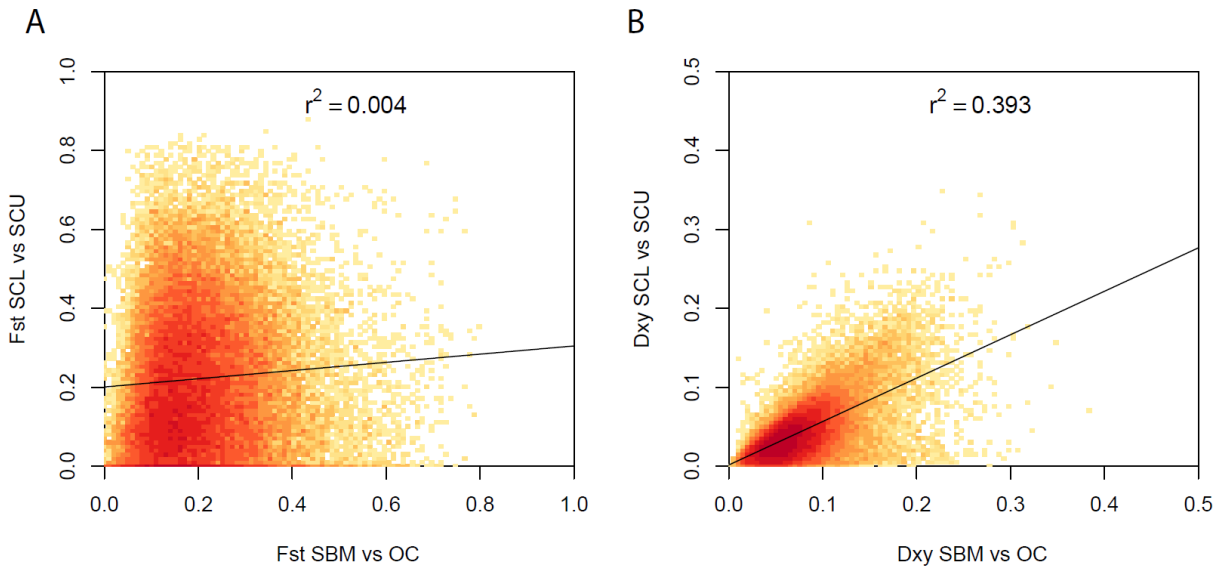


Figure S 2.9: Heatplot of correlation of (A) F_{st} and (B) D_{xy} between pairwise comparisons. The warmer the color, the more overlap between points. Correlation was stronger between absolute (D_{xy}) than relative (F_{st}) divergence.

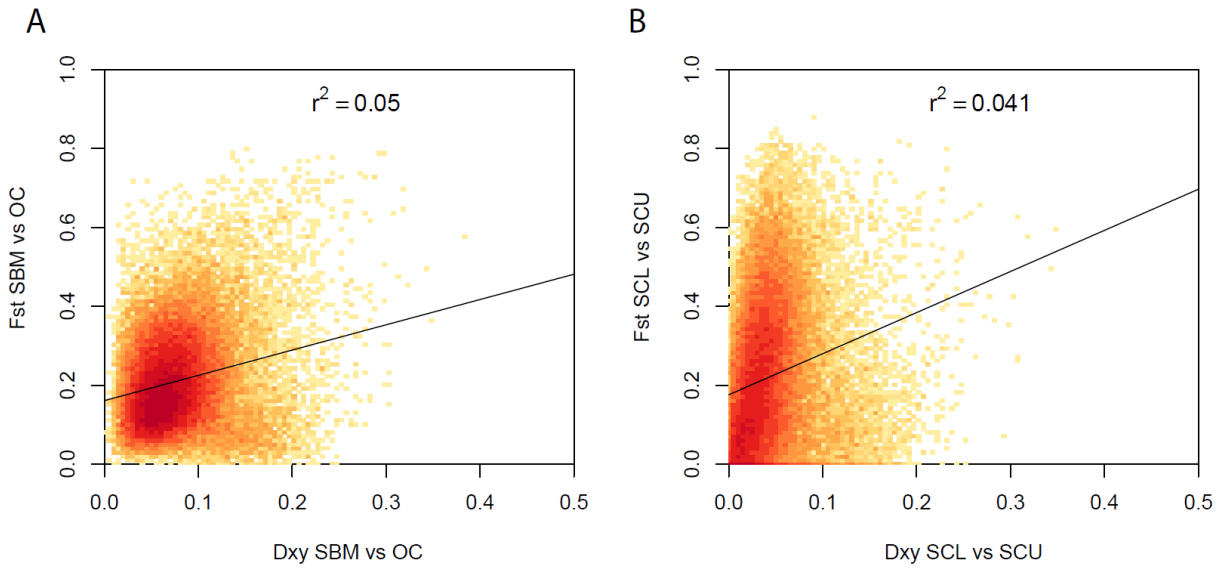


Figure S 2.10: Heatplot of correlation between F_{st} and D_{xy} for each pairwise comparison: (A) SBM vs OC; (B) SCL vs SCU. D_{xy} was corrected for mutation rate by dividing it with the mean of the divergence of each population with the outgroup (*G. nipponicus*). Overall, there was little correlation between both metrics for each pairwise comparison.

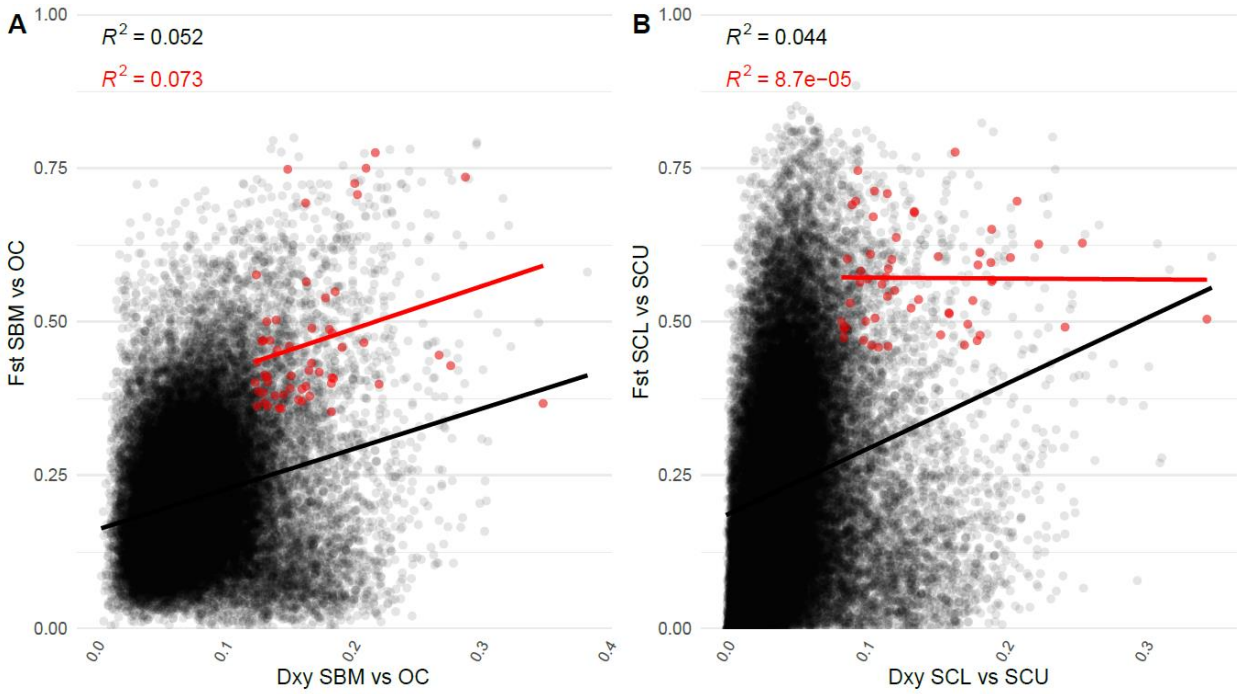


Figure S 2.11: Scatter plot of correlation between F_{st} and D_{xy} for each pairwise comparison: (A) SBM vs OC; (B) SCL vs SCU. Red circles represent the shared 55 windows within the 90th percentile of F_{st} and D_{xy} for both pairwise comparisons. There was a marginal improvement in the correlation of the 55 windows in the first comparison (A), but not in the second (B).

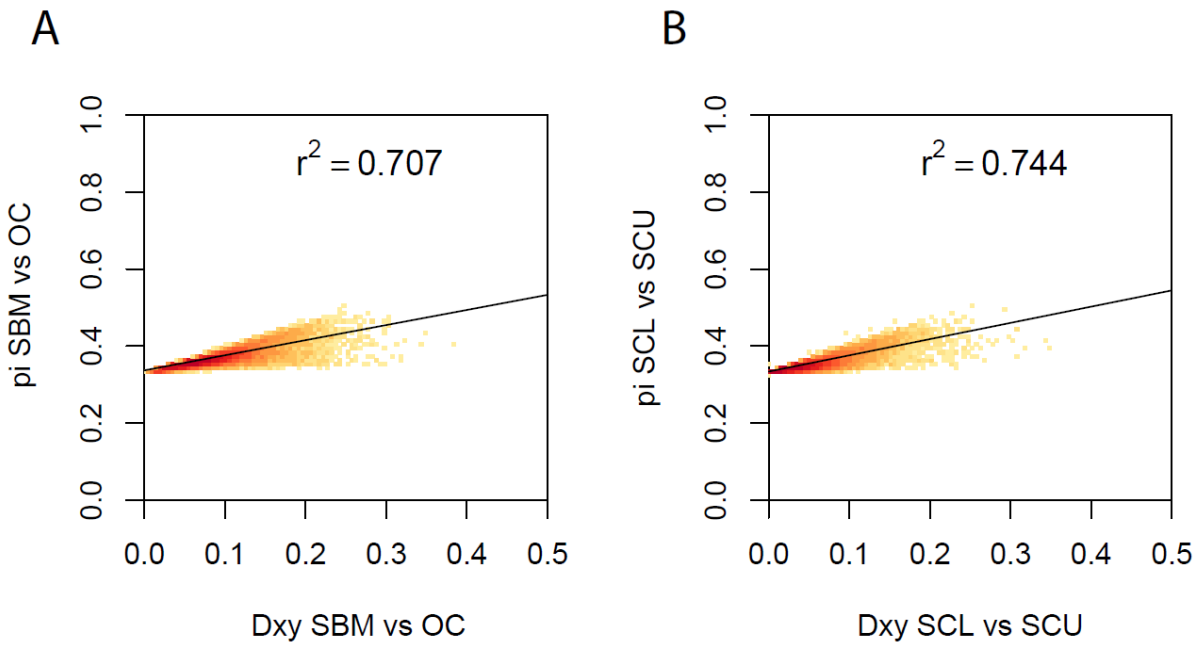


Figure S 2.12: Heatplot of correlation between *pi* and D_{xy} for each pairwise comparison: (A) SBM vs OC; (B) SCL vs SCU. The high correlation of nucleotide diversity and absolute divergences suggests that most of the divergence accumulated in each population is likely driven by mutations and low recombination rates.

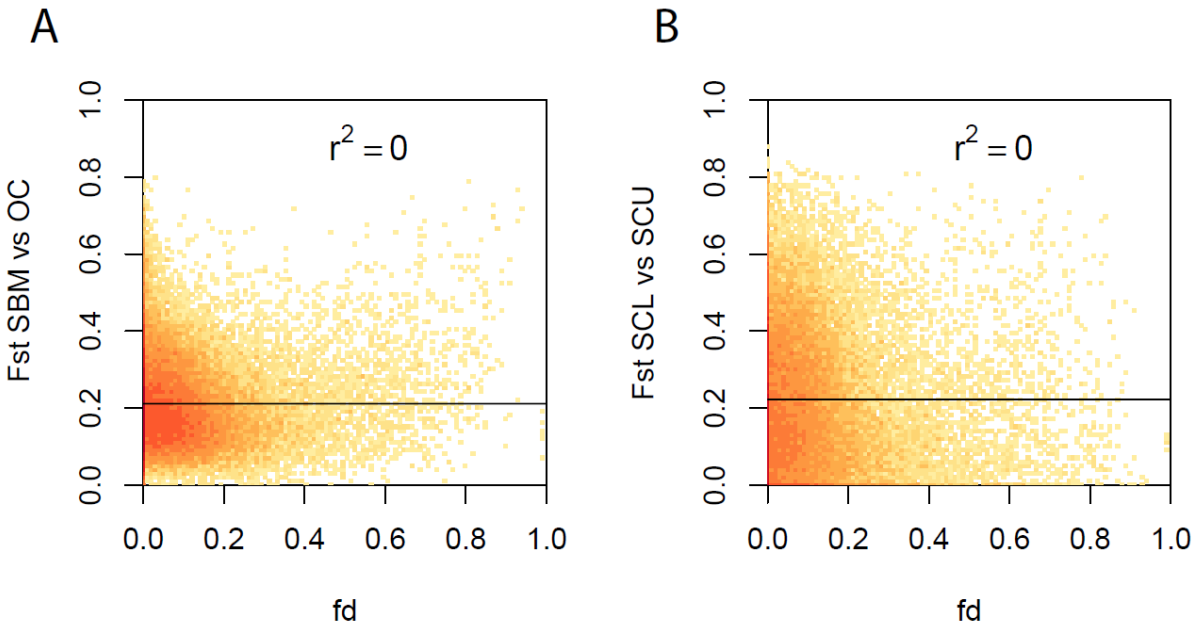


Figure S 2.13: Heatplot of correlation between D_{xy} and f_d for each pairwise comparison: (A) SBM vs OC; (B) SCL vs SCU. There is no correlation between relative divergence in each pairwise comparison and regions of excess allele sharing between unarmored populations.

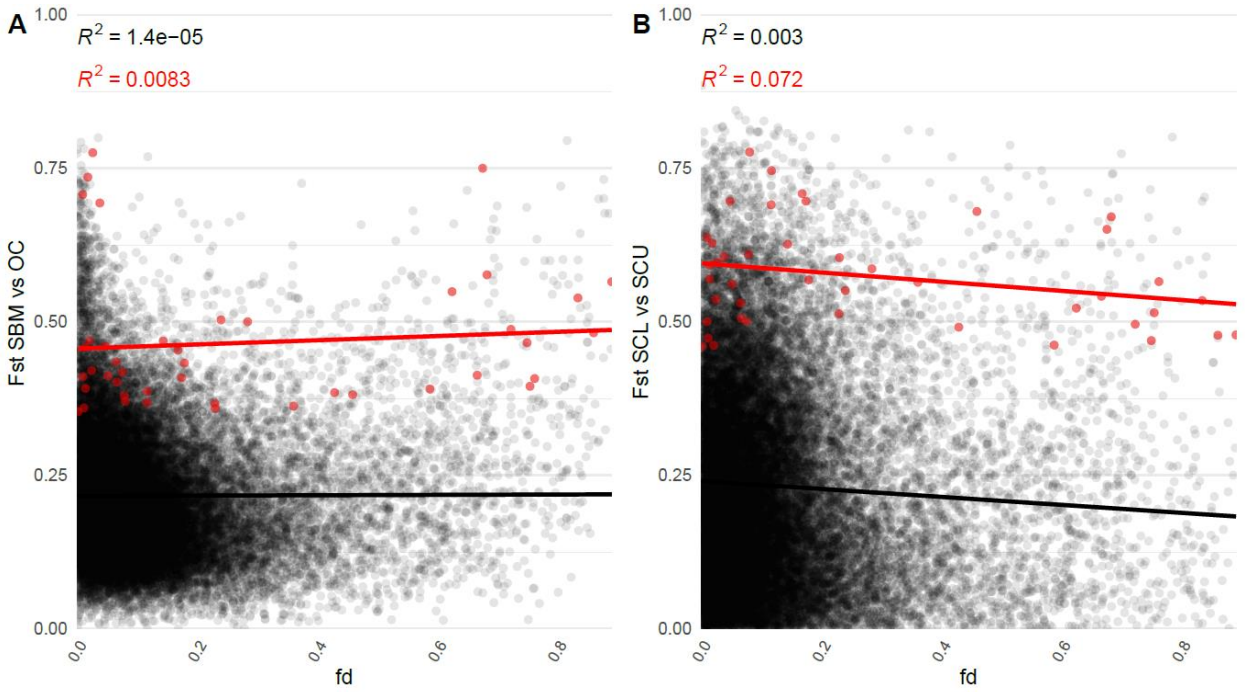


Figure S 2.14: Scatter plot of correlation between F_{st} and f_d for each pairwise comparison: (A) SBM vs OC; (B) SCL vs SCU. Red circles represent the shared 55 windows within the 90th percentile of F_{st} and D_{xy} for both pairwise comparisons. There was a marginal improvement in the correlation of the 55 windows in both comparisons.

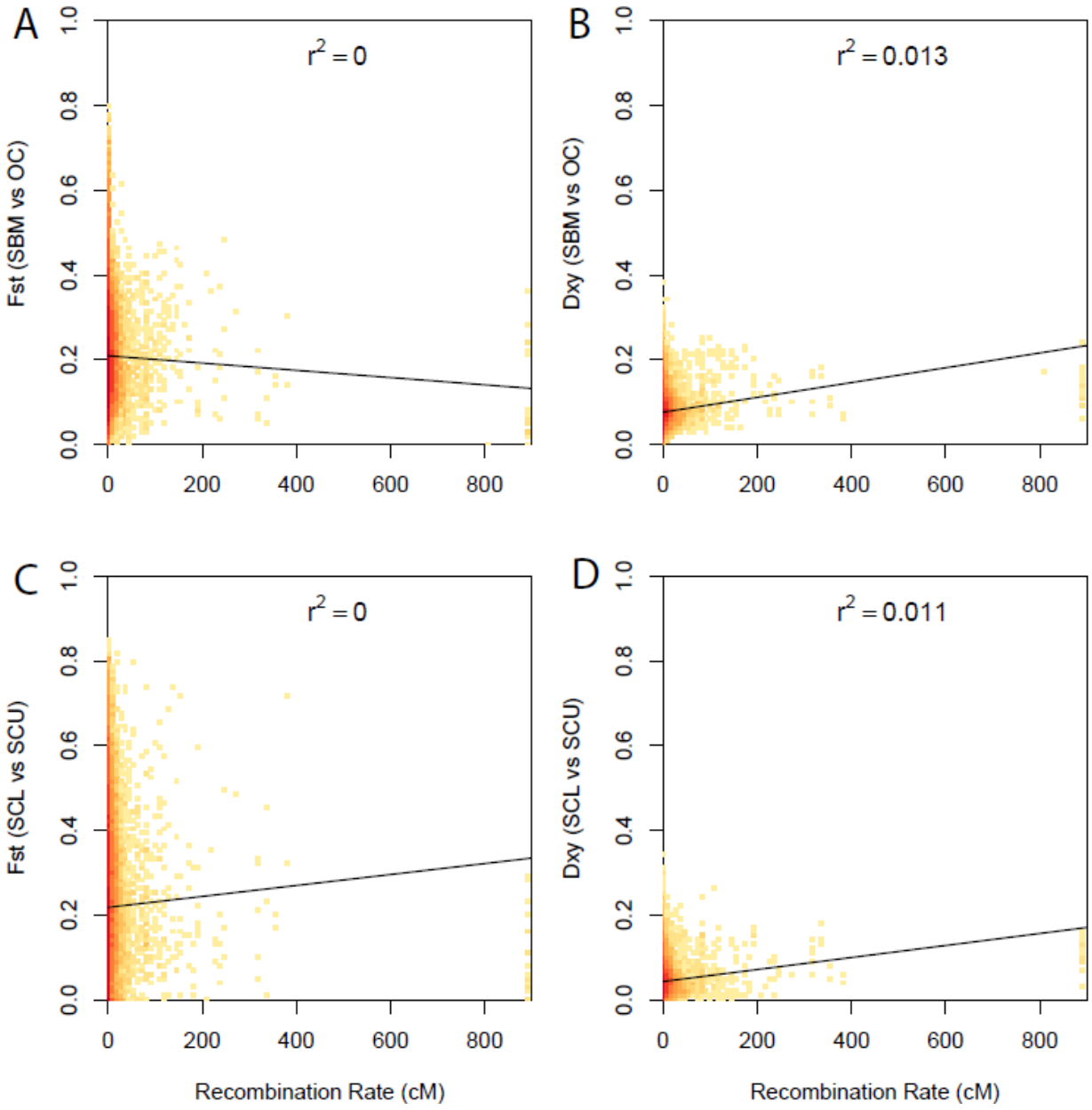


Figure S 2.15: Heatplot of correlation between recombination rates and divergence metrics (F_{st} and D_{xy}) for each pairwise comparison: (A-B) SBM vs OC; (C-D) SCL vs SCU. There is no correlation between either relative or absolute divergence to recombination rates. However, windows of recombination do not necessarily match windows of divergence since they were based on different reference genomes.

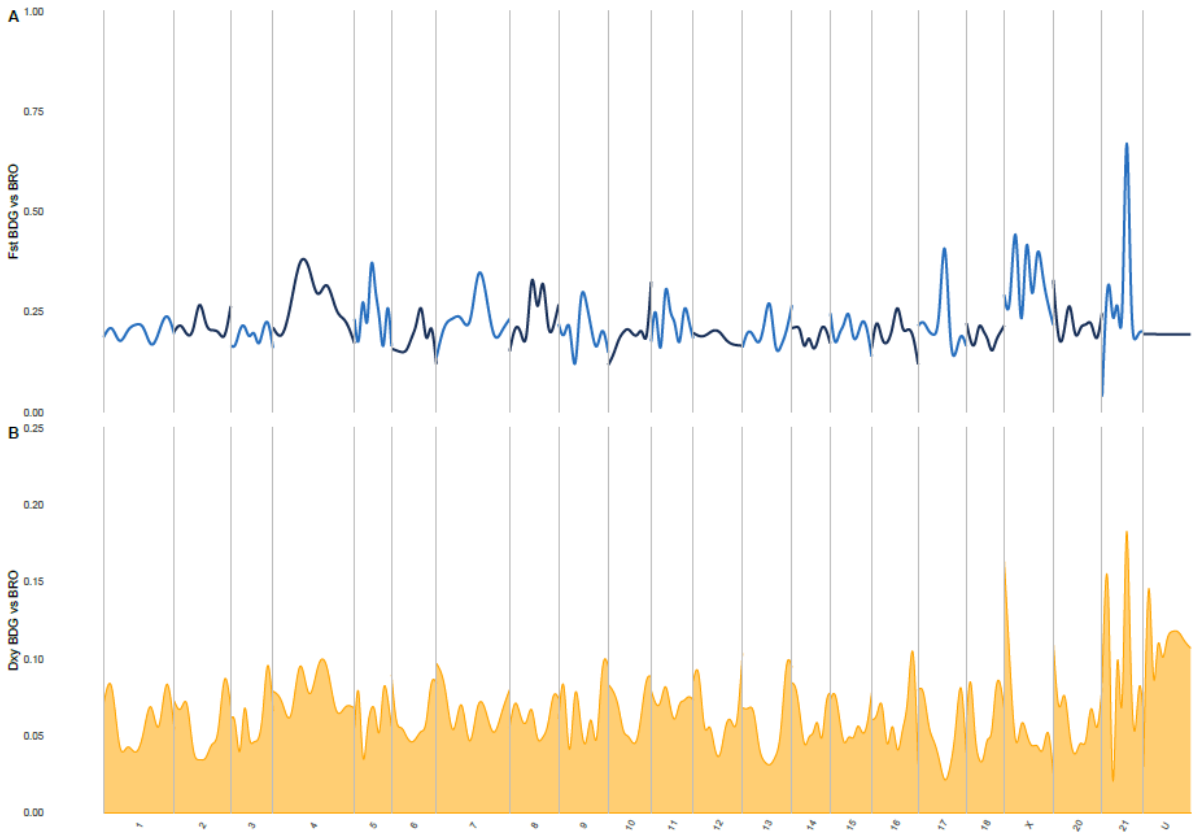


Figure S 2.16: Genome-wide sliding windows approach using the pairwise comparison between fully armored, Bodega Bay (BDG), and low armored, Bocana El Rosario (BRO) individuals. We were able to capture the major divergence peak on chromosome four associated with the *Eda* gene and plate number change (details on Figure S 2.17). There is another peak on chromosome 21, where a supergene cluster is located that is also associated with major trait divergence in marine-freshwater transition of *G. aculeatus* (Reid et al., 2021).

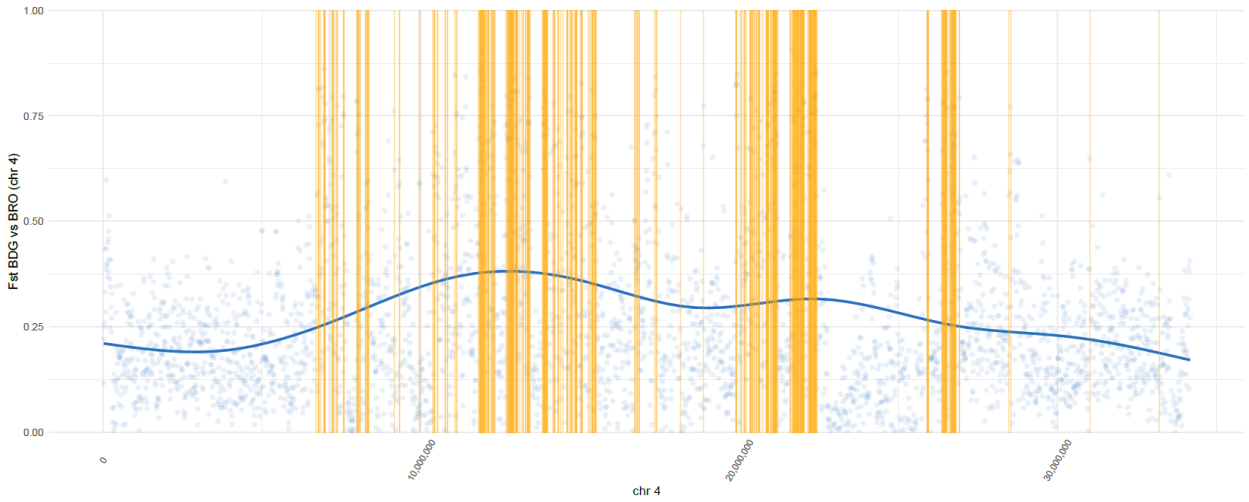


Figure S 2.17: Close up of chromosome four from sliding window approach on the pairwise comparison between BDG and BRO. Yellow bars denote windows where the 90th percentile F_{st} and D_{xy} values overlap. The location of the *Eda* gene is between positions 12,783,708-12,793,934, where the major F_{st} peak is located.

Bibliography

Alberdi, A., Aizpurua, O., Gilbert, M. T. P., & Bohmann, K. (2018). Scrutinizing key steps for reliable metabarcoding of environmental samples. *Methods in Ecology and Evolution*, *9*(1), 134–147. https://doi.org/10.1111/2041-210X.12849

Ballard, J., Pezda, J., Spencer, D., & Plantinga, A. (n.d.). *An Economic Valuation of Southern California Coastal Wetlands*. http://scwrp.org/wp-content/uploads/2017/06/SoCalWetlands_FinalReport.pdf

Barrett, R. D., Rogers, S. M., & Schluter, D. (2008). Natural selection on a major armor gene in threespine stickleback. *Science*, *322*(5899), 255–257.

Baumgartner, J. V. (1986). The genetics of differentiation in a stream population of the threespine stickleback, Gasterosteus aculeatus. *Heredity*, *57*(2), 199–208.

Beichman, A. C., Phung, T. N., & Lohmueller, K. E. (2017). Comparison of Single Genome and Allele Frequency Data Reveals Discordant Demographic Histories. *G3: Genes, Genomes, Genetics*, *7*(11), 3605–3620. https://doi.org/10.1534/g3.117.300259

Bell, M. A. (1976). Evolution of Phenotypic Diversity in Gasterosteus Aculeatus Superspecies on the Pacific Coast of North America. *Systematic Biology*, *25*(3), 211–227. https://doi.org/10.2307/2412489

Bell, M. A. (1978). Fishes of the Santa Clara River system, southern California. *Natural History Museum of Los Angeles County Contributions in Science*, *295*, 1–20.

Bell, M. A. (2001). Lateral plate evolution in the threespine stickleback: Getting nowhere fast. In A. P. Hendry & M. T. Kinnison (Eds.), *Microevolution Rate, Pattern, Process* (Vol. 8, pp. 445–461). Springer.

Bell, M. A., & Foster, S. A. (1994). *The Evolutionary Biology of the Threespine Stickleback.pdf*. Oxford University Press.

Bell, M. A., Orti, G., Walker, J. A., & Koenings, J. P. (1993). Evolution of Pelvic Reduction in Threespine Stickleback Fish: A Test of Competing Hypotheses. *Evolution*, *47*(3), 906. https://doi.org/10.2307/2410193

Bell, M. A., & Richkind, K. E. (1981). Clinal Variation of Lateral Plates in Threespine Stickleback Fish. *The American Naturalist*, *117*(2), 113–132. https://doi.org/10.1086/283693

Benjamin, A., May, B., O'Brien, J., & Finger, A. J. (2016). Conservation Genetics of an Urban Desert Fish, the Arroyo Chub. *Transactions of the American Fisheries Society*, *145*(2), 277–286. https://doi.org/10.1080/00028487.2015.1121925

Benke, A. C. (1990). A Perspective on America's Vanishing Streams. *Journal of the North American Benthological Society*, *9*(1), 77–88. https://doi.org/10.2307/1467936

Bergstrom, C. A. (2002). Fast-start swimming performance and reduction in lateral plate number in threespine stickleback. *Canadian Journal of Zoology*, *80*(2), 207–213. https://doi.org/10.1139/z01-226

Bohmann, K., Chua, P., Holman, L. E., & Lynggaard, C. (2021). DNAqua-Net conference unites participants from around the world with the quest to standardize and implement DNA-based aquatic biomonitoring. *Environmental DNA*, *3*(5), 884–888. https://doi.org/10.1002/edn3.207

Buth, D. G., Crabtree, C. B., Orton, R. D., & Rainboth, W. J. (1984). Genetic differentiation between the freshwater subspecies of Gasterosteus aculeatus in Southern California.

Biochemical Systematics and Ecology, *12*(4), 423–432. https://doi.org/10.1016/0305-1978(84)90075-9

Buxton, A. S., Groombridge, J. J., & Griffiths, R. A. (2017). Is the detection of aquatic environmental DNA influenced by substrate type? *PLOS ONE*, *12*(8), e0183371. https://doi.org/10.1371/journal.pone.0183371

Callahan, B. J., McMurdie, P. J., Rosen, M. J., Han, A. W., Johnson, A. J. A., & Holmes, S. P. (2016). DADA2: High-resolution sample inference from Illumina amplicon data. *Nature Methods*, *13*(7), 581–583. https://doi.org/10.1038/nmeth.3869

Campos, P. F., & Gilbert, T. M. P. (2012). DNA Extraction from Formalin-Fixed Material. In B. Shapiro & M. Hofreiter (Eds.), *Ancient DNA: Methods and Protocols* (pp. 81–85). Humana Press. https://doi.org/10.1007/978-1-61779-516-9_11

Caroe, C., & Bohmann, K. (2020). Tagsteady: A metabarcoding library preparation protocol to avoid false assignment of sequences to samples. *BioRxiv*.

Chen, H. (2018). *VennDiagram: Generate High-Resolution Venn and Euler Plots* (1.6.20). https://CRAN.R-project.org/package=VennDiagram

Colosimo, P. F., Hosemann, K. E., Balabhadra, S., Villarreal, G., Dickson, M., Grimwood, J., Schmutz, J., Myers, R. M., Schluter, D., & Kingsley, D. M. (2005). Widespread parallel evolution in sticklebacks by repeated fixation of ectodysplasin alleles. *Science*, *307*(5717), 1928–1933.

Colosimo, P. F., Peichel, C. L., Nereng, K., Blackman, B. K., Shapiro, M. D., Schluter, D., & Kingsley, D. M. (2004). The genetic architecture of parallel armor plate reduction in threespine sticklebacks. *PLoS Biol*, *2*(5), e109.

Conte, G. L., Arnegard, M. E., Best, J., Chan, Y. F., Jones, F. C., Kingsley, D. M., Schluter, D., & Peichel, C. L. (2015). Extent of QTL Reuse During Repeated Phenotypic Divergence of Sympatric Threespine Stickleback. *Genetics*, *201*(3), 1189–1200. https://doi.org/10.1534/genetics.115.182550

Cresko, W. A., Amores, A., Wilson, C., Murphy, J., Currey, M., Phillips, P., Bell, M. A., Kimmel, C. B., & Postlethwait, J. H. (2004). Parallel genetic basis for repeated evolution of armor loss in Alaskan threespine stickleback populations. *Proceedings of the National Academy of Sciences*, *101*(16), 6050–6055. https://doi.org/10.1073/pnas.0308479101

Cruickshank, T. E., & Hahn, M. W. (2014). Reanalysis suggests that genomic islands of speciation are due to reduced diversity, not reduced gene flow. *Molecular Ecology*, *23*(13), 3133–3157. https://doi.org/10.1111/mec.12796

Culver, G. B., & Hubbs, C. L. (1917). The fishes of the Santa Ana system of streams in Southern California. *Lorguinia*, *1*, 82–83.

Curd, E., Gomer, J., Kandlikar, G., Gold, Z., Ogden, M., & Shi, B. (2018). *The Anacapa Toolkit*. https://github.com/limey-bean/Anacapa

Curd, E., Kandlikar, G., & Gomer, J. (2018). *CRUX: Creating Reference libraries Using eXisting tools*. https://github.com/limey-bean/CRUX_Creating-Reference-libraries-Using-eXisting-tools

Cuvier, G., & Valenciennes, M. (1829). *Histoire naturelle des poissons*. Chez F. G. Levrault; [etc., etc.]. https://www.biodiversitylibrary.org/item/54024

Danecek, P., Auton, A., Abecasis, G., Albers, C. A., Banks, E., DePristo, M. A., Handsaker, R. E., Lunter, G., Marth, G. T., Sherry, S. T., McVean, G., & Durbin, R. (2011). The variant

call format and VCFtools. *Bioinformatics*, *27*(15), 2156–2158. https://doi.org/10.1093/bioinformatics/btr330

Danecek, P., Bonfield, J. K., Liddle, J., Marshall, J., Ohan, V., Pollard, M. O., Whitwham, A., Keane, T., McCarthy, S. A., Davies, R. M., & Li, H. (2021). Twelve years of SAMtools and BCFtools. *GigaScience*, *10*(2), giab008. https://doi.org/10.1093/gigascience/giab008

Deagle, B. E., Jones, F. C., Absher, D. M., Kingsley, D. M., & Reimchen, T. E. (2013a). Phylogeography and adaptation genetics of stickleback from the H aida G waii archipelago revealed using genome-wide single nucleotide polymorphism genotyping. *Molecular Ecology*, *22*(7), 1917–1932.

Deagle, B. E., Jones, F. C., Absher, D. M., Kingsley, D. M., & Reimchen, T. E. (2013b). Phylogeography and adaptation genetics of stickleback from the Haida Gwaii archipelago revealed using genome-wide single nucleotide polymorphism genotyping. *Molecular Ecology*, *22*(7), 1917–1932. https://doi.org/10.1111/mec.12215

Deiner, K., Bik, H. M., Mächler, E., Seymour, M., Lacoursière-Roussel, A., Altermatt, F., Creer, S., Bista, I., Lodge, D. M., & De Vere, N. (2017). Environmental DNA metabarcoding: Transforming how we survey animal and plant communities. *Molecular Ecology*, *26*(21), 5872–5895.

Deiner, K., Walser, J.-C., Mächler, E., & Altermatt, F. (2015). Choice of capture and extraction methods affect detection of freshwater biodiversity from environmental DNA. *Biological Conservation*, *183*, 53–63.

Dejean, T., Valentini, A., Miquel, C., Taberlet, P., Bellemain, E., & Miaud, C. (2012). Improved detection of an alien invasive species through environmental DNA barcoding: The example of the American bullfrog *Lithobates catesbeianus*: *Alien invasive species*

detection using eDNA. Journal of Applied Ecology, *49*(4), 953–959. https://doi.org/10.1111/j.1365-2664.2012.02171.x

Dell'Anno, A., & Corinaldesi, C. (2004). Degradation and Turnover of Extracellular DNA in Marine Sediments: Ecological and Methodological Considerations Degradation and Turnover of Extracellular DNA in Marine Sediments: Ecological and Methodological Considerations. *Applied and Environmental Microbiology*, *70*(7), 4384–4386. https://doi.org/10.1128/AEM.70.7.4384

Des Roches, S., Bell, M. A., & Palkovacs, E. P. (2020). Climate-driven habitat change causes evolution in Threespine Stickleback. *Global Change Biology*, *26*(2), 597–606.

Doi, H., Uchii, K., Matsuhashi, S., Takahara, T., Yamanaka, H., & Minamoto, T. (2017). Isopropanol precipitation method for collecting fish environmental DNA. *Limnology and Oceanography: Methods*, *15*(2), 212–218. https://doi.org/10.1002/lom3.10161

Dolby, G. A., Bedolla, A. M., Bennett, S. E. K., & Jacobs, D. K. (2020). Global physical controls on estuarine habitat distribution during sea level change: Consequences for genetic diversification through time. *Global and Planetary Change*, *187*, 103128. https://doi.org/10.1016/j.gloplacha.2020.103128

Dolby, G. A., Ellingson, R. A., Findley, L. T., & Jacobs, D. K. (2018). How sea level change mediates genetic divergence in coastal species across regions with varying tectonic and sediment processes. *Molecular Ecology*, *27*(4), 994–1011. https://doi.org/10.1111/mec.14487

Dolby, G. A., Hechinger, R., Ellingson, R. A., Findley, L. T., Lorda, J., & Jacobs, D. K. (2016). Sea-level driven glacial-age refugia and post-glacial mixing on subtropical coasts, a

palaeohabitat and genetic study. *Proceedings of the Royal Society B: Biological Sciences*, *283*(1843), 20161571.

Earl, D. A., Louie, K. D., Bardeleben, C., Swift, C. C., & Jacobs, D. K. (2010). Rangewide microsatellite phylogeography of the endangered tidewater goby, a genetically subdivided coastal fish with limited marine dispersal. *Conservation Genetics*, *11*, 103–104. https://doi.org/10.1007/s10592-009-0008-9

Enciso-Romero, J., Pardo-Díaz, C., Martin, S. H., Arias, C. F., Linares, M., McMillan, W. O., Jiggins, C. D., & Salazar, C. (2017). Evolution of novel mimicry rings facilitated by adaptive introgression in tropical butterflies. *Molecular Ecology*, *26*(19), 5160–5172. https://doi.org/10.1111/mec.14277

Eriksson, B. K., Yanos, C., Bourlat, S. J., Donadi, S., Fontaine, M. C., Hansen, J. P., Jakubavičiūtė, E., Kiragosyan, K., Maan, M. E., Merilä, J., Austin, Å. N., Olsson, J., Reiss, K., Sundblad, G., Bergström, U., & Eklöf, J. S. (2021). Habitat segregation of plate phenotypes in a rapidly expanding population of three-spined stickleback. *Ecosphere*, *12*(6), 1–18. https://doi.org/10.1002/ecs2.3561

Esling, P., Lejzerowicz, F., & Pawlowski, J. (2015). Accurate multiplexing and filtering for high-throughput amplicon-sequencing. *Nucleic Acids Research*, *43*(5), 2513–2524. https://doi.org/10.1093/nar/gkv107

Evanno, G., Regnaut, S., & Goudet, J. (2005). Detecting the number of clusters of individuals using the software structure: A simulation study. *Molecular Ecology*, *14*(8), 2611–2620. https://doi.org/10.1111/j.1365-294X.2005.02553.x

Ewing, G. B., & Jensen, J. D. (2016). The consequences of not accounting for background selection in demographic inference. *Molecular Ecology*, *25*(1), 135–141.

Faber, P. M., Keller, E., Sands, A., & Massey, B. M. (1989). *The Ecology of Riparian Habitats of the Southern California Coastal Region: A Community Profile: Vol. 7.27*. U.S. Department of the Interior, Fish and Wildlife Service, Research and Development, National Wetlands Research Center.

Fang, B., Kemppainen, P., Momigliano, P., Feng, X., & Merilä, J. (2020). On the causes of geographically heterogeneous parallel evolution in sticklebacks. *Nature Ecology & Evolution*, *4*(8), 1105–1115.

Fang, B., Merilä, J., Ribeiro, F., Alexandre, C. M., & Momigliano, P. (2018). Worldwide phylogeny of three-spined sticklebacks. *Molecular Phylogenetics and Evolution*, *127*, 613–625.

Faribanks, H. W., & Carey, E. P. (1910). Glaciation in the San Bernardino Range, California. *Science*, *31*(784), 32–33.

FASTX-Toolkit. (n.d.). Retrieved January 11, 2018, from http://hannonlab.cshl.edu/fastx_toolkit/

Ficetola, G. F., Miaud, C., Pompanon, F., & Taberlet, P. (2008). Species detection using environmental DNA from water samples. *Biology Letters*, *4*(4), 423–425. https://doi.org/10.1098/rsbl.2008.0118

Fitak, R. R. (2021). OptM: Estimating the optimal number of migration edges on population trees using Treemix. *Biology Methods and Protocols*, *6*(1), bpab017. https://doi.org/10.1093/biomethods/bpab017

Flint, L. E., Brandt, J., Christensen, A. H., Flint, A. L., Hevesi, J. A., Jachens, R., Kulongoski, J. T., Martin, P., & Sneed, M. (2012). Geohydrology of Big Bear Valley, California: Phase 1--geologic framework, recharge, and preliminary assessment of the source and age of groundwater. In *Geohydrology of Big Bear Valley, California: Phase 1—Geologic*

framework, recharge, and preliminary assessment of the source and age of groundwater (USGS Numbered Series No. 2012–5100; Scientific Investigations Report, Vols. 2012–5100, p. 130). U.S. Geological Survey. https://doi.org/10.3133/sir20125100

Flint, L. E., Underwood, E. C., Flint, A. L., & Hollander, A. D. (2019). Characterizing the Influence of Fire on Hydrology in Southern California. *Natural Areas Journal*, *39*(1), 108–121.

Freedman, A. H., Gronau, I., Schweizer, R. M., Vecchyo, D. O.-D., Han, E., Silva, P. M., Galaverni, M., Fan, Z., Marx, P., Lorente-Galdos, B., Beale, H., Ramirez, O., Hormozdiari, F., Alkan, C., Vilà, C., Squire, K., Geffen, E., Kusak, J., Boyko, A. R., … Novembre, J. (2014). Genome Sequencing Highlights the Dynamic Early History of Dogs. *PLOS Genetics*, *10*(1), e1004016. https://doi.org/10.1371/journal.pgen.1004016

French, J. J., & Busby, M. W. (1974). Flood-hazard study, 100-year stage for Baldwin Lake, San Bernardino County, California. In *Flood-hazard study, 100-year stage for Baldwin Lake, San Bernardino County, California* (USGS Numbered Series No. 74–26; Water-Resources Investigations Report, Vols. 74–26). U.S. Geological Survey. https://doi.org/10.3133/wri7426

Gao, X., Lin, H., Revanna, K., & Dong, Q. (2017). A Bayesian taxonomic classification method for 16S rRNA gene sequences with improved species-level accuracy. *BMC Bioinformatics*, *18*(1), 247. https://doi.org/10.1186/s12859-017-1670-4

Gbif.Org. (2022). *Occurrence Download* (p. 170487) [Darwin Core Archive]. The Global Biodiversity Information Facility. https://doi.org/10.15468/DL.HTJ3HT

Girard, C. F. (1854). *Descriptions of new fishes, collected by Dr. AL Heermann, naturalist attached to the survey of the Pacific railroad route, under Lieut. RS Williamson, USA. 7*, 129–140.

Glover, K. C., Chaney, A., Kirby, M. E., Patterson, W. P., & MacDonald, G. M. (2020). Southern California vegetation, wildfire, and erosion had nonlinear responses to climatic forcing during marine isotope stages 5–2 (120–15 ka). *Paleoceanography and Paleoclimatology*, *35*(2), 1–15.

Glover, K. C., George, J., Heusser, L., & MacDonald, G. M. (2021). West Coast vegetation shifts as a response to climate change over the past 130,000 years: Geographic patterns and process from pollen data. *Physical Geography*, 1–19.

Glover, K. C., MacDonald, G. M., Kirby, M. E., Rhodes, E. J., Stevens, L., Silveira, E., Whitaker, A., & Lydon, S. (2017). Evidence for orbital and North Atlantic climate forcing in alpine Southern California between 125 and 10 ka from multi-proxy analyses of Baldwin Lake. *Quaternary Science Reviews*, *167*, 47–62.

Goldberg, C. S., Turner, C. R., Deiner, K., Klymus, K. E., Thomsen, P. F., Murphy, M. A., Spear, S. F., McKee, A., Oyler-McCance, S. J., & Cornman, R. S. (2016). Critical considerations for the application of environmental DNA methods to detect aquatic species. *Methods in Ecology and Evolution*, *7*(11), 1299–1307.

Gumprecht, B. (1999). *The Los Angeles River: Its Life, Death, and Possible Rebirth* (2001st ed.). Johns Hopkins University Press.

Guo, Y., Ye, F., Sheng, Q., Clark, T., & Samuels, D. C. (2014). Three-stage quality control strategies for DNA re-sequencing data. *Briefings in Bioinformatics*, *15*(6), 879–889. https://doi.org/10.1093/bib/bbt069

Hagen, D. W., & Gilbertson, L. G. (1973). Selective predation and the intensity of selection acting upon the lateral plates of threespine sticklebacks. *Heredity*, *30*(3), 273–287.

Haglund, T., & Buth, D. (1988). Allozymes of the unarmored threespine stickleback (Gasterosteus aculeatus williamsoni) and identification of the Shay Creek population. *Isozyme Bull*, *21*, 196.

Han, F., Lamichhaney, S., Grant, B. R., Grant, P. R., Andersson, L., & Webster, M. T. (2017). Gene flow, ancient polymorphism, and ecological adaptation shape the genomic landscape of divergence among Darwin's finches. *Genome Research*, *27*(6), 1004–1015. https://doi.org/10.1101/gr.212522.116

Harper, L. R., Buxton, A. S., Rees, H. C., Bruce, K., Brys, R., Halfmaerten, D., Read, D. S., Watson, H. V., Sayer, C. D., & Jones, E. P. (2019). Prospects and challenges of environmental DNA (eDNA) monitoring in freshwater ponds. *Hydrobiologia*, *826*(1), 25–41.

Hendry, A. P. (2013). Key questions in the genetics and genomics of eco-evolutionary dynamics. *Heredity*, *111*(6), 456–466. https://doi.org/10.1038/hdy.2013.75

Hendy, I. (2009). A fresh perspective on the Cordilleran Ice Sheet. *The Geological Society of America*, *19*(4–5), 60. https://doi.org/10.1130/0091

Heusser, L. (1998). Direct correlation of millennial-scale changes in western North American vegetation and climate with changes in the California Current System over the past ~60 kyr. *Paleoceanography*, *13*(3), 252–262. https://doi.org/10.1029/98PA00670

Heusser, L. E., Kirby, M. E., & Nichols, J. E. (2015). Pollen-based evidence of extreme drought during the last Glacial (32.6–9.0 ka) in coastal southern California. *Quaternary Science Reviews*, *126*, 242–253.

Higuchi, M., Sakai, H., & Goto, A. (2014). A new threespine stickleback, Gasterosteus nipponicus sp. nov.(Teleostei: Gasterosteidae), from the Japan Sea region. *Ichthyological Research*, *61*(4), 341–351.

Hinlo, R., Gleeson, D., Lintermans, M., & Furlan, E. (2017). Methods to maximise recovery of environmental DNA from water samples. *PloS One*, *12*(6), e0179251.

Hohenlohe, P. A., & Magalhaes, I. S. (2020). The Population Genomics of Parallel Adaptation: Lessons from Threespine Stickleback. In M. F. Oleksiak & O. P. Rajora (Eds.), *Population Genomics: Marine Organisms* (pp. 249–276). Springer International Publishing. https://doi.org/10.1007/13836_2019_67

Indjeian, V. B., Kingman, G. A., Jones, F. C., Guenther, C. A., Grimwood, J., Schmutz, J., Myers, R. M., & Kingsley, D. M. (2016). Evolving new skeletal traits by cis-regulatory changes in bone morphogenetic proteins. *Cell*, *164*(1–2), 45–56.

Jacobs, D. K., Haney, T. A., & Louie, K. D. (2004). Genes, diversity, and geologic process on the Pacific coast. *Annual Review of Earth and Planetary Sciences*, *32*(1), 601–652. https://doi.org/10.1146/annurev.earth.32.092203.122436

Jacobs, D. K., Stein, E. D., & Longcore, T. (2011a). Classification of California Estuaries Based on Natural Closure Patterns: Templates for Restoration and Management Management. *Technical Report, August*, 1–72.

Jacobs, D. K., Stein, E. D., & Longcore, T. (2011b). Classification of California Estuaries Based on Natural Closure Patterns: Templates for Restoration and Management Management. *Technical Report, August*, 1–72.

Jacobs, D., Stein, E. D., & Longcore, T. (2011). *Classification of California Estuaries Based on Natural Closure Patterns: Templates for Restoration and Management* (619.a; p. 72).

Jombart, T., & Ahmed, I. (2011). adegenet 1.3-1: New tools for the analysis of genome-wide SNP data. *Bioinformatics*, *27*(21), 3070–3071. https://doi.org/10.1093/bioinformatics/btr521

Jones, D. H., & John, A. W. G. (1978). A three-spined stickleback, Gasterosteus aculeatus L. from the North Atlantic. *Journal of Fish Biology*, *13*(2), 231–236. https://doi.org/10.1111/j.1095-8649.1978.tb03430.x

Jones, F. C., Grabherr, M. G., Chan, Y. F., Russell, P., Mauceli, E., Johnson, J., Swofford, R., Pirun, M., Zody, M. C., & White, S. (2012). The genomic basis of adaptive evolution in threespine sticklebacks. *Nature*, *484*(7392), 55–61.

Jones, F. C., Grabherr, M. G., Chan, Y. F., Russell, P., Mauceli, E., Johnson, J., Swofford, R., Pirun, M., Zody, M. C., White, S., Birney, E., Searle, S., Schmutz, J., Grimwood, J., Dickson, M. C., Myers, R. M., Miller, C. T., Summers, B. R., Knecht, A. K., … Kingsley, D. M. (2012). The genomic basis of adaptive evolution in threespine sticklebacks. *Nature*, *484*(7392), 55–61. https://doi.org/10.1038/nature10944

Jónsson, H., Ginolhac, A., Schubert, M., Johnson, P. L. F., & Orlando, L. (2013). MapDamage2.0: Fast approximate Bayesian estimates of ancient DNA damage parameters. *Bioinformatics*, *29*(13), 1682–1684. https://doi.org/10.1093/bioinformatics/btt193

Kandlikar, G. (2020). *ranacapa: Utility Functions and "shiny" App for Simple Environmental DNA Visualizations and Analyses* (0.1.0). https://github.com/gauravsk/ranacapa

Kelly, R. P., Shelton, A. O., & Gallego, R. (2019). Understanding PCR Processes to Draw Meaningful Conclusions from Environmental DNA Studies. *Scientific Reports*, *9*(1), 12133. https://doi.org/10.1038/s41598-019-48546-x

Kingman, G. A. R., Vyas, D. N., Jones, F. C., Brady, S. D., Chen, H. I., Reid, K., Milhaven, M., Bertino, T. S., Aguirre, W. E., & Heins, D. C. (2021). Predicting future from past: The genomic basis of recurrent and rapid stickleback evolution. *Science Advances*, *7*(25), eabg5285.

Kirby, M. E., Feakins, S. J., Bonuso, N., Fantozzi, J. M., & Hiner, C. A. (2013). Latest Pleistocene to Holocene hydroclimates from Lake Elsinore, California. *Quaternary Science Reviews*, *76*, 1–15. https://doi.org/10.1016/j.quascirev.2013.05.023

Kircher, M., Sawyer, S., & Meyer, M. (2012). Double indexing overcomes inaccuracies in multiplex sequencing on the Illumina platform. *Nucleic Acids Research*, *40*(1), e3–e3. https://doi.org/10.1093/nar/gkr771

Konijnendijk, N., Shikano, T., Daneels, D., Volckaert, F. A., & Raeymaekers, J. A. (2015). Signatures of selection in the three-spined stickleback along a small-scale brackish water–freshwater transition zone. *Ecology and Evolution*, *5*(18), 4174–4186.

Kopelman, N. M., Mayzel, J., Jakobsson, M., Rosenberg, N. A., & Mayrose, I. (2015). Clumpak: A program for identifying clustering modes and packaging population structure inferences across K. *Molecular Ecology Resources*, *15*(5), 1179–1191. https://doi.org/10.1111/1755-0998.12387

Kumar, G., Farrell, E., Reaume, A. M., Eble, J. A., & Gaither, M. R. (2022). One size does not fit all: Tuning eDNA protocols for high- and low-turbidity water sampling. *Environmental DNA*, *4*(1), 167–180. https://doi.org/10.1002/edn3.235

Kwambana, B. A., Mohammed, N. I., Jeffries, D., Barer, M., Adegbola, R. A., & Antonio, M. (2011). Differential effects of frozen storage on the molecular detection of bacterial taxa

that inhabit the nasopharynx. *BMC Clinical Pathology*, *11*(1), 2. https://doi.org/10.1186/1472-6890-11-2

Lai, Y.-T., Yeung, C. K., Omland, K. E., Pang, E.-L., Hao, Y., Liao, B.-Y., Cao, H.-F., Zhang, B.-W., Yeh, C.-F., & Hung, C.-M. (2019). Standing genetic variation as the predominant source for adaptation of a songbird. *Proceedings of the National Academy of Sciences*, *116*(6), 2152–2157.

Langmead, B., & Salzberg, S. L. (2012). Fast gapped-read alignment with Bowtie 2. *Nature Methods*, *9*(4), 357–359. https://doi.org/10.1038/nmeth.1923

Laramie, M. B., Pilliod, D. S., Goldberg, C. S., & Strickler, K. M. (2015). Environmental DNA sampling protocol—Filtering water to capture DNA from aquatic organisms. *U.S Geological Survey Techniques and Methods*, *Book 2*(Chapter A13), 15 p. https://doi.org/10.3133/TM2A13

Larsson, A. J. M., Stanley, G., Sinha, R., Weissman, I. L., & Sandberg, R. (2018). Computational correction of index switching in multiplexed sequencing libraries. *Nature Methods*, *15*(5), 305–307. https://doi.org/10.1038/nmeth.4666

Leidy, R. A., & Moyle, P. B. (2021). Keeping up with the status of freshwater fishes: A California (USA) perspective. *Conservation Science and Practice*, *3*(8). https://doi.org/10.1111/csp2.474

Levick, L. R., Goodrich, D. C., Hernandez, M., Fonseca, J., Semmens, D. J., Stromberg, J. C., Tluczek, M., Leidy, R. A., Scianni, M., & Guertin, D. P. (2008). *The ecological and hydrological significance of ephemeral and intermittent streams in the arid and semi-arid American Southwest*. Environmental Protection Agency, Office of Research and Development.

Levy-Booth, D. J., Campbell, R. G., Gulden, R. H., Hart, M. M., Powell, J. R., Klironomos, J. N., Pauls, K. P., Swanton, C. J., Trevors, J. T., & Dunfield, K. E. (2007). Cycling of extracellular DNA in the soil environment. *Soil Biology and Biochemistry*, *39*(12), 2977–2991. https://doi.org/10.1016/j.soilbio.2007.06.020

Li, H. (2013a). *Aligning sequence reads, clone sequences and assembly contigs with BWA-MEM*. *00*(00), 1–3. https://doi.org/10.48550/arXiv.1303.3997

Li, H. (2013b). *Aligning sequence reads, clone sequences and assembly contigs with BWA-MEM*. *00*(00), 1–3. https://doi.org/arXiv:1303.3997 [q-bio.GN]

Li, H., & Durbin, R. (2011). Inference of human population history from individual whole-genome sequences. *Nature*, *475*(7357), 493–496. https://doi.org/10.1038/nature10231

Li, J., Lawson Handley, L.-J., Read, D. S., & Hänfling, B. (2018). The effect of filtration method on the efficiency of environmental DNA capture and quantification via metabarcoding. *Molecular Ecology Resources*, *18*(5), 1102–1114.

Liang, Z., & Keeley, A. (2013). Filtration Recovery of Extracellular DNA from Environmental Water Samples. *Environmental Science & Technology*, *47*(16), 9324–9331. https://doi.org/10.1021/es401342b

Lipson, M. (2020). Applying f_4 -statistics and admixture graphs: Theory and examples. *Molecular Ecology Resources*, *20*(6), 1658–1667. https://doi.org/10.1111/1755-0998.13230

Liu, S., Hansen, M. M., & Jacobsen, M. W. (2016). Region-wide and ecotype-specific differences in demographic histories of threespine stickleback populations, estimated from whole genome sequences. *Molecular Ecology*, *25*(20), 5187–5202. https://doi.org/10.1111/mec.13827

Love, M. I., Huber, W., & Anders, S. (2014). Moderated estimation of fold change and dispersion for RNA-seq data with DESeq2. *Genome Biology*, *15*(12), 550. https://doi.org/10.1186/s13059-014-0550-8

Maier, R., Flegontov, P., Flegontova, O., Changmai, P., & Reich, D. (2022). *On the limits of fitting complex models of population history to genetic data* (p. 2022.05.08.491072). bioRxiv. https://doi.org/10.1101/2022.05.08.491072

Majaneva, M., Diserud, O. H., Eagle, S. H., Boström, E., Hajibabaei, M., & Ekrem, T. (2018). Environmental DNA filtration techniques affect recovered biodiversity. *Scientific Reports*, *8*(1), 1–11.

Mäkinen, H. S., Cano, J. M., & Merilä, J. (2006). Genetic relationships among marine and freshwater populations of the European three-spined stickleback (Gasterosteus aculeatus) revealed by microsatellites. *Molecular Ecology*, *15*(6), 1519–1534.

Malcolm, J. R. (1992). *Supporting information for a petition to list as endangered or threatened: Shay Creek stickleback, Gasterosteus sp* (P. B. Moyle & R. M. Yoshiyama, Eds.; pp. 213–222).

Marques, D. A., Jones, F. C., Di Palma, F., Kingsley, D. M., & Reimchen, T. E. (2018). Experimental evidence for rapid genomic adaptation to a new niche in an adaptive radiation. *Nature Ecology & Evolution*, *2*(7), 1128–1138. https://doi.org/10.1038/s41559-018-0581-8

Martin, M. (2013). Cutadapt removes adapter sequences from high-throughput sequencing reads. *EMBnet.Journal*, *17*(1), 10–12.

Martin, S. H., Davey, J. W., & Jiggins, C. D. (2015). Evaluating the Use of ABBA–BABA Statistics to Locate Introgressed Loci. *Molecular Biology and Evolution*, *32*(1), 244–257. https://doi.org/10.1093/molbev/msu269

Masters, P. M. (2006). Holocene sand beaches of southern California: ENSO forcing and coastal processes on millennial scales. *Palaeogeography, Palaeoclimatology, Palaeoecology*, *232*(1), 73–95.

Mather, N., Traves, S. M., & Ho, S. Y. W. (2020). A practical introduction to sequentially Markovian coalescent methods for estimating demographic history from genomic data. *Ecology and Evolution*, *10*(1), 579–589. https://doi.org/10.1002/ece3.5888

Mattern, M. Y. (2004). Molecular phylogeny of the Gasterosteidae: The importance of using multiple genes. *Molecular Phylogenetics and Evolution*, *30*(2), 366–377.

Mazet, O., Rodríguez, W., Grusea, S., Boitard, S., & Chikhi, L. (2016). On the importance of being structured: Instantaneous coalescence rates and human evolution—Lessons for ancestral population size inference? *Heredity*, *116*(4), 362–371. https://doi.org/10.1038/hdy.2015.104

Mazzarella, A. B., Boessenkool, S., Østbye, K., Vøllestad, L. A., & Trucchi, E. (2016). Genomic signatures of the plateless phenotype in the threespine stickleback. *Ecology and Evolution*, *6*(10), 3161–3173. https://doi.org/10.1002/ece3.2072

McKenna, A., Hanna, M., Banks, E., Sivachenko, A., Cibulskis, K., Kernytsky, A., Garimella, K., Altshuler, D., Gabriel, S., Daly, M., & DePristo, M. A. (2010). The Genome Analysis Toolkit: A MapReduce framework for analyzing next-generation DNA sequencing data. *Genome Research*, *20*(9), 1297–1303. https://doi.org/10.1101/gr.107524.110

McMurdie, P. J., & Holmes, S. (2013). phyloseq: An R Package for Reproducible Interactive Analysis and Graphics of Microbiome Census Data. *PLoS ONE*, *8*(4), e61217. https://doi.org/10.1371/journal.pone.0061217

McMurdie, P. J., & Holmes, S. (2014). Waste Not, Want Not: Why Rarefying Microbiome Data Is Inadmissible. *PLoS Computational Biology*, *10*(4), e1003531. https://doi.org/10.1371/journal.pcbi.1003531

Meier, J. I., Marques, D. A., Wagner, C. E., Excoffier, L., & Seehausen, O. (2018). Genomics of Parallel Ecological Speciation in Lake Victoria Cichlids. *Molecular Biology and Evolution*, *35*(6), 1489–1506. https://doi.org/10.1093/molbev/msy051

Meirmans, P. G. (2020). GenoDive version 3.0: Easy-to-use software for the analysis of genetic data of diploids and polyploids. *Molecular Ecology Resources*, *20*(4), 1126–1131. https://doi.org/10.1111/1755-0998.13145

Mendenhall, W. C. (1908). Ground waters and irrigation enterprises in the foothill belt, southern California. In *Water Supply Paper* (No. 219; Water Supply Paper, p. 180). https://doi.org/10.3133/wsp219

Miller, R. R. (1961). *Man and the changing fish fauna of the American southwest*. *46*, 365–404.

Miller, R. R., & Hubbs, C. L. (1969). Systematics of Gasterosteus aculeatus, with Particular Reference to Intergradation and Introgression along the Pacific Coast of North America: A Commentary on a Recent Contribution. *Copeia*, *1*(1), 52–69.

Minckley, W. L., & Marsh, P. C. (2009). *Inland Fishes of the Greater Southwest: Chronicle of a Vanishing Biota*. University of Arizona Press.

Morris, M. R., Bowles, E., Allen, B. E., Jamniczky, H. A., & Rogers, S. M. (2018). Contemporary ancestor? Adaptive divergence from standing genetic variation in Pacific marine threespine stickleback. *BMC Evolutionary Biology*, *18*(1), 1–21.

Mount, J. F. (1995). *California Rivers and Streams: The Conflict Between Fluvial Process and Land Use*. University of California Press.

Moyle, P. B. (2002). *Inland Fishes of California: Revised and Expanded*. University of California Press.

Moyle, P. B., Quiñones, R. M., Katz, J. V., & Weaver, J. (2015). *Fish species of special concern in California* (p. 842). California Department of Fish and Wildlife. https://wildlife.ca.gov/Conservation/SSC/Fishes

Nadachowska-Brzyska, K., Burri, R., Smeds, L., & Ellegren, H. (2016). PSMC analysis of effective population sizes in molecular ecology and its application to black-and-white Ficedula flycatchers. *Molecular Ecology*, *25*(5), 1058–1072. https://doi.org/10.1111/mec.13540

Nath, S., Shaw, D. E., & White, M. A. (2020). Improved contiguity of the threespine stickleback genome using long-read sequencing. *BioRxiv*, 2020.06.30.170787. https://doi.org/10.1101/2020.06.30.170787

Nico, L., Fuller, P., & Neilson, M. (2022). *Acanthogobius flavimanus (Temminck and Schlegel, 1845)*. U.S. Geological Survey, Nonindigenous Aquatic Species Database. https://nas.er.usgs.gov/queries/FactSheet.aspx?speciesID=707

Noor, M. a. F., & Bennett, S. M. (2009). Islands of speciation or mirages in the desert? Examining the role of restricted recombination in maintaining species. *Heredity*, *103*(6), Article 6. https://doi.org/10.1038/hdy.2009.151

O'Brown, N. M., Summers, B. R., Jones, F. C., Brady, S. D., & Kingsley, D. M. (2015a). A recurrent regulatory change underlying altered expression and Wnt response of the stickleback armor plates gene EDA. *Elife, 4*, e05290.

O'Brown, N. M., Summers, B. R., Jones, F. C., Brady, S. D., & Kingsley, D. M. (2015b). A recurrent regulatory change underlying altered expression and Wnt response of the stickleback armor plates gene EDA. *ELife*, *4*(4), 1–17. https://doi.org/10.7554/eLife.05290

Oksanen, J., Blanchet, F. G., Friendly, M., Kindt, R., Legendre, P., McGlinn, D., Minchin, P. R., O'Hara, R. B., Simpson, G. L., Solymos, P., Stevens, M. H. H., Szoecs, E., & Wagner, H. (2019). *vegan: Community Ecology Package* (2.5-6). https://CRAN.R-project.org/package=vegan

Orsi, J. P. (2004). *Hazardous metropolis: Flooding and urban ecology in Los Angeles* [Ph.D., The University of Wisconsin - Madison]. https://www.proquest.com/docview/304537402/abstract/68D98B12099C4AE4PQ/1

Owen, L. A., Finkel, R. C., Minnich, R. A., & Perez, A. E. (2003). Extreme southwestern margin of late Quaternary glaciation in North America: Timing and controls. *Geology*, *31*(8), 729–732.

Paccard, A., Wasserman, B. A., Hanson, D., Astorg, L., Durston, D., Kurland, S., Apgar, T. M., El-Sabaawi, R. W., Palkovacs, E. P., & Hendry, A. P. (2018). Adaptation in temporally variable environments: Stickleback armor in periodically breaching bar-built estuaries. *Journal of Evolutionary Biology*, *31*(5), 735–752.

Patterson, N., Moorjani, P., Luo, Y., Mallick, S., Rohland, N., Zhan, Y., Genschoreck, T., Webster, T., & Reich, D. (2012). Ancient Admixture in Human History. *Genetics, 192*(3), 1065–1093. https://doi.org/10.1534/genetics.112.145037

Pawlowski, J., Bruce, K., Panksep, K., Aguirre, F. I., Amalfitano, S., Apothéloz-Perret-Gentil, L., Baussant, T., Bouchez, A., Carugati, L., Cermakova, K., Cordier, T., Corinaldesi, C., Costa, F. O., Danovaro, R., Dell'Anno, A., Duarte, S., Eisendle, U., Ferrari, B. J. D., Frontalini, F., … Fazi, S. (2022). Environmental DNA metabarcoding for benthic monitoring: A review of sediment sampling and DNA extraction methods. *Science of The Total Environment*, *818*, 151783. https://doi.org/10.1016/j.scitotenv.2021.151783

Peichel, C. L., & Marques, D. A. (2017). The genetic and molecular architecture of phenotypic diversity in sticklebacks. *Philosophical Transactions of the Royal Society B: Biological Sciences, 372*(1713), 20150486. https://doi.org/10.1098/rstb.2015.0486

Perkins, T. L., Clements, K., Baas, J. H., Jago, C. F., Jones, D. L., Malham, S. K., & McDonald, J. E. (2014). Sediment Composition Influences Spatial Variation in the Abundance of Human Pathogen Indicator Bacteria within an Estuarine Environment. *PLoS ONE, 9*(11), e112951. https://doi.org/10.1371/journal.pone.0112951

Pickrell, J., & Pritchard, J. (2012). Inference of population splits and mixtures from genome-wide allele frequency data. *Nature Precedings*, 1–1.

Pietramellara, G., Ascher, J., Borgogni, F., Ceccherini, M. T., Guerri, G., & Nannipieri, P. (2009). Extracellular DNA in soil and sediment: Fate and ecological relevance. *Biology and Fertility of Soils, 45*(3), 219–235. https://doi.org/10.1007/s00374-008-0345-8

Pilliod, D. S., Goldberg, C. S., Arkle, R. S., Waits, L. P., & Richardson, J. (2013). Estimating occupancy and abundance of stream amphibians using environmental DNA from filtered

water samples. *Canadian Journal of Fisheries and Aquatic Sciences*, *70*(8), 1123–1130. https://doi.org/10.1139/cjfas-2013-0047

Port, J. A., O'Donnell, J. L., Romero-Maraccini, O. C., Leary, P. R., Litvin, S. Y., Nickols, K. J., Yamahara, K. M., & Kelly, R. P. (2016). Assessing vertebrate biodiversity in a kelp forest ecosystem using environmental DNA. *Molecular Ecology*, *25*(2), 527–541. https://doi.org/10.1111/mec.13481

Pringle, C. M., Freeman, M. C., & Freeman, B. J. (2000). Regional Effects of Hydrologic Alterations on Riverine Macrobiota in the New World: Tropical-Temperate Comparisons: The massive scope of large dams and other hydrologic modifications in the temperate New World has resulted in distinct regional trends of biotic impoverishment. While neotropical rivers have fewer dams and limited data upon which to make regional generalizations, they are ecologically vulnerable to increasing hydropower development and biotic patterns are emerging. *BioScience*, *50*(9), 807–823. https://doi.org/10.1641/0006-3568(2000)050[0807:REOHAO]2.0.CO;2

Pritchard, J. K., Stephens, M., & Donnelly, P. (2000). Inference of Population Structure Using Multilocus Genotype Data. *Genetics*, *155*(2), 945 LP – 959.

Propst, D. L., Williams, J. E., Bestgen, K. R., & Hoagstrom, C. W. (2021). *Standing between Life and Extinction: Ethics and Ecology of Conserving Aquatic Species in North American Deserts*. University of Chicago Press.

R Core Team. (2018a). *R: A language and environment for statistical computing*. R Foundation for Statistical Computing. https://www.R-project.org/

R Core Team. (2018b). *R: A language and environment for statistical computing*. R Foundation for Statistical Computing. https://www.R-project.org/

Raj, A., Stephens, M., & Pritchard, J. K. (2014). FastSTRUCTURE: Variational Inference of Population Structure in Large SNP Data Sets. *Genetics*, *197*(2), 573–589. https://doi.org/10.1534/genetics.114.164350

Ravinet, M., Faria, R., Butlin, R. K., Galindo, J., Bierne, N., Rafajlović, M., Noor, M. a. F., Mehlig, B., & Westram, A. M. (2017). Interpreting the genomic landscape of speciation: A road map for finding barriers to gene flow. *Journal of Evolutionary Biology*, *30*(8), 1450–1477. https://doi.org/10.1111/jeb.13047

Ravinet, M., Prodöhl, P. A., & Harrod, C. (2013). Parallel and nonparallel ecological, morphological and genetic divergence in lake–stream stickleback from a single catchment. *Journal of Evolutionary Biology*, *26*(1), 186–204.

Rees, H. C., Maddison, B. C., Middleditch, D. J., Patmore, J. R. M., & Gough, K. C. (2014). REVIEW: The detection of aquatic animal species using environmental DNA – a review of eDNA as a survey tool in ecology. *Journal of Applied Ecology*, *51*(5), 1450–1459. https://doi.org/10.1111/1365-2664.12306

Reid, K., Bell, M. A., & Veeramah, K. R. (2021). Threespine Stickleback: A Model System For Evolutionary Genomics. *Annual Review of Genomics and Human Genetics*, *22*(1), 357–383. https://doi.org/10.1146/annurev-genom-111720-081402

Reimchen, T. E. (1994). Predators and evolution in threespine stickleback. In M. A. Bell & S. A. Foster (Eds.), *Evolution of the threespine stickleback* (pp. 240–273). Oxford University Press.

Reimchen, T. E. (2000). Predator handling failures of lateral plate morphs in Gasterosteus aculeatus: Functional implications for the ancestral plate condition. *Behaviour*, *137*(7–8), 1081–1096. https://doi.org/10.1163/156853900502448

Ricciardi, A., & Rasmussen, J. B. (1999). Extinction Rates of North American Freshwater Fauna. *Conservation Biology*, *13*(5), 1220–1222. https://doi.org/10.1046/j.1523-1739.1999.98380.x

Richmond, J. Q., Backlin, A. R., Galst-Cavalcante, C., O'Brien, J. W., & Fisher, R. N. (2018). Loss of dendritic connectivity in southern California's urban riverscape facilitates decline of an endemic freshwater fish. *Molecular Ecology*, *27*(2), 369–386. https://doi.org/10.1111/mec.14445

Richmond, J. Q., Jacobs, D. K., Backlin, A. R., Swift, C. C., Dellith, C., & Fisher, R. N. (2015). Ephemeral stream reaches preserve the evolutionary and distributional history of threespine stickleback in the Santa Clara and Ventura River watersheds of southern California. *Conservation Genetics*, *16*(1), 85–101. https://doi.org/10.1007/s10592-014-0643-7

Robinson, J. A., Brown, C., Kim, B. Y., Lohmueller, K. E., & Wayne, R. K. (2018). Purging of Strongly Deleterious Mutations Explains Long-Term Persistence and Absence of Inbreeding Depression in Island Foxes. *Current Biology*, *28*(21), 3487-3494.e4. https://doi.org/10.1016/j.cub.2018.08.066

Robson, H. L. A., Noble, T. H., Saunders, R. J., Robson, S. K. A., Burrows, D. W., & Jerry, D. R. (2016). Fine-tuning for the tropics: Application of eDNA technology for invasive fish detection in tropical freshwater ecosystems. *Molecular Ecology Resources*, *16*(4), 922–932. https://doi.org/10.1111/1755-0998.12505

Rollins, J. L. (2017). Body-size and growth-rate divergence among populations of threespine stickleback (*Gasterosteus aculeatus*) in Cook Inlet, Alaska, USA. *Canadian Journal of Zoology*, *95*(11), 877–884. https://doi.org/10.1139/cjz-2017-0092

RStudio Team. (2020a). *RStudio: Integrated Development for R*. RStudio, PBC. http://www.rstudio.com/

RStudio Team. (2020b). *RStudio: Integrated Development for R*. RStudio, PBC. http://www.rstudio.com/

Ruiz-Campos, G., Camarena-Rosales, F., González-Acosta, A. F., Maeda-Martinez, A. M., García de León, F. J., Varela-Romero, A., & Andreu-Soler, A. (2014). Estatus actual de conservación de seis especies de peces dulceacuícolas de la península de Baja California, México. *Revista Mexicana de Biodiversidad*, *85*(4), 1235–1248. https://doi.org/10.7550/rmb.43747

Ruiz-Campos, G., & González-Acosta, A. F. (in press). *Peces de las aguas continentales de Baja California, México*. Universidad Autónoma de Baja California.

Sales, N. G., Wangensteen, O. S., Carvalho, D. C., & Mariani, S. (2019). Influence of preservation methods, sample medium and sampling time on eDNA recovery in a neotropical river. *Environmental DNA*, *1*(2), edn3.14. https://doi.org/10.1002/edn3.14

Samuels, D. C., Wang, J., Ye, F., He, J., Levinson, R. T., Sheng, Q., Zhao, S., Capra, J. A., Shyr, Y., Zheng, W., & Guo, Y. (2016). Heterozygosity Ratio, a Robust Global Genomic Measure of Autozygosity and Its Association with Height and Disease Risk. *Genetics*, *204*(3), 893–904. https://doi.org/10.1534/genetics.116.189936

Santoro, A. E., Dupont, C. L., Richter, R. A., Craig, M. T., Carini, P., McIlvin, M. R., Yang, Y., Orsi, W. D., Moran, D. M., & Saito, M. A. (2015). Genomic and proteomic characterization of "Candidatus Nitrosopelagicus brevis": An ammonia-oxidizing archaeon from the open ocean. *Proceedings of the National Academy of Sciences*, *112*(4), 1173–1178. https://doi.org/10.1073/pnas.1416223112

Sard, N. M., Herbst, S. J., Nathan, L., Uhrig, G., Kanefsky, J., Robinson, J. D., & Scribner, K. T. (2019). Comparison of fish detections, community diversity, and relative abundance using environmental DNA metabarcoding and traditional gears. *Environmental DNA*, *1*(4), 368–384. https://doi.org/10.1002/edn3.38

Schaarschmidt, F., & Gerhard, D. (2019). *PairwiseCI: Confidence Intervals for Two Sample Comparisons* (0.1-27). https://CRAN.R-project.org/package=pairwiseCI

Schiffels, S., & Durbin, R. (2014). Inferring human population size and separation history from multiple genome sequences. *Nature Genetics*, *46*(8), 919–925. https://doi.org/10.1038/ng.3015

Schluter, D. (2000). *The Ecology of Adaptive Radiation*. OUP Oxford.

Schluter, D., & Nagel, L. M. (1995). Parallel Speciation by Natural Selection. *The American Naturalist*, *146*(2), 292–301. https://doi.org/10.1086/285799

Schnell, I. B., Bohmann, K., & Gilbert, M. T. P. (2015). Tag jumps illuminated—Reducing sequence-to-sample misidentifications in metabarcoding studies. *Molecular Ecology Resources*, *15*(6), 1289–1303. https://doi.org/10.1111/1755-0998.12402

Schrider, D. R., Shanku, A. G., & Kern, A. D. (2016). Effects of linked selective sweeps on demographic inference and model selection. *Genetics*, *204*(3), 1207–1223.

SCWRP. (2018). *Wetlands on the Edge: The Future of Southern California's Wetlands: Regional Strategy 2018* (p. 142). California State Coastal Conservancy. scwrp.databasin.org

Sekar, R., Kaczmarsky, L. T., & Richardson, L. L. (2009). Effect of Freezing on PCR Amplification of 16S rRNA Genes from Microbes Associated with Black Band Disease

of Corals. *Applied and Environmental Microbiology*, *75*(8), 2581–2584. https://doi.org/10.1128/AEM.01500-08

Shaffer, H. B., Fellers, G. M., Randal Voss, S., Oliver, J. C., & Pauly, G. B. (2004). Species boundaries, phylogeography and conservation genetics of the red-legged frog (Rana aurora/draytonii) complex. *Molecular Ecology*, *13*(9), 2667–2677. https://doi.org/10.1111/j.1365-294X.2004.02285.x

Shanfelter, A. F., Archambeault, S. L., & White, M. A. (2019). Divergent fine-scale recombination landscapes between a freshwater and marine population of threespine stickleback fish. *Genome Biology and Evolution*, *11*(6), 1573.

Sharp, R. P., Allen, C. R., & Meier, M. F. (1959). Pleistocene glaciers on southern California mountains. *American Journal of Science*, *257*(2), 81–94.

Shirazi, S., Meyer, R. S., & Shapiro, B. (2021). Revisiting the effect of PCR replication and sequencing depth on biodiversity metrics in environmental DNA metabarcoding. *Ecology and Evolution*, *11*(22), 15766–15779. https://doi.org/10.1002/ece3.8239

Smart, A. S., Weeks, A. R., Rooyen, A. R., Moore, A., McCarthy, M. A., & Tingley, R. (2016). Assessing the cost-efficiency of environmental DNA sampling. *Methods in Ecology and Evolution*, *7*(11), 1291–1298. https://doi.org/10.1111/2041-210X.12598

Stein, E. D., Cayce, K., Salomon, M., Bram, D. L., De Mello, D., Grossinger, R., & Dark, S. (2014). *Wetlands of the Southern California Coast: Historical Extent and Change Over Time* (SFEI Report 720; SCCWRP Technical Report 826; p. 58). Southern California Coastal Water Research Project and San Francisco Estuary Institute. https://www.caltsheets.org/socal/download.html

Stephenson, J. R., & Calcarone, G. M. (1999). *Southern California Mountains and Foothills Assessment: Habitat and Species Conservation Issues*. U.S. Department of Agriculture, Forest Service, Pacific Southwest Research Station.

Stiller, J., da Fonseca, R. R., Alfaro, M. E., Faircloth, B. C., Wilson, N. G., & Rouse, G. W. (2021). Using ultraconserved elements to track the influence of sea-level change on leafy seadragon populations. *Molecular Ecology*, *30*(6), 1364–1380.

Suomalainen, L.-R., Reunanen, H., Ijäs, R., Valtonen, E. T., & Tiirola, M. (2006). Freezing Induces Biased Results in the Molecular Detection of *Flavobacterium columnare*. *Applied and Environmental Microbiology*, *72*(2), 1702–1704. https://doi.org/10.1128/AEM.72.2.1702-1704.2006

Swift, C. C., Haglund, T. R., Ruiz, M., & Fisher, R. N. (1993a). The Status and Distribution of the Freshwater Fishes of Southern California. *Bulletin of the Southern California Academy of Sciences*, *92*(3), 101–167.

Swift, C. C., Haglund, T. R., Ruiz, M., & Fisher, R. N. (1993b). The Status and Distribution of the Freshwater Fishes of Southern California. *Bulletin of the Southern California Academy of Sciences*, *92*(3), 101–167.

Swift, C. C., Spies, B., Ellingson, R. A., & Jacobs, D. K. (2016). A New Species of the Bay Goby Genus Eucyclogobius, Endemic to Southern California: Evolution, Conservation, and Decline. *PloS One*, *11*(7), e0158543. https://doi.org/10.1371/journal.pone.0158543

Taberlet, P., Coissac, E., Pompanon, F., Brochmann, C., & Willerslev, E. (2012). Towards next-generation biodiversity assessment using DNA metabarcoding: NEXT-GENERATION DNA METABARCODING. *Molecular Ecology*, *21*(8), 2045–2050. https://doi.org/10.1111/j.1365-294X.2012.05470.x

Taberlet, P., Prud'Homme, S. M., Campione, E., Roy, J., Miquel, C., Shehzad, W., Gielly, L., Rioux, D., Choler, P., Clément, J.-C., Melodelima, C., Pompanon, F., & Coissac, E. (2012). Soil sampling and isolation of extracellular DNA from large amount of starting material suitable for metabarcoding studies: EXTRACTION OF EXTRACELLULAR DNA FROM SOIL. *Molecular Ecology*, *21*(8), 1816–1820. https://doi.org/10.1111/j.1365-294X.2011.05317.x

Takahara, T., Minamoto, T., & Doi, H. (2015). Effects of sample processing on the detection rate of environmental DNA from the Common Carp (Cyprinus carpio). *Biological Conservation*, *183*, 64–69. https://doi.org/10.1016/j.biocon.2014.11.014

Takahata, N., & Nei, M. (1985). Gene Genealogy and Variance of Interpopulational Nucleotide Differences. *Genetics*, *110*(2), 325–344.

Thomsen, P. F., & Willerslev, E. (2015). Environmental DNA – An emerging tool in conservation for monitoring past and present biodiversity. *Biological Conservation*, *183*, 4–18. https://doi.org/10.1016/j.biocon.2014.11.019

Torti, A., Lever, M. A., & Jørgensen, B. B. (2015). Origin, dynamics, and implications of extracellular DNA pools in marine sediments. *Marine Genomics*, *24*, 185–196. https://doi.org/10.1016/j.margen.2015.08.007

Tsuji, S., Takahara, T., Doi, H., Shibata, N., & Yamanaka, H. (2019). The detection of aquatic macroorganisms using environmental DNA analysis—A review of methods for collection, extraction, and detection. *Environmental DNA*, *1*(2), 99–108. https://doi.org/10.1002/edn3.21

Turba, R., Richmond, J. Q., Fitz-Gibbon, S., Morselli, M., Fisher, R. N., Swift, C. C., Ruiz-Campos, G., Backlin, A. R., Dellith, C., & Jacobs, D. K. (2022). Genetic structure and

historic demography of endangered unarmoured threespine stickleback at southern latitudes signals a potential new management approach. *Molecular Ecology*, mec.16722. https://doi.org/10.1111/mec.16722

Turner, C. R., Barnes, M. A., Xu, C. C. Y., Jones, S. E., Jerde, C. L., & Lodge, D. M. (2014). Particle size distribution and optimal capture of aqueous macrobial eDNA. *Methods in Ecology and Evolution*, *5*(7), 676–684. https://doi.org/10.1111/2041-210X.12206

Turner, C. R., Miller, D. J., Coyne, K. J., & Corush, J. (2014). Improved methods for capture, extraction, and quantitative assay of environmental DNA from Asian bigheaded carp (hypophthalmichthys spp.). *PLoS ONE*, *9*(12), 1–20. https://doi.org/10.1371/journal.pone.0114329

Turner, C. R., Uy, K. L., & Everhart, R. C. (2015). Fish environmental DNA is more concentrated in aquatic sediments than surface water. *Biological Conservation*, *183*, 93–102. https://doi.org/10.1016/j.biocon.2014.11.017

USFWS. (2021). *Unarmored threespine stickleback (Gasterosteus aculeatus williamsoni) 5-Year Review: Summary and Evaluation*. Ventura Fish and Wildlife Office. https://ecos.fws.gov/ecp/species/7002

van der Loos, L. M., & Nijland, R. (2021). Biases in bulk: DNA metabarcoding of marine communities and the methodology involved. *Molecular Ecology*, *30*(13), 3270–3288. https://doi.org/10.1111/mec.15592

Van Wormer, S. R. (1991). A History of Flood Control in the Los Angeles County Drainage Area. *Southern California Quarterly*, *73*(1), 55–94. https://doi.org/10.2307/41171560

Wasserman, B. A., Paccard, A., Apgar, T. M., Des Roches, S., Barrett, R. D. H., Hendry, A. P., & Palkovacs, E. P. (2020). Ecosystem size shapes antipredator trait evolution in estuarine threespine stickleback. *Oikos*, *129*(12), 1795–1806. https://doi.org/10.1111/oik.07482

WeatherSpark.com. (n.d.). *Historical Weather Summer 2018 at Point Mugu Naval Air Warfare Center*. WeatherSpark.Com. Retrieved June 8, 2022, from https://weatherspark.com/h/s/145310/2018/1/Historical-Weather-Summer-2018-at-Point-Mugu-Naval-Air-Warfare-Center;-California;-United-States#Figures-Rainfall

Weir, B. S., & Cockerham, C. C. (1984). Estimating F-Statistics for the Analysis of Population Structure. *Evolution*, *38*(6), 1358–1370. https://doi.org/10.2307/2408641

Wickham, H. (2016). *ggplot2: Elegant Graphics for Data Analysis*.

Williams, K. E., Huyvaert, K. P., & Piaggio, A. J. (2017). Clearing muddied waters: Capture of environmental DNA from turbid waters. *PLOS ONE*, *12*(7), e0179282. https://doi.org/10.1371/journal.pone.0179282

Wolf, J. B. W., & Ellegren, H. (2017). Making sense of genomic islands of differentiation in light of speciation. *Nature Reviews Genetics*, *18*(2), Article 2. https://doi.org/10.1038/nrg.2016.133

Zheng, X., Levine, D., Shen, J., Gogarten, S. M., Laurie, C., & Weir, B. S. (2012). A high-performance computing toolset for relatedness and principal component analysis of SNP data. *Bioinformatics*, *28*(24), 3326–3328.

Zinger, L., Lionnet, C., Benoiston, A.-S., Donald, J., Mercier, C., & Boyer, F. (2020). *metabaR: An R package for the evaluation and improvement of DNA metabarcoding data quality* [Preprint]. Bioinformatics. https://doi.org/10.1101/2020.08.28.271817

Chapter 3 . To freeze or to scoop? Dealing with the turbid waters of California's coastal lagoons

Abstract

Coastal lagoons are an important habitat for endemic and threatened species in California that have suffered impacts from urbanization and increased drought. Environmental DNA has been promoted as a way to aid in the monitoring of biological communities, but much remains to be understood on the biases introduced by different protocols meant to overcome challenges presented by unique systems under study. Turbid water is one methodologic challenge to eDNA recovery in these systems as it quickly clogs filters, preventing timely processing of samples. We investigated biases in community composition produced by two solutions to overcome slow filtration due to turbidity: freezing of water prior to filtration (for storage purposes and long-term processing), and use of sediment (as opposed to water samples). Bias assessments of community composition in downstream eDNA analysis was conducted for two sets of primers, 12S (fish) and 16S (bacteria and archaea). Our results show that freezing water prior to filtration had no effects on community composition for either primer, even when using a filter of larger pore size (3 μm), and therefore it is a viable approach in this system for comparison of water borne fish, bacteria and archaea. However, the 16S primer showed significantly different community composition in sediments compared to water samples, although still recovering eDNA of organisms from the water column. Sediment sample replicates were heterogeneous, and therefore increasing the number of replicates would be recommended for similar habitats.

Introduction

Coastal lagoons in California are the numerically dominant form of coastal wetland (D. K. Jacobs et al., 2011b; Stein et al., 2014) and are important in many other Mediterranean climates and subtropical environments. These lagoons are characterized by seasonal and episodic breaching (opening of the lagoon to the sea, usually by stream flow) and closure (isolation of the lagoon by a high sandbar), which provide a suite of ecological services: from groundwater infiltration to support of unique biodiversity (Ballard et al., n.d.). This system serves as important habitat and nursery for endemic and endangered fishes and amphibians, such as the steelhead (*Oncorhynchus mykiss*), red-legged frog (*Rana aurora draytonii*), and the tidewater goby (*Eucyclogobius newberryi*) (Earl et al., 2010; Shaffer et al., 2004; Swift et al., 1993b, 2016). Thus, California lagoons are spatially and temporally variable systems with unique biodiversity and biodiversity assessment challenges.

Coastal lagoons have been drastically reduced in numbers along the California coastline, driven mostly by the impact of coastal land use for transport structures, agriculture, and development. These are further exacerbated by ongoing changes in the hydrological cycles due to climate change (SCWRP, 2018). While these sites are critical for endangered species conservation, they are also subject to frequent invasion and their response to environmental variation is poorly documented. However, monitoring of this habitat can be limited by a variety of issues, ranging from limited human power and access to challenges driven by the natural complexity and dynamism of these lagoons.

The use of environmental DNA (eDNA) has been advocated as an alternative for monitoring communities and target species (Thomsen & Willerslev, 2015), and can overcome and

complement certain field limitations from traditional methods (e.g. seining, trapping). On-site collection can be relatively fast, and therefore allow field workers to cover more ground. It can also recover the DNA signal of species that are rare, cryptic and/or hard to capture by traditional methods, and being non-intrusive, it offers an alternative when working with endangered species for which permits are necessary (Deiner et al., 2017; Dejean et al., 2012; Sard et al., 2019). In addition, metabarcoding approaches allow the investigation of multiple species from a single collection (Taberlet, Coissac, et al., 2012).

Nevertheless, it is important to recognize that this approach also brings its own limitations and biases (van der Loos & Nijland, 2021). In some circumstances, eDNA sampling can be more expensive than traditional, more established methods (Smart et al., 2016). Since there are no voucher specimens from collections, contamination is a major issue that needs to be addressed early on, following best practices in the field (Goldberg et al., 2016). The lack of voucher specimens also leads to an overdependence on the use of barcodes and genetic databases for taxonomic identification, which introduces another set of biases, from misidentification to lack of species representation (Taberlet, Coissac, et al., 2012). Other challenges arise from the non-universality of sampling methods and downstream processing, with the probability of detection varying depending on the species and their density, as well as the type of environment, which affects rates of DNA degradation (Deiner et al., 2015; Rees et al., 2014; Williams et al., 2017).

Coastal lagoons can vary in their environmental qualities quite drastically. One major challenge is the high and variable turbidity of the water. High turbidity usually occurs when lagoons are closed to the ocean by a sandbar and driven by organic and inorganic matter. In this case, filtering water on-site becomes a problem. Filtration is a widespread method for handling water samples (Laramie et al., 2015; Tsuji et al., 2019). Set volumes of water are run through a

small filter to concentrate DNA before extractions. However, high concentration of fine sediment or organic matter in water quickly obstructs filters, making the filtration process time-consuming (although it could actually aid recovery by binding DNA to suspended particles: Kumar et al., 2022; Liang & Keeley, 2013; Torti et al., 2015).

To overcome this issue, some stakeholders have relied on a tiered filtration step (prefiltration) to reduce particles and avoid clogging filters (Tsuji et al., 2019), but this approach increases costs, labor and opportunities for potential contamination (J. Li et al., 2018; Majaneva et al., 2018; Robson et al., 2016). The use of filters of bigger pore sizes, up to 20 µm, has been previously tested and in cases of turbid waters is generally preferred, but requires filtering larger volumes of water to capture the same amount of DNA recovered in smaller pore size filters (Robson et al., 2016; Turner, Barnes, et al., 2014).

Freezing water for storage purposes prior to filtration can mitigate the issue of slow filtration in the field and allow it to be done in batches in the laboratory at a later time, but this type of sample storage might introduce bias on DNA capture and community composition (Kwambana et al., 2011; Sekar et al., 2009). Cells can disrupt and extrude their DNA in the environment, an issue that has been demonstrated in certain cases (e.g. Suomalainen et al., 2006), which would then make it easier for it to pass through the filter pores. In the case of turbid waters, increasing the pore size of filters to speed the filtration process could worsen this problem by letting DNA in solution flow through the pores more easily.

When dealing with turbid waters, some stakeholders have opted to use the centrifugation approach (e.g. Williams et al., 2017). Extracellular DNA (i.e. DNA not contained within a cell wall) can be bound to particles (Torti et al., 2015) and consequently be captured and detected more easily following centrifugation of particles into pellets. However, the amount of water used is

limited by centrifuge size, usually around 15-30 mL per replicate (Doi et al., 2017; Ficetola et al., 2008), which might limit recovery of diluted DNA (Deiner et al., 2015).

Processing sediment samples may be preferable to processing highly turbid water samples. However, it is important to understand how DNA recovery from these different media compare to one another. Turner et al. (2015) and Perkins et al. (2014) have shown that sediment can have a higher concentration of fish eDNA and some bacteria, respectively. This could be related to the organic-particle binding and sinking properties, and a longer DNA persistence in sediment compared to water samples. However, as is the case with water samples, there is no consensus on the rate of degradation of eDNA in soil and sediment (Dell'Anno & Corinaldesi, 2004; Levy-Booth et al., 2007; Torti et al., 2015), and this will depend on multiple local biotic and abiotic factors. In addition, biological communities will naturally differ between water column and sediments, even though we expect some level of overlap due to both DNA sinking and suspension.

Previous work have been done comparing different approaches to processing eDNA, such as filtration and storage methods (Hinlo et al., 2017; Takahara et al., 2015), including some work on turbid waters (Kumar et al., 2022; Robson et al., 2016; Williams et al., 2017), and comparisons between water and sediment eDNA recovery (Sales et al., 2019; Turner et al., 2015). But results have been contradictory, or limited to looking at just DNA concentration, or at a single targeted species.

The goal of the present study is to compare how freezing water prior to filtration and using water versus sediment samples induce and/or exacerbate biases in taxa detection for a set of universal primers targeting different biological communities–12S (fish) and 16S (bacteria and archaea)—in coastal lagoons. By understanding the biases introduced when processing environmental samples, we will be able to inform decisions regarding experimental design for

monitoring such a dynamic and challenging habitat, which has invaluable importance for the maintenance of ecosystem services for both wild and urban populations. We expect these results will be of interest relative to eDNA sampling in other aquatic systems as well.

Material and Methods

Site - Topanga Lagoon

To determine the variability of species detection for each protocol, water and sediment samples were collected from a south-facing coastal lagoon in southern California, located in Malibu, a stretch of coast that runs from Santa Monica to Point Mugu. This lagoon is part of the Topanga State Park and is currently undergoing plannings for restoration. It is the only lagoon on this stretch of coast that still harbors a stable population of tidewater goby (*E. newberryi*), a federally endangered species, and is relatively less impacted than other lagoons in the same region. The endangered southern steelhead trout (*O. mykiss*) is also found in this system during anadromy when the lagoon is breached. Due to the presence of these species, Topanga lagoon has been periodically surveyed by the Jacobs' lab members and collaborators such as researchers at the Resource Conservation District of The Santa Monica Mountains (RCDSMM), and therefore its macrobiota is regularly studied, especially the fish fauna. The lagoon was sampled on September 6th, 2018, at the end of the Summer season, and as is typical of this time of the year, the weather was dry with no record of precipitation since June (WeatherSpark.com, n.d.). The lagoon was closed to the ocean by a sandbar and the water was murky (Figure 3.1), which in the author's experience, such turbidity slowed filtration and easily clogged 0.45 μm cellulose nitrate filters.

Protocols and samples

A sterilized water jug was used to collect a single water sample in the lagoon, at a mid-point between the mouth margin and the road bridge (Figure 3.1). The sample was then placed on ice and brought to the laboratory (~1 hr car ride). This method of "grab-and-hold" has proven to be similarly effective as on-site filtration in a previous study (Pilliod et al., 2013). Once in the laboratory, the total volume was divided in three batches for each treatment: (i) centrifugation followed by filtration of supernatant (5 replicates of 50 mL falcon tube) (Doi et al., 2017); (ii) pre-freezing followed by double filtration (5 replicates of 500 mL Nalgene bottles) (Turner, Miller, et al., 2014); and (iii) no freezing followed by double filtration on the same day of collection (5 replicates of 500 mL Nalgene bottles) (Turner, Miller, et al., 2014).

For the pre-freezing protocol, water bottles were frozen at -20 °C for 3 days before thawing for filtration. Double filtration for both pre-freezing and no-freezing treatments was done through cellulose nitrate filters, firstly on a 3 µm pore size filter, then followed by a 0.45 µm pore size using an adapted vacuum pump in the pre-PCR room of the laboratory (Figure S 3.1). The centrifugation protocol also included a second stage filtration of the supernatant using a 0.45 µm pore size filter. Here, we will focus only on the results from the first filtration step of the water filtration protocol. More details on that are explained further in the Supplementary Information section.

Surficial sediment was collected in triplicates at the same location where water was sampled (5 replicates of triplicate 2 mL cryotubes, 15 tubes total), following instructions as defined by the CALeDNA program (https://ucedna.com/methods-for-researchers). These were also kept on ice during field work and stored in a -80°C freezer upon arrival at the laboratory until DNA

extractions. Results from sediment samples were compared against both filtration protocols: (1) pre-freezing followed by filtration; (2) no freezing followed by filtration.

DNA Extraction

DNA from sediments and filters were extracted following the PowerSoil extraction protocol. Filters were chopped into thin strips before being added to the bead tubes, and sediment triplicates were pooled in small batches to reach 0.25-0.3 g before processing. We used the soil extraction kit on the filters as well to reduce potential PCR inhibition caused by the water turbidity (Kumar et al., 2022), but also to limit the number of variables in the research design by adding another extraction protocol.

Contamination best practices

Care was taken to avoid contamination both in the field and the lab. Before collection, bottles and water jug were cleaned and bleached and then handled with clean gloves on site. Extractions and PCR were done in a separate pre-PCR room. Utensils and bench top were cleaned with 10% bleach, followed by 70% ethanol. Forceps and scissors for handling filters were seared and cleaned with bleach and ethanol after dealing with each sample. PCR reagents were prepared in a clean, PCR-free, positive pressure hood. Sediment samples were collected with new 2 mL cryotubes and following field protocol as recommended by the CALeDNA program. Blanks were made for the field collection, laboratory filtration and PCR (5 blanks in total) and included in the library for sequencing.

Sequencing

Library preparation followed CALeDNA protocols (https://ucedna.com/methods-for-researchers). Metabarcode libraries were generated for bacteria and archaea (16S rRNA), fish (12S rRNA) and metazoans (CO1). Sequences for each primer can be found at Table 1. All libraries consisted of triplicate PCR reactions. PCR products were visualized using gel electrophoresis, and for each barcode, PCR triplicates were pooled by sample. After bead cleaning, all markers were pooled by sample and tagged for sequencing (single indexing). Libraries were pooled and run on a MiSeq SBS Sequencing v3 in a pair-end 2x300 bp format [Technology Center for Genomics & Bioinformatics (TCGB), UCLA] with a target sequencing depth of 25,000 reads/sample/metabarcode. Two sequencing runs were conducted, but the CO1 primer was still below the sequencing depth threshold and therefore its results will not be discussed here (see Figure S 3.2-Figure S 3.3). For each run, our library was pooled with different samples from different collaborators to maximize efficiency of the sequencing run.

Bioinformatics and data pre-processing

Sequence data was bioinformatically processed in Hoffman2, the High Performance Computing cluster at UC Los Angeles, using the Anacapa Toolkit (Curd, Gomer, et al., 2018) with default settings. Briefly, reads are demultiplexed and trimmed for adapters (cutadapt, Martin, 2013) and low-quality reads (FastX Toolkit, *FASTX-Toolkit*, n.d.). Dada2 (Callahan et al., 2016) is used to denoise, dereplicate, merge and remove chimeras, and the resulting clean Amplicon Sequence Variants (ASVs) have their taxonomy assigned using Bowtie2 (Langmead & Salzberg, 2012), matched to a custom reference library (CRUX, Curd, Kandlikar, et al., 2018). Confidence levels are determined by the BLCA algorithm (Gao et al., 2017) to generate a table of best taxonomic

hits, from super-kingdom to species level. The pipeline was designed to process not only paired, but also unmerged and unpaired reads.

Taxonomic tables with a bootstrap confidence cutoff score of 0.6 were used for downstream analyses. Except when noted, all bioinformatic analyses mentioned beyond this point were performed using R v.3.6.2 (R Core Team, 2018b) in RStudio v.1.2.1335 (RStudio Team, 2020b). Decontamination was done separately for each primer set and each run (since the dataset was pooled with different combinations of samples for sequencing). We used the package metabaR (Zinger et al., 2020) to lower tag-jumping and remove contaminants through detection of ASVs whose relative abundance is highest in negative controls. We also ran a modification of the gruinard pipeline (https://github.com/zjgold/gruinard_decon), including only steps 4 (site occupancy modeling) and 5 (dissimilarity between replicates), since previous steps were redundant with the metabaR decontamination steps. Lastly, taxa classified as "Not_found", "Unclassified", "*Canis lupus*", "*Bos taurus*", and "*Homo sapiens*" were removed from the final tables before being merged and used in downstream analyses.

Diversity analysis

We used the laboratory's own sampling record and the Global Biodiversity Information Facility database (Gbif.Org, 2022) to manually check the 12S primer final taxonomic table. The number of species captured by each treatment was visualized using Venn Diagrams (package VennDiagram, Chen, 2018). Species rarefaction curves were made for each metabarcode to inspect the level of species saturation for each protocol replicate. The slope of each curve was calculated using the rareslope function in the vegan package (Oksanen et al., 2019), and the confidence interval for each protocol was calculated using pairwiseCI (Schaarschmidt & Gerhard, 2019) with

confidence level at 95%. Rarefaction curves were plotted using the ggrare function from the ranacapa package (using step = 5).

Differential abundance

The raw dataset was analyzed using DESeq2 to look at differential abundance between protocols (Love et al., 2014). The default testing framework was used (test = "Wald", fitType = "parametric"), which includes the Benjamini-Hochberg multiple inference correction. The sfType option was defined as poscounts since this estimator is able to handle zeros. The log2 fold change of each pairwise comparison for which there were significant differences in abundances was plotted.

Beta diversity

For the beta diversity analysis, samples were standardized by using either the eDNA index (Kelly et al., 2019) or by rarefying them as a way to equalize sequencing effort and minimize stochasticity and bias. For the eDNA index, we followed the Wisconsin double standardization method in the vegan package. The custom_rarefaction function in the R package ranacapa (Kandlikar, 2020) was used to rarefy the dataset with 10 replicates.

For the 12S primer, samples were rarefied to 20 000 reads. Three sediment samples were excluded due to very low read numbers (<100). For the 16S, samples were rarefied to 15,000 and one sediment sample that had ~5,000 reads was excluded. The number of reads per taxa for each protocol replicate was plotted using the phyloseq package (McMurdie & Holmes, 2013), for both the raw and rarefied dataset.

The rarefied dataset followed a Constrained Analysis of Principal Coordinates (CAP) using the capscale function in vegan and Bray-Curtis distance. This ordination method, which can be used with non-Euclidean dissimilarity indices, explains the ordination of assemblage composition based on species abundances. The difference in community composition for each treatment was then analyzed using a PERMANOVA and Bray-Curtis dissimilarity, followed by a pairwise PERMANOVA comparison (all with the vegan package). P-values were adjusted using the FDR (False Discovery Rate) approach.

Results

Sequencing

The first run generated a total of 6 407 371 reads: 3 817 216 reads for the 12S primer, 2,393,627 for 16S, and 196,528 for CO1. In the second run there were a total of 9,088,496 reads: 6,685,673 reads for the 12S metabarcode, 1,904,283 reads for 16S and 498,540 for the CO1. For the 12S and 16S primers, we were able to reach our threshold of 25,000 reads/sample in most cases, while that was not the case for all except one sample of the CO1 primer. Because of this limitation on the number of reads/sample, the CO1 metabarcode will not be discussed further in the main paper (but check the Supplemental Information for more details).

Bioinformatics and data pre-processing

The number of reads per sample after decontamination and combining both runs is illustrated in Figure S 3.3. We manually checked the final taxonomic tables of each separate run for the 12S primer to look for signs of contamination and evaluate how well the bioinformatic decontamination steps worked (metabaR and gruinard). The taxonomic tables for the 12S primer have substantially

less species than the 16S, and the local fish fauna is relatively well known, making the process more tractable.

For the run that was pooled with samples from Palmyra Atoll, the output still retained some tropical reef and pelagic fish and elasmobranch species that are not found in coastal lagoons in California. We can expect that tag-jumping contamination is also present in the other sequencing runs and primers as well. Interestingly, eight out of 28 of those tropical species (ca. 28%) were found exclusively on the sediment samples and not the water samples (e.g. *Acanthurus achilles, Scarus altipinnis, Lutjanus russellii*).

Barplots for both the raw and rarefied dataset (Figure S 3.3-Figure S 3.4, respectively) show that sediment replicates had greater variability amongst themselves, both in number of reads and community composition, compared to the replicates of either water protocols. Water replicates were more consistent within and between protocols, and had an overall higher number of reads than the sediment samples.

Diversity

After the decontamination steps (metabaR and gruinard) and removing specific, uninformative ASVs (as listed above), the total number of species assigned to 12S was 39, distributed in 20 orders and 22 families. Of these 39 species, only four had been previously recorded for the site (Table S 3.1). For 16S, the total number of taxa assigned to species was 2,625, distributed in 45 phyla and 335 families.

We have also noticed some dubious taxonomic assignments. For example, for the 12S primer, we had one hit for *Fundulus diaphanus*, which is a species of killifish native to the northeast of North America. However, the californian species *F. parvipinnis* has been previously

documented in Topanga by lab members sampling at the site. Similarly, there were two hits for *Phoxinus phoxinus*, which has a European distribution with a closely related North American counterpart, *P. eos*, although this species has not been identified in collections from Topanga lagoon. Another dubious identification occurred for two species of *Odontesthes*, *O. incisa* and *O. smitti*, which were among the most abundant hits in our dataset but are native to the southwest Atlantic. These two species, however, are relatives of topsmelt (*Atherinops affinis*), commonly found in coastal lagoons and estuaries in California (Table S 3.1).

The Venn Diagram (Figure 3.2) shows that even though sediment samples had lower numbers of reads overall (Figure S 3.2-Figure S 3.3), they had the highest number of species recovered (12S primer: N=27, 19 unique; 16S primer: N=1,929, 1,178 unique). The species overlap between protocols for the 12S was only 1.2% (n=1), and for the 16S primer it was 3.5% (n=402).

Species rarefaction curves also show that sediment samples are further from reaching saturation compared to water samples, both for 12S and 16S primers (Figure 3.3), although there was more variation between the replicates for the 12S sediment samples. For 12S primer, there is a significant difference in the slope of the species curves between the sediment and no freezing protocols (Figure 3.4), while for 16S, all pairwise comparisons between protocols showed significant differences.

Differential Abundance

For the 12S primer, there was no significant difference between species abundance for any of the protocols' pairwise comparisons. For the 16S primer, there was no significant difference in comparison between the water protocols (pre- and no freezing). However, there were significant

differences in the pairwise comparisons of water samples and sediment samples (Figure 3.5, Table 3.2-Table 3.3). The top five differentially abundant species in the water protocols were representatives of the families Aphanizomenonaceae, Comamonadaceae and Flavobacteriaceae (in both pre- and no freezing); plus Hemiselmidaceae and Geminigeraceae (pre-freezing protocol only). These comprise groups of cyanobacteria (Aphanizomenonaceae) and algae (Hemiselmidaceae and Geminigeraceae), as well as environmental bacteria (Comamonadaceae and Flavobacteriaceae).

The most differentially abundant species found in the sediment were representatives of the families Catenulaceae, Fragilariaceae and an archaea assigned to the Thaumarchaeota phylum (both pre- and no freezing); plus Woeseiaceae and Elphidiidae (no freezing protocol only); and Anaerolineaceae and Desulfobacteraceae (pre-freezing protocol only). These comprise groups of diatoms (Catenulaceae and Fragilariaceae), environmental bacteria (Woeseiaceae, Anaerolineaceae and Desulfobacteraceae) and archaea (Thaumarchaeota), and foraminiferans (Elphidiidae).

Beta diversity

When using the eDNA index, the CAP analysis for the 12S primer showed that many of the species driving the differences in assemblage composition were the tropical species that are coming from the tag-jumping contamination (Figure S 3.5). For example, we see overrepresentation in the sediment samples of *Stegastes nigricans* and *Caranx melampygus*; and in the no freezing water samples, *Sphyraena barracuda*. Nevertheless, we also see some other species that are known to be found in the lagoon, such as the *Eucyclogobius newberryi*, being mostly overrepresented in the water samples compared to the sediments; and *Gila orcutii*, overrepresented in the no freezing

protocol. Two species of dubious taxonomic assignment are also overrepresented in the sediment: *Phoxinus phoxinus* (as discussed in the previous 'Diversity' section); and *Acanthogobius flavimanus*, which is a species of goby native to Asia, but that has been recorded previously in California estuaries (Nico et al., 2022). The PERMANOVA results were not significant ($p = 0.067$).

For the rarefied dataset, the CAP analysis was not able to recover any differences in assemblage composition for the 12S primer for any of the protocols (Figure 3.6). One sediment replicate is driving most of the difference (CAP1=86%) with the overrepresentation of many tropical species, likely tag-jump contaminants. The PERMANOVA results were at the threshold of significance ($p = 0.05$), but the pairwise test was not significant for any protocol comparison (Table 3.2). The lack of significant differences between water and sediment samples could have been driven by the loss of three sediment replicates when rarefying the dataset.

For the rarefied 16S primer dataset, the different protocols showed significant differences in assemblage composition. The first axis explains most of the total variation (CAP1=86%), with the tidewater goby being the most underrepresented in the sediment compared to the water samples, especially in the no freezing protocol (Figure 3.6). Sediment samples were also slightly overrepresented by a few other species compared to water samples. One of them was identified as Candidatus *Nitrosopelagicus brevis*, which is a species of ammonia-oxidizing archaea (Thaumarchaeota) found mainly in the epi- and upper mesopelagic environments of the open oceans (Santoro et al., 2015). There are also two species of *Monomorphina*, (*M. pyrum* and *M. pseudonordstedti*) that belong to the Euglenaceae family, a group of eukaryotic flagellates found in freshwater environments. Lastly, there is *Elphidium williamsoni*, a foraminifera belonging to the family Elphidiidae found in tidal flats of the North Sea. CAP2 is representing the remaining

variation (14%) found between the water protocols, with the most distinguishing species being the *Guillardia theta*, a species of flagellate algae belonging to the family Geminigeraceae, overrepresented in the pre-freezing protocol. The PERMANOVA result was significant for the 16S primer ($p = 0.001$), as well as for all the pairwise comparisons (Table 3.2).

The species represented in the rarefied dataset differ from the ones found when using the eDNA index for the 16S primer. Most of the community assemblage difference (CAP1=85%) is driven by differences between water and sediment samples, with six species being underrepresented in the latter: *Burkholderiales bacterium* TP637, *Curvibacter* sp. UKPF8, beta proteobacterium Mzo1, *Diaphorobacter ruginosibacter*, *Stella humosa* and *Verminephrobacter aporrectodeae*. All of them, with the exception of the last one, *V. aporrectodeae*, were also found as significantly different in the DeSeq2 analysis. The PERMANOVA result was also significant in this case ($p = 0.001$), as well as for all the pairwise comparisons (Table 3.3).

Discussion

Standardized protocols to process eDNA are under development (e.g. Bohmann et al., 2021), but to implement these efficiently it is necessary to compare biases in taxa detection associated with different protocols. Here, we have explored the detection biases in community composition introduced by freezing water samples prior to filtration (for storage purposes), and the use of sediment samples as an alternative to sampling turbid waters. We find that pre-freezing water does not affect the recovery of community composition either for the 12S and 16S primers, compared to the no freezing protocol. This is the case even when filters of larger pore size (3 μm) are used. Sediment samples recovered eDNA from organisms that inhabit the water column, however, due

to high variability among replicates in read abundance, we suggest increasing the number of biological replicates in the field.

Tag-jumping contamination

Contamination concerns are usually centered around pre-sequencing, during the field and wet laboratory work. These are of fundamental importance and care should be taken by sterilizing equipment and using negative controls. However, previous literature shows that the sequencing phase can be another source of contamination, generating up to 10% of contaminated reads by tag-jumping (Larsson et al., 2018; Schnell et al., 2015), which can skew analyses of taxa abundance and composition towards the rare taxa. There are ways to help minimize this issue by making use of dual indexing (Kircher et al., 2012)—although see Caroe and Bohmann (2020) for a library approach without dual indexing—, and amplification positive controls. The latter can be used to track the rate and level of contamination after sequencing to guide read cutoffs on samples (Deiner et al., 2017; Port et al., 2016).

Bioinformatics and data pre-processing

We relied on a bioinformatic approach developed by the metabaR package, adapted from Esling et al. (2015), to reduce the issue of contamination from tag-jumping, since it does not rely solely on the use of positive controls (which we lacked in this analysis) to make the estimated cutoff thresholds. However, after manually checking the fish dataset (12S primer), the final taxonomic tables still contained reads assigned to taxa that are not found in coastal lagoons in California (Table S 3.1). Some of it might be contamination from tag-jumping, although we cannot rule out the possibility that for a few of these species the eDNA could have come from local aquaria, as some are known in the pet trade (e.g. *Acanthurus achilles*). We also cannot disregard the

limitations of the reference database, especially related to the absence of estuarine and lagoonal taxa that may lead to dubious assignments to non-local related species. Due to inability to completely remove potential tag-jump contaminants from the dataset, we can expect a bias towards the rare taxa that will inflate diversity metrics in our samples for all primer sets.

Sediment samples generally showed higher variability among replicates compared to water samples for both primer sets, both in number of reads and community composition (Figure S 3.3-Figure S 3.4). The greater consistency of water replicates is an artifact of the single source for the water samples (the large jug), while sediment replicates were done by individually sampling the bottom of the lagoon. Although replicates were done a few centimeters apart, the bottom of the lagoon appears to have small-scale heterogeneity. The spatial variation of soil and sediment samples is recognized in the literature (Perkins et al., 2014; Taberlet, Prud'Homme, et al., 2012), and can be caused by sediment composition but also by the flow dynamic and distribution of eDNA in the water column. While this variability has been shown to occur for water samples as well in lentic environments (Harper et al., 2019 and references therein), the heterogeneity of water replicates in this system still requires further investigation.

Sediment samples also had an overall lower number of reads compared to water samples for both primer sets (Figure S 3.3). The lower number of reads seems to go against the expectations that eDNA can be more concentrated in sediments (Dell'Anno & Corinaldesi, 2004; Harper et al., 2019; Turner et al., 2015). This could be due to a few issues, some of which may interact. First, it could be related to a faster degradation and/or turn-over rates of eDNA in the sediment, which are determined by the soil and eDNA characteristics, as well as enzymatic and microbial activities (Levy-Booth et al., 2007; Pietramellara et al., 2009; Torti et al., 2015). The overall lower abundance of eDNA in the sediments could also be driven by increased inhibition (Buxton et al.,

2017; Pawlowski et al., 2022). Even though we used a specific soil extraction kit for both sediment and filtered water samples, the purification steps in the protocol could still not have been enough to reduce inhibition in the sediment as well as for the water samples. Lastly, this could have been driven by the much lower volume of sediment used: 0.25-0.3 g versus 500 mL for water samples.

There is also the fact that this type of environment is affected by scouring (purging of sediment to the ocean) during high precipitation events and increased flow of freshwater. However, since the sediment collection was done out of the rainy season and the lagoon was closed by a sandbar with no signs of scouring, we are confident that this was not a factor that could have caused the decreased ability to recover eDNA from the sediments. Therefore, we expect that this difference in read abundance between sediment and water samples would be more related to the other factors mentioned above, such as eDNA degradation and turn-over rates, inhibition, and different process volumes. Considering both the high variability and the lower sequencing throughput of the sediment replicates, we advise using a modified sampling protocol, e.g. the one developed by Taberlet, Prud'Homme, et al. (2012) that includes increasing the number of replicates and mixing larger volumes before processing.

Diversity

Considering that contamination through tag-jumping could be inflating the numbers of rare species in the dataset, the steepness and lack of a plateau for many of the species rarefaction curves could be artificial. This is especially evident for the 12S primer, since we were able to manually investigate the taxonomy tables (Figure 3.2-Figure 3.3). However, this lack of a plateau is an expected outcome from environmental samples (Alberdi et al., 2018), and has been shown to occur more acutely in a coastal lagoon in California when compared to other environments in California

(Shirazi et al., 2021)—albeit the authors were looking specifically at plants and fungi. The high number of species recovered from the sediment for the 16S primer (Figure 3.2) is likely driven by the recovery of a rich and complex sediment biota that is not paralleled in the water column.

The low taxonomic assignment to the species level for some of the dubious fish species found in our dataset, e.g *Phoxinus phoxinus, Odontesthes* spp. and *Sebastes pachycephalus*, also highlight the need to expand barcoding efforts to the local estuarine taxa to improve reference databases. On the other hand, *Fundulus diaphanus*, the northeastern killifish, did receive a few high taxonomic scores at the species level, which merit further consideration for biomonitoring of coastal lagoons in the region.

Pre-freezing water prior to filtration had an effect on the species curves of the 16S primer dataset, but not on the 12S. This could be explained by how differently eDNA molecules are found in the environment for these two different groups of organisms, and how freezing and thawing water would impact them. In the case of the fish fauna, the DNA that is shed from the organisms would be either found within cells, or adsorbed to colloids (Liang & Keeley, 2013; Torti et al., 2015; Turner, Barnes, et al., 2014). Even if cell walls were to disintegrate from the freezing and thawing process, they could still release intact mitochondria (which range from 1-8 μm in length) that could still be captured by our 3 μm pore size filters. On the other hand, bacteria and archaea, which are prokaryotic and often single celled organisms, would have their DNA released directly to the medium and pass through the larger pore size filters (>0.2 μm). Nevertheless, this freezing effect on cell walls has been shown to not always occur and likely be species-dependent (Sekar et al., 2009; Suomalainen et al., 2006).

Differential abundance

Pre-freezing water did not introduce any significant bias in species abundance compared to the "grab-and-hold", no freezing protocol, for any of the primer sets, even when using larger pore size filters (3 μm). Our results differ from other reports, where it was shown that freezing had differential effects on detection and relative abundance of different prokaryotic taxa (Kwambana et al., 2011; Sekar et al., 2009; Suomalainen et al., 2006). This could have been due to several reasons. First, the lack of effect pre-freezing had on community composition could be related to water properties of coastal lagoons that would have promoted the retention of DNA in the cellulose filters used in this analysis. Liang and Keeley (2013) have shown that presence and size of colloids, and the strength of ionic components, have an effect on increasing the binding affinities of DNA to the filters, especially the mixed cellulose esters filters (MCE). Another important aspect to consider is that the 'nominal' size of cellulose filters does not necessarily correspond to their 'effective' size. MCE filters do not have a uniform pore size like polycarbonate and nylon filters; rather, they are characterized by a 'tortuous flow path' from which particles are trapped more easily (Turner, Barnes, et al., 2014). This property of cellulose filters likely worked to our advantage, but also causes cellulose filters to be more susceptible to clogging than others.

Due to eDNA precipitation and resuspension, we expect to capture some community overlap between water and surficial sediment samples, however abundances should be different following the origin and fate of the eDNA in the environment and the processes acting on it throughout (Torti et al., 2015). Not surprisingly, with the DeSeq2 analysis, we see more algae (Hemiselmidaceae and Geminigeraceae) and cyanobacteria (Aphanizomenonaceae) in the water samples, and statistically higher representation of presumptively benthic diatoms (Catenulaceae and Fragilariaceae) and foraminiferans (Elphidiidae) in the sediment. In addition, the types of

environmental bacteria most abundant in the sediments were typical of soil and sediments elsewhere. Of particular note are those from anoxic environments (e.g. Anaerolineaceae and Desulfobacteraceae) as lagoon sediments are often dark and sulfide-rich.

The family Flavobacteriaceae was overrepresented in the water samples relative to the sediment, both in the pre- and no freezing protocols. In this family, there are important pathogens of fish and humans that belong to the genus *Flavobacterium*. (Suomalainen et al., 2006) found that *F. columnare* was more susceptible to having its cell walls disrupted to freezing due to high amounts of DNAases, lyases and proteases, likely connected to its pathogenicity, which then led to lower rates of DNA recovery. The species found in our dataset was *F. johnsoniae*, a species not known to be pathogenic–albeit with low species taxonomic score. Given that there was no difference in abundance for this species in our pre- and no freezing protocols, different from the results for the pathogenic species, *F. columnare*, this might relate to a true non-pathogenic species. However, considering that the endangered northern tidewater goby often achieves high abundance in this lagoon, more detailed assessment of the *Flavobacterium* species inhabiting this site would be of interest.

The other species assignment that draws our attention is the archea Candidatus *Nitrosopelagicus brevis* (Thaumarchaeota), which is significantly more abundant in sediment than water samples. As mentioned earlier, this is a pelagic species, normally found in the open ocean worldwide. Although coastal lagoons are subject to marine input, the relatively high concentration in sediment is unexpected and merits inquiry, especially considering that the confidence in its taxonomic assignment was low across reads. Likely, this represents a new environmental archaea that is abundant in coastal lagoon sediments.

Beta diversity

McMurdie and Holmes (2014) recommends against rarefying datasets due to the risk of removing true, rare ASVs. However, in our case, where we were unable to completely remove tag-jumping contaminants, this pre-process could help alleviate some of the noise caused by contaminants. Nevertheless, the CAP and PERMANOVA results on both the rarefied and standardized (eDNA index) dataset mostly corroborate some of our previous findings with the DeSeq2 analysis (*Differential abundance* section), showing significant differences in assemblage composition for the 16S primer, but not the 12S primer.

For the rarefied 16S primer dataset, all the species that were over- and underrepresented by CAP and PERMANOVA analyses were the same as those found by DeSeq2, such as *Guillardia theta* (Geminigeraceae), which was overrepresented in the pre-freezing protocol compared to the no freezing protocol. In addition, the species of foraminifera, *Elphidium williamsoni* (Elphidiidae) and the archea Candidatus *Nitrosopelagicus brevis* (Thaumarchaeota) were found to be overrepresented in sediment samples compared to water samples for both freezing protocols. The CAP results on the 16S primer dataset standardized using the eDNA index (Figure S 3.5) showed different species as underrepresented in the sediment compared to water samples but those also showed up as significantly differentially represented in the DeSeq2 analysis, with the exception of one, *Verminephrobacter aporrectodeae*.

Interestingly, the CAP analysis was also able to capture the underrepresentation of tidewater gobies (*E. newberryi*) in sediment samples on the 16S primer when compared to the no freezing protocol (Figure 3.6B). This reinforces the idea discussed earlier (*Bioinformatics and data pre-processing* section) that fish eDNA, at least in this environment, is less concentrated in the

sediment than in the water column, which contradicts other findings from the literature (Perkins et al., 2014; Turner et al., 2015). But it is worth noting that this underrepresentation of fish eDNA in the sediment was found not significant for the 12S primer, though, and there could be some bias related to how these two genes behave and degrade differently in the environment for the fish fauna.

Lessons Learned

Here is a list of recommendations and best practices for eDNA sampling and analysis in coastal environments that we have learned throughout this work and believe will be useful for others working in similar environments with turbid water and highly heterogeneous sediment/soil:

1. Filtered water samples had an overall higher number of reads compared to sediment for both primer sets. Therefore, we recommend the use of this protocol as it will increase chances of species detection;
2. If using sediment samples, we recommend increasing the number of replicates and mixing larger volumes before processing for DNA extractions (as in Taberlet, Prud'Homme, et al., 2012);
3. Pre-freezing water samples prior to filtration are an effective long-term storage solution and, at least for 3 μm pore size filters, it did not introduce bias in community composition compared to no freezing;
4. The use of dual-indexing and positive controls during library preparation will help minimize and address cross-contamination from tag-jumping, as is now widely recognized in many best-practice protocols (e.g. Deiner et al., 2017; Goldberg et al., 2016);

5. Although rarefying the dataset is not recommended (McMurdie & Holmes, 2014), we recognize that it can aid in reducing the noise of contaminants from your dataset, as long as they are rare. Otherwise, the use of eDNA index (Kelly et al., 2019) can be an alternative to standardize your dataset.

Conclusions

In this work, we assessed environmental DNA protocols for use in coastal lagoons, a highly dynamic habitat at the intersection of terrestrial, freshwater and marine environments. Pre-freezing water combined with the use of larger pore size filters (at least up to 3 μm) is a viable alternative for storage and processing of turbid water samples and, at least in the case of coastal lagoons, can work for the investigation of both fish (12S, MiFish) and bacteria and archaea (16S) communities. However, the use of sediment samples as an alternative to processing water samples should be done with caution, and at minimum the number of biological replicates should be increased to more than the five used in this work. Also, while sediment samples were able to recover eDNA from organisms commonly found in the water column, such as the tidewater goby, this was achieved during a period of relatively long lagoon closure, when there was no recent scouring of sediments to the ocean.

While we expect these guidelines to be helpful in the development of strategies to use eDNA as a monitoring resource in similar environments, protocol testing is still strongly advised whenever possible, especially when working in a new system. Much work is necessary to understand the full potential eDNA brings for the conservation and restoration of endangered species and habitats.

Tables

Table 3.1: Detailed information of the primers used.

Primer	Targets	Forward Primer	Reverse Primer	Reference
12S	Fish	GTCGGTAAAACTCGTG CCAGC	CATAGTGGGGTATCTA ATCCCAGTTTG	Miya *et al.* 2015
16S	Bacteria and archaea	GTGYCAGCMGCCGCGG TAA	GGACTACNVGGGTWT CTAAT	Caporaso *et al.*, 2012 (F: 515F and R: 806R)
CO1	Animals	GGWACWGGWTGAACW GTWTAYCCYCC	TANACYTCnGGRTGNC CRAARAAYCA	Leray *et al*. 2013

Table 3.2: Pairwise PERMANOVA (rarefied dataset) between all three protocols: pre- and no freezing water prior to filtration and sediment samples. P.adjusted is the adjusted p-value after FDR correction.

Primer	Comparison	F.Model	R^2	p.value	p.adjusted
12S	No freezing vs Pre-freezing	2.07252	0.20576	0.151	0.297
	No freezing vs Sediment	3.56051	0.41592	0.297	0.297
	Pre-freezing vs Sediment	2.25713	0.31102	0.224	0.297
16S	No freezing vs Pre-freezing	10.3356	0.56369	0.008	0.012
	No freezing vs Sediment	12.1022	0.63355	0.012	0.012
	Pre-freezing vs Sediment	12.5474	0.6419	0.008	0.012

Table 3.3: Pairwise PERMANOVA (eDNA index dataset) between all three protocols: pre- and no freezing water prior to filtration and sediment samples. P.adjusted is the adjusted p-value after FDR correction.

Primer	Comparison	F.Model	R^2	p.value	p.adjusted
16S	No freezing vs Pre-freezing	1.479053	0.156034	0.007	0.016
	No freezing vs Sediment	5.965368	0.427154	0.011	0.016
	Pre-freezing vs Sediment	6.514592	0.448831	0.016	0.016

Figures

Figure 3.1: Photo of Topanga lagoon taken on August 22nd, 2018, a few weeks after collection. There was no record of precipitation for the previous three months and the lagoon was closed to the ocean by a sandbar. There was also no sign of recent waves topping over the sandbar and reaching the lagoon.

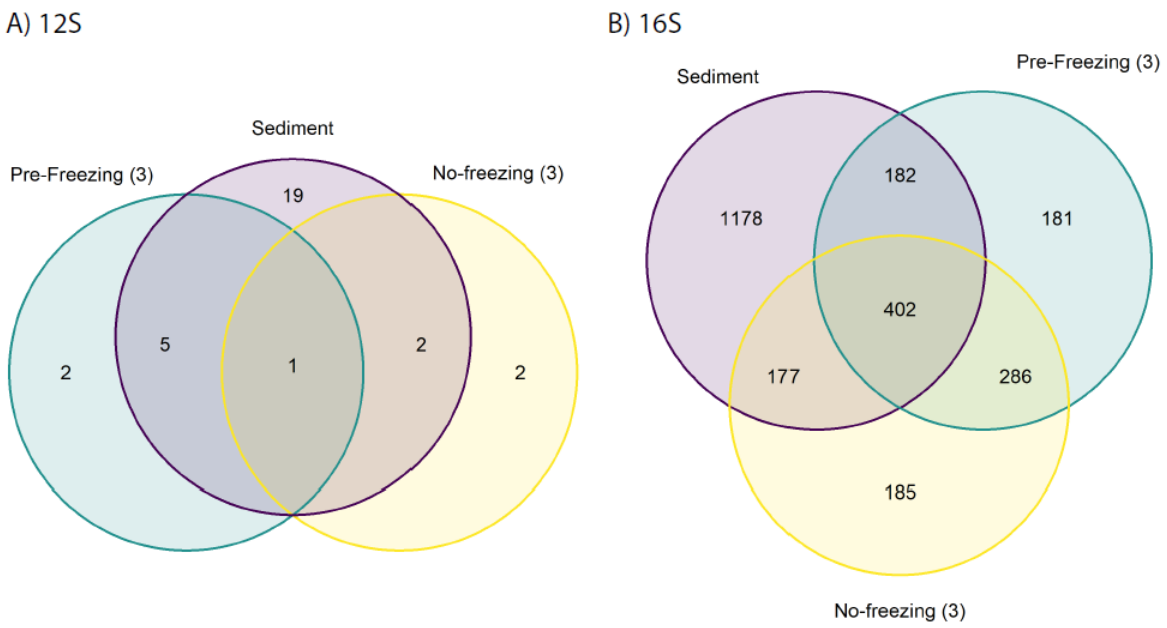


Figure 3.2: Venn diagrams of A) 12S and B) 16S primers showing the number of species found at and between each protocol. Sediment samples showed the highest number of unique species for both primers, although for the 12S dataset, about 28% are the result of contamination from tag-jumping.

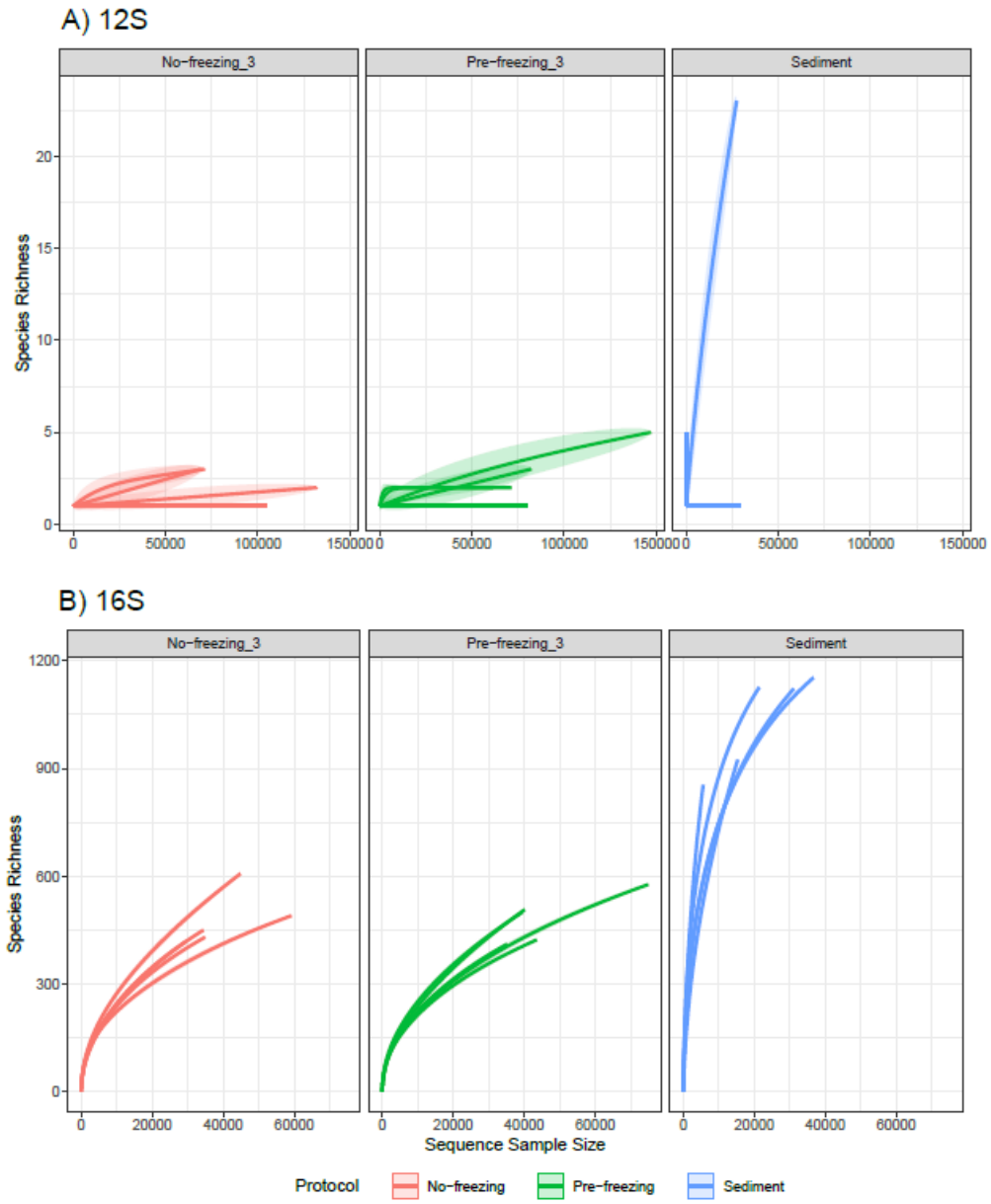


Figure 3.3: Species rarefaction curves based on sequencing effort for each protocol. A) 12S primer; B) 16S primer. With the exception of the water samples for the 12S primer, none of the curves have reached a plateau, although we expect the high diversity seen for the 12S sediment samples be due to contamination from tag-jumping.

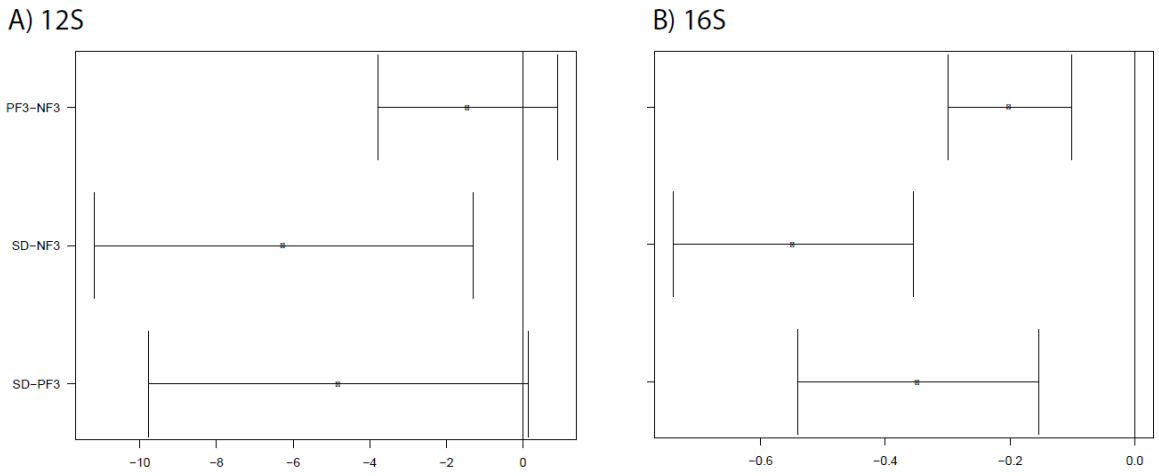


Figure 3.4: Confidence interval (CI) for slopes of rarefaction curves (Fig. 3) for each pairwise comparison of the different protocols. Only the comparison between pre- versus no freezing water samples, and pre-freezing versus sediment samples for the 12S primer (A) have come out non significant. The remaining comparisons showed significant differences between rarefaction slopes.

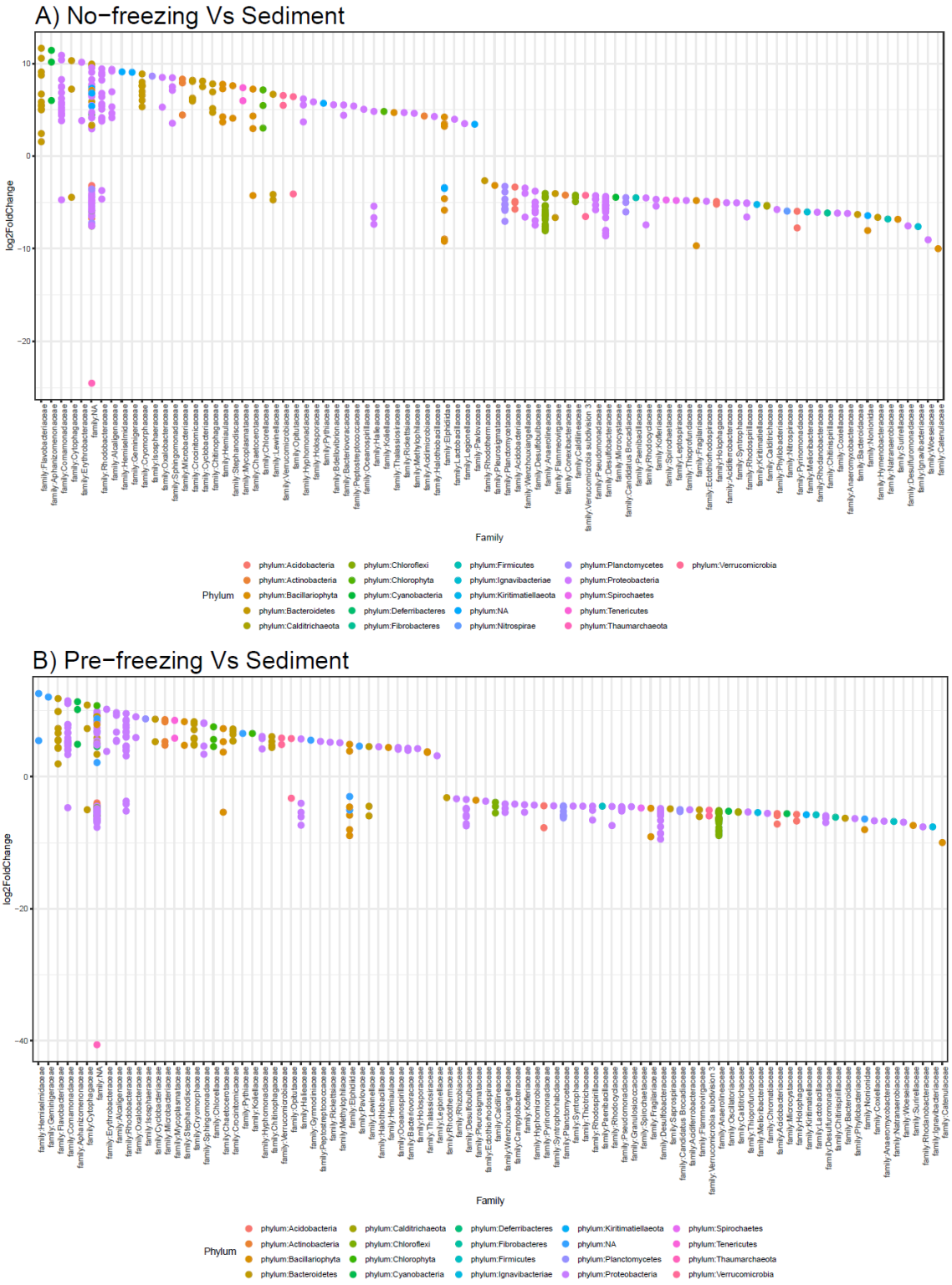


Figure 3.5: Plots of log2fold change of families of bacteria and archaea (16S primer) for the pairwise comparison between A) no freezing versus sediment; and B) pre-freezing versus sediment. Circles are colored by phylum. Species present above zero are overrepresented in the pre- or no freezing protocol, and species below the zero threshold are overrepresented in the sediments.

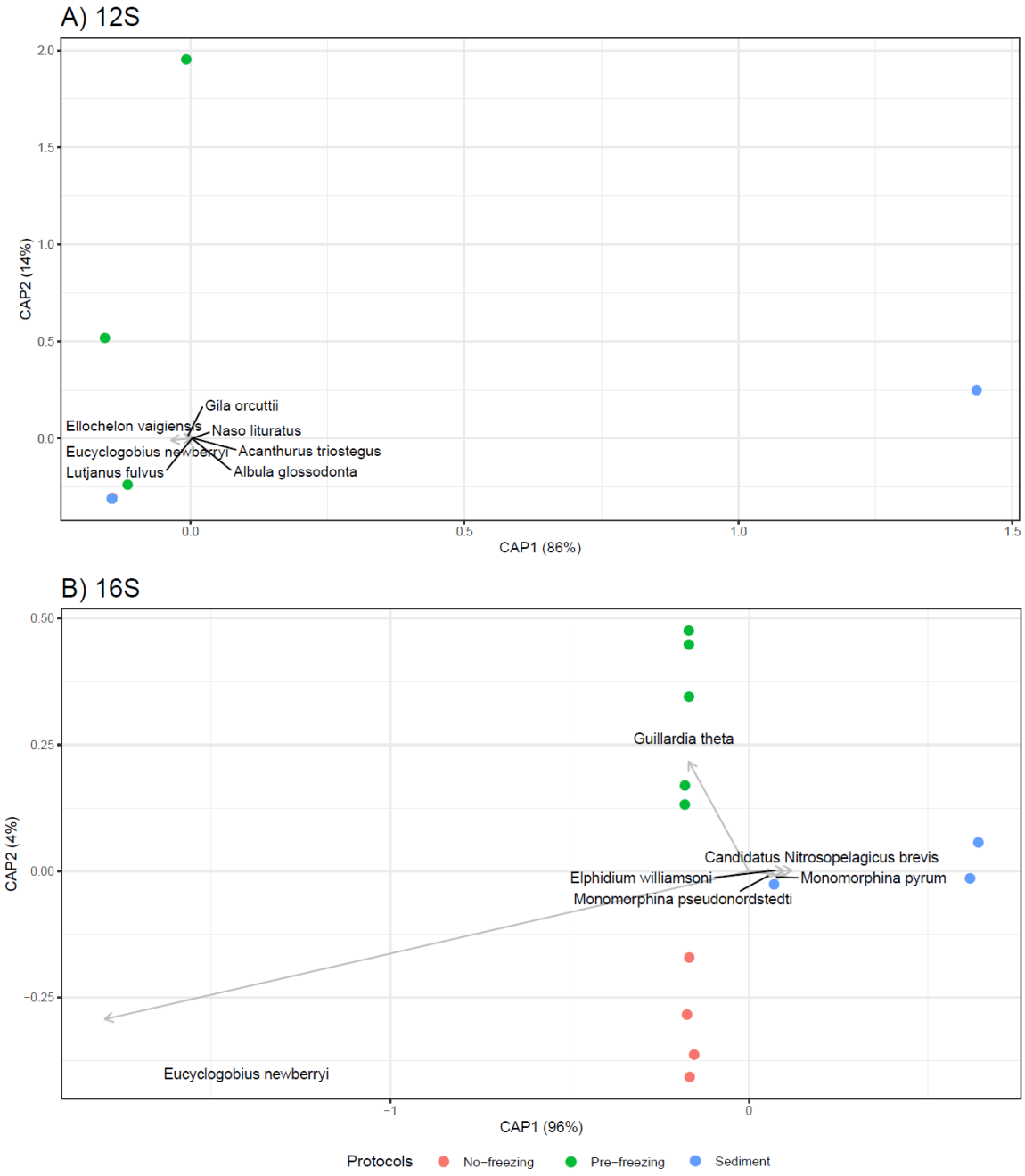


Figure 3.6: Constrained Analysis of Principal Coordinates (CAP) of A) 12S and B) 16S primer rarefied datasets. Circles are colored by protocol.

Supplementary Information

Excluded dataset

For the sake of clarity and conciseness, we decided to exclude from the main paper the sequencing results from the centrifugation protocol and the second stage filtration using the 0.45 µm pore size filter. In a preliminary investigation, the substantially lower sequence throughput from these protocols were reducing the power of the combined rarefied dataset. Besides, our assumption is that the bias driven by the water freezing protocol would still be captured by the first stage filtration. Some preliminary results of the centrifugation and second stage filtration protocols are shown in Figure S2, and the available raw read data for all protocols is provided at (https://gitlab.com/rturba/coastal-lagoon-edna).

CO1 Sequencing

It is hard to pinpoint exactly what was the issue that led to the substantive lower throughput of the CO1 primer compared to the other two. By following the CALeDNA protocol, PCR products were cleaned using magnetic beads, checked in agarose gel for size range, and insert size was taken into account when pooling libraries. Therefore, considering that we saw no bias at the quality check step when filtering reads, some error must have occurred when doing the bench work that affected all samples for this specific primer. We suspect it could have happened when calculating the cleaned PCR product concentrations using the fluorometer plate reader, since the CO1 primer was the only one that was read in a separate plate from the 12S and 16S primers.

Supplementary Tables

(See attached supplemental material)

Table S 3.1: List of fish species for the 12S primer dataset after the decontamination pipeline. Gbif = species listed in the GBIF database (Gbif.Org, 2022); Lab collection = species listed in the laboratory collection database; Taxonomy = taxonomic level match between the eDNA results and the other databases.

Supplementary Figures

Figure S 3.1: Adapted vacuum pump in the pre-PCR room.

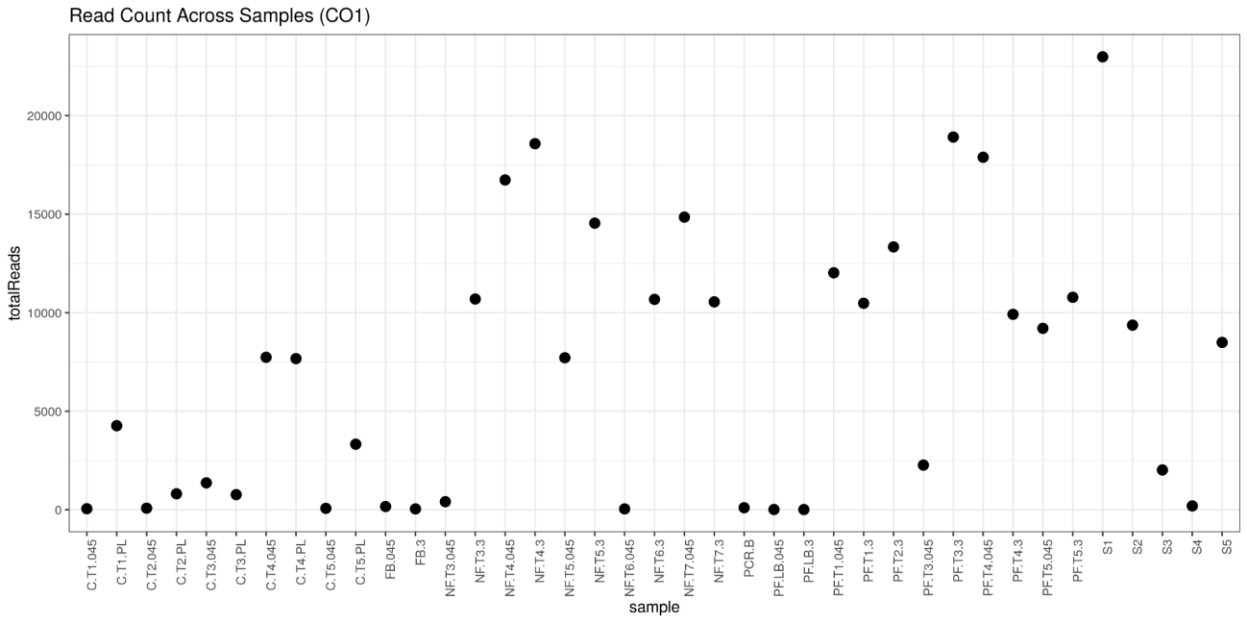
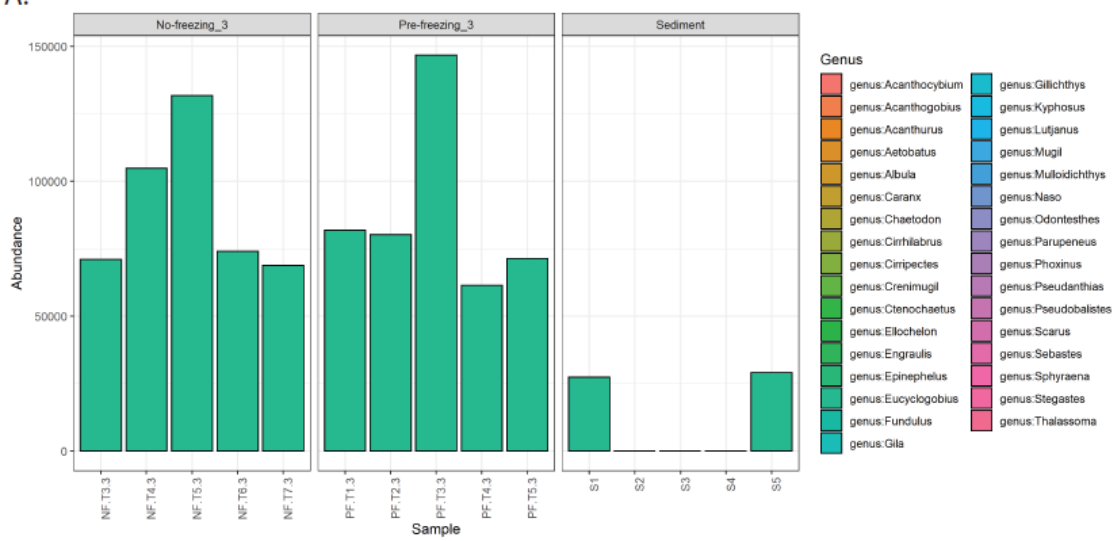
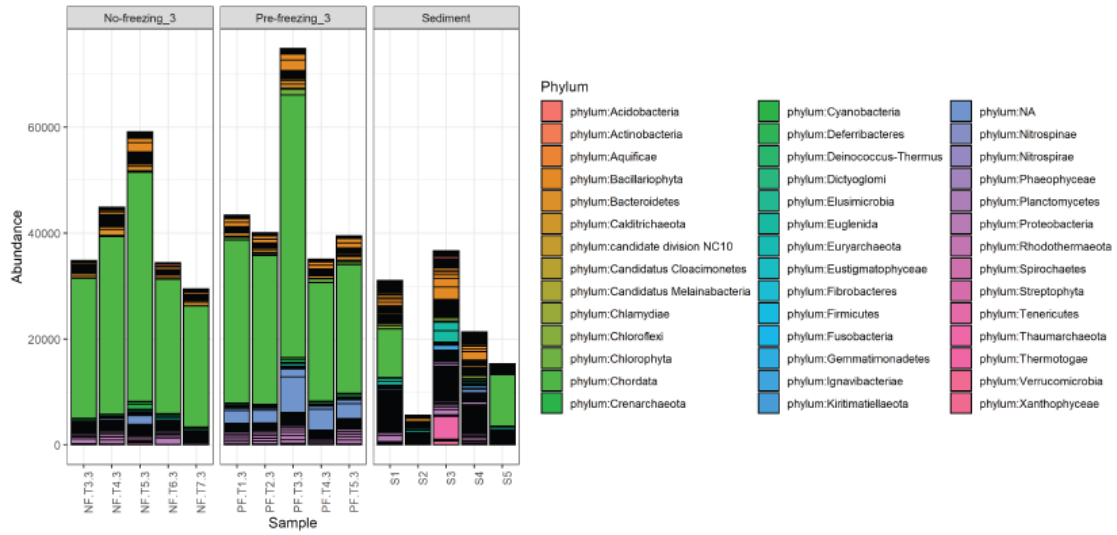


Figure S 3.2: Total number of reads per sample per barcode. C: centrifugation; NF: no freezing; PF: pre-freezing; S: sediment. Numbers at the end of the protocols relate to the size of the filter pores (3 µm and 0.45 µm). PL relates to the extracted pellet from the centrifugation protocol.

A.


No-freezing_3
Pre-freezing_3
Sediment
Abundance
150000
100000
50000
0
NF.T3.3
NF.T4.3
NF.T5.3
NF.T6.3
NF.T7.3
PF.T1.3
PF.T2.3
PF.T3.3
PF.T4.3
PF.T5.3
S1
S2
S3
S4
S5
Sample
Genus
genus:Acanthocybium
genus:Acanthogobius
genus:Acanthurus
genus:Aetobatus
genus:Albula
genus:Caranx
genus:Chaetodon
genus:Cirrhilabrus
genus:Cirripectes
genus:Crenimugil
genus:Ctenochaetus
genus:Ellochelon
genus:Engraulis
genus:Epinephelus
genus:Eucyclogobius
genus:Fundulus
genus:Gila
genus:Gillichthys
genus:Kyphosus
genus:Lutjanus
genus:Mugil
genus:Mulloidichthys
genus:Naso
genus:Odontesthes
genus:Parupeneus
genus:Phoxinus
genus:Pseudanthias
genus:Pseudobalistes
genus:Scarus
genus:Sebastes
genus:Sphyraena
genus:Stegastes
genus:Thalassoma

B.


No-freezing_3
Pre-freezing_3
Sediment
Abundance
60000
40000
20000
0
NF.T3.3
NF.T4.3
NF.T5.3
NF.T6.3
NF.T7.3
PF.T1.3
PF.T2.3
PF.T3.3
PF.T4.3
PF.T5.3
S1
S2
S3
S4
S5
Sample
Phylum
phylum:Acidobacteria
phylum:Actinobacteria
phylum:Aquificae
phylum:Bacillariophyta
phylum:Bacteroidetes
phylum:Calditrichaeota
phylum:candidate division NC10
phylum:Candidatus Cloacimonetes
phylum:Candidatus Melainabacteria
phylum:Chlamydiae
phylum:Chloroflexi
phylum:Chlorophyta
phylum:Chordata
phylum:Crenarchaeota
phylum:Cyanobacteria
phylum:Deferribacteres
phylum:Deinococcus-Thermus
phylum:Dictyoglomi
phylum:Elusimicrobia
phylum:Euglenida
phylum:Euryarchaeota
phylum:Eustigmatophyceae
phylum:Fibrobacteres
phylum:Firmicutes
phylum:Fusobacteria
phylum:Gemmatimonadetes
phylum:Ignavibacteriae
phylum:Kiritimatiellaeota
phylum:NA
phylum:Nitrospinae
phylum:Nitrospirae
phylum:Phaeophyceae
phylum:Planctomycetes
phylum:Proteobacteria
phylum:Rhodothermaeota
phylum:Spirochaetes
phylum:Streptophyta
phylum:Tenericutes
phylum:Thaumarchaeota
phylum:Thermotogae
phylum:Verrucomicrobia
phylum:Xanthophyceae

C.

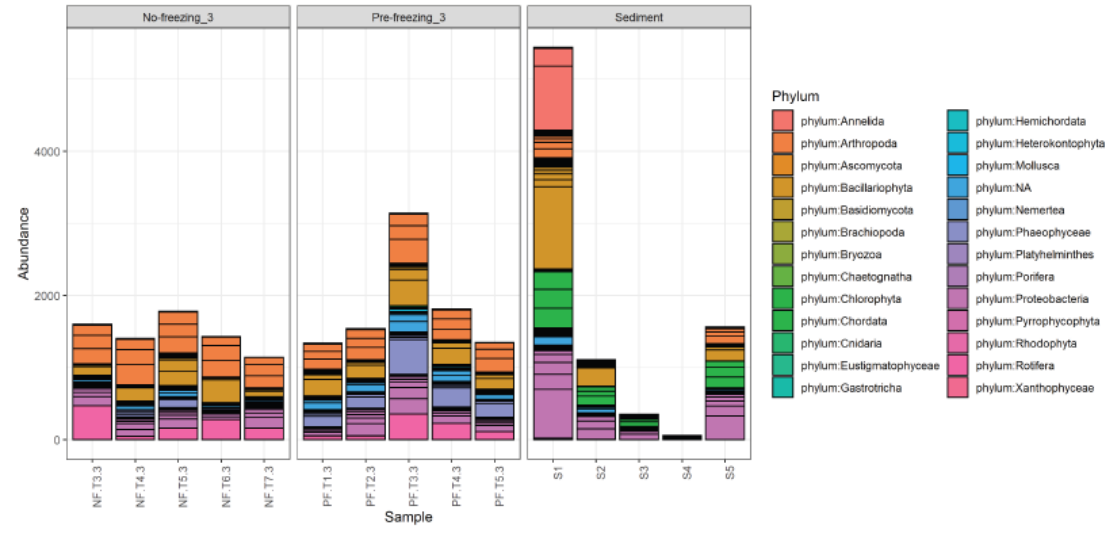

No-freezing_3
Pre-freezing_3
Sediment
Abundance
4000
2000
0
NF.T3.3
NF.T4.3
NF.T5.3
NF.T6.3
NF.T7.3
PF.T1.3
PF.T2.3
PF.T3.3
PF.T4.3
PF.T5.3
S1
S2
S3
S4
S5
Sample
Phylum
phylum:Annelida
phylum:Arthropoda
phylum:Ascomycota
phylum:Bacillariophyta
phylum:Basidiomycota
phylum:Brachiopoda
phylum:Bryozoa
phylum:Chaetognatha
phylum:Chlorophyta
phylum:Chordata
phylum:Cnidaria
phylum:Eustigmatophyceae
phylum:Gastrotricha
phylum:Hemichordata
phylum:Heterokontophyta
phylum:Mollusca
phylum:NA
phylum:Nemertea
phylum:Phaeophyceae
phylum:Platyhelminthes
phylum:Porifera
phylum:Proteobacteria
phylum:Pyrrophycophyta
phylum:Rhodophyta
phylum:Rotifera
phylum:Xanthophyceae

Figure S 3.3: Barplots of genus (A: 12S) and phylum (B: 16S; and C: CO1) read abundance. Sample code the same as Figure S2.

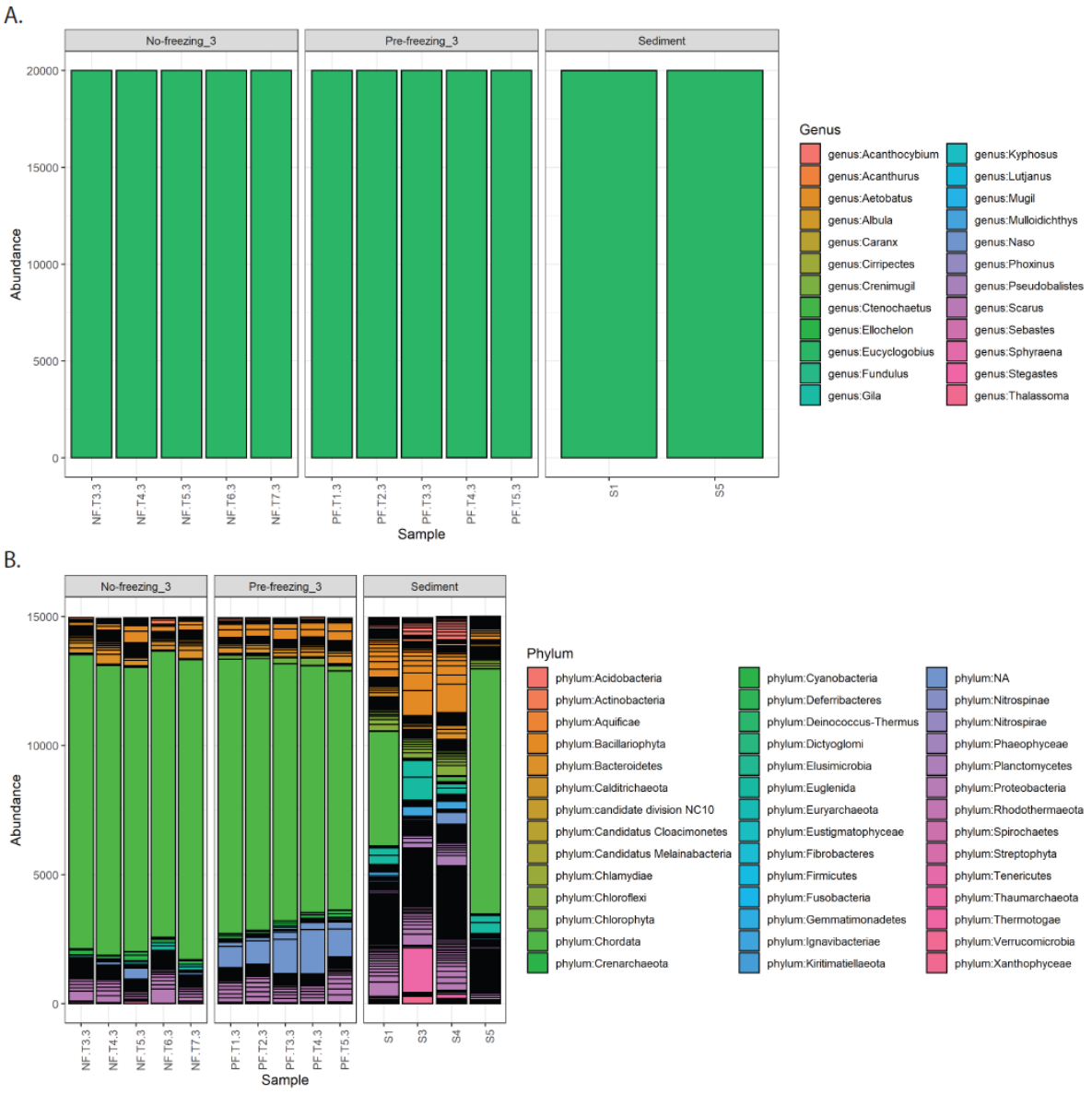


Figure S 3.4: Barplots of rarefied dataset of genus (A: 12S primer) and phylum (B: 16S primer) abundance for each protocol: no freezing, pre-freezing and sediment. Sample code the same as Figure S 3.2.

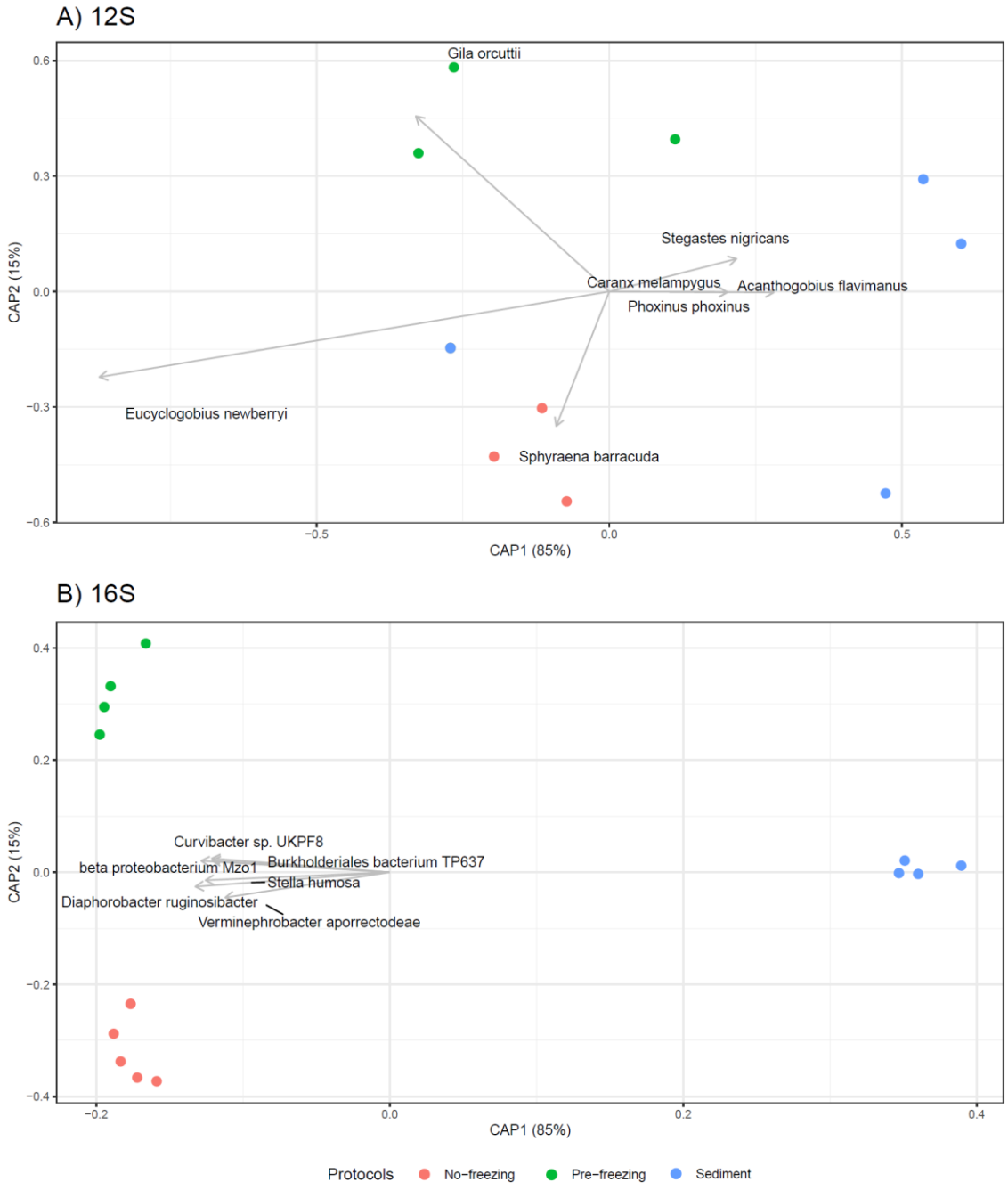


Figure S 3.5: Constrained Analysis of Principal Coordinates (CAP) of A) 12S and B) 16S primer datasets standardized via Wisconsin double standardization (eDNA index). It can be noted that in the 12S dataset, the tropical species suspected to be contaminants from tag-jumping are driving the community assemblage differences between protocols, which is different from the rarefied dataset (Figure 3.6).

Bibliography

Alberdi, A., Aizpurua, O., Gilbert, M. T. P., & Bohmann, K. (2018). Scrutinizing key steps for reliable metabarcoding of environmental samples. *Methods in Ecology and Evolution*, *9*(1), 134–147. https://doi.org/10.1111/2041-210X.12849

Ballard, J., Pezda, J., Spencer, D., & Plantinga, A. (n.d.). *An Economic Valuation of Southern California Coastal Wetlands*. http://scwrp.org/wp-content/uploads/2017/06/SoCalWetlands_FinalReport.pdf

Barrett, R. D., Rogers, S. M., & Schluter, D. (2008). Natural selection on a major armor gene in threespine stickleback. *Science*, *322*(5899), 255–257.

Baumgartner, J. V. (1986). The genetics of differentiation in a stream population of the threespine stickleback, Gasterosteus aculeatus. *Heredity*, *57*(2), 199–208.

Beichman, A. C., Phung, T. N., & Lohmueller, K. E. (2017). Comparison of Single Genome and Allele Frequency Data Reveals Discordant Demographic Histories. *G3: Genes, Genomes, Genetics*, *7*(11), 3605–3620. https://doi.org/10.1534/g3.117.300259

Bell, M. A. (1976). Evolution of Phenotypic Diversity in Gasterosteus Aculeatus Superspecies on the Pacific Coast of North America. *Systematic Biology*, *25*(3), 211–227. https://doi.org/10.2307/2412489

Bell, M. A. (1978). Fishes of the Santa Clara River system, southern California. *Natural History Museum of Los Angeles County Contributions in Science*, *295*, 1–20.

Bell, M. A. (2001). Lateral plate evolution in the threespine stickleback: Getting nowhere fast. In A. P. Hendry & M. T. Kinnison (Eds.), *Microevolution Rate, Pattern, Process* (Vol. 8, pp. 445–461). Springer.

Bell, M. A., & Foster, S. A. (1994). *The Evolutionary Biology of the Threespine Stickleback.pdf*. Oxford University Press.

Bell, M. A., Orti, G., Walker, J. A., & Koenings, J. P. (1993). Evolution of Pelvic Reduction in Threespine Stickleback Fish: A Test of Competing Hypotheses. *Evolution*, *47*(3), 906. https://doi.org/10.2307/2410193

Bell, M. A., & Richkind, K. E. (1981). Clinal Variation of Lateral Plates in Threespine Stickleback Fish. *The American Naturalist*, *117*(2), 113–132. https://doi.org/10.1086/283693

Benjamin, A., May, B., O'Brien, J., & Finger, A. J. (2016). Conservation Genetics of an Urban Desert Fish, the Arroyo Chub. *Transactions of the American Fisheries Society*, *145*(2), 277–286. https://doi.org/10.1080/00028487.2015.1121925

Benke, A. C. (1990). A Perspective on America's Vanishing Streams. *Journal of the North American Benthological Society*, *9*(1), 77–88. https://doi.org/10.2307/1467936

Bergstrom, C. A. (2002). Fast-start swimming performance and reduction in lateral plate number in threespine stickleback. *Canadian Journal of Zoology*, *80*(2), 207–213. https://doi.org/10.1139/z01-226

Bohmann, K., Chua, P., Holman, L. E., & Lynggaard, C. (2021). DNAqua-Net conference unites participants from around the world with the quest to standardize and implement DNA-based aquatic biomonitoring. *Environmental DNA*, *3*(5), 884–888. https://doi.org/10.1002/edn3.207

Buth, D. G., Crabtree, C. B., Orton, R. D., & Rainboth, W. J. (1984). Genetic differentiation between the freshwater subspecies of Gasterosteus aculeatus in Southern California.

Biochemical Systematics and Ecology, *12*(4), 423–432. https://doi.org/10.1016/0305-1978(84)90075-9

Buxton, A. S., Groombridge, J. J., & Griffiths, R. A. (2017). Is the detection of aquatic environmental DNA influenced by substrate type? *PLOS ONE*, *12*(8), e0183371. https://doi.org/10.1371/journal.pone.0183371

Callahan, B. J., McMurdie, P. J., Rosen, M. J., Han, A. W., Johnson, A. J. A., & Holmes, S. P. (2016). DADA2: High-resolution sample inference from Illumina amplicon data. *Nature Methods*, *13*(7), 581–583. https://doi.org/10.1038/nmeth.3869

Campos, P. F., & Gilbert, T. M. P. (2012). DNA Extraction from Formalin-Fixed Material. In B. Shapiro & M. Hofreiter (Eds.), *Ancient DNA: Methods and Protocols* (pp. 81–85). Humana Press. https://doi.org/10.1007/978-1-61779-516-9_11

Caroe, C., & Bohmann, K. (2020). Tagsteady: A metabarcoding library preparation protocol to avoid false assignment of sequences to samples. *BioRxiv*.

Chen, H. (2018). *VennDiagram: Generate High-Resolution Venn and Euler Plots* (1.6.20). https://CRAN.R-project.org/package=VennDiagram

Colosimo, P. F., Hosemann, K. E., Balabhadra, S., Villarreal, G., Dickson, M., Grimwood, J., Schmutz, J., Myers, R. M., Schluter, D., & Kingsley, D. M. (2005). Widespread parallel evolution in sticklebacks by repeated fixation of ectodysplasin alleles. *Science*, *307*(5717), 1928–1933.

Colosimo, P. F., Peichel, C. L., Nereng, K., Blackman, B. K., Shapiro, M. D., Schluter, D., & Kingsley, D. M. (2004). The genetic architecture of parallel armor plate reduction in threespine sticklebacks. *PLoS Biol*, *2*(5), e109.

Conte, G. L., Arnegard, M. E., Best, J., Chan, Y. F., Jones, F. C., Kingsley, D. M., Schluter, D., & Peichel, C. L. (2015). Extent of QTL Reuse During Repeated Phenotypic Divergence of Sympatric Threespine Stickleback. *Genetics*, *201*(3), 1189–1200. https://doi.org/10.1534/genetics.115.182550

Cresko, W. A., Amores, A., Wilson, C., Murphy, J., Currey, M., Phillips, P., Bell, M. A., Kimmel, C. B., & Postlethwait, J. H. (2004). Parallel genetic basis for repeated evolution of armor loss in Alaskan threespine stickleback populations. *Proceedings of the National Academy of Sciences*, *101*(16), 6050–6055. https://doi.org/10.1073/pnas.0308479101

Cruickshank, T. E., & Hahn, M. W. (2014). Reanalysis suggests that genomic islands of speciation are due to reduced diversity, not reduced gene flow. *Molecular Ecology*, *23*(13), 3133–3157. https://doi.org/10.1111/mec.12796

Culver, G. B., & Hubbs, C. L. (1917). The fishes of the Santa Ana system of streams in Southern California. *Lorguinia*, *1*, 82–83.

Curd, E., Gomer, J., Kandlikar, G., Gold, Z., Ogden, M., & Shi, B. (2018). *The Anacapa Toolkit*. https://github.com/limey-bean/Anacapa

Curd, E., Kandlikar, G., & Gomer, J. (2018). *CRUX: Creating Reference libraries Using eXisting tools*. https://github.com/limey-bean/CRUX_Creating-Reference-libraries-Using-eXisting-tools

Cuvier, G., & Valenciennes, M. (1829). *Histoire naturelle des poissons*. Chez F. G. Levrault; [etc., etc.]. https://www.biodiversitylibrary.org/item/54024

Danecek, P., Auton, A., Abecasis, G., Albers, C. A., Banks, E., DePristo, M. A., Handsaker, R. E., Lunter, G., Marth, G. T., Sherry, S. T., McVean, G., & Durbin, R. (2011). The variant

call format and VCFtools. *Bioinformatics*, *27*(15), 2156–2158. https://doi.org/10.1093/bioinformatics/btr330

Danecek, P., Bonfield, J. K., Liddle, J., Marshall, J., Ohan, V., Pollard, M. O., Whitwham, A., Keane, T., McCarthy, S. A., Davies, R. M., & Li, H. (2021). Twelve years of SAMtools and BCFtools. *GigaScience*, *10*(2), giab008. https://doi.org/10.1093/gigascience/giab008

Deagle, B. E., Jones, F. C., Absher, D. M., Kingsley, D. M., & Reimchen, T. E. (2013a). Phylogeography and adaptation genetics of stickleback from the H aida G waii archipelago revealed using genome-wide single nucleotide polymorphism genotyping. *Molecular Ecology*, *22*(7), 1917–1932.

Deagle, B. E., Jones, F. C., Absher, D. M., Kingsley, D. M., & Reimchen, T. E. (2013b). Phylogeography and adaptation genetics of stickleback from the Haida Gwaii archipelago revealed using genome-wide single nucleotide polymorphism genotyping. *Molecular Ecology*, *22*(7), 1917–1932. https://doi.org/10.1111/mec.12215

Deiner, K., Bik, H. M., Mächler, E., Seymour, M., Lacoursière-Roussel, A., Altermatt, F., Creer, S., Bista, I., Lodge, D. M., & De Vere, N. (2017). Environmental DNA metabarcoding: Transforming how we survey animal and plant communities. *Molecular Ecology*, *26*(21), 5872–5895.

Deiner, K., Walser, J.-C., Mächler, E., & Altermatt, F. (2015). Choice of capture and extraction methods affect detection of freshwater biodiversity from environmental DNA. *Biological Conservation*, *183*, 53–63.

Dejean, T., Valentini, A., Miquel, C., Taberlet, P., Bellemain, E., & Miaud, C. (2012). Improved detection of an alien invasive species through environmental DNA barcoding: The example of the American bullfrog *Lithobates catesbeianus*: *Alien invasive species*

detection using eDNA. Journal of Applied Ecology, *49*(4), 953–959. https://doi.org/10.1111/j.1365-2664.2012.02171.x

Dell'Anno, A., & Corinaldesi, C. (2004). Degradation and Turnover of Extracellular DNA in Marine Sediments: Ecological and Methodological Considerations Degradation and Turnover of Extracellular DNA in Marine Sediments: Ecological and Methodological Considerations. *Applied and Environmental Microbiology*, *70*(7), 4384–4386. https://doi.org/10.1128/AEM.70.7.4384

Des Roches, S., Bell, M. A., & Palkovacs, E. P. (2020). Climate-driven habitat change causes evolution in Threespine Stickleback. *Global Change Biology*, *26*(2), 597–606.

Doi, H., Uchii, K., Matsuhashi, S., Takahara, T., Yamanaka, H., & Minamoto, T. (2017). Isopropanol precipitation method for collecting fish environmental DNA. *Limnology and Oceanography: Methods*, *15*(2), 212–218. https://doi.org/10.1002/lom3.10161

Dolby, G. A., Bedolla, A. M., Bennett, S. E. K., & Jacobs, D. K. (2020). Global physical controls on estuarine habitat distribution during sea level change: Consequences for genetic diversification through time. *Global and Planetary Change*, *187*, 103128. https://doi.org/10.1016/j.gloplacha.2020.103128

Dolby, G. A., Ellingson, R. A., Findley, L. T., & Jacobs, D. K. (2018). How sea level change mediates genetic divergence in coastal species across regions with varying tectonic and sediment processes. *Molecular Ecology*, *27*(4), 994–1011. https://doi.org/10.1111/mec.14487

Dolby, G. A., Hechinger, R., Ellingson, R. A., Findley, L. T., Lorda, J., & Jacobs, D. K. (2016). Sea-level driven glacial-age refugia and post-glacial mixing on subtropical coasts, a

palaeohabitat and genetic study. *Proceedings of the Royal Society B: Biological Sciences*, *283*(1843), 20161571.

Earl, D. A., Louie, K. D., Bardeleben, C., Swift, C. C., & Jacobs, D. K. (2010). Rangewide microsatellite phylogeography of the endangered tidewater goby, a genetically subdivided coastal fish with limited marine dispersal. *Conservation Genetics*, *11*, 103–104. https://doi.org/10.1007/s10592-009-0008-9

Enciso-Romero, J., Pardo-Díaz, C., Martin, S. H., Arias, C. F., Linares, M., McMillan, W. O., Jiggins, C. D., & Salazar, C. (2017). Evolution of novel mimicry rings facilitated by adaptive introgression in tropical butterflies. *Molecular Ecology*, *26*(19), 5160–5172. https://doi.org/10.1111/mec.14277

Eriksson, B. K., Yanos, C., Bourlat, S. J., Donadi, S., Fontaine, M. C., Hansen, J. P., Jakubavičiūtė, E., Kiragosyan, K., Maan, M. E., Merilä, J., Austin, Å. N., Olsson, J., Reiss, K., Sundblad, G., Bergström, U., & Eklöf, J. S. (2021). Habitat segregation of plate phenotypes in a rapidly expanding population of three-spined stickleback. *Ecosphere*, *12*(6), 1–18. https://doi.org/10.1002/ecs2.3561

Esling, P., Lejzerowicz, F., & Pawlowski, J. (2015). Accurate multiplexing and filtering for high-throughput amplicon-sequencing. *Nucleic Acids Research*, *43*(5), 2513–2524. https://doi.org/10.1093/nar/gkv107

Evanno, G., Regnaut, S., & Goudet, J. (2005). Detecting the number of clusters of individuals using the software structure: A simulation study. *Molecular Ecology*, *14*(8), 2611–2620. https://doi.org/10.1111/j.1365-294X.2005.02553.x

Ewing, G. B., & Jensen, J. D. (2016). The consequences of not accounting for background selection in demographic inference. *Molecular Ecology*, *25*(1), 135–141.

Faber, P. M., Keller, E., Sands, A., & Massey, B. M. (1989). *The Ecology of Riparian Habitats of the Southern California Coastal Region: A Community Profile: Vol. 7.27*. U.S. Department of the Interior, Fish and Wildlife Service, Research and Development, National Wetlands Research Center.

Fang, B., Kemppainen, P., Momigliano, P., Feng, X., & Merilä, J. (2020). On the causes of geographically heterogeneous parallel evolution in sticklebacks. *Nature Ecology & Evolution*, *4*(8), 1105–1115.

Fang, B., Merilä, J., Ribeiro, F., Alexandre, C. M., & Momigliano, P. (2018). Worldwide phylogeny of three-spined sticklebacks. *Molecular Phylogenetics and Evolution*, *127*, 613–625.

Faribanks, H. W., & Carey, E. P. (1910). Glaciation in the San Bernardino Range, California. *Science*, *31*(784), 32–33.

FASTX-Toolkit. (n.d.). Retrieved January 11, 2018, from http://hannonlab.cshl.edu/fastx_toolkit/

Ficetola, G. F., Miaud, C., Pompanon, F., & Taberlet, P. (2008). Species detection using environmental DNA from water samples. *Biology Letters*, *4*(4), 423–425. https://doi.org/10.1098/rsbl.2008.0118

Fitak, R. R. (2021). OptM: Estimating the optimal number of migration edges on population trees using Treemix. *Biology Methods and Protocols*, *6*(1), bpab017. https://doi.org/10.1093/biomethods/bpab017

Flint, L. E., Brandt, J., Christensen, A. H., Flint, A. L., Hevesi, J. A., Jachens, R., Kulongoski, J. T., Martin, P., & Sneed, M. (2012). Geohydrology of Big Bear Valley, California: Phase 1--geologic framework, recharge, and preliminary assessment of the source and age of groundwater. In *Geohydrology of Big Bear Valley, California: Phase 1—Geologic*

framework, recharge, and preliminary assessment of the source and age of groundwater (USGS Numbered Series No. 2012–5100; Scientific Investigations Report, Vols. 2012–5100, p. 130). U.S. Geological Survey. https://doi.org/10.3133/sir20125100

Flint, L. E., Underwood, E. C., Flint, A. L., & Hollander, A. D. (2019). Characterizing the Influence of Fire on Hydrology in Southern California. *Natural Areas Journal*, *39*(1), 108–121.

Freedman, A. H., Gronau, I., Schweizer, R. M., Vecchyo, D. O.-D., Han, E., Silva, P. M., Galaverni, M., Fan, Z., Marx, P., Lorente-Galdos, B., Beale, H., Ramirez, O., Hormozdiari, F., Alkan, C., Vilà, C., Squire, K., Geffen, E., Kusak, J., Boyko, A. R., … Novembre, J. (2014). Genome Sequencing Highlights the Dynamic Early History of Dogs. *PLOS Genetics*, *10*(1), e1004016. https://doi.org/10.1371/journal.pgen.1004016

French, J. J., & Busby, M. W. (1974). Flood-hazard study, 100-year stage for Baldwin Lake, San Bernardino County, California. In *Flood-hazard study, 100-year stage for Baldwin Lake, San Bernardino County, California* (USGS Numbered Series No. 74–26; Water-Resources Investigations Report, Vols. 74–26). U.S. Geological Survey. https://doi.org/10.3133/wri7426

Gao, X., Lin, H., Revanna, K., & Dong, Q. (2017). A Bayesian taxonomic classification method for 16S rRNA gene sequences with improved species-level accuracy. *BMC Bioinformatics*, *18*(1), 247. https://doi.org/10.1186/s12859-017-1670-4

Gbif.Org. (2022). *Occurrence Download* (p. 170487) [Darwin Core Archive]. The Global Biodiversity Information Facility. https://doi.org/10.15468/DL.HTJ3HT

Girard, C. F. (1854). *Descriptions of new fishes, collected by Dr. AL Heermann, naturalist attached to the survey of the Pacific railroad route, under Lieut. RS Williamson, USA. 7*, 129–140.

Glover, K. C., Chaney, A., Kirby, M. E., Patterson, W. P., & MacDonald, G. M. (2020). Southern California vegetation, wildfire, and erosion had nonlinear responses to climatic forcing during marine isotope stages 5–2 (120–15 ka). *Paleoceanography and Paleoclimatology*, *35*(2), 1–15.

Glover, K. C., George, J., Heusser, L., & MacDonald, G. M. (2021). West Coast vegetation shifts as a response to climate change over the past 130,000 years: Geographic patterns and process from pollen data. *Physical Geography*, 1–19.

Glover, K. C., MacDonald, G. M., Kirby, M. E., Rhodes, E. J., Stevens, L., Silveira, E., Whitaker, A., & Lydon, S. (2017). Evidence for orbital and North Atlantic climate forcing in alpine Southern California between 125 and 10 ka from multi-proxy analyses of Baldwin Lake. *Quaternary Science Reviews*, *167*, 47–62.

Goldberg, C. S., Turner, C. R., Deiner, K., Klymus, K. E., Thomsen, P. F., Murphy, M. A., Spear, S. F., McKee, A., Oyler-McCance, S. J., & Cornman, R. S. (2016). Critical considerations for the application of environmental DNA methods to detect aquatic species. *Methods in Ecology and Evolution*, *7*(11), 1299–1307.

Gumprecht, B. (1999). *The Los Angeles River: Its Life, Death, and Possible Rebirth* (2001st ed.). Johns Hopkins University Press.

Guo, Y., Ye, F., Sheng, Q., Clark, T., & Samuels, D. C. (2014). Three-stage quality control strategies for DNA re-sequencing data. *Briefings in Bioinformatics*, *15*(6), 879–889. https://doi.org/10.1093/bib/bbt069

Hagen, D. W., & Gilbertson, L. G. (1973). Selective predation and the intensity of selection acting upon the lateral plates of threespine sticklebacks. *Heredity*, *30*(3), 273–287.

Haglund, T., & Buth, D. (1988). Allozymes of the unarmored threespine stickleback (Gasterosteus aculeatus williamsoni) and identification of the Shay Creek population. *Isozyme Bull*, *21*, 196.

Han, F., Lamichhaney, S., Grant, B. R., Grant, P. R., Andersson, L., & Webster, M. T. (2017). Gene flow, ancient polymorphism, and ecological adaptation shape the genomic landscape of divergence among Darwin's finches. *Genome Research*, *27*(6), 1004–1015. https://doi.org/10.1101/gr.212522.116

Harper, L. R., Buxton, A. S., Rees, H. C., Bruce, K., Brys, R., Halfmaerten, D., Read, D. S., Watson, H. V., Sayer, C. D., & Jones, E. P. (2019). Prospects and challenges of environmental DNA (eDNA) monitoring in freshwater ponds. *Hydrobiologia*, *826*(1), 25–41.

Hendry, A. P. (2013). Key questions in the genetics and genomics of eco-evolutionary dynamics. *Heredity*, *111*(6), 456–466. https://doi.org/10.1038/hdy.2013.75

Hendy, I. (2009). A fresh perspective on the Cordilleran Ice Sheet. *The Geological Society of America*, *19*(4–5), 60. https://doi.org/10.1130/0091

Heusser, L. (1998). Direct correlation of millennial-scale changes in western North American vegetation and climate with changes in the California Current System over the past ~60 kyr. *Paleoceanography*, *13*(3), 252–262. https://doi.org/10.1029/98PA00670

Heusser, L. E., Kirby, M. E., & Nichols, J. E. (2015). Pollen-based evidence of extreme drought during the last Glacial (32.6–9.0 ka) in coastal southern California. *Quaternary Science Reviews*, *126*, 242–253.

Higuchi, M., Sakai, H., & Goto, A. (2014). A new threespine stickleback, Gasterosteus nipponicus sp. nov.(Teleostei: Gasterosteidae), from the Japan Sea region. *Ichthyological Research*, *61*(4), 341–351.

Hinlo, R., Gleeson, D., Lintermans, M., & Furlan, E. (2017). Methods to maximise recovery of environmental DNA from water samples. *PloS One*, *12*(6), e0179251.

Hohenlohe, P. A., & Magalhaes, I. S. (2020). The Population Genomics of Parallel Adaptation: Lessons from Threespine Stickleback. In M. F. Oleksiak & O. P. Rajora (Eds.), *Population Genomics: Marine Organisms* (pp. 249–276). Springer International Publishing. https://doi.org/10.1007/13836_2019_67

Indjeian, V. B., Kingman, G. A., Jones, F. C., Guenther, C. A., Grimwood, J., Schmutz, J., Myers, R. M., & Kingsley, D. M. (2016). Evolving new skeletal traits by cis-regulatory changes in bone morphogenetic proteins. *Cell*, *164*(1–2), 45–56.

Jacobs, D. K., Haney, T. A., & Louie, K. D. (2004). Genes, diversity, and geologic process on the Pacific coast. *Annual Review of Earth and Planetary Sciences*, *32*(1), 601–652. https://doi.org/10.1146/annurev.earth.32.092203.122436

Jacobs, D. K., Stein, E. D., & Longcore, T. (2011a). Classification of California Estuaries Based on Natural Closure Patterns: Templates for Restoration and Management Management. *Technical Report, August*, 1–72.

Jacobs, D. K., Stein, E. D., & Longcore, T. (2011b). Classification of California Estuaries Based on Natural Closure Patterns: Templates for Restoration and Management Management. *Technical Report, August*, 1–72.

Jacobs, D., Stein, E. D., & Longcore, T. (2011). *Classification of California Estuaries Based on Natural Closure Patterns: Templates for Restoration and Management* (619.a; p. 72).

Jombart, T., & Ahmed, I. (2011). adegenet 1.3-1: New tools for the analysis of genome-wide SNP data. *Bioinformatics*, *27*(21), 3070–3071. https://doi.org/10.1093/bioinformatics/btr521

Jones, D. H., & John, A. W. G. (1978). A three-spined stickleback, Gasterosteus aculeatus L. from the North Atlantic. *Journal of Fish Biology*, *13*(2), 231–236. https://doi.org/10.1111/j.1095-8649.1978.tb03430.x

Jones, F. C., Grabherr, M. G., Chan, Y. F., Russell, P., Mauceli, E., Johnson, J., Swofford, R., Pirun, M., Zody, M. C., & White, S. (2012). The genomic basis of adaptive evolution in threespine sticklebacks. *Nature*, *484*(7392), 55–61.

Jones, F. C., Grabherr, M. G., Chan, Y. F., Russell, P., Mauceli, E., Johnson, J., Swofford, R., Pirun, M., Zody, M. C., White, S., Birney, E., Searle, S., Schmutz, J., Grimwood, J., Dickson, M. C., Myers, R. M., Miller, C. T., Summers, B. R., Knecht, A. K., … Kingsley, D. M. (2012). The genomic basis of adaptive evolution in threespine sticklebacks. *Nature*, *484*(7392), 55–61. https://doi.org/10.1038/nature10944

Jónsson, H., Ginolhac, A., Schubert, M., Johnson, P. L. F., & Orlando, L. (2013). MapDamage2.0: Fast approximate Bayesian estimates of ancient DNA damage parameters. *Bioinformatics*, *29*(13), 1682–1684. https://doi.org/10.1093/bioinformatics/btt193

Kandlikar, G. (2020). *ranacapa: Utility Functions and "shiny" App for Simple Environmental DNA Visualizations and Analyses* (0.1.0). https://github.com/gauravsk/ranacapa

Kelly, R. P., Shelton, A. O., & Gallego, R. (2019). Understanding PCR Processes to Draw Meaningful Conclusions from Environmental DNA Studies. *Scientific Reports*, *9*(1), 12133. https://doi.org/10.1038/s41598-019-48546-x

Kingman, G. A. R., Vyas, D. N., Jones, F. C., Brady, S. D., Chen, H. I., Reid, K., Milhaven, M., Bertino, T. S., Aguirre, W. E., & Heins, D. C. (2021). Predicting future from past: The genomic basis of recurrent and rapid stickleback evolution. *Science Advances*, *7*(25), eabg5285.

Kirby, M. E., Feakins, S. J., Bonuso, N., Fantozzi, J. M., & Hiner, C. A. (2013). Latest Pleistocene to Holocene hydroclimates from Lake Elsinore, California. *Quaternary Science Reviews*, *76*, 1–15. https://doi.org/10.1016/j.quascirev.2013.05.023

Kircher, M., Sawyer, S., & Meyer, M. (2012). Double indexing overcomes inaccuracies in multiplex sequencing on the Illumina platform. *Nucleic Acids Research*, *40*(1), e3–e3. https://doi.org/10.1093/nar/gkr771

Konijnendijk, N., Shikano, T., Daneels, D., Volckaert, F. A., & Raeymaekers, J. A. (2015). Signatures of selection in the three-spined stickleback along a small-scale brackish water–freshwater transition zone. *Ecology and Evolution*, *5*(18), 4174–4186.

Kopelman, N. M., Mayzel, J., Jakobsson, M., Rosenberg, N. A., & Mayrose, I. (2015). Clumpak: A program for identifying clustering modes and packaging population structure inferences across K. *Molecular Ecology Resources*, *15*(5), 1179–1191. https://doi.org/10.1111/1755-0998.12387

Kumar, G., Farrell, E., Reaume, A. M., Eble, J. A., & Gaither, M. R. (2022). One size does not fit all: Tuning eDNA protocols for high- and low-turbidity water sampling. *Environmental DNA*, *4*(1), 167–180. https://doi.org/10.1002/edn3.235

Kwambana, B. A., Mohammed, N. I., Jeffries, D., Barer, M., Adegbola, R. A., & Antonio, M. (2011). Differential effects of frozen storage on the molecular detection of bacterial taxa

that inhabit the nasopharynx. *BMC Clinical Pathology*, *11*(1), 2. https://doi.org/10.1186/1472-6890-11-2

Lai, Y.-T., Yeung, C. K., Omland, K. E., Pang, E.-L., Hao, Y., Liao, B.-Y., Cao, H.-F., Zhang, B.-W., Yeh, C.-F., & Hung, C.-M. (2019). Standing genetic variation as the predominant source for adaptation of a songbird. *Proceedings of the National Academy of Sciences*, *116*(6), 2152–2157.

Langmead, B., & Salzberg, S. L. (2012). Fast gapped-read alignment with Bowtie 2. *Nature Methods*, *9*(4), 357–359. https://doi.org/10.1038/nmeth.1923

Laramie, M. B., Pilliod, D. S., Goldberg, C. S., & Strickler, K. M. (2015). Environmental DNA sampling protocol—Filtering water to capture DNA from aquatic organisms. *U.S Geological Survey Techniques and Methods*, *Book 2*(Chapter A13), 15 p. https://doi.org/10.3133/TM2A13

Larsson, A. J. M., Stanley, G., Sinha, R., Weissman, I. L., & Sandberg, R. (2018). Computational correction of index switching in multiplexed sequencing libraries. *Nature Methods*, *15*(5), 305–307. https://doi.org/10.1038/nmeth.4666

Leidy, R. A., & Moyle, P. B. (2021). Keeping up with the status of freshwater fishes: A California (USA) perspective. *Conservation Science and Practice*, *3*(8). https://doi.org/10.1111/csp2.474

Levick, L. R., Goodrich, D. C., Hernandez, M., Fonseca, J., Semmens, D. J., Stromberg, J. C., Tluczek, M., Leidy, R. A., Scianni, M., & Guertin, D. P. (2008). *The ecological and hydrological significance of ephemeral and intermittent streams in the arid and semi-arid American Southwest*. Environmental Protection Agency, Office of Research and Development.

Levy-Booth, D. J., Campbell, R. G., Gulden, R. H., Hart, M. M., Powell, J. R., Klironomos, J. N., Pauls, K. P., Swanton, C. J., Trevors, J. T., & Dunfield, K. E. (2007). Cycling of extracellular DNA in the soil environment. *Soil Biology and Biochemistry*, *39*(12), 2977–2991. https://doi.org/10.1016/j.soilbio.2007.06.020

Li, H. (2013a). *Aligning sequence reads, clone sequences and assembly contigs with BWA-MEM*. *00*(00), 1–3. https://doi.org/10.48550/arXiv.1303.3997

Li, H. (2013b). *Aligning sequence reads, clone sequences and assembly contigs with BWA-MEM*. *00*(00), 1–3. https://doi.org/arXiv:1303.3997 [q-bio.GN]

Li, H., & Durbin, R. (2011). Inference of human population history from individual whole-genome sequences. *Nature*, *475*(7357), 493–496. https://doi.org/10.1038/nature10231

Li, J., Lawson Handley, L.-J., Read, D. S., & Hänfling, B. (2018). The effect of filtration method on the efficiency of environmental DNA capture and quantification via metabarcoding. *Molecular Ecology Resources*, *18*(5), 1102–1114.

Liang, Z., & Keeley, A. (2013). Filtration Recovery of Extracellular DNA from Environmental Water Samples. *Environmental Science & Technology*, *47*(16), 9324–9331. https://doi.org/10.1021/es401342b

Lipson, M. (2020). Applying f_4 -statistics and admixture graphs: Theory and examples. *Molecular Ecology Resources*, *20*(6), 1658–1667. https://doi.org/10.1111/1755-0998.13230

Liu, S., Hansen, M. M., & Jacobsen, M. W. (2016). Region-wide and ecotype-specific differences in demographic histories of threespine stickleback populations, estimated from whole genome sequences. *Molecular Ecology*, *25*(20), 5187–5202. https://doi.org/10.1111/mec.13827

Love, M. I., Huber, W., & Anders, S. (2014). Moderated estimation of fold change and dispersion for RNA-seq data with DESeq2. *Genome Biology*, *15*(12), 550. https://doi.org/10.1186/s13059-014-0550-8

Maier, R., Flegontov, P., Flegontova, O., Changmai, P., & Reich, D. (2022). *On the limits of fitting complex models of population history to genetic data* (p. 2022.05.08.491072). bioRxiv. https://doi.org/10.1101/2022.05.08.491072

Majaneva, M., Diserud, O. H., Eagle, S. H., Boström, E., Hajibabaei, M., & Ekrem, T. (2018). Environmental DNA filtration techniques affect recovered biodiversity. *Scientific Reports*, *8*(1), 1–11.

Mäkinen, H. S., Cano, J. M., & Merilä, J. (2006). Genetic relationships among marine and freshwater populations of the European three-spined stickleback (Gasterosteus aculeatus) revealed by microsatellites. *Molecular Ecology*, *15*(6), 1519–1534.

Malcolm, J. R. (1992). *Supporting information for a petition to list as endangered or threatened: Shay Creek stickleback, Gasterosteus sp* (P. B. Moyle & R. M. Yoshiyama, Eds.; pp. 213–222).

Marques, D. A., Jones, F. C., Di Palma, F., Kingsley, D. M., & Reimchen, T. E. (2018). Experimental evidence for rapid genomic adaptation to a new niche in an adaptive radiation. *Nature Ecology & Evolution*, *2*(7), 1128–1138. https://doi.org/10.1038/s41559-018-0581-8

Martin, M. (2013). Cutadapt removes adapter sequences from high-throughput sequencing reads. *EMBnet.Journal*, *17*(1), 10–12.

Martin, S. H., Davey, J. W., & Jiggins, C. D. (2015). Evaluating the Use of ABBA–BABA Statistics to Locate Introgressed Loci. *Molecular Biology and Evolution*, *32*(1), 244–257. https://doi.org/10.1093/molbev/msu269

Masters, P. M. (2006). Holocene sand beaches of southern California: ENSO forcing and coastal processes on millennial scales. *Palaeogeography, Palaeoclimatology, Palaeoecology*, *232*(1), 73–95.

Mather, N., Traves, S. M., & Ho, S. Y. W. (2020). A practical introduction to sequentially Markovian coalescent methods for estimating demographic history from genomic data. *Ecology and Evolution*, *10*(1), 579–589. https://doi.org/10.1002/ece3.5888

Mattern, M. Y. (2004). Molecular phylogeny of the Gasterosteidae: The importance of using multiple genes. *Molecular Phylogenetics and Evolution*, *30*(2), 366–377.

Mazet, O., Rodríguez, W., Grusea, S., Boitard, S., & Chikhi, L. (2016). On the importance of being structured: Instantaneous coalescence rates and human evolution—Lessons for ancestral population size inference? *Heredity*, *116*(4), 362–371. https://doi.org/10.1038/hdy.2015.104

Mazzarella, A. B., Boessenkool, S., Østbye, K., Vøllestad, L. A., & Trucchi, E. (2016). Genomic signatures of the plateless phenotype in the threespine stickleback. *Ecology and Evolution*, *6*(10), 3161–3173. https://doi.org/10.1002/ece3.2072

McKenna, A., Hanna, M., Banks, E., Sivachenko, A., Cibulskis, K., Kernytsky, A., Garimella, K., Altshuler, D., Gabriel, S., Daly, M., & DePristo, M. A. (2010). The Genome Analysis Toolkit: A MapReduce framework for analyzing next-generation DNA sequencing data. *Genome Research*, *20*(9), 1297–1303. https://doi.org/10.1101/gr.107524.110

McMurdie, P. J., & Holmes, S. (2013). phyloseq: An R Package for Reproducible Interactive Analysis and Graphics of Microbiome Census Data. *PLoS ONE*, *8*(4), e61217. https://doi.org/10.1371/journal.pone.0061217

McMurdie, P. J., & Holmes, S. (2014). Waste Not, Want Not: Why Rarefying Microbiome Data Is Inadmissible. *PLoS Computational Biology*, *10*(4), e1003531. https://doi.org/10.1371/journal.pcbi.1003531

Meier, J. I., Marques, D. A., Wagner, C. E., Excoffier, L., & Seehausen, O. (2018). Genomics of Parallel Ecological Speciation in Lake Victoria Cichlids. *Molecular Biology and Evolution*, *35*(6), 1489–1506. https://doi.org/10.1093/molbev/msy051

Meirmans, P. G. (2020). GenoDive version 3.0: Easy-to-use software for the analysis of genetic data of diploids and polyploids. *Molecular Ecology Resources*, *20*(4), 1126–1131. https://doi.org/10.1111/1755-0998.13145

Mendenhall, W. C. (1908). Ground waters and irrigation enterprises in the foothill belt, southern California. In *Water Supply Paper* (No. 219; Water Supply Paper, p. 180). https://doi.org/10.3133/wsp219

Miller, R. R. (1961). *Man and the changing fish fauna of the American southwest*. *46*, 365–404.

Miller, R. R., & Hubbs, C. L. (1969). Systematics of Gasterosteus aculeatus, with Particular Reference to Intergradation and Introgression along the Pacific Coast of North America: A Commentary on a Recent Contribution. *Copeia*, *1*(1), 52–69.

Minckley, W. L., & Marsh, P. C. (2009). *Inland Fishes of the Greater Southwest: Chronicle of a Vanishing Biota*. University of Arizona Press.

Morris, M. R., Bowles, E., Allen, B. E., Jamniczky, H. A., & Rogers, S. M. (2018). Contemporary ancestor? Adaptive divergence from standing genetic variation in Pacific marine threespine stickleback. *BMC Evolutionary Biology*, *18*(1), 1–21.

Mount, J. F. (1995). *California Rivers and Streams: The Conflict Between Fluvial Process and Land Use*. University of California Press.

Moyle, P. B. (2002). *Inland Fishes of California: Revised and Expanded*. University of California Press.

Moyle, P. B., Quiñones, R. M., Katz, J. V., & Weaver, J. (2015). *Fish species of special concern in California* (p. 842). California Department of Fish and Wildlife. https://wildlife.ca.gov/Conservation/SSC/Fishes

Nadachowska-Brzyska, K., Burri, R., Smeds, L., & Ellegren, H. (2016). PSMC analysis of effective population sizes in molecular ecology and its application to black-and-white Ficedula flycatchers. *Molecular Ecology*, *25*(5), 1058–1072. https://doi.org/10.1111/mec.13540

Nath, S., Shaw, D. E., & White, M. A. (2020). Improved contiguity of the threespine stickleback genome using long-read sequencing. *BioRxiv*, 2020.06.30.170787. https://doi.org/10.1101/2020.06.30.170787

Nico, L., Fuller, P., & Neilson, M. (2022). *Acanthogobius flavimanus (Temminck and Schlegel, 1845)*. U.S. Geological Survey, Nonindigenous Aquatic Species Database. https://nas.er.usgs.gov/queries/FactSheet.aspx?speciesID=707

Noor, M. a. F., & Bennett, S. M. (2009). Islands of speciation or mirages in the desert? Examining the role of restricted recombination in maintaining species. *Heredity*, *103*(6), Article 6. https://doi.org/10.1038/hdy.2009.151

O'Brown, N. M., Summers, B. R., Jones, F. C., Brady, S. D., & Kingsley, D. M. (2015a). A recurrent regulatory change underlying altered expression and Wnt response of the stickleback armor plates gene EDA. *Elife*, *4*, e05290.

O'Brown, N. M., Summers, B. R., Jones, F. C., Brady, S. D., & Kingsley, D. M. (2015b). A recurrent regulatory change underlying altered expression and Wnt response of the stickleback armor plates gene EDA. *ELife*, *4*(4), 1–17. https://doi.org/10.7554/eLife.05290

Oksanen, J., Blanchet, F. G., Friendly, M., Kindt, R., Legendre, P., McGlinn, D., Minchin, P. R., O'Hara, R. B., Simpson, G. L., Solymos, P., Stevens, M. H. H., Szoecs, E., & Wagner, H. (2019). *vegan: Community Ecology Package* (2.5-6). https://CRAN.R-project.org/package=vegan

Orsi, J. P. (2004). *Hazardous metropolis: Flooding and urban ecology in Los Angeles* [Ph.D., The University of Wisconsin - Madison]. https://www.proquest.com/docview/304537402/abstract/68D98B12099C4AE4PQ/1

Owen, L. A., Finkel, R. C., Minnich, R. A., & Perez, A. E. (2003). Extreme southwestern margin of late Quaternary glaciation in North America: Timing and controls. *Geology*, *31*(8), 729–732.

Paccard, A., Wasserman, B. A., Hanson, D., Astorg, L., Durston, D., Kurland, S., Apgar, T. M., El-Sabaawi, R. W., Palkovacs, E. P., & Hendry, A. P. (2018). Adaptation in temporally variable environments: Stickleback armor in periodically breaching bar-built estuaries. *Journal of Evolutionary Biology*, *31*(5), 735–752.

Patterson, N., Moorjani, P., Luo, Y., Mallick, S., Rohland, N., Zhan, Y., Genschoreck, T., Webster, T., & Reich, D. (2012). Ancient Admixture in Human History. *Genetics, 192*(3), 1065–1093. https://doi.org/10.1534/genetics.112.145037

Pawlowski, J., Bruce, K., Panksep, K., Aguirre, F. I., Amalfitano, S., Apothéloz-Perret-Gentil, L., Baussant, T., Bouchez, A., Carugati, L., Cermakova, K., Cordier, T., Corinaldesi, C., Costa, F. O., Danovaro, R., Dell'Anno, A., Duarte, S., Eisendle, U., Ferrari, B. J. D., Frontalini, F., … Fazi, S. (2022). Environmental DNA metabarcoding for benthic monitoring: A review of sediment sampling and DNA extraction methods. *Science of The Total Environment*, *818*, 151783. https://doi.org/10.1016/j.scitotenv.2021.151783

Peichel, C. L., & Marques, D. A. (2017). The genetic and molecular architecture of phenotypic diversity in sticklebacks. *Philosophical Transactions of the Royal Society B: Biological Sciences, 372*(1713), 20150486. https://doi.org/10.1098/rstb.2015.0486

Perkins, T. L., Clements, K., Baas, J. H., Jago, C. F., Jones, D. L., Malham, S. K., & McDonald, J. E. (2014). Sediment Composition Influences Spatial Variation in the Abundance of Human Pathogen Indicator Bacteria within an Estuarine Environment. *PLoS ONE, 9*(11), e112951. https://doi.org/10.1371/journal.pone.0112951

Pickrell, J., & Pritchard, J. (2012). Inference of population splits and mixtures from genome-wide allele frequency data. *Nature Precedings*, 1–1.

Pietramellara, G., Ascher, J., Borgogni, F., Ceccherini, M. T., Guerri, G., & Nannipieri, P. (2009). Extracellular DNA in soil and sediment: Fate and ecological relevance. *Biology and Fertility of Soils, 45*(3), 219–235. https://doi.org/10.1007/s00374-008-0345-8

Pilliod, D. S., Goldberg, C. S., Arkle, R. S., Waits, L. P., & Richardson, J. (2013). Estimating occupancy and abundance of stream amphibians using environmental DNA from filtered

water samples. *Canadian Journal of Fisheries and Aquatic Sciences*, *70*(8), 1123–1130. https://doi.org/10.1139/cjfas-2013-0047

Port, J. A., O'Donnell, J. L., Romero-Maraccini, O. C., Leary, P. R., Litvin, S. Y., Nickols, K. J., Yamahara, K. M., & Kelly, R. P. (2016). Assessing vertebrate biodiversity in a kelp forest ecosystem using environmental DNA. *Molecular Ecology*, *25*(2), 527–541. https://doi.org/10.1111/mec.13481

Pringle, C. M., Freeman, M. C., & Freeman, B. J. (2000). Regional Effects of Hydrologic Alterations on Riverine Macrobiota in the New World: Tropical-Temperate Comparisons: The massive scope of large dams and other hydrologic modifications in the temperate New World has resulted in distinct regional trends of biotic impoverishment. While neotropical rivers have fewer dams and limited data upon which to make regional generalizations, they are ecologically vulnerable to increasing hydropower development and biotic patterns are emerging. *BioScience*, *50*(9), 807–823. https://doi.org/10.1641/0006-3568(2000)050[0807:REOHAO]2.0.CO;2

Pritchard, J. K., Stephens, M., & Donnelly, P. (2000). Inference of Population Structure Using Multilocus Genotype Data. *Genetics*, *155*(2), 945 LP – 959.

Propst, D. L., Williams, J. E., Bestgen, K. R., & Hoagstrom, C. W. (2021). *Standing between Life and Extinction: Ethics and Ecology of Conserving Aquatic Species in North American Deserts*. University of Chicago Press.

R Core Team. (2018a). *R: A language and environment for statistical computing*. R Foundation for Statistical Computing. https://www.R-project.org/

R Core Team. (2018b). *R: A language and environment for statistical computing*. R Foundation for Statistical Computing. https://www.R-project.org/

Raj, A., Stephens, M., & Pritchard, J. K. (2014). FastSTRUCTURE: Variational Inference of Population Structure in Large SNP Data Sets. *Genetics*, *197*(2), 573–589. https://doi.org/10.1534/genetics.114.164350

Ravinet, M., Faria, R., Butlin, R. K., Galindo, J., Bierne, N., Rafajlović, M., Noor, M. a. F., Mehlig, B., & Westram, A. M. (2017). Interpreting the genomic landscape of speciation: A road map for finding barriers to gene flow. *Journal of Evolutionary Biology*, *30*(8), 1450–1477. https://doi.org/10.1111/jeb.13047

Ravinet, M., Prodöhl, P. A., & Harrod, C. (2013). Parallel and nonparallel ecological, morphological and genetic divergence in lake–stream stickleback from a single catchment. *Journal of Evolutionary Biology*, *26*(1), 186–204.

Rees, H. C., Maddison, B. C., Middleditch, D. J., Patmore, J. R. M., & Gough, K. C. (2014). REVIEW: The detection of aquatic animal species using environmental DNA – a review of eDNA as a survey tool in ecology. *Journal of Applied Ecology*, *51*(5), 1450–1459. https://doi.org/10.1111/1365-2664.12306

Reid, K., Bell, M. A., & Veeramah, K. R. (2021). Threespine Stickleback: A Model System For Evolutionary Genomics. *Annual Review of Genomics and Human Genetics*, *22*(1), 357–383. https://doi.org/10.1146/annurev-genom-111720-081402

Reimchen, T. E. (1994). Predators and evolution in threespine stickleback. In M. A. Bell & S. A. Foster (Eds.), *Evolution of the threespine stickleback* (pp. 240–273). Oxford University Press.

Reimchen, T. E. (2000). Predator handling failures of lateral plate morphs in Gasterosteus aculeatus: Functional implications for the ancestral plate condition. *Behaviour*, *137*(7–8), 1081–1096. https://doi.org/10.1163/156853900502448

Ricciardi, A., & Rasmussen, J. B. (1999). Extinction Rates of North American Freshwater Fauna. *Conservation Biology*, *13*(5), 1220–1222. https://doi.org/10.1046/j.1523-1739.1999.98380.x

Richmond, J. Q., Backlin, A. R., Galst-Cavalcante, C., O'Brien, J. W., & Fisher, R. N. (2018). Loss of dendritic connectivity in southern California's urban riverscape facilitates decline of an endemic freshwater fish. *Molecular Ecology*, *27*(2), 369–386. https://doi.org/10.1111/mec.14445

Richmond, J. Q., Jacobs, D. K., Backlin, A. R., Swift, C. C., Dellith, C., & Fisher, R. N. (2015). Ephemeral stream reaches preserve the evolutionary and distributional history of threespine stickleback in the Santa Clara and Ventura River watersheds of southern California. *Conservation Genetics*, *16*(1), 85–101. https://doi.org/10.1007/s10592-014-0643-7

Robinson, J. A., Brown, C., Kim, B. Y., Lohmueller, K. E., & Wayne, R. K. (2018). Purging of Strongly Deleterious Mutations Explains Long-Term Persistence and Absence of Inbreeding Depression in Island Foxes. *Current Biology*, *28*(21), 3487-3494.e4. https://doi.org/10.1016/j.cub.2018.08.066

Robson, H. L. A., Noble, T. H., Saunders, R. J., Robson, S. K. A., Burrows, D. W., & Jerry, D. R. (2016). Fine-tuning for the tropics: Application of eDNA technology for invasive fish detection in tropical freshwater ecosystems. *Molecular Ecology Resources*, *16*(4), 922–932. https://doi.org/10.1111/1755-0998.12505

Rollins, J. L. (2017). Body-size and growth-rate divergence among populations of threespine stickleback (*Gasterosteus aculeatus*) in Cook Inlet, Alaska, USA. *Canadian Journal of Zoology*, *95*(11), 877–884. https://doi.org/10.1139/cjz-2017-0092

RStudio Team. (2020a). *RStudio: Integrated Development for R*. RStudio, PBC. http://www.rstudio.com/

RStudio Team. (2020b). *RStudio: Integrated Development for R*. RStudio, PBC. http://www.rstudio.com/

Ruiz-Campos, G., Camarena-Rosales, F., González-Acosta, A. F., Maeda-Martinez, A. M., García de León, F. J., Varela-Romero, A., & Andreu-Soler, A. (2014). Estatus actual de conservación de seis especies de peces dulceacuícolas de la península de Baja California, México. *Revista Mexicana de Biodiversidad*, *85*(4), 1235–1248. https://doi.org/10.7550/rmb.43747

Ruiz-Campos, G., & González-Acosta, A. F. (in press). *Peces de las aguas continentales de Baja California, México*. Universidad Autónoma de Baja California.

Sales, N. G., Wangensteen, O. S., Carvalho, D. C., & Mariani, S. (2019). Influence of preservation methods, sample medium and sampling time on eDNA recovery in a neotropical river. *Environmental DNA*, *1*(2), edn3.14. https://doi.org/10.1002/edn3.14

Samuels, D. C., Wang, J., Ye, F., He, J., Levinson, R. T., Sheng, Q., Zhao, S., Capra, J. A., Shyr, Y., Zheng, W., & Guo, Y. (2016). Heterozygosity Ratio, a Robust Global Genomic Measure of Autozygosity and Its Association with Height and Disease Risk. *Genetics*, *204*(3), 893–904. https://doi.org/10.1534/genetics.116.189936

Santoro, A. E., Dupont, C. L., Richter, R. A., Craig, M. T., Carini, P., McIlvin, M. R., Yang, Y., Orsi, W. D., Moran, D. M., & Saito, M. A. (2015). Genomic and proteomic characterization of "Candidatus Nitrosopelagicus brevis": An ammonia-oxidizing archaeon from the open ocean. *Proceedings of the National Academy of Sciences*, *112*(4), 1173–1178. https://doi.org/10.1073/pnas.1416223112

Sard, N. M., Herbst, S. J., Nathan, L., Uhrig, G., Kanefsky, J., Robinson, J. D., & Scribner, K. T. (2019). Comparison of fish detections, community diversity, and relative abundance using environmental DNA metabarcoding and traditional gears. *Environmental DNA*, *1*(4), 368–384. https://doi.org/10.1002/edn3.38

Schaarschmidt, F., & Gerhard, D. (2019). *PairwiseCI: Confidence Intervals for Two Sample Comparisons* (0.1-27). https://CRAN.R-project.org/package=pairwiseCI

Schiffels, S., & Durbin, R. (2014). Inferring human population size and separation history from multiple genome sequences. *Nature Genetics*, *46*(8), 919–925. https://doi.org/10.1038/ng.3015

Schluter, D. (2000). *The Ecology of Adaptive Radiation*. OUP Oxford.

Schluter, D., & Nagel, L. M. (1995). Parallel Speciation by Natural Selection. *The American Naturalist*, *146*(2), 292–301. https://doi.org/10.1086/285799

Schnell, I. B., Bohmann, K., & Gilbert, M. T. P. (2015). Tag jumps illuminated—Reducing sequence-to-sample misidentifications in metabarcoding studies. *Molecular Ecology Resources*, *15*(6), 1289–1303. https://doi.org/10.1111/1755-0998.12402

Schrider, D. R., Shanku, A. G., & Kern, A. D. (2016). Effects of linked selective sweeps on demographic inference and model selection. *Genetics*, *204*(3), 1207–1223.

SCWRP. (2018). *Wetlands on the Edge: The Future of Southern California's Wetlands: Regional Strategy 2018* (p. 142). California State Coastal Conservancy. scwrp.databasin.org

Sekar, R., Kaczmarsky, L. T., & Richardson, L. L. (2009). Effect of Freezing on PCR Amplification of 16S rRNA Genes from Microbes Associated with Black Band Disease

of Corals. *Applied and Environmental Microbiology*, *75*(8), 2581–2584. https://doi.org/10.1128/AEM.01500-08

Shaffer, H. B., Fellers, G. M., Randal Voss, S., Oliver, J. C., & Pauly, G. B. (2004). Species boundaries, phylogeography and conservation genetics of the red-legged frog (Rana aurora/draytonii) complex. *Molecular Ecology*, *13*(9), 2667–2677. https://doi.org/10.1111/j.1365-294X.2004.02285.x

Shanfelter, A. F., Archambeault, S. L., & White, M. A. (2019). Divergent fine-scale recombination landscapes between a freshwater and marine population of threespine stickleback fish. *Genome Biology and Evolution*, *11*(6), 1573.

Sharp, R. P., Allen, C. R., & Meier, M. F. (1959). Pleistocene glaciers on southern California mountains. *American Journal of Science*, *257*(2), 81–94.

Shirazi, S., Meyer, R. S., & Shapiro, B. (2021). Revisiting the effect of PCR replication and sequencing depth on biodiversity metrics in environmental DNA metabarcoding. *Ecology and Evolution*, *11*(22), 15766–15779. https://doi.org/10.1002/ece3.8239

Smart, A. S., Weeks, A. R., Rooyen, A. R., Moore, A., McCarthy, M. A., & Tingley, R. (2016). Assessing the cost-efficiency of environmental DNA sampling. *Methods in Ecology and Evolution*, *7*(11), 1291–1298. https://doi.org/10.1111/2041-210X.12598

Stein, E. D., Cayce, K., Salomon, M., Bram, D. L., De Mello, D., Grossinger, R., & Dark, S. (2014). *Wetlands of the Southern California Coast: Historical Extent and Change Over Time* (SFEI Report 720; SCCWRP Technical Report 826; p. 58). Southern California Coastal Water Research Project and San Francisco Estuary Institute. https://www.caltsheets.org/socal/download.html

Stephenson, J. R., & Calcarone, G. M. (1999). *Southern California Mountains and Foothills Assessment: Habitat and Species Conservation Issues*. U.S. Department of Agriculture, Forest Service, Pacific Southwest Research Station.

Stiller, J., da Fonseca, R. R., Alfaro, M. E., Faircloth, B. C., Wilson, N. G., & Rouse, G. W. (2021). Using ultraconserved elements to track the influence of sea-level change on leafy seadragon populations. *Molecular Ecology*, *30*(6), 1364–1380.

Suomalainen, L.-R., Reunanen, H., Ijäs, R., Valtonen, E. T., & Tiirola, M. (2006). Freezing Induces Biased Results in the Molecular Detection of *Flavobacterium columnare*. *Applied and Environmental Microbiology*, *72*(2), 1702–1704. https://doi.org/10.1128/AEM.72.2.1702-1704.2006

Swift, C. C., Haglund, T. R., Ruiz, M., & Fisher, R. N. (1993a). The Status and Distribution of the Freshwater Fishes of Southern California. *Bulletin of the Southern California Academy of Sciences*, *92*(3), 101–167.

Swift, C. C., Haglund, T. R., Ruiz, M., & Fisher, R. N. (1993b). The Status and Distribution of the Freshwater Fishes of Southern California. *Bulletin of the Southern California Academy of Sciences*, *92*(3), 101–167.

Swift, C. C., Spies, B., Ellingson, R. A., & Jacobs, D. K. (2016). A New Species of the Bay Goby Genus Eucyclogobius, Endemic to Southern California: Evolution, Conservation, and Decline. *PloS One*, *11*(7), e0158543. https://doi.org/10.1371/journal.pone.0158543

Taberlet, P., Coissac, E., Pompanon, F., Brochmann, C., & Willerslev, E. (2012). Towards next-generation biodiversity assessment using DNA metabarcoding: NEXT-GENERATION DNA METABARCODING. *Molecular Ecology*, *21*(8), 2045–2050. https://doi.org/10.1111/j.1365-294X.2012.05470.x

Taberlet, P., Prud'Homme, S. M., Campione, E., Roy, J., Miquel, C., Shehzad, W., Gielly, L., Rioux, D., Choler, P., Clément, J.-C., Melodelima, C., Pompanon, F., & Coissac, E. (2012). Soil sampling and isolation of extracellular DNA from large amount of starting material suitable for metabarcoding studies: EXTRACTION OF EXTRACELLULAR DNA FROM SOIL. *Molecular Ecology*, *21*(8), 1816–1820. https://doi.org/10.1111/j.1365-294X.2011.05317.x

Takahara, T., Minamoto, T., & Doi, H. (2015). Effects of sample processing on the detection rate of environmental DNA from the Common Carp (Cyprinus carpio). *Biological Conservation*, *183*, 64–69. https://doi.org/10.1016/j.biocon.2014.11.014

Takahata, N., & Nei, M. (1985). Gene Genealogy and Variance of Interpopulational Nucleotide Differences. *Genetics*, *110*(2), 325–344.

Thomsen, P. F., & Willerslev, E. (2015). Environmental DNA – An emerging tool in conservation for monitoring past and present biodiversity. *Biological Conservation*, *183*, 4–18. https://doi.org/10.1016/j.biocon.2014.11.019

Torti, A., Lever, M. A., & Jørgensen, B. B. (2015). Origin, dynamics, and implications of extracellular DNA pools in marine sediments. *Marine Genomics*, *24*, 185–196. https://doi.org/10.1016/j.margen.2015.08.007

Tsuji, S., Takahara, T., Doi, H., Shibata, N., & Yamanaka, H. (2019). The detection of aquatic macroorganisms using environmental DNA analysis—A review of methods for collection, extraction, and detection. *Environmental DNA*, *1*(2), 99–108. https://doi.org/10.1002/edn3.21

Turba, R., Richmond, J. Q., Fitz-Gibbon, S., Morselli, M., Fisher, R. N., Swift, C. C., Ruiz-Campos, G., Backlin, A. R., Dellith, C., & Jacobs, D. K. (2022). Genetic structure and

historic demography of endangered unarmoured threespine stickleback at southern latitudes signals a potential new management approach. *Molecular Ecology*, mec.16722. https://doi.org/10.1111/mec.16722

Turner, C. R., Barnes, M. A., Xu, C. C. Y., Jones, S. E., Jerde, C. L., & Lodge, D. M. (2014). Particle size distribution and optimal capture of aqueous macrobial eDNA. *Methods in Ecology and Evolution*, *5*(7), 676–684. https://doi.org/10.1111/2041-210X.12206

Turner, C. R., Miller, D. J., Coyne, K. J., & Corush, J. (2014). Improved methods for capture, extraction, and quantitative assay of environmental DNA from Asian bigheaded carp (hypophthalmichthys spp.). *PLoS ONE*, *9*(12), 1–20. https://doi.org/10.1371/journal.pone.0114329

Turner, C. R., Uy, K. L., & Everhart, R. C. (2015). Fish environmental DNA is more concentrated in aquatic sediments than surface water. *Biological Conservation*, *183*, 93–102. https://doi.org/10.1016/j.biocon.2014.11.017

USFWS. (2021). *Unarmored threespine stickleback (Gasterosteus aculeatus williamsoni) 5-Year Review: Summary and Evaluation*. Ventura Fish and Wildlife Office. https://ecos.fws.gov/ecp/species/7002

van der Loos, L. M., & Nijland, R. (2021). Biases in bulk: DNA metabarcoding of marine communities and the methodology involved. *Molecular Ecology*, *30*(13), 3270–3288. https://doi.org/10.1111/mec.15592

Van Wormer, S. R. (1991). A History of Flood Control in the Los Angeles County Drainage Area. *Southern California Quarterly*, *73*(1), 55–94. https://doi.org/10.2307/41171560

Wasserman, B. A., Paccard, A., Apgar, T. M., Des Roches, S., Barrett, R. D. H., Hendry, A. P., & Palkovacs, E. P. (2020). Ecosystem size shapes antipredator trait evolution in estuarine threespine stickleback. *Oikos*, *129*(12), 1795–1806. https://doi.org/10.1111/oik.07482

WeatherSpark.com. (n.d.). *Historical Weather Summer 2018 at Point Mugu Naval Air Warfare Center*. WeatherSpark.Com. Retrieved June 8, 2022, from https://weatherspark.com/h/s/145310/2018/1/Historical-Weather-Summer-2018-at-Point-Mugu-Naval-Air-Warfare-Center;-California;-United-States#Figures-Rainfall

Weir, B. S., & Cockerham, C. C. (1984). Estimating F-Statistics for the Analysis of Population Structure. *Evolution*, *38*(6), 1358–1370. https://doi.org/10.2307/2408641

Wickham, H. (2016). *ggplot2: Elegant Graphics for Data Analysis*.

Williams, K. E., Huyvaert, K. P., & Piaggio, A. J. (2017). Clearing muddied waters: Capture of environmental DNA from turbid waters. *PLOS ONE*, *12*(7), e0179282. https://doi.org/10.1371/journal.pone.0179282

Wolf, J. B. W., & Ellegren, H. (2017). Making sense of genomic islands of differentiation in light of speciation. *Nature Reviews Genetics*, *18*(2), Article 2. https://doi.org/10.1038/nrg.2016.133

Zheng, X., Levine, D., Shen, J., Gogarten, S. M., Laurie, C., & Weir, B. S. (2012). A high-performance computing toolset for relatedness and principal component analysis of SNP data. *Bioinformatics*, *28*(24), 3326–3328.

Zinger, L., Lionnet, C., Benoiston, A.-S., Donald, J., Mercier, C., & Boyer, F. (2020). *metabaR: An R package for the evaluation and improvement of DNA metabarcoding data quality* [Preprint]. Bioinformatics. https://doi.org/10.1101/2020.08.28.271817