UNIVERSITY OF CALIFORNIA
SANTA CRUZ

GENOMICS OF CALIFORNIA AND SEA OF CORTEZ FISHES: SPECIATION, POPULATION GENOMICS, AND LOCAL ADAPTATION

A dissertation submitted in partial satisfaction
of the requirements for the degree of

DOCTOR OF PHILOSOPHY

in

ECOLOGY AND EVOLUTIONARY BIOLOGY

by

Allyson C. Salazar Sawkins

August 2026

The Dissertation of Allyson C. Salazar Sawkins is approved:

_____
Professor Giacomo Bernardi, chair

_____
Professor Peter T. Raimondi

_____
Professor Cheryl A. Logan

_____
Devon E. Pearse, Ph.D.

_____
Donald Smith
Interim Vice Provost and Dean of Graduate Studies

Table of Contents

List of Figures & Tables

CHAPTER 1

CHAPTER 2

CHAPTER 3

Abstract

GENOMICS OF CALIFORNIA AND SEA OF CORTEZ FISHES: SPECIATION, POPULATION GENOMICS, AND LOCAL ADAPTATION

by Allyson C. Salazar Sawkins

Marine populations were long expected to exhibit weak genetic structure because many species possess pelagic larvae, large population sizes, and few barriers to dispersal. However, strong population structure is shown to be widespread across marine taxa even among species with high dispersal potential. Understanding how population divergence arises and persists in such dynamic systems remains a central challenge in evolutionary biology. This dissertation examines how historical isolation, environmental gradients, and selection shape population structure in marine fishes using physiological and genomic approaches. **In Chapter 1**, I examine population-level variation in thermal physiology between Pacific and Sea of Cortez populations of the longjaw mudsucker (*Gillichthys mirabilis*). Although broad thermal performance curves were largely similar, Sea of Cortez individuals showed lower thermal sensitivity and higher acute thermal tolerance after warm pre-conditioning, patterns that mirror a more thermally variable environment, though adaptation cannot be confirmed from this comparison alone. **In Chapter 2**, I use

low-coverage whole-genome resequencing to assess genomic divergence between Pacific and Sea of Cortez populations of *G. mirabilis*. I find exceptionally high genomic differentiation, extensive fixed differences, and reciprocal phylogenetic monophyly between lineages, consistent with long-term allopatric isolation. Genome scans identify outlier loci enriched for nonsynonymous substitutions, suggesting selection acted alongside drift to shape divergence between these lineages. **In Chapter 3**, I investigate population genomic patterns in the barred surfperch (*Amphistichus argenteus*), a nearshore species lacking a pelagic larval stage. Genomic data reveal three distinct coastal clades and a highly divergent Santa Rosa Island lineage, shaped by isolation by distance and geographic breaks near major habitat discontinuities. The largest coastal break is across the Palos Verdes Peninsula, where candidate genes under selection are enriched for functions related to spermatogenesis and sperm motility, suggesting this break is maintained by selection on reproductive compatibility. Together, these results show contemporary population structure reflects the cumulative effects of historical isolation, ecological context, and selection across spatial and temporal scales. This work provides a comparative framework for understanding how population divergence arises and persists in marine environments and informs predictions of responses to environmental change.

To my little field mate, I will forever cherish these memories with you.

Elkhorn Slough, CA, USA

Puerto Peñasco, Sonora, Mexico

Acknowledgements

Advisor and Committee

I'd like to thank my advisor, Giacomo Bernardi, for welcoming me into his lab and for the years of mentorship that shaped this work. It has been the utmost pleasure to learn from him. Throughout my doctorate, I struggled to believe I could be a scientist, but Giacomo never hesitated to cheer me on and advocate for me. They say that, "people may not remember what you say, but they'll remember how you made them feel". Giacomo made me feel like I could do this. I am also deeply grateful to my committee, Pete Raimondi, Cheryl Logan, and Devon Pearse, for their time, questions, and insightful feedback at every stage. This dissertation is far stronger for their input.

Funding Sources

This work was made possible by the support of the NSF Graduate Research Fellowship, Eugene Cota-Robles Fellowship, Institute of Marine Sciences at the University of California, Santa Cruz, Dr. Earl H. Myers and Ethel M. Myers Oceanographic and Marine Biology Trust, and Activating Innovative Graduate Research Award and the Graduate Dean's Travel Grant from University of California, Santa Cruz. I am thankful for the fieldwork, conference, and research opportunities this funding provided.

Undergraduate Mentors

To my undergraduate mentors who inspired me to go down this path, Erica Jarvis Mason and Andrew Thompson, I thought of you often as I mentored my own

students. I'm so grateful for your thoughtful mentorship and encouragement, which led me to where I am today. I'll spend the rest of my career trying to pay forward the compassion and care you gave me. I would also like to thank Noelle Bowlin, John Hyde, Matt Craig, Brice Semmens, and Dovi Kacev for your support and encouragement during my senior thesis.

Undergraduate Mentees

To all of my undergraduate mentees past and present, I cannot thank you enough for your time and efforts in making these projects possible. To Shiqi (Emma) Yang and Raven Long, my most dedicated fieldmates, it was the utmost pleasure working together. I will cherish our memories of collecting fishes all over this state for life.

Colleagues

To my labmates in the Bernardi Lab past and present, you made this experience much richer. To Eric Witte and Shinji Yamamoto, I thank you for your unwavering support and encouragement. I'll miss our antics and mischievousness. To Natasha Hinojosa, my productivity partner, office mate, and close friend, I couldn't have survived the last year of the doctorate without you. It is with sincere gratitude that I thank you for being there alongside me to overcome it.

Destin's family and friends

To Destin's family and friends, I'd like to thank you for supporting and caring for me while I've been far away from home. The years have gone by fast, but are filled with warm memories of our shenanigans. I'd like to give a special thanks to

Mason and Marel, Destin's friends who I've spent every weekend with for the past year. In the most difficult time of my doctorate, my weekends were a reprieve, filled with laughter and lightheartedness.

Friends and family

To Binh Phun, I won't soon forget how you were there at a moment's notice to fish together at sunrise on a Sunday morning, or go to the Sea of Cortez with me to place temperature loggers. You're a true friend. To Viridiana Castro, my cohort-mate and close friend, I want to sincerely thank you for your support in the doctorate both as a friend and a fieldmate. You graciously gave up your limited personal time to attend fieldwork trips with me all over this state (even a Channel Island). You played an invaluable role in these projects, and it was a joy seeing you fall in love with fishing. To my childhood friends Gaby Pacheco and Izzy Marchionne, thank you for being some of my biggest cheerleaders from the very start. No matter which states we're living in, I always feel your love and support.

To my Aunt Gina and Uncle Tom, you've always been there for me to lean on when I needed it most. Thank you for always being a phone call away. To Lulu and Papa, thank you for always going above and beyond to help me anyway you could. From sweeping up dead fish for me, to doing my laundry unprompted while I was out in the field, your love and support has meant the world to me. To my parents, I'd like to thank you for always encouraging me to pursue my passions and perhaps more importantly, instilling optimism in me. It is because of you that I want to go out in the world and try my hardest.

To Destin, my significant other, I'd like to thank you for your support through my scientific career. Our relationship first began in college when I was just starting to work in a lab, which means you've witnessed first-hand my journey into becoming a scientist. You, more than anyone, heard every fear and every doubt. I will never forget how, when it all felt insurmountable, you were there to empower me to achieve it. This is as much mine as it is yours.

Introduction

The diversity of life histories, dispersal strategies, and habitats among marine fishes makes them particularly well suited for studying the mechanisms underlying population divergence and speciation (Bernardi, 2013). Marine environments have long been viewed as settings in which high dispersal opposes population structure, particularly for fishes whose pelagic larval stages may be transported over large distances by currents. Under such conditions, extensive gene flow is expected to homogenize populations and limit divergence. Yet many studies show that strong population structure can be found even in the absence of obvious physical barriers (Palumbi, 1994; Selkoe & Toonen, 2011). This apparent paradox motivates a central question in evolutionary biology: how do populations diverge and maintain differentiation in dynamic, seemingly connected marine environments? Simple predictors of connectivity such as pelagic larval duration or dispersal potential explain only a portion of observed variation in population structure. Instead, divergence often reflects the combined effects of historical vicariance, dispersal, demography, and selection.

Geographic isolation remains a powerful driver of divergence, particularly when barriers to gene flow persist over evolutionary time. In such cases, populations can accumulate differences through genetic drift, local adaptation, or a combination, potentially leading to incipient or complete speciation. Divergence may be further accelerated when isolation coincides with strong environmental gradients, such that distinct selective pressures act on populations occupying different habitats. However,

relatively few marine systems combine clear barriers to gene flow with pronounced environmental heterogeneity, limiting our ability to disentangle the relative contributions of isolation and selection.

Environmental gradients not only shape genomic patterns of divergence, but also generate measurable differences in organismal performance and physiological tolerance (Bernatchez et al., 2024; Whitehead, 2012). In ectothermic organisms such as fishes, temperature is a key environmental variable influencing metabolism, stress responses, and survival (Hill et al., 2012; Somero, 2020). Populations occupying contrasting thermal regimes may differ in physiological tolerance, reflecting plastic responses, genetically based adaptation, or a combination of both. Documenting population-level variation in thermal physiology provides a critical phenotypic context for interpreting genomic divergence.

Advances in population genomics now provide powerful tools for investigating the genomic basis of divergence across geographic and ecological contexts. Whole-genome data reveal that divergence is often heterogeneous, with a subset of loci contributing disproportionately to population differentiation, a pattern sometimes referred to as “islands of speciation” (Nosil et al., 2009). Such patterns are consistent with selection maintaining divergence at functionally important regions despite ongoing or episodic gene flow. Combining genome-wide scans for selection with multiple population structure analyses provides a framework for investigating mechanisms underlying population divergence.

Rather than relying on a single system, resolving these questions requires comparative frameworks that span variation in geography, environment, and dispersal strategy. Marine fishes are especially informative because they encompass both isolated and connected populations and a wide range of dispersal strategies, making it possible to examine how these differences influence population divergence.

In this dissertation, I examine how population structure is generated and maintained in marine fishes using two systems that span contrasting biogeographic regions, life histories, and environmental gradients. I use whole-genome data to investigate population genomic patterns across both obvious and subtle barriers to gene flow. By integrating genomic patterns with physiological evidence, this work connects environmental variation, life history, and historical isolation to evolutionary divergence, with broader implications for evolutionary theory and applied management.

Chapter 1: Divergent thermal physiology in isolated Pacific and Sea of Cortez lineages of the longjaw mudsucker (*Gillichthys mirabilis*)

ABSTRACT

Marine fishes with allopatric distributions are rare due to the scarcity of hard barriers to gene flow, yet such systems provide valuable models for studying speciation and adaptive evolution. The longjaw mudsucker (*Gillichthys mirabilis*) is a disjunct species with a Pacific lineage along the coasts of California and Baja California and an isolated lineage in the northern Sea of Cortez. In this study, we compare the thermal physiology of disjunct lineages of *G. mirabilis* in response to acute heat stress to characterize physiological variation between these isolated lineages. Fish were collected from Estero la Choya, Sonora, Mexico (SOC; Sea of Cortez) and Elkhorn Slough, California, USA (PAC; Pacific coast). Two experiments were conducted following ambient common-garden acclimation: (1) intermittent respirometry to construct thermal performance curves (TPCs); (2) critical thermal maxima (CT_{max}) measured in ambient-acclimated and 25°C pre-conditioned fish. TPCs revealed significantly higher performance rates (r_{max}) in PAC (322.6 mg O_2 kg^{-1} h^{-1} ± 38.6) than SOC (267.9 mg O_2 kg^{-1} h^{-1} ± 29.4; $p = 0.004$), indicating higher performance capacity but potentially higher thermal sensitivity in the Pacific lineage. In contrast, after pre-conditioning to 25°C, CT_{max} was significantly higher in SOC than PAC (+0.73°C, $p = 0.0015$). CT_{max} increased significantly in SOC ($\Delta = 1.2$°C, $p = 0.0016$), but not in PAC ($\Delta = 0.3$°C, $p = 0.13$). Together, these results suggest divergence in thermal physiology between the two lineages. Although a two-lineage comparison cannot by itself distinguish local adaptation from genetic drift or founder

effects, the greater thermal tolerance and plasticity of the SOC lineage align with the more extreme thermal environment it occupies.

INTRODUCTION

Environmental temperature is a fundamental driver of physiological performance in ectothermic organisms because metabolic and biochemical processes scale with temperature (Schulte, 2015). Organisms that experience large thermal variation provide valuable systems for studying physiological responses to environmental change (Verberk et al., 2026). Estuarine fishes frequently encounter rapid thermal fluctuations due to shallow water, tidal exchange, and strong diel heating and cooling (Madeira et al., 2012). Many estuarine species exhibit broad thermal tolerances and pronounced physiological plasticity, making them useful models for examining the mechanisms that allow organisms to cope with environmental variability (Verberk et al., 2026).

In the marine environment, absolute barriers to gene flow are rare (Cowen & Sponaugle, 2009; Palumbi, 1994). When populations become geographically isolated, they provide valuable opportunities to examine how spatial separation shapes evolutionary divergence (Futuyma & Kirkpatrick, 2017; Schluter, 2001). Isolation can expose populations to distinct genetic drift dynamics and selective pressures, potentially leading to local adaptation and incipient speciation (Butlin & Faria, 2024). Studying how geographically isolated lineages evolve into distinct species is a central question in evolutionary biology (Nosil, 2012; Nosil et al., 2009). A disjunct species

exhibits a discontinuous geographic distribution, with populations separated by regions of unsuitable habitat that limit gene flow and frequently result in allopatry. Such systems provide natural opportunities for studying how environmental heterogeneity and geographic isolation shape evolutionary trajectories (Losos & Ricklefs, 2009; Palumbi, 1994).

The longjaw mudsucker, *Gillichthys mirabilis,* exemplifies this phenomenon. *G. mirabilis* is an intertidal goby that inhabits shallow soft-bottom habitats in estuaries and sloughs (Allen, 2006; Weisel, 1947). Pacific populations occur along the coasts of California and Baja California, while an isolated lineage occurs in the northern Sea of Cortez, also called the Gulf of California (Allen, 2006). These distributions are maintained in allopatry, as the species is absent around the Cape (Cabo San Lucas region) (Bernardi et al., 2003). *G. mirabilis* is one of 19 disjunct fish species in the Baja California system, a region that provides exceptional opportunities to explore how geographic separation influences divergence and adaptation (Bernardi et al., 2003). Molecular studies estimate divergence between the Pacific and Sea of Cortez lineages at approximately 0.76 to 2.3 million years ago (Huang & Bernardi, 2001), consistent with Pleistocene sea-level fluctuations and geological changes. Genetic studies find high levels of sequence divergence with a high degree of fixed differences between Pacific and Sea of Cortez lineages, indicating prolonged isolation (Bernardi et al., 2003; Garcia et al., 2020; Huang & Bernardi, 2001). Recent genome-wide analyses reinforce this picture, recovering reciprocally monophyletic Pacific and Sea of Cortez clades, an absence of

contemporary gene flow, and pairwise differentiation (F_{ST} up to 0.55) comparable to that between recognized species (Salazar Sawkins & Bernardi, in review).

G. mirabilis inhabits a broad range of thermal environments, from temperate estuaries of central California to the seasonally extreme estuaries of the northern Sea of Cortez. Annual estuarine temperatures in Bodega Bay, CA range from approximately 8 to 19 °C (Sorte & Stachowicz, 2011), while in Puerto Peñasco, MX, temperatures range from 5 to 36 °C (Place & Hofmann, 2001). This extensive thermal variation classifies *G. mirabilis* as an extreme eurytherm, defined as a species experiencing environmental temperature variation exceeding 20 °C (Logan & Buckley, 2015). Past research demonstrates that *G. mirabilis* exhibits substantial thermal plasticity at both molecular and organismal levels. Acclimation modifies thresholds for heat shock protein induction (HSP70, HSP90) across tissues, indicating that cellular stress-response pathways are tuned to the thermal environment (Dietz, 1994; Dietz & Somero, 1992). At the organismal level, acclimation produces large shifts in upper thermal tolerance, with CT_{max} increasing with higher acclimation temperatures (Logan & Somero, 2010). Heat stress elicits sequential activation of stress-response genes, indicating that *G. mirabilis* can dynamically modulate cellular defenses depending on thermal intensity (Logan & Somero, 2011). These findings provide a framework for interpreting how acclimation and acute thermal stress influence stress-response expression in *G. mirabilis*.

Variation in thermal physiology of *G. mirabilis* between Pacific and Sea of Cortez lineages is not yet well characterized. Barlow (1961) compared oxygen

consumption across several localities in the Pacific and Sea of Cortez, finding that populations within the Sea of Cortez exhibited consistently lower Q_{10} values than their Pacific counterparts, potentially reflecting adaptation to the more variable climate. Combined with evidence of plasticity at cellular, molecular, and organismal levels, these results position *G. mirabilis* as a model for examining how physiological plasticity influences thermal limits and local adaptation. Building on this, we address three key questions: (1) Do Pacific and Sea of Cortez lineages differ in critical thermal maxima or other measures of thermal tolerance? (2) Are there lineage-specific differences in metabolic rate or thermal sensitivity (e.g., Q_{10})? (3) Does short-term warm acclimation ("pre-conditioning") shift critical thermal maxima differently between lineages? In this study, we address these questions to better understand how environmental heterogeneity and isolation shape thermal physiology and stress-response mechanisms in Pacific and Sea of Cortez lineages of *G. mirabilis*.

METHODS

Specimen Collection

G. mirabilis were collected from two sites: Estero la Choya, Sonora, MEX (31.34°N, 113.61°W) in July 2024 and Elkhorn Slough, California, USA (36.82°N, 121.78°W) in July-August 2024 (Figures 1 and 2). Hereafter, these sites will be referred to as SOC (Sea of Cortez) and PAC (Pacific), respectively. Fish were captured using baited minnow traps (CDFW permit #S221030003-22181) and transported to Long Marine Labs, Santa Cruz, CA in aerated, insulated containers.

Animals were held in flow-through seawater aquaria in a common garden ambient temperature (15.2 ± 1.3 °C) for > 4 wks. Animals were fed a diet of brine shrimp three times per week *ad libitum*. To minimize post-digestive increase in metabolic rate, animals were fasted for 48 h prior to an experiment. All animal care and use protocols for these experiments were approved by the UC Santa Cruz Institutional Animal Care and Use Committee (protocol ID: Berng2310).

Environmental temperatures at each collection site were monitored using HOBO temperature loggers for at least five months prior to animal collection (Table 1, Figure 3), recording data every five minutes. Each logger was secured to a cinder block with zip ties and placed in the center of a channel, where it remained continuously submerged throughout the measurement period. Loggers were positioned at the same locations from which fish were collected.

Map Generation

To provide geographical and environmental context to the observed physiological variation, we overlaid our sampling sites onto an oceanographic map that was generated using E.U. Copernicus Marine Service Information products: 'SST_GLO_SST_L4_NRT_OBSERV-ATIONS_010_001' (1/20th degree grid) (European Union-Copernicus Marine Service, 2015) for sea surface temperatures and 'GLOBAL_ANALYSISFORECAST_PHY_001_024' (1/12th degree horizontal resolution) (European Union-Copernicus Marine Service, 2016) for surface currents (Figure 2). In both daily data sets, daily data through 2023 were averaged. High

resolution shorelines were imported from the Natural Earth dataset in RStudio using the ne_states function of rnaturalearth 1.0.1 package (Massicotte P & South A, 2025).

Intermittent Respirometry

To assess the differences in thermal performance in two isolated lineages, we performed an acute temperature ramp on individuals while measuring respiration rate (MO_2) using a single-chamber Loligo Systems Intermittent-Flow Respirometry setup. Fish were acclimated to common garden ambient temperature (15.2 ± 1.3 °C) > 4 wks. Following the acclimation period, nine individuals per locality (PAC: n = 9, SOC: n = 9) were tested, and each fish was weighed immediately prior to its trial. Average mass was 7.99 ± 1.60 g for PAC fish and 8.55 ± 1.40 g for SOC fish.

The respirometry system was placed in a large, insulated tank with the common garden ambient seawater, a temperature logger, oxygen bubbler, recirculating pump, and four probe heaters. Seawater was chilled to 10 °C prior to experimentation using double-bagged ice. The tank was covered, and ambient lighting minimized to reduce stress and avoid confounding effects on respirometry measurements. Individuals were exposed to a thermal ramp starting at 10°C, increasing at +4°C per hour starting until the critical thermal maximum was reached. Due to the species' extensive eurythermal range, each trial required ~8 hr to reach CT_{max}, and no more than one trial was run per day. Intermittent respirometry cycles consisted of 120 s flush, 60 s wait, and 120 s measurement periods. Thermal performance curve trials for PAC individuals were conducted between September 19 and October 25, 2024, and SOC trials were conducted between October 17 and November 11, 2024 (Figure S2).

Respirometry experiments were conducted over multiple days, during which ambient seawater temperature in the wet laboratory exhibited slight day-to-day variation, ranging from 12.48 to 17.37°C on the day of trials (Figure S2). To evaluate whether day-to-day variation in ambient seawater temperature influenced estimates of thermal optima (T_{opt}), we examined the relationship between T_{opt} and seawater temperature on the morning of each respirometry trial, as well as averages of ambient seawater temperature 72 h, 1 wk, and 2 wk prior to each trial (Figure S3). These relationships were visualized separately for each lineage using locality-specific linear regressions, with regression equations and coefficients of determination displayed on the plot. Variation in short-term thermal history may partially account for within-lineage differences in parameter estimates.

TPC Modeling, Parameters, and Bootstrapping

MO_2 data were analyzed using the rTPC package in R (Padfield et al., 2021). Multiple candidate thermal performance curve models were fit, varying in functional form and biological interpretation. Although simpler phenomenological models (e.g., modified Gaussian, Jöhnk) yielded lower AIC values, we retained the Sharpe–Schoolfield high-temperature model (Schoolfield et al., 1981) because it provides a mechanistic representation of enzyme activation and high-temperature deactivation, yielding biologically interpretable parameters relevant to thermal sensitivity and limits. Thermal performance parameters (Table 2, 3) were estimated for each individual using the Sharpe–Schoolfield high-temperature model (Schoolfield et al., 1981) in rTPC (Padfield et al., 2021) where model outputs

included breadth, critical thermal maximum (CT_{max}), optimal temperature (T_{o}□□), activation energy (e), deactivation energy (eh), maximum performance rate (r□$_{ax}$), skewness, and thermal safety margin**.** Q_{10} values were calculated separately from model-predicted metabolic rates as described below.

To quantify uncertainty in the estimated parameters and model predictions, we performed residual bootstrapping of the fitted Sharpe–Schoolfield curves using the Boot() function from the car package in R (Fox & Weisberg, 2019). For each curve, residuals from the original fit were resampled with replacement and the model was refitted to each resampled dataset. Predicted metabolic rates across a standardized temperature range were calculated for each bootstrap replicate, and 95% confidence intervals at each temperature were computed as the 2.5th and 97.5th percentiles of the bootstrap predictions. The original fitted curve was plotted alongside these confidence intervals with raw data overlaid for visualization of both model central tendency and uncertainty (Figure 4). For each TPC, we conducted 500 bootstrap replicates (R = 500) using the case resampling method. Bootstrap confidence intervals were computed using the bias-corrected and accelerated (BCa) method, which accounts for both bias and skew in the bootstrap distributions. Lineage-level comparisons of derived thermal performance parameters were conducted using the fitted model estimates (Figures 5 and S1).

Thermal Sensitivity (Q_{10})

Temperature sensitivity of metabolic rate was quantified using Q_{10} coefficients calculated from model-predicted metabolic rates between 10 and 36°C. For each

curve, we fit a linear regression of the natural logarithm of predicted metabolic rate against temperature where the slope of this relationship (*b*) was used to compute Q_{10} as: This formula estimates the proportional change in metabolic rate associated with a 10°C increase in temperature. Q_{10} values were computed independently for each TPC and grouped by lineage (PAC and SOC). Lineage-level differences were assessed using a two-sample Welch's t-test. Finally, to assess whether our sample size limited detection of differences in Q_{10} between lineages, we calculated Cohen's d and estimated achieved and prospective power using the pwr.t.test() function from the R `pwr` package.

Critical Thermal Maxima (CT_{max}) Experiments

To assess the differences in CT_{max} in two isolated lineages, we performed experiments on new cohorts of fish collected on the same dates as described above. All fish were acclimated to common garden ambient temperature (13.5 ± 1.8°C) for eight months prior to experimentation. In Experiment 1, fish were exposed to an acute thermal ramp and temperature at loss of equilibrium (LOE) recorded. In Experiment 2, fish were first pre-conditioned to 25°C prior to an acute thermal ramp and temperature at LOE recorded. CT_{max} was defined as the temperature at which a fish could no longer maintain dorsoventral orientation for at least one minute (Currie et al., 1998). Fish were weighed following experimentation.

CT_{max} Experiment 1: Common Garden Ambient

Eight individuals were tested from each locality (PAC: n = 8, SOC: n = 8; mean body mass: PAC 3.67 ± 1.44 g, SOC 4.43 ± 1.09 g). All fish were experimented

on simultaneously using a split-tank to separate either lineage. The experiment was conducted in a single day over the course of ~8 hr on April 29, 2025. All fish were placed in a single split tank with a temperature logger, oxygen bubbler, recirculating pump, and four probe heaters. Fish were exposed to a temperature heat ramp at a rate of +4°C h^{-1} starting at the ambient common garden temperature until reaching the critical thermal maximum (Figure 6A). CT_{max} was analyzed using Welch's two-sample t-tests to evaluate lineage effects.

CT_{max} Experiment 2: 25°C Pre-Conditioning

A new cohort of fish from the same collection dates was tested (PAC: n = 7, SOC: n = 8; mean body mass: PAC 3.70 ± 2.43 g, SOC 4.38 ± 0.83 g) over April 30 – May 1, 2025. All fish were experimented on simultaneously using a split-tank to separate either lineage and the experiment was conducted in ~26 hr. Fish were removed from the common garden ambient temperature and placed into an insulated tank with a temperature logger, oxygen bubbler, recirculating pump, and four probe heaters. Tank temperature was increased to 25 °C (+1 °C h^{-1}), and fish were pre-conditioned at this temperature for 12 h. Following pre-conditioning, the thermal ramp increased temperature at +4 °C h^{-1} until fish reached their critical thermal maximum (Figure 6B).

The goal of the 12-hour pre-conditioning was to evaluate whether short-term exposure to elevated temperature could alter CT_{max} through acute plasticity. Although 12 hours is generally insufficient for a complete acclimation response, this design allowed us to investigate how either lineage responds to ecologically relevant rapid,

short-term shifts in temperature. CT_{max} was analyzed using Welch's two-sample t-tests to evaluate the effects of lineage and pre-conditioning temperature (Figures 6 and 7; Tables S1 and S2).

RESULTS

Thermal Performance Curves (TPCs)

We tested whether isolated lineages SOC and PAC differed in thermal performance parameters by measuring respiration rate across a range of temperatures. In Figure 4, TPCs are shown for each individual tested. The parameters for the TPCs are visualized in Figures 5, S1, and summarized in Table 3. There was a significant difference in the two lineages in r_{max}, the maximum performance rate (SOC: 267.9 mg $O_2\,kg^{-1}\,h^{-1} \pm 29.4$, PAC: 322.6 mg $O_2\,kg^{-1}\,h^{-1} \pm 38.6$; $p = 0.004$). The differences in breadth ($p = 0.099$), CT_{max} ($p = 0.098$), and thermal safety margin ($p = 0.10$), and Q_{10} ($p = 0.186$) for SOC and PAC were not significant, but the trends suggest a possible lineage effect. To assess whether our sample size limited detection of differences in Q_{10} between lineages, we conducted a post-hoc power analysis. Mean metabolic rates of SOC and PAC (n = 9 per group) were compared with a two-sample t-test, yielding a moderate effect size (Cohen's d = 0.65, 95% CI: –1.59 to 0.31). Based on this effect, achieved power was 0.25, indicating low probability of detecting a true difference. A prospective analysis suggested ~38 individuals per group would be needed for 80% power at $\alpha = 0.05$. These results suggest the nonsignificant difference likely reflects limited sample size rather than absence of a biological effect.

Critical Thermal Maxima Experiments (CT_{max})

We tested whether lineages SOC and PAC differed in terms of their critical thermal maxima and plasticity by carrying out two experiments. For both experiments, fish were acclimated to common garden ambient temperature (13.5°C ± 1.8) for eight months. In the first experiment, temperature was increased from common garden ambient temperature until LOE at a rate of +4°C h^{-1}. Temperature at LOE was measured for every fish. In the second experiment, tank temperature was ramped up to pre-conditioning temperature at a rate of +1°C h^{-1} until temperature reached 25°C, and fish were pre-conditioned to 25°C for 12 h. Then, temperature was increased from common garden ambient temperature until LOE at a rate of +4°C h^{-1}. Temperature at LOE was measured for every fish. After acclimation to common garden ambient temperature, there was no significant difference in CT_{max} between SOC and PAC (SOC: 36.29∞C ± 0.67, PAC: 36.44∞C ± 0.41; $p = 0.59$) (Figure 6A). However, after a 12-hour pre-conditioning period at 25°C, SOC displayed significantly higher critical thermal maxima than PAC (SOC: 37.44∞C ± 0.43, PAC: 36.71∞C ± 0.21; $p = 0.0015$) (Figure 6B). Short-term thermal history significantly affected CT_{max} within lineages. In PAC, the change was not significant ($PAC_{ambient}$: 36.44∞C ± 0.41, $PAC_{25°C}$: 36.71∞C ± 0.21; $p = 0.13$) (Figure 7A), while in SOC, CT_{max} increased significantly with after a 12 hour hold at 25°C ($SOC_{ambient}$: 36.29∞C ± 0.67, $SOC_{25°C}$: 37.44∞C ± 0.43; $p = 0.0016$) (Figure 7B).

DISCUSSION

Our design compares two lineages, and inferences about adaptation drawn from such comparisons warrant explicit caution. As Garland and Adolph (1994) argue, in a comparison of only two groups the environmental contrast of interest is statistically confounded with all other ways the groups may differ, including genetic drift, founder effects, and unmeasured environmental variables. Thus, the effective number of independent units for an evolutionary inference is two regardless of how many individuals are measured within each lineage. Consequently, while our per-lineage sample sizes provide power to detect phenotypic differences between PAC and SOC, they do not permit us to attribute those differences specifically to thermal selection, nor to establish which character state is ancestral versus derived. We therefore frame this study as a physiological characterization of an under-described allopatric system and a source of testable hypotheses, rather than a formal test of local adaptation.

Several features of the system nonetheless mitigate these concerns. Comparisons among populations within a species carry an advantage over interspecific comparisons. Observed differences are more readily related to the present environments the populations occupy and are far less confounded by characters inherited from a distant common ancestor (Garland and Adolph, 1994). The two lineages can be related only as a simple bifurcation, removing any ambiguity about internal phylogeny. The PAC and SOC lineages are reproductively isolated and strongly genetically differentiated (Huang and Bernardi, 2001; Bernardi et al., 2003; Salazar Sawkins & Bernardi, in review), so restricted gene flow makes any retained

physiological divergence more interpretable than it would be among freely interbreeding populations. Critically, our common-garden acclimation isolates genetically based differentiation from acclimatization to field conditions, directly addressing a central difficulty in studies of physiological differentiation among populations (Garland and Adolph, 1991). We regard the most robust path forward as expanding sampling to additional populations within each region and pairing physiology with molecular mechanism, such as the killifish (*Fundulus heteroclitus*), which has become a model for population physiology (Crawford et al., 2020)

Thermal Performance Curves (TPCs)

Our results broadly align with early observations by Barlow, who reported physiological differences between Pacific and Sea of Cortez lineages of *G. mirabilis* in how metabolism scales with temperature (Barlow, 1961). In that study, Sea of Cortez fish consistently exhibited lower Q_{10} values than their Pacific counterparts. We observed the same directional pattern: PAC fish had a higher Q_{10} than SOC fish (1.34 vs. 1.25), indicating a steeper metabolic response to warming. Although this difference was not statistically significant ($p = 0.19$), the higher value observed in Pacific fish is in the direction expected if they were more metabolically sensitive to increasing temperature.

Given the low power of this comparison, however, we do not interpret it as evidence of a true difference and treat the TPC comparisons as inconclusive. With only nine individuals per locality, the achieved power of our test was low (0.25) despite a moderate effect size (Cohen's d = 0.65), indicating a substantial risk of Type

II error. A prospective power analysis suggests that approximately 38 individuals per locality would be required to detect an effect of this magnitude with adequate confidence, highlighting the importance of larger sample sizes in future comparative studies.

Notably, the physiological patterns we observed are consistent with expectations based on environmental conditions. Individuals from SOC tended to exhibit lower Q_{10} and r_{max} values, suggesting reduced thermal sensitivity and more stable metabolic performance across the tested temperature range. Such stability can be advantageous in environments characterized by large and rapid temperature fluctuations, where maintaining consistent physiological performance may be more important than maximizing metabolic output (Kern et al., 2015; Logan & Buckley, 2015). A similar pattern has been documented in the mummichog, *Fundulus heteroclitus*, where myofibrillar ATPase activity exhibits low thermal sensitivity above approximately 12°C. This closely matches the summer tidal conditions experienced by these fish, where temperatures can fluctuate by up to 15°C within a single day (Sidell et al., 1983).

In contrast, the higher Q_{10} observed in PAC fish reflects greater thermal sensitivity; fish from this lineage had to increase respiration rates higher to cope with the same temperatures as their Sea of Cortez counterparts. This is consistent with their environment, which is more thermally stable. In environments with a narrower thermal range, higher thermal sensitivity may be advantageous by enabling rapid metabolic adjustments that optimize growth and activity during short-term favorable

conditions and allow quick responses to acute thermal stress (Kontopoulos et al., 2020).

A limitation of our study was a lack of acclimation treatments to compare how acclimation impacts respiration and thermal performance. Low differentiation in the TPCs may stem from the relatively low TPC acclimation temperature (15.2 ± 1.3°C). Differences in CT_{max} were likewise undetectable under ambient acclimation (13.5 ± 1.8°C), emerging only after 25°C pre-conditioning. Numerous studies have shown that acclimation strongly influences thermal performance, shifting the onset of the cellular stress response (Dietz, 1994; Dietz & Somero, 1992; Logan & Somero, 2011), extending thermal limits (Logan & Somero, 2010), and elucidating lineage-specific differences (Barlow, 1961). Warmer acclimation temperatures often reshape the position and shape of thermal performance curves (Schulte et al., 2011), potentially exposing differences that remain hidden under cooler conditions.

Consistent with this, we observed notable differences in the thermal limits (CT_{max}) of the two lineages when groups were pre-conditioned to 25°C (Figures 6 and 7). These results suggest that while broad metabolic responses to temperature may appear similar between disjunct *G. mirabilis* lineages, lineage-specific physiological traits become more apparent under warm-acclimation conditions that more closely reflect the unique environmental conditions of each habitat.

Critical Thermal Maxima Experiments (CT_{max})

Upper thermal tolerance was similar between lineages when fish were maintained at common-garden ambient temperature (13.5 °C ± 1.8) for eight months,

with no significant difference in CT_{max} between SOC and PAC individuals. However, lineage-specific differences emerged when fish were briefly exposed to warmer conditions. Following a 12-hour pre-conditioning period at 25 °C, SOC individuals exhibited significantly higher CTmax than PAC individuals (37.44 °C vs. 36.71 °C, $p = 0.0015$). In addition, SOC fish displayed greater plasticity in thermal tolerance: CT_{max} increased significantly after pre-conditioning in SOC (+1.2 °C), whereas PAC fish showed only a modest and non-significant increase (+0.3 °C). Together, these results suggest that fish from the Sea of Cortez possess both higher upper thermal limits and greater short-term plasticity in response to warming.

The emergence of these differences only after warm exposure highlights the importance of acclimation history in shaping thermal tolerance. Under cool common-garden conditions, physiological differences between lineages may remain masked. Once individuals experience warmer temperatures, however, lineage-specific physiology became more apparent. Similar patterns have been reported in other ectotherms, where population-level differences in CT_{max} are revealed only under warmer acclimation treatments (Bugg et al., 2020; Waterbury et al., 2024).

The physiological differences we observed correspond closely with the contrasting thermal environments experienced by the two isolated lineages. Temperature loggers deployed at our field sites showed that Estero La Choya (SOC) experiences both higher maximum temperatures and greater overall thermal variability than the PAC site. Temperatures at the SOC site spanned a total range of 24.15 °C, compared with 19.26 °C at PAC (Table 1). Moreover, SOC habitats

experienced more rapid short-term temperature changes, with faster warming and slightly faster cooling rates at minute timescales. Such rapid fluctuations characterize the type of thermal regime often associated with physiological systems that respond quickly to acute thermal stress (Kern et al., 2015; Logan & Buckley, 2015).

Both habitats experience rapid temperature fluctuations, but SOC exhibits a wider thermal range and faster short-term changes. Such conditions are commonly associated with physiological mechanisms for responding quickly to acute warming: induction of heat shock proteins, activation of molecular chaperones, or rapid metabolic adjustments (Kültz, 2005; Logan & Somero, 2011). PAC, while also dynamic, experiences slightly narrower and less extreme temperature fluctuations, where such rapid-response mechanisms may be less critical. A lower investment in highly inducible cellular stress responses may be energetically economical in more stable thermal environments, where moderate metabolic responses are sufficient (Angilletta, 2009; Huey & Kingsolver, 1993). Populations inhabiting warmer, more variable habitats often display traits that enhance tolerance to acute stress while maintaining metabolic stability, as seen in Atlantic killifish (*Fundulus heteroclitus*), where warm-associated populations display higher CT_{max} and reduced thermal sensitivity (DeLiberto et al., 2022).

The higher CT_{max} and stronger short-term plasticity observed in SOC fish may arise from rapidly inducible cellular stress responses. When exposed to stressful temperatures, SOC fish may more quickly activate protective mechanisms such as heat shock proteins, molecular chaperones, or metabolic adjustments. These

responses could be mediated by flexible gene expression (Logan & Somero, 2011), post-translational modifications (phosphorylation, ubiquitination) (Xiao et al., 2023), or shifts in enzyme activity that stabilize metabolism under acute warming (Solovyev et al., 2021). Developmental or epigenetic effects may also contribute to lineage-specific differences (Anastasiadi et al., 2017; Scott & Johnston, 2012). While these mechanisms were not directly measured here, they provide plausible explanations for the observed patterns and could be explored in future molecular or genomic studies.

The patterns observed here are consistent with previous work on *G. mirabilis*. Early work by Barlow (1961) documented ecological and morphological divergence between Pacific and Sea of Cortez lineages, suggesting that long-term geographic isolation could drive physiological differentiation. Subsequent studies revealed substantial thermal plasticity in this species, including acclimation-induced shifts in the induction thresholds of HSP70 and HSP90 (Dietz, 1994; Dietz & Somero, 1992). Our results extend this body of work by showing that, following short-term warm pre-conditioning, Sea of Cortez fish both tolerate higher acute temperatures and exhibit stronger plastic responses to warming.

CONCLUSION

Our study provides initial evidence that isolated Pacific and Sea of Cortez lineages of *G. mirabilis* differ in thermal physiology in ways consistent with the thermal characteristics of their respective habitats. The Sea of Cortez lineage (SOC)

showed higher acute thermal tolerance and more pronounced plasticity following brief warm pre-conditioning, potentially reflecting a distinctive cellular stress response (e.g., via molecular mechanisms such as rapidly inducible HSP expression). In contrast, the Pacific lineage (PAC) exhibited higher maximum metabolic performance in our thermal performance curves but limited plasticity in CT_{max}. These differences are consistent with two thermal strategies that balance trade-offs between metabolic performance and the energetic costs of rapid plasticity. Notably, even short-term exposure to elevated temperatures elicited lineage-specific differences, underscoring the potential for rapid cellular stress responses to buffer against extreme heat events such as heat waves, which are predicted to increase in frequency and intensity under climate change (Frölicher & Laufkötter, 2018). Overall, these contrasting responses suggest that the SOC lineage may show greater short-term resilience to acute thermal perturbation. Because these inferences rest on a single pair of lineages, we treat them as hypotheses to be tested with additional populations and paired with molecular mechanisms, rather than as conclusive evidence of local adaptation.

FIGURES

Figure 1.1 - Lineage-specific morphology. A) Specimen collected from Elkhorn Slough, California, USA (PAC) in July 2024. B) Specimen collected from Estero la Choya, Sonora, Mexico (SOC) in July 2024.

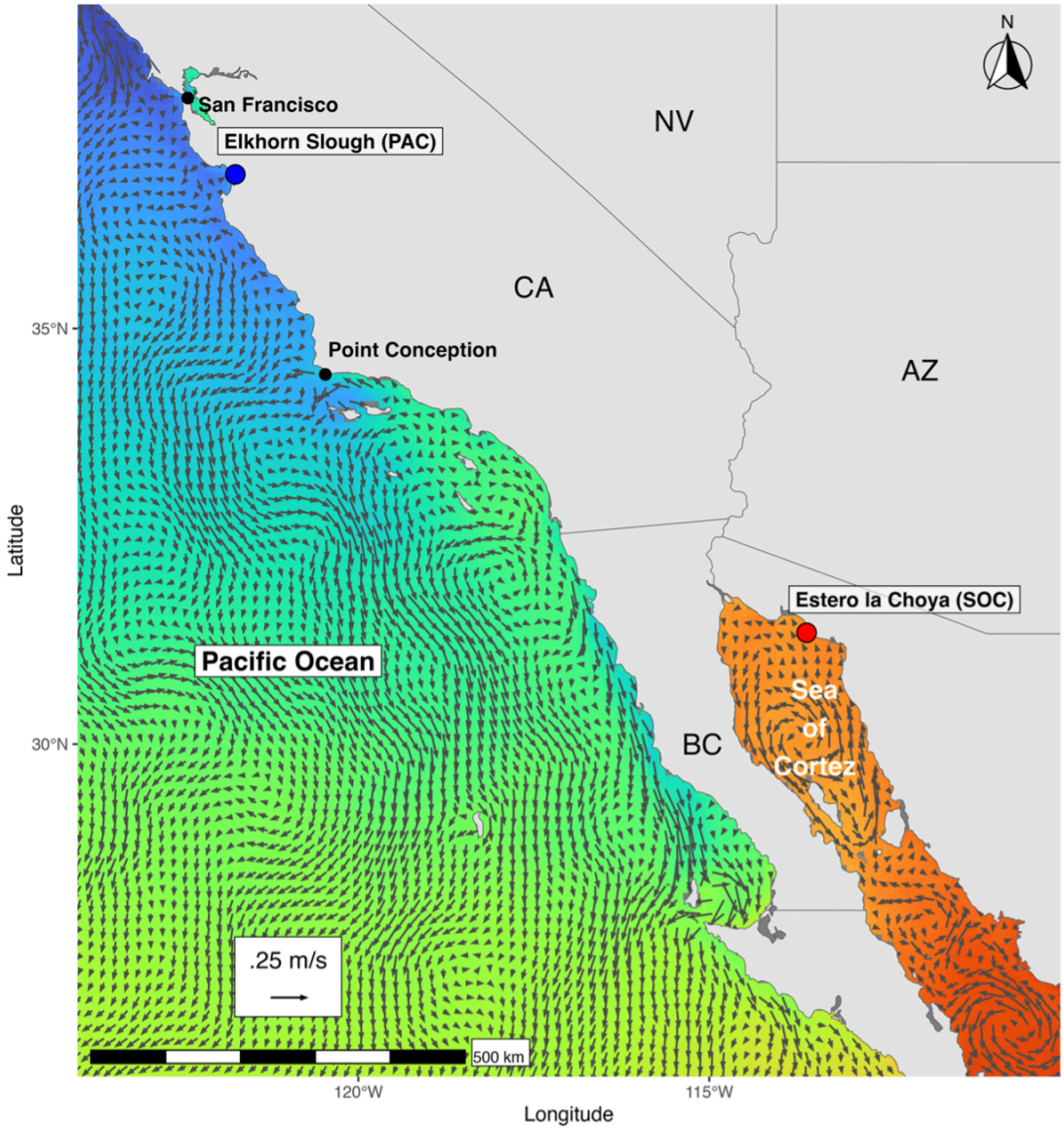

Figure 1.2 - Geographic and Oceanographic Context of Sampling Sites. The two collection localities (Elkhorn Slough, CA, USA (PAC) = blue, Estero la Choya, Sonora, MEX (SOC) = red) are overlaid on mean sea surface temperatures and surface currents. SST data were from SST_GLO_SST_L4_NRT_OBSERVATIONS_010_001 (1/20° grid; Copernicus Marine Service, 2015) and currents from GLOBAL_ANALYSISFORECAST_PHY_001_024 (1/12° grid; Copernicus Marine Service, 2016), averaged through 2023. Shorelines were imported from Natural Earth using rnaturalearth 1.0.1 (Massicotte & South, 2025).

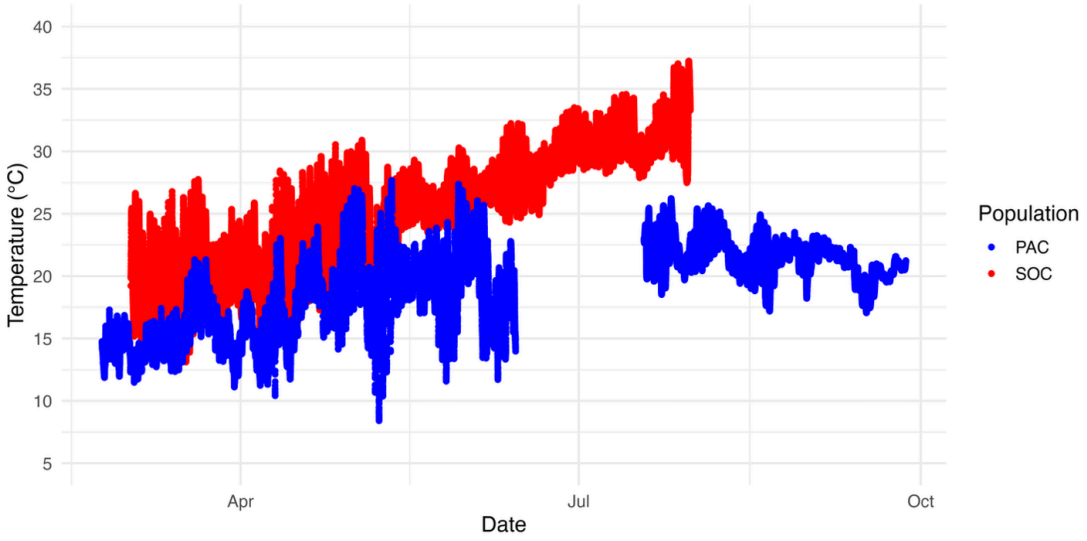

Figure 1.3 - Environmental Temperature. In situ temperature records collected using HOBO data loggers deployed at two field locations: Elkhorn Slough, California, USA (PAC; blue) and Estero La Choya, Sonora, Mexico (SOC; red). Water temperature (°C) is shown on the y-axis and date on the x-axis. Measurements were recorded every 5 minutes from February 23 to June 13, 2024 and July 18 to September 26, 2024 at the PAC site, and from March 2 to July 30, 2024 at the SOC site. A one-month data gap at the PAC site reflects loss of a logger, which was subsequently replaced.

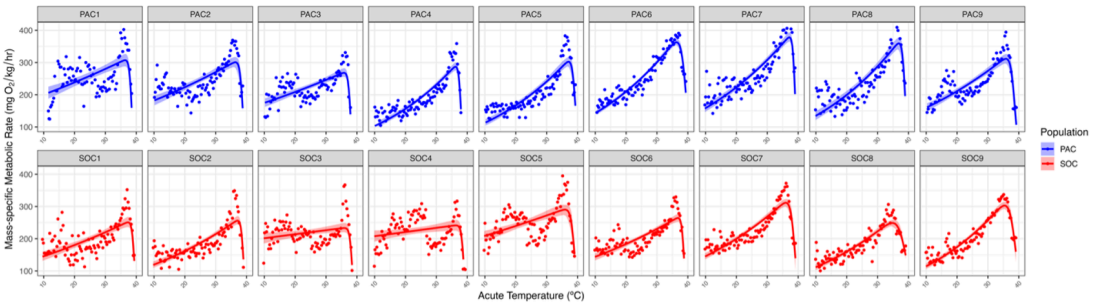

Figure 1.4 - Acute thermal performance curves (TPCs). Measured acute TPCs for PAC (n=9) and SOC (n=9) lineages after acclimation to common garden ambient temperature (15.2 ± 1.3°C) for > 4 wks. Each panel represents a single individual, and points denote individual measurements of mass‑specific respiration rate obtained using a Loligo intermittent‑flow respirometry system. The solid line within each panel shows the fitted Sharpe–Schoolfield high‑temperature model (Schoolfield et al., 1981), while the shaded ribbon represents 95% confidence intervals of model predictions derived from non-parametric bootstrap resampling.

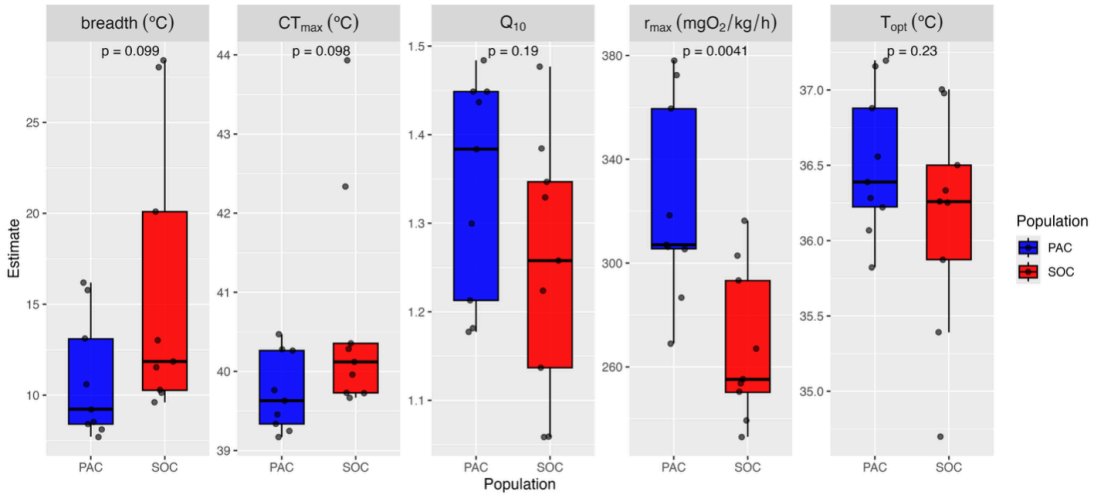

Figure 1.5 - Thermal performance curve (TPC) parameters. Parameter estimates for breadth, critical thermal maximum (CT_{max}), Q_{10}, r_{max}, and thermal optimum (T_{opt}) derived from the Sharpe–Schoolfield high-temperature model (*sharpeschoolhigh_1981*) are shown for each lineage following acclimation to common garden ambient temperature (15.2 ± 1.3°C) for > 4 wks. The PAC lineage in blue is shown and the SOC lineage in red. Welch's two-sample *t*-tests were used to assess differences in each parameter between lineages; corresponding *p*-values are reported above each panel. Definitions of all parameters are provided in Table 2 and values are reported in Table 3.

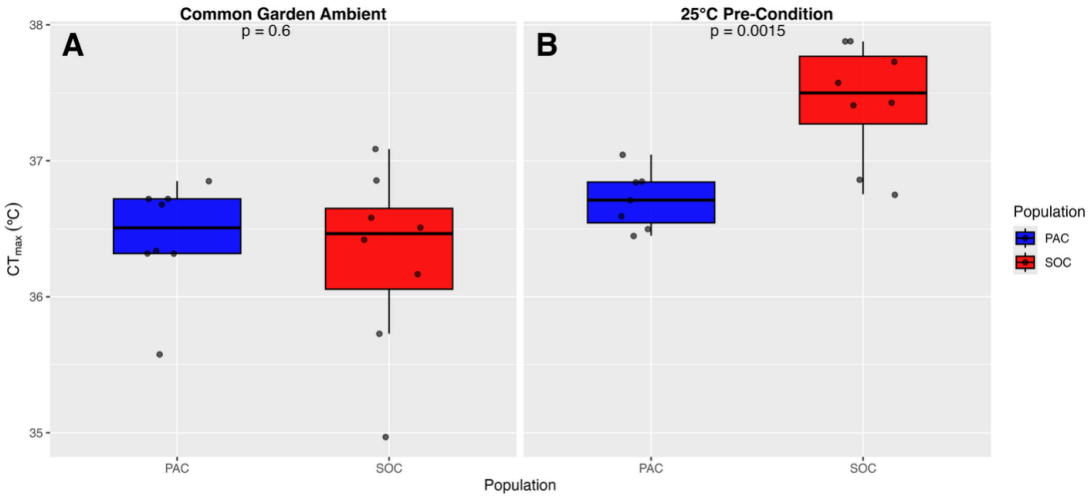

Figure 1.6 - Critical thermal maximum (CT_{max}) experiments. (A) Boxplot showing observed CT_{max} values for individuals from the SOC (n = 8) and PAC (n = 8) lineages following acclimation to common garden ambient temperature (13.5 ± 1.8°C) for 8 months. (B) Boxplot showing observed CT_{max} values for individuals from the SOC (n = 8) and PAC (n = 7) lineages following pre-conditioning at 25 °C for 12 h. In both experiments, temperature was increased from the pre-conditioning temperature at a rate of +4 °C h^{-1} until loss of equilibrium (LOE), defined as the inability to maintain dorsoventral orientation for 60 s (Currie et al., 1998). Individual observations are shown as grey points. Differences between lineages were assessed using Welch's two-tailed t-tests; corresponding p-values are reported above each panel.

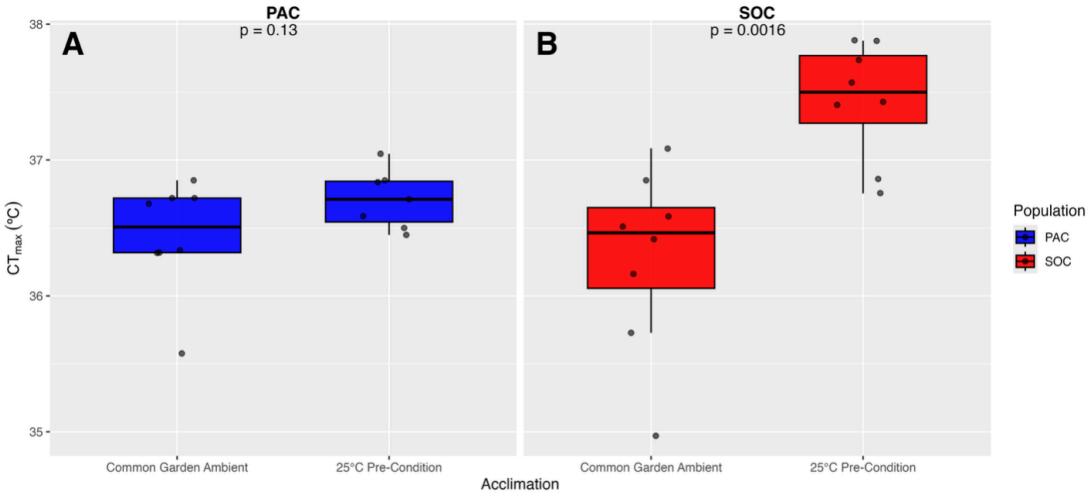

Figure 1.7 - CTmax: Short-term thermal history. (A) Boxplot showing observed CT_{max} values for individuals from the SOC lineage before and after a 12 h pre-conditioning at 25 °C. (B) Boxplot showing observed CT_{max} values for individuals from the PAC lineage before and after a 12 h pre-conditioning at 25 °C. Individual observations are shown in grey. Differences between pre- and post-conditioning were assessed using Welch's two-tailed t-tests, and corresponding p-values are reported above each panel.

TABLES

Table 1.1 - Environmental Temperature Statistics. Maximum temperature, minimum temperature, total temperature range, and maximum +/- rate of change on ΔºC min^{-1} and ΔºC hr^{-1} timescales are reported for each location using the data displayed in Fig. 3. Dates of temperature collection: March 2 – Jul 30, 2024 (SOC); Feb 23 – June 13, 2024 and July 18 – Sept 26, 2024 (PAC). Measurements taken every 5 minutes, except for a 1-month period at PAC site when a logger was lost and subsequently replaced.

Location	Max. Temp	Min. Temp	Range	Max. Warming	Max. Cooling	Max. Warming	Max. Cooling
SOC	37.23	13.08	24.15	0.68	-0.57	6.31	-6.26
PAC	27.67	8.41	19.26	0.27	-0.44	6.40	-7.80

Table 1.2 - Parameter Definitions. rTPC Model sharpeschoolhigh_1981 parameter definitions are reported. Parameter estimates from Sharpe–Schoolfield thermal performance models. All parameters were obtained directly from model fits except Q_{10} values which were calculated post hoc from model-predicted rates.

Parameter	Definition
Breadth	The range of temperatures over which performance remains above 80% of r_{max}.
CT_{max}	Critical thermal maximum; the highest temperature at which the organism can maintain performance before mortality. Calculated by predicting beyond the maximum value in the dataset where the value closest to 0 is then extracted.
e	Activation Energy; Describes the exponential rate of increase in performance with temperature below T_{opt}.
eh	Deactivation energy; Describes the steepness of the performance decline above T_{opt}.
Q_{10}	Temperature coefficient; Unitless factor that describes the change in rate of respiration over every 10ºC increase in temperature.
r_{max}	Maximum performance rate; the peak of the thermal performance curve.
Skewness	Degree of asymmetry in the TPC. A negative skewness indicates the TPC is left skewed, the drop after the optimum is steeper than the rise up to the optimum. A positive skewness means that the TPC is right skewed and a value of 0 would mean the curve is symmetrical around the optimum.
Thermal Safety Margin	Thermal safety margin is calculated as: CT_{max} - T_{opt}.
T_{opt}	Optimum temperature (c) is calculated by predicting over the temperature range using the previously estimated parameters and determining the temperature where the largest rate value (r_{max}) occurs.

Table 1.3 - TPC Parameters. Mean, standard deviation, and p-value are reported for each lineage (SOC and PAC) at each parameter following acclimation to common garden ambient temperature (15.2 ± 1.3 °C) for > 4 wks. The p-values reported are calculated from two-tailed Welch's t-test.

Parameter	**mean (SOC)**	**mean (PAC)**	**sd (SOC)**	**sd (PAC)**	***p***
Breadth (°C)	15.9	10.9	7.7	3.3	0.099
CT_{max} (°C)	40.7	39.7	1.5	0.5	0.098
e (eV)	0.2	0.2	0.1	0.1	0.269
eh (eV)	2.6	2.6	0.8	0.5	0.935
Q_{10}	1.25	1.34	0.15	0.12	0.186
r_{max} (mg O_2 kg^{-1} h^{-1})	267.9	322.6	29.4	38.6	0.004**
Skewness	-2.5	-2.4	0.9	0.5	0.949
Thermal Safety Margin (°C)	4.5	3.2	2.1	0.2	0.100
T_{opt} (°C)	36.1	36.5	0.7	0.5	0.234

Chapter 2: Population genomics and speciation of a Baja California disjunct species, the longjaw mudsucker (*Gillichthys mirabilis*)

ABSTRACT

Allopatric isolation can drive genomic divergence, but distinguishing population structure from speciation remains challenging in marine systems. The longjaw mudsucker (*Gillichthys mirabilis*) is one of 19 Baja California disjunct species, exhibiting a pronounced population disjunction between the Pacific coast and the Sea of Cortez. This system provides a natural test of how prolonged geographic isolation drives divergence and speciation. We generated a reference genome and low-coverage whole-genome data for 113 individuals from 14 localities spanning the Pacific coast and two localities in the northern Sea of Cortez. Genome-wide analyses reveal extreme divergence between Pacific and Sea of Cortez populations (F_{ST} = 0.45–0.55), with reciprocal monophyly and no evidence of admixture, indicating long-term isolation. In contrast, Pacific populations show weak, clinal structure shaped by limited dispersal and habitat discontinuity, with a notable exception at Elkhorn Slough, a highly divergent site that may represent a historical refugium. Across the genome, we identify thousands of outlier loci, including protein-coding variants enriched for nonsynonymous substitutions (dN/dS = 1.41), implicating selection alongside drift. Among numerous candidate loci, genes with functions tied to environmental stress (i.e., protein homeostasis and ion transport) stand out as biologically plausible targets of selection, consistent with thermal and osmotic contrasts between regions. Together, these results reveal deeply diverged evolutionary lineages and suggest that Pacific and Sea of Cortez populations represent distinct

species, highlighting how prolonged allopatry and environmental heterogeneity drive genomic divergence in marine taxa.

INTRODUCTION

Marine fishes offer exceptional opportunities for studying mechanisms driving population divergence and speciation because they span a wide range of life histories, dispersal strategies, and environmental gradients (Bernardi, 2013). A central goal in speciation research is to identify conditions under which geographically separated populations accumulate sufficient genetic divergence to generate species (Rundle & Nosil, 2005; The Marie Curie SPECIATION Network, 2012). In marine systems, such conditions are often thought to be uncommon because the apparent absence of physical barriers promotes extensive dispersal and gene flow, which can homogenize populations and oppose local adaptation (Palumbi, 1992, 1994; Rocha & Bowen, 2008; Sanford & Kelly, 2011).

Species with disjunct distributions provide powerful natural experiments for studying speciation, as they are shaped by rare but persistent barriers to gene flow (Avise, 2000; Bernardi, 2013). Two classic examples include geological barriers such as the rise of the Isthmus of Panama, which created transisthmian geminate species (Knowlton et al., 1993; Lessios, 2008), and vast oceanic barriers that separate eastern and western Pacific faunas (Lessios & Robertson, 2006; Rocha et al., 2007). In such systems, prolonged geographic isolation prevents gene flow, allowing divergence to accumulate through drift, local adaptation, or both.

Selection can accelerate divergence between isolated populations, even over relatively short evolutionary timescales (Nosil, 2012; Schluter, 2001). When barriers to gene flow coincide with different selective pressures, isolated populations can rapidly diverge, promoting allopatric speciation (Nosil, 2012; Rundle & Nosil, 2005). Classic examples include adaptive divergence in sticklebacks across contrasting habitats (McKinnon & Rundle, 2002; Rundle et al., 2000) and repeated evolution of Atlantic mollies in extreme environments (e.g., sulphidic caves) (Kelley et al., 2012; Plath et al., 2007). However, relatively few marine fish systems combine clear barriers to gene flow with strong environmental gradients (Funk et al., 2006; Wolf et al., 2010), limiting our ability to disentangle the roles of isolation and selection in driving divergence. Moreover, it is often unclear whether such divergence represents transient population structure or early stages of speciation (Bernardi, 2013; Coyne & Orr, 2004; Palumbi, 1994).

The Pacific–Baja California biogeographic disjunction represents a rare natural experiment for examining divergence in marine fishes. Approximately 19 fish species exhibit disjunct distributions between the temperate Pacific coast and the northern Sea of Cortez, a region characterized by strong thermal and environmental contrasts (Bernardi et al., 2003; Huang & Bernardi, 2001). Fish communities on the Pacific and Sea of Cortez sides of the Baja California Peninsula experience different thermal regimes, reflecting distinct oceanographic provinces (Spalding et al., 2007). Previous work on Pacific–Baja California disjunct species has documented varying degrees of morphological, ecological, and genetic differentiation, suggesting that

isolation across this region promotes divergence (Bernardi, 2014; Bernardi et al., 2003; Garcia et al., 2020; Schinske et al., 2010; Terry et al., 2000). However, no studies have leveraged whole-genome data to assess the genomic basis and evolutionary significance of this divergence.

Here, we focus on the longjaw mudsucker (*Gillichthys mirabilis*), a eurythermal estuarine goby with a Pacific–Sea of Cortez disjunct distribution and broad Pacific-coast range spanning approximately 10° of latitude (Eschmeyer & Herald, 1999). This species occupies highly variable thermal environments and spans one of the steepest temperature gradients experienced by coastal fishes in the eastern Pacific (Place & Hofmann, 2001), making it a powerful system for examining divergence across both geographic and environmental barriers. Previous studies using mitochondrial and reduced-representation genomic markers have revealed strong population structure in the longjaw mudsucker, particularly between Pacific and Sea of Cortez populations (Bernardi et al., 2003; Garcia et al., 2020; Huang & Bernardi, 2001). However, the extent to which this structure reflects genome-wide divergence, local adaptation, or speciation remains unresolved. Building off earlier work, several key questions remain: (1) To what extent is previously observed genetic differentiation (e.g., mtDNA, RADseq) mirrored across the full nuclear genome?; (2) Is there evidence for local adaptation, particularly the presence of genes under strong selection across environmental gradients?; and (3) What is the genetic relationship between Pacific and Sea of Cortez populations, and do patterns of divergence support their classification as speciated lineages? In this study, we address these questions

using low-coverage whole-genome resequencing to characterize population structure, genomic differentiation, and signatures of selection in the longjaw mudsucker across its Pacific–Sea of Cortez range.

METHODS

Reference Genome

We collected an adult male specimen by baited minnow trap from Elkhorn Slough, CA, USA (36.855338 °N, 121.754302 °W) for reference genome sequencing and assembly (CDFW permit S-221030003-24080-001). We sexed the specimen by dissection and sampled fin tissue for DNA extraction, after which the specimen was preserved in 95% ethanol and stored at -20°C. Genomic DNA was extracted from fin tissue using the Qiagen DNeasy Blood & Tissue Kit and high molecular weight was confirmed on a 0.7% agarose gel. Genomic DNA was then used to prepare libraries for Nanopore long-read sequencing. We prepared two libraries for two separate Nanopore sequencing runs using the SQK-LSK109 chemistry kit and protocol. Libraries were sequenced on a Nanopore MinION device using two R9.4.1 flow cells. In total, we obtained 26.2 Gb of raw sequence data (11,148,238 raw reads), with an N50 of 4,338 kb.

Base-calling of Nanopore reads was done using Guppy (Oxford Nanopore Technologies, Oxford, UK; version 6.4.6), and only reads passing default quality filters were retained for downstream analyses. Basecalled reads from each sequencing run were concatenated and merged across runs when applicable. Read quality and

length distributions were assessed using NanoStat (De Coster et al., 2018) and FastQ utilities (Cock et al., 2010). Adapter and primer sequences were removed using Porechop (https://github.com/rrwick/Porechop), and quality control was repeated post-trimming using FastQC (Andrews, 2010). Reads were filtered by length and quality using NanoFilt (De Coster et al., 2018) to generate two datasets: (i) reads ≥ 1000 bp with Q ≥ 3 and (ii) reads ≥ 500 bp with Q ≥ 5. De novo genome assemblies were generated for both datasets using wtdbg2 (Ruan & Li, 2020), and consensus sequences were produced with wtpoa-cns (Ruan & Li, 2020).

Assembly quality was evaluated using assembly-stats (https://github.com /sanger-pathogens/assembly-stats), and completeness was assessed with BUSCO (v5.2.2) using the actinopterygii_odb10 database (Manni et al., 2021). For polishing, reads were mapped back to the assembly using minimap2 (Li, 2018), and consensus correction was performed with Racon (Vaser et al., 2017). Post-assembly contamination screening was conducted using Kraken2 (Wood et al., 2019), and contigs classified as bacterial were removed from the final assembly. Final assemblies were re-evaluated using assembly-stats to assess improvements in contiguity and completeness.

Population Genomics Sampling

We collected fish by baited minnow trap, seine, or hand net at 16 sites, covering the range of the species (Table 1) (CDFW permit S-221030003-24080-001). Several additional sites were surveyed using the same methods, but *G. mirabilis* was not encountered (e.g., Pescadero Marsh, Goleta Slough, and Tijuana Slough).

Specimens were sampled for fin clips, photographed, and returned to the environment. Fin clips were immediately placed in 95% ethanol and kept as cool as possible in the field (ambient temperature, on ice, or in a refrigerator) until reaching the laboratory where they were stored at −20°C.

DNA Extraction and Library Preparation

We extracted genomic DNA from all samples using the Promega Wizard Genomic DNA Purification Kit and confirmed DNA quality on a 0.7% agarose gel. Sequencing libraries were prepared from genomic DNA using half reactions of the NEBNext UltraExpress FS DNA Library Prep Kit (New England Biolabs) and unique dual indexes provided by the sequencing facility. We incubated the samples for 20 min during the enzymatic fragmentation step and amplified the libraries for 5 cycles using dual unique indexes. We eluted each library in 15 uL of 0.1xTE. Prepared libraries were quantified via Qubit (Invitrogen), fragment analyzed (Agilent), pooled, and then sequenced (2x150 bp) on Illumina NovaSeq 6000 (S4) lanes at the Vincent J. Coates Genomics Sequencing Lab at the University of California, Berkeley with a target depth of 6x genome-wide coverage.

Several of the analyses described were implemented using a published workflow from Toy et al., available on GitHub (https://github.com/Arthur-Oules/EJA_lcwgs_GenPop).

Map Generation

To contextualize patterns of genetic variation geographically and environmentally, sampling locations were projected onto an oceanographic base map

constructed from E.U. Copernicus Marine Service Information products. Sea surface temperature data were obtained from 'SST_GLO_SST_L4_NRT_OBSERVATIONS_010_001' (1/20th degree grid) (European Union-Copernicus Marine Service, 2015), and surface current data from 'GLOBAL_ANALYSISFORECAST_PHY_001_024' (1/12th degree horizontal resolution) (European Union-Copernicus Marine Service, 2016) (Figure 1B). For both data sets, daily values for 2023 were averaged to produce annual mean surfaces. High resolution coastlines were imported from the Natural Earth dataset in RStudio using the ne_states function of rnaturalearth 1.0.1 package (Massicotte P & South A, 2025).

Mapping and Variant Calling

We performed mapping and variant calling using the assembled reference genome described in Section 2.1. We used the snpArcher workflow (Mirchandani et al., 2024), an accelerated workflow for variant calling, to generate high-quality variant calls for downstream analysis. Reads were first quality-filtered and adapter-trimmed with fastp (Chen et al., 2018), then aligned to the concatenated reference genome using bwa mem (Li, 2013) with recommended parameters. Variant calling for each individual was performed using Sentieon Haplotyper (Kendig et al., 2019), and joint genotyping was subsequently conducted with Sentieon Genotyper to generate the combined VCF (variant call format). Sites that failed GATK hard filtering thresholds defined in the snpArcher workflow were removed. We generated a second SNP dataset by applying minor allele frequency (MAF) filtering and removing multiallelic sites and indels. To reduce noise from rare variants, we used

bcftools (Danecek et al., 2021) to retain only variants with MAF ≥ 0.01. With the exception of Manhattan plots and genes under selection, the following methods and results are described using the MAF-filtered, biallelic SNP data set.

Population F_{ST} and Hamming Distances

To compute Wright's F_{ST} estimates between each pair of populations using Weir and Cockerham's estimate of F_{ST} (Weir & Cockerham, 1984), we used the --fst function of the PLINK v2.00a6 package (Chang et al., 2015). We also quantified pairwise Hamming distances among individuals using the --distance function in PLINK 1.9.0 beta 7 (Chang et al., 2015), based on allele count differences. Pairwise Hamming distances were then derived by averaging distances between individuals from different populations and correcting for within-population distances.

Analysis of Population Structure and Ancestry

Population structure was evaluated using principal components analysis (PCA) at the whole genome level using PCADAPT 4.4.0 (Privé et al., 2020). After converting the file to the ped file format using the –export function of PLINK v2.00a6 (Chang et al., 2015), it was converted to the geno file format for ancestry (admixture) analysis in sNMF using the ped2geno function of LEA 3.16.0 (Frichot & François, 2015). We used the snmf function from ped2geno to estimate individual ancestry coefficients and allele frequencies using Non-Negative Matrix Factorization algorithms and least-squares optimization algorithms. This approach produces comparable results to STRUCTURE– or ADMIXTURE–based analyses with similar underlying assumptions, while relaxing the Hardy-Weinberg equilibrium assumption

and offering faster computing time and resources for large datasets (Frichot et al., 2014).

Model fit was assessed using the cross-entropy criterion calculated by snmf, which evaluated the accuracy of masked genotype prediction and provides a basis for selecting the number of ancestral populations (*K*) (Frichot et al., 2014). We performed 30 repeated runs of sNMF for a range of *K* values from 2 to 8 and compared the computed cross-entropy (Figure S2). The α regularization parameter was left at its default value of 10, as its effects are most pronounced in smaller datasets (<10,000 SNPs). The run and *K* value minimizing the cross-entropy were used for downstream analyses (Figure 1C).

Isolation-by-Distance (IBD)

We evaluated patterns of isolation by distance (IBD) by testing the relationship between genetic and geographic divergence among individuals sampled along the Pacific Coast. Geographic distances among sampling sites were calculated as great-circle distances using the distm(…, fun = distVincentyEllipsoid) function from the R package geosphere (Hijmans, 2010). Genome-wide genetic differentiation was estimated in 10 kb windows using Hudson's FST, implemented in scikit-allel v1.3.6 via the allel.windowed_hudson_fst() function (Miles, 2015). In addition, principal component (PC) scores from the PCA were used to construct an alternative genetic distance matrix in ordination space. Pairwise Cavalli-Sforza and Edwards chord distances were also calculated using the dist.genpop() function (method = 2) in the adegenet v2.1.10 package (Jombart, 2008). Relationships between pairwise and

chord distances were evaluated using a defined a linear model with the base R function stats::lm() (e.g., $F_{ST} \sim m * \text{Physical Distance} + b$). Statistical significance of isolation by distance was tested using the mantel.randtest function from the adegenet 2.1.10 package (Jombart, 2008).

Triangle Plots

To distinguish between alternative explanations for spatial genetic structure, we used triangle plots (Wiens & Colella, 2025), which visualize hybrid index against interclass heterozygosity. In a two-population system, hybrid index reflects the proportion of ancestry derived from each parental group, while interclass heterozygosity measures the proportion of loci carrying alleles from both groups. Under this framework, parental individuals are expected to fall at opposite corners of the triangle (hybrid indices of 0 and 1, and zero interclass heterozygosity), whereas F1 hybrids and subsequent backcrosses are expected to plot above the Hardy–Weinberg expectation boundary. Deviations from Hardy–Weinberg equilibrium can result in individuals falling below this threshold (Wiens & Colella, 2024).

Simulation results have shown that this approach can distinguish isolation by distance (IBD) from neutral diffusion (ND) when introgression is sufficiently recent (Wiens & Colella, 2025). Under IBD, hybrid index and interclass heterozygosity are expected to exhibit a negative linear relationship, whereas ND produces a curved distribution within the triangle space.

Triangle plots were generated using the triangulaR package (Wiens & Colella, 2024). We used the alleleFreqDiff() function to generate datasets filtered by allele frequency difference thresholds (δ) of 1, 0.75, and 0.5 between parental populations (n=8 (ELK) and n=8 (ECO)), identifying ancestry-informative markers (i.e., markers uniquely fixed in each parental populations). This resulted in 168,725 , 41,133, and 5,120 retained sites, respectively. Hybrid index and interclass heterozygosity were then calculated for each dataset using the hybridIndex() function.

Phylogenetic Analysis

We constructed a maximum-likelihood (ML) phylogeny by converting the pruned VCF into a single sequence alignment using vcf2phylip.py (Ortiz, 2019). The resulting FASTA alignment was used to infer an ML tree in IQ-TREE v2.2.2.6 (Minh et al., 2020) under the GTR+G4 substitution model, with branch support assessed using 1000 replicates of the ultrafast bootstrap approximation (UFBoot; Hoang et al., 2018). The ML consensus tree was visualized and annotated in FigTree (https://github.com/rambaut/figtree).

Outlier Loci Analysis

To identify candidate loci under selection, we converted the VCF file to the bed file format using the --make-bed function of the PLINK v2.00a6 package (Chang et al., 2015), we used PCADAPT 4.4.0 package (Privé et al., 2020) to first perform a PCA on the centered and scaled genotype matrix. Mahalanobis distance (distance from any point to the main distribution) is computed for each SNP using the *K* first Principal Components (PCs), squared and divided by the constant genomic inflation

factor. The resulting value is a test statistic that should follow a Chi-squared distribution with *K* degrees of freedom and allow the computation of *p*-values. We chose the *K* number of components to compute *p*-values following Cattell's rule (components accounting for most of the dataset's variance are retained) and checking biplots of PCs to ensure contribution of retained PCs to data structure. The negative log of the *p*-values produced by PCADAPT were mapped along the genome to create Manhattan plots (Figure 4). In some comparisons, Manhattan plots appear visually noisy due to the presence of a small number of loci containing missing genotype calls. Removal of all loci with any missing data substantially reduced this background noise while preserving the overall pattern of outlier peaks (Figure S5), indicating that the observed signal is robust to missing data. We used the Bonferroni method to correct *p*-values for multiple testing, and a cutoff for candidate loci was set at 1E-300. We then took read windows of 500bp centered around each outlier and extracted the sequences written as a FASTA file for downstream analysis.

Functional Analyses

To annotate the genomic regions found under selection using PCADAPT, nucleotide BLAST searches were conducted against a mudskipper, *Periophthalmus magnuspinnatus* (GenBank accession no. GCA_009829125.3), reference database using BLASTn (NCBI BLAST+). A maximum of five target sequences were retained per query and four CPU threads. Output was generated in tabular format (outfmt 6), including alignment statistics: percent identity, alignment length, e-value, and bit score. BLAST hits were filtered to retain only high-confidence matches (e-value <=

1e-10, percent identity >= 80%, alignment length >= 200bp). For each query window, the best hit was selected based on lowest e-value and highest big score. The resulting set of best hits represents putative functional annotations for each outlier region. When protein-coding matching sequences were found, they were classified using KEGG assignments (Kanehisa et al., 2007).

Outlier SNPs: dN/dS Analysis

To assess whether fixed genetic differences between Pacific and Gulf populations alter protein sequences, we performed a dN/dS analysis on single nucleotide polymorphisms (SNPs) located in coding regions. Genome annotation from the *Periophthalmus magnuspinnatus* (GenBank accession no. GCA_009829125.3) reference genome was transferred to the *Gillichthys mirabilis* reference assembly using Liftoff v1.6.3 (Shumate & Salzberg, 2021), generating a .gff3 file containing coding sequence (CDS) coordinates. Outlier loci identified using PCADAPT were intersected with the filtered VCF, and only SNPs (i.e., excluding indels) were retained. SNPs fixed within Pacific and Sea of Cortez populations (allele frequencies of 0/0 or 1/1) were identified using bcftools (Danecek et al., 2021). Fixed differences between populations were defined as sites where Pacific and Sea of Cortez populations were fixed for alternate alleles.

To determine which fixed differences fell within coding regions, SNP coordinates were intersected with CDS features from the Liftoff annotation. For SNPs located within CDS, a custom python script was used to determine codon context and functional effect. The script reconstructed CDS sequences for each gene using

reference genome sequence and CDS coordinates, accounting for gene strand orientation. For genes with multiple transcript isoforms, the longest CDS transcript was used as the representative model. At each SNP position, Pacific and Sea of Cortez alleles were substituted into the reference codon, translating using the standard genetic code, and compared. Sites that could not be classified (e.g., incomplete codons) were excluded.

RESULTS

Reference Genome of G. mirabilis

The final reference genome assembly had a total length of 700 Mb across 5,169 contigs, with a mean contig length of 135 kb, a maximum contig length of 3.8 Mb, and an N50 of 483 kb. Coverage was on average 37x. The assembly had a GC content of 40.8%. Assembly completeness was assessed using BUSCO v5.2.2 (Simão et al., 2015) with the *actinopterygii_odb10* dataset (n = 3,640 genes) in genome mode with MetaEuk as the gene predictor. The assembly contained 95.1% complete BUSCOs (93.4% single-copy, 1.7% duplicated), 1.2% fragmented, and 3.7% missing. This Whole Genome Shotgun project has been deposited at DDBJ/ENA/GenBank under the accession JBWLTA000000000.

Sampling, Loci, and Polymorphism Statistics

We sampled 113 individuals of *Gillichthys mirabilis* from Bodega Bay (California, USA) to the Estero Coyote (Baja California, Mexico) in the Pacific, as well as two sites in the northern Sea of Cortez (Figure 1B, Table 1). Low coverage

genomic sequencing of our samples resulted in an average 6.3X coverage per individual. In total, the dataset comprised 2,573,755 variant sites following initial GATK best practices filters (Van Der Auwera et al., 2013). After restricting the dataset to biallelic loci using PLINK v2.00a6 package (Chang et al., 2015), 908,249 sites were removed, leaving 1,665,506 SNPs. We further filtered the dataset as described in the Materials and Methods section, retaining 1,424,602 biallelic SNPs. Population genomic analyses were performed using this MAF–filtered, biallelic SNP dataset unless otherwise stated.

Population Structure

Population structure was first assessed using a Principal Components Analysis (PCA). We visualized the first two components, which account for 35.0% (PC1) and 4.0% (PC2) of the total variance (Figure 1A). PC1 clearly separated the Sea of Cortez sites from the Pacific sites, indicating deep divergence between regions. PC2 captured latitudinal structure among Pacific sites, with individuals and sampling locations forming a largely continuous north-south cline. Elkhorn Slough, however, deviated from this pattern, exhibiting the highest PC2 values despite its intermediate latitude relative to other Pacific sites. Individuals clustered by sampling site, though partial overlap among populations was observed among central Pacific sites and between two Sea of Cortez sites. Overall, these results support strong genetic differentiation between the Sea of Cortez and Pacific, coupled with more moderate, clinal population structure found along the Pacific coast.

These patterns were mirrored by the ancestry analysis, where the most likely number of clusters based on cross-entropy minimization corresponded to four ancestral genetic populations (Figures 1C, S2, S3). One cluster was exclusive to the Sea of Cortez and showed little to no shared ancestry with Pacific populations, consistent with strong regional isolation. The remaining three genetic clusters formed a gradient of ancestry proportions along the Pacific coast, broadly corresponding to the greater San Francisco Bay Area, the central and southern California coast, and the northern to central Baja California region. Based on these ancestry patterns, we hereafter refer to Pacific sampling locations as belonging to northern (Bodega Bay to Elkhorn Slough), central (Morro Bay to San Diego Bay), and southern (San Quintín to Estero Coyote) regions. Individuals with mixed ancestry proportions were observed and may reflect some level of current gene flow or the retention of ancient polymorphisms. Consistent with PCA results, individuals from Elkhorn Slough, CA deviated from the clinal pattern and displayed near-fixation for a single ancestral population, while adjacent populations showed only partial membership to that population (Figure 1C).

Isolation by Distance, Neutral Diffusion, and Admixture

Consistent with the patterns described above, genetic structure of *G. mirabilis* is characterized by a strong divergence between Pacific and Sea of Cortez populations, with Pacific individuals further partitioned into three ancestral clades. Average pairwise F_{ST} estimates between the Sea of Cortez population and each Pacific clade were high (Northern: 0.55; Central: 0.50; Southern: 0.45), indicating substantial

genomic differentiation. In contrast, differentiation among Pacific clades was markedly lower (North–South and 0.15; North–Central: 0.08; Central–Southern: 0.06). Comparisons among any of the Pacific locations do not exceed 0.25 (Figure 5).

We assessed isolation by distance (IBD) among Pacific populations. A weak correlation between genetic and geographic distance was detected, but it was not statistically significant (Mantel test $p = 0.089$; Figure 2A). Notably, individuals from Elkhorn Slough exhibit near-fixation for a single ancestral population despite occupying an intermediate geographic position, potentially weakening the overall IBD signal. While a weak signal of correlation was found between genetic and geographic distances, which is generally considered to be IBD, it is important to discern if the pattern is the result of true IBD of a single population, or of admixture between historically allopatric groups followed by neutral outward diffusion of admixed genotypes (Wiens & Colella, 2024). Triangle plots of our data do not clearly match simulated patterns expected for either history, but there is a linear trend of decreasing interclass heterozygosity moving from the south (e.g., Estero Coyote, B.C.) to the north (e.g., Elkhorn Slough) (Figure 2B), indicative of a northward expansion and IBD (Wiens & Colella, 2025).

Evolutionary Relationships of Populations

We constructed a maximum likelihood tree of individuals (Figure 3). We observe monophyletic grouping of the Sea of Cortez population (100% UFBoot) that is clearly separated from remaining populations, with Guerrero Negro being the most closely related Pacific locality. Pacific samples form a single large clade with internal

structure. From Guerrero Negro to Santa Barbara, individuals form a continuum rather than separated groups, which could be due to IBD or ongoing gene flow. Morro Bay and the northern Pacific population show separation from other Pacific populations. One Newport Bay individual clusters more closely to the southern Pacific localities, which could indicate recent dispersal or incomplete lineage sorting.

Outlier Loci and Signals of Selection

We identified outlier loci using PCADAPT, a non-F_{ST} based method shown to be more appropriate for species with strongly structured populations (Toy et al., 2025). Principal components (PCA), ADMIXTURE, and phylogenetic analyses (Figures 1A, 1B, 3) consistently resolved two major clades, alongside finer-scale clinal structuring within Pacific populations. Genome-wide Manhattan plots provide further insight into these patterns (Figure 4). When comparing all Pacific and Sea of Cortez populations, we discovered 12,015 outlier loci (out of 1,665,506 loci), of which 507 loci mapped to 444 unique protein coding genes (Table S1), with several genes containing multiple outlier loci. Out of those 444 protein coding genes, 400 were annotated with known functional pathways (Figure S4), with the largest proportions mapping to genetic information processing (22.6%), environmental information processing (16.6%), cellular signaling and processes (12.5%), and genetic information processing pathways (11.4%).

Among the 507 outlier loci mapping to protein coding genes, 171 sites occurred within annotated coding regions. Of these, 169 sites were successfully classified as synonymous or nonsynonymous substitutions. 99 sites (58.6%) were

nonsynonymous, and 70 sites (41.4%) were synonymous, corresponding to a dN/dS ratio of approximately 1.41:1, consistent with potential functional differentiation in highly divergent loci.

Genes Under Directional Selection

Among outlier loci intersecting coding regions, nonsynonymous variants are of greatest interest because unlike synonymous changes, they alter the amino acid sequence of the encoded protein. Therefore, they represent candidates for functionally significant divergence between Pacific and Sea of Cortez populations. Examination of the genes harboring these nonsynonymous variants revealed several recurring biological categories: neurological and synaptic signaling (*gabra1*, *gabra5*, *grm5b*, *shank3a*), ion and solute transport (*slc12a4*, *slc45a2*, *slc7a1a*), chromatin remodeling (*chd8*, *raraa*, *arid1ab*), protein quality control and ER stress response (*uggt2*, *p4hb*, *arel1*), and cytoskeletal dynamics (*mical3a*, *kif18a*, *rock2a*). Notably, two genes harbored multiple nonsynonymous substitutions within the same coding sequence: *mical3a* and *tut7*, suggesting these loci may have experienced repeated or sustained divergent selection. The *mical3a* gene encodes a multifunctional protein involved in guiding neuron growth, coordinating cell movement, and regulating the cytoskeleton (Xue et al., 2010). The *tut7* gene encodes an enzyme that regulates gene expression by tagging specific microRNAs for degradation (small regulatory RNAs that silence target genes). In doing so, *tut7* influences the expression of genes controlling cell differentiation and body patterning during development (Thornton et al., 2014). Among nonsynonymous variants, five involved stop codon changes: three were stop

codon losses (*mrps34*, *shank3a*, *LOC117380625*) and two were stop codon gains (*hmcn2*, *prr36a*) in the Sea of Cortez population relative to the Pacific. The remaining 70 synonymous variants, while not altering amino acid sequence, may reflect linked selection or regulatory divergence at these loci.

DISCUSSION

Speciation is increasingly viewed as a continuum of divergence rather than a discrete transition between population differentiation and fully isolated species (Feder et al., 2012; Nosil, 2012). This perspective is particularly relevant in allopatric systems, where prolonged geographic isolation will lead to substantial genomic divergence, complicating species delimitation (Nosil & Feder, 2012). Genomic studies have shown that divergence under isolation is often heterogeneous across the genome, with some regions exhibiting pronounced differentiation (e.g., islands of speciation), while others remain comparatively conserved (Nosil & Feder, 2012; Wolf & Ellegren, 2017). Such patterns can arise through the combined and sometimes competing effects of genetic drift, which acts genome-wide and is amplified in small or isolated populations, and local adaptation, which targets specific loci or genomic regions associated with environmental differences (Nosil et al., 2009). For marine organisms, high dispersal and connectivity are often assumed to limit population divergence, making cases of strong geographic structure particularly informative for understanding speciation processes (Faria et al., 2021).

In allopatric marine systems, prolonged geographic isolation can promote divergence through a combination of genetic drift and local adaptation, yet distinguishing deep population structure from speciation remains challenging, particularly in marine vertebrates, where clear phenotypic or reproductive barriers are often absent (Butlin & Faria, 2024). The longjaw mudsucker, *Gillichthys mirabilis*, presents a compelling system in this context, with disjunct populations between the Pacific coast and the Sea of Cortez isolated since the late Miocene (Bernardi et al., 2003). Limited dispersal and strong geographic and environmental discontinuities associated with the Baja California Peninsula created conditions for long-term divergence and have made this system well suited to evaluate how historical isolation shapes genomic differentiation, as well as investigating whether deeply structured allopatric populations represent genetically divergent species.

Allopatry and Dispersal

Species with allopatric distributions are expected to exhibit high F_{ST} due to restricted gene flow and the effects of genetic drift (Holsinger & Weir, 2009). Our results are consistent with this expectation. PCA and phylogenetic analyses reveal strong genetic differentiation between allopatric populations, with deep differentiation consistent with long-term isolation. Phylogenetic relationships further inform the historical origin of this isolation. Guerrero Negro represents the earliest-diverging lineage within the Pacific clade (Fig. 3), rather than the southernmost Pacific site of Estero Coyote. This pattern is consistent with predictions of the midpeninsular seaway hypothesis, which proposes that during the Miocene several seaways were

present along the peninsula (Riddle et al., 2000), allowing marine species to move between the Gulf of California and the Pacific coast before subsequent isolation. Further support is found in Manhattan plots (Fig. 4), which show abundant genome-wide outlier loci and substantial fixed differences between Pacific and Sea of Cortez populations.

Along the Pacific coast, *Gillichthys mirabilis* exhibits genetic structure broadly consistent with isolation by distance (Fig. 2). The longjaw mudsucker is a benthic species exhibiting high site fidelity, completing its life cycle within a single estuary (Weisel, 1947; Yoklavich et al., 1992). Adjacent sampling sites showed relatively low differentiation (pairwise F_{ST} as low as 0.01), however, notable exceptions occurred. Elkhorn Slough and Alviso slough exhibited high differentiation ($F_{ST} = 0.25$) despite close geographic proximity, a pattern that may reflect the more isolated, interior position of Alviso Slough within San Francisco Bay. This is consistent with phylogenetic results indicating more recent common ancestry between Elkhorn Slough and the more coastal Bodega Bay and Tomales Bay localities. However, the pronounced genetic distinctiveness of Elkhorn Slough despite its proximity to other sites suggests an additional process beyond simple geographic isolation. Near-fixation of a single ancestral population (Figure 1C) supports a refugium-like scenario, in which persistent habitat in Monterey Bay during Pleistocene sea-level fluctuations preserved that ancestral lineage.

Given the strong association of *G. mirabilis* with demersal estuarine habitats, we predicted that long stretches of coastline lacking suitable habitat would act as

barriers to dispersal. Indeed, populations separated by such habitat discontinuities, such as Elkhorn Slough and Morro Bay (F_{ST} = 0.23), were among the most genetically distinct. Conversely, regions with greater estuarine habitat availability exhibited lower overall levels of F_{ST} (e.g., Los Angeles–Orange County region), likely because increased availability of nearby suitable habitat facilitates larval export from estuaries and subsequent settlement in adjacent habitats.

Most marine fishes are characterized by high dispersal potential due to pelagic larval stages that facilitate extensive gene flow and reduce opportunities for population divergence (Cowen & Sponaugle, 2009), with ocean currents further enhancing connectivity among populations, often obscuring the effects of geographic isolation (White et al., 2010). The longjaw mudsucker exhibits a coastal, benthic ecology with limited adult movement and strong association with estuarine habitats. Due to limited adult dispersal, gene flow is thought to be facilitated by their planktonic larval stage (Barlow, 1963; Saha et al., 2025).

Larval Dispersal and Retention. While pelagic larval duration has not been explicitly studied for *G. mirabilis*, other ecologically similar gobies have PLD of ~14-39 days (Beldade et al., 2007; Maeda et al., 2022). However, several lines of evidence suggest effective dispersal may be rare. Larval entrainment surveys at coastal power plants consistently capture *G. mirabilis* larvae at within-bay sites (e.g., Morro Bay and San Diego Bay), but not at the adjacent coastal site at Diablo Canyon Power Plant, located approximately 15 miles south of the Morro Bay estuary mouth (Raimondi, 2011). Additionally, California Cooperative Oceanic Fisheries

Investigations (CalCOFI) surveys, which take place four times per year since 1951, have captured *G. mirabilis* larvae in plankton tows on only three occasions between 1975 and 2008 at nearshore localities off California and Baja California (NOAA SWFSC, 2024). While larval fish distributions are often highly patchy and influenced by fine-scale oceanography, these observations suggest limited offshore presence. We infer that planktonic larvae are frequently present within bays and estuaries and are likely exported via tidal outflow, with successful settlement depending on the probability of encountering nearby suitable habitat, which is strongly influenced by spatial proximity, consistent with the rarity of offshore detections in CalCOFI surveys and the observed spatial structure in genomic differentiation.

Postlarval settlement filtering. Nektonic postlarvae have been observed to have heavily pigmented coloration at a relatively small size (8–12mm), potentially increasing predation risk compared to more translucent taxa (Barlow, 1963). Behavioral observations indicate that postlarvae resist passive transport by descending and holding position when current strength increases (Barlow, 1963). Moreover, successful recruitment of juveniles requires a suite of environmental conditions, such as appropriate salinity, soft sediment habitat for burrowing, and thermal conditions; these requirements likely impose a strong post-settlement bottleneck. Given that patchy, discontinuous distribution of estuarine habitat along the California coast, these constraints likely restrict gene flow even when larval dispersal occurs. Although rare long-distance dispersal events may occur during big storms, the high genetic differentiation observed between the north and south distribution of

Pacific *G. mirabilis* (F_{ST} = 0.15) suggests that such events contribute little to ongoing gene flow.

Life-history and demographic constraints. Gillichthys mirabilis is relatively short-lived for a marine fish, with a maximum recorded age of approximately six years (McGourty et al., 2009), implying high generational turnover. Under restricted connectivity, such life-history traits facilitate the accumulation of genetic divergence over comparatively few generations. The density-dependent life history of *G. mirabilis* may further limit effective gene flow: males establish and defense breeding territories, and recruitment is therefore constrained in stable populations, with young-of-the-year occupying marginal habitats until territory becomes available (e.g., when older individuals die) (McGourty et al., 2009). Phylogenetic analyses further indicate population structure is shaped strongly by habitat availability and oceanographic and environmental discontinuities than by simple geographic distance.

Natural Selection

Allopatrically distributed species are expected to diverge inevitably from genetic drift, resulting in high genome-wide F_{ST} values between populations, as is observed in longjaw mudsuckers. Under such conditions, identifying outlier loci can be difficult because the neutral baseline is very high. Despite this, we observed numerous outlier loci, many with fixed differences between Pacific and Sea of Cortez populations, suggesting that divergence at these loci exceeds what drift alone would predict. Codon-based analyses further support this interpretation. SNPs within coding regions exhibit elevated dN/dS ratios between Pacific and Sea of Cortez lineages,

consistent with positive selection rather than neutral divergence alone. The Sea of Cortez presents a markedly different environmental regime from the Pacific coast, characterized by hotter and more variable water temperatures (Allen, 2006; Lavín & Marinone, 2003; Sanchez-Cabeza et al., 2022), elevated salinity (Beron-Vera & Ripa, 2002; Lynn, 1967), greater tidal amplitude (Roden & Groves, 1959), and reduced oceanic connectivity (Lavín & Marinone, 2003; Marinone et al., 2008). These environmental differences provide a plausible ecological basis for the pattern of protein-coding divergence observed here.

Temperature. Of high ecological significance are genes involved in protein quality control and ER stress response (*uggt2*, *p4hb*, *arel1*). The endoplasmic reticulum is acutely sensitive to elevated temperatures (Malhotra & Kaufman, 2007). Proteins are highly temperature-sensitive, and clear patterns of adaptive variation have been discovered in structural and functional properties of proteins from species adapted to different temperatures (Somero, 2020). Divergence in the proteins that monitor and correct misfolding under heat stress would be expected in populations experiencing chronic thermal differences. Indeed, protein processing in the endoplasmic reticulum represents one of the most significantly enriched pathways in transcriptomic responses to heat stress in fish (Huang et al., 2018; Zhao et al., 2022). Genes *uggt2*, *p4hb*, and *arel1* are ER-resident components of the unfolded protein response and oxidative protein folding machinery that form an integral part of the cellular response to thermally induced protein misfolding. Notably, *p4hb* encodes a protein disulfide isomerase — the same enzyme family as PDIA3 and PDIA4, which

are among the protein-folding genes most strongly upregulated in warm-acclimated *G. mirabilis* gill tissue (Logan & Somero, 2010). That divergence at *p4hb* specifically, rather than at the more commonly studied canonical HSPs, may reflect that constitutive oxidative folding capacity represents an axis of thermal adaptation distinguishing chronically warm Sea of Cortez populations from their Pacific counterparts.

Osmoregularity. Among the most ecologically compelling candidate genes are members of the solute carrier superfamily (*slc12a4*, *slc45a2*, *slc7a1a*), which mediate ion and solute transport across epithelial membranes (Verri et al., 2012) and represent strong candidates for osmoregulatory adaptation. Selection on solute carrier genes across salinity boundaries has been documented repeatedly, including Atlantic cod adapting to the low-salinity Baltic Sea (Berg et al., 2015), killifish populations diverging across osmotic gradients (Brennan et al., 2015), and watersnakes inhabiting inland versus coastal saline habitats (Rautsaw et al., 2021), suggesting that this gene family may represent a convergent genomic target of osmoregulatory adaptation. The solute carrier family has been identified as a shared selected gene family across at least eight species of teleost fishes adapting to different salinity regimes, with selection appearing to act at the level of core osmoregulatory pathways rather than being constrained to any specific genes or loci (Velotta et al., 2022). Together, the functional identity of these candidate genes is consistent with the environmental differences between the two systems, lending ecological credibility to the genomic patterns observed here.

Beyond amino acid substitutions, five fixed nonsynonymous variants involved stop codon changes. While stop codon changes can reflect pseudogenization (Albalat & Cañestro, 2016), they may also represent functional isoform diversification (Lewis et al., 2003), particularly when occurring in large multidomain proteins (Vogel et al., 2004), such as *hmcn2* . In sum, the elevated dN/dS ratio, the functional identity of candidate genes, and the presence of radical amino acid changes among outlier loci collectively suggest that divergence between Pacific and Sea of Cortez populations reflects not only inevitable accumulation of neutral differences under allopatry, but also the signature of selection acting on ecologically relevant protein-coding sequence. This pattern, combined with the extreme levels of genomic differentiation documented here, raises the question of whether these populations represent more than simply divergent conspecifics.

Speciation in Longjaw Mudsucker

This study revealed a high magnitude of genetic differentiation between allopatric populations of *Gillichthys mirabilis*. Analyses of ancestry proportions, PCA, and F_{ST} consistently indicate long-term absence of gene flow, deep coalescent times, and genome-wide divergence. Pairwise F_{ST} values were extremely high between Pacific and Sea of Cortez populations (as high as $F_{ST} = 0.55$), exceeding levels typically reported among conspecific populations and comparable to those observed between recognized species in marine fishes (Gandra et al., 2021). Consistent with this pattern, Manhattan plots revealed a large number of fixed differences and outlier loci across the genome.

Together, these results support long-term separation of Pacific and Sea of Cortez lineages. Divergence is not solely attributable to neutral processes: several outlier loci occur in genes with ecologically relevant functions, suggesting a role for both drift and local adaptation. Within the Pacific coast, genetic differentiation is lower but structured (F_{ST} = 0.01–0.25), indicating heterogeneous connectivity among estuarine populations. This pattern is consistent with restricted dispersal across discontinuous coastal habitats and patchy estuarine habitat, which together limit exchange despite pelagic larval stages. Life-history traits further reinforce this structure, including short lifespan (~6 years; McGourty et al., 2009), limited adult dispersal, postlarval settlement filtering, and male territorial brooding. These features collectively reduce effective gene flow and increase the influence of drift while maintaining opportunities for local selection.

Taken together, the absence of gene flow, reciprocal phylogenetic monophyly, extreme genetic differentiation, and the abundance of outlier and fixed loci, alongside evidence for selection on ecologically relevant genes, support the interpretation that Pacific and Sea of Cortez populations represent deeply diverged evolutionary lineages. While formal taxonomic revision is beyond the scope of this study, these patterns are consistent with long-term isolation and may warrant their recognition as distinct species.

CONCLUSION

This study demonstrates extreme genomic divergence between Pacific and Sea of Cortez populations of *Gillichthys mirabilis*, characterized by deep phylogenetic separation, genome-wide differentiation, and a lack of contemporary gene flow. Together with signals of putative local adaptation and life-history constraints that limit dispersal, these results indicate long-term allopatric isolation. Disjunct marine systems such as this reveal how geographic and ecological barriers interact to generate and maintain divergence across the speciation continuum, even in taxa with potential for high dispersal. Future work can build on results presented here, taking advantage of this unique system and advances in low-coverage whole genome sequencing (lcWGS), to investigate whether common genomic signatures are shared across disjunct marine species. High-throughput genomic approaches in such systems offer critical insight into fundamental speciation processes.

FIGURES

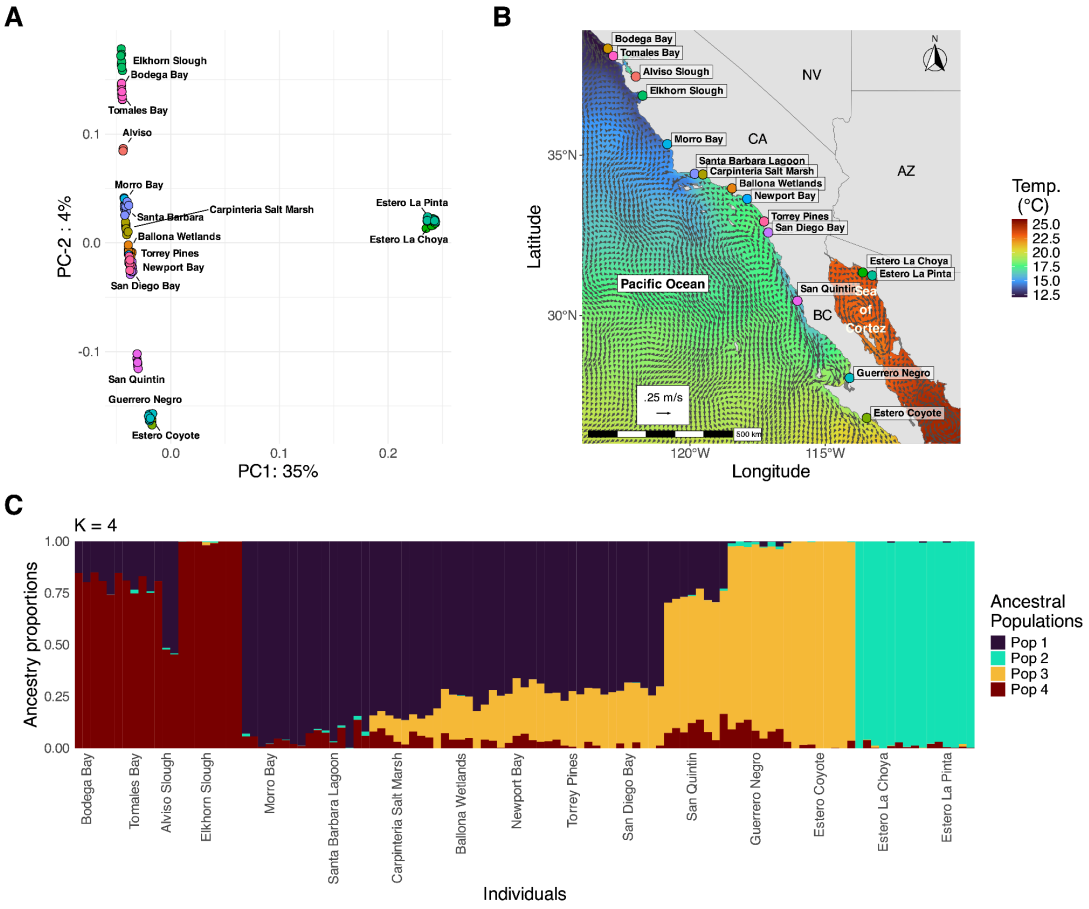

Figure 2.1 - Genetic population structure of the Longjaw mudsucker, *Gillichthys mirabilis*. Panel A) Principle Components Analysis (PCA) of 113 individuals' whole genomes of *G. mirabilis*, colored by sampling location. Panel B) Map of the study region showing sampling locations used in this study and yearly average current and sea surface temperatures. Panel C) Ancestry plot divided into four genetic clusters. Each vertical bar represents an individual, populations are indicated on the x-axis and are placed from north to south along the Pacific coast. Sea of Cortez populations are placed on the far right side of the plot.

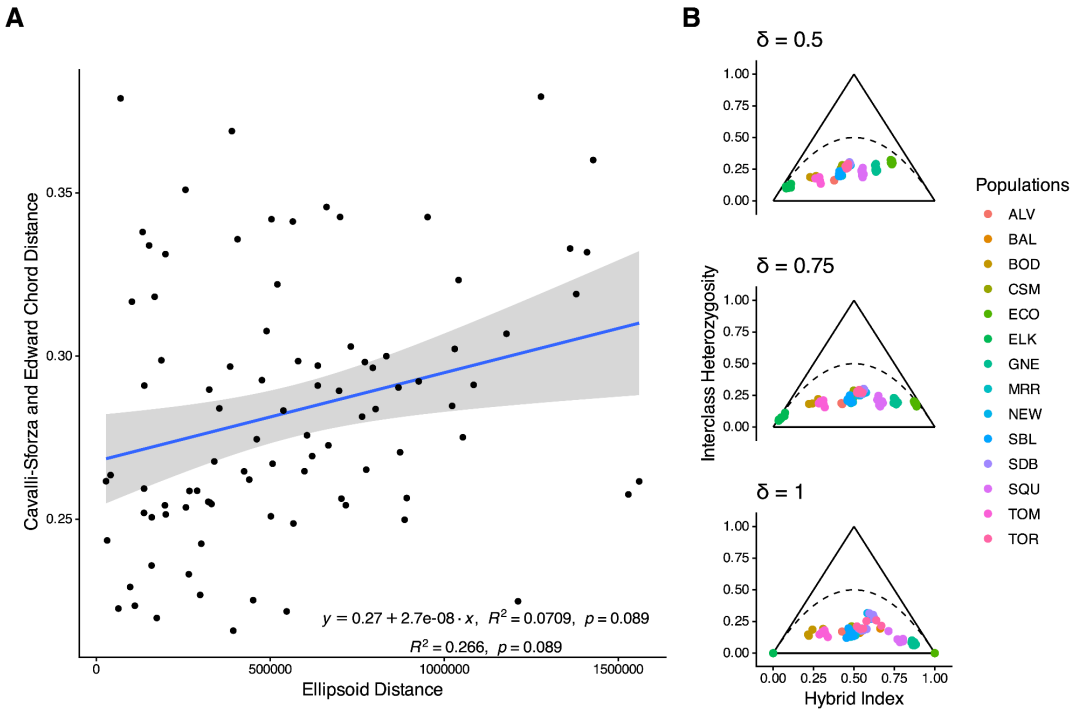

Figure 2.2 - Isolation by Distance in Pacific populations of the Longjaw mudsucker, *Gillichthys mirabilis*. Panel A). Mantel test of the correlation between the geographical distance matrix and the genetic distance matrix between Pacific populations. First line at the bottom right shows linear regression fit and associated R^2 and p-value, second line shows Mantel test R^2 and p-value. Panel B). Triangle plots using the same populations and colour codes as in Figure 1. Triangle plots with Elkhorn Slough and Estero Coyote set as parental population, allele frequency thresholds (δ) of 0.5, 0.75 and 1 (top to bottom) allow the selection of Ancestry-Informative Markers. Upper part of the triangle over the dotted lines represents the expected position of hybrids and backcrossed individuals under Hardy–Weinberg equilibrium (Wiens and Colella 2024a, 2024b). The mantel test displays strong correlation between the geographical and genetic distance between populations which is characteristic of an isolation by distance (IBD) scenario. The linear relationship between Interclass heterozygosity and hybrid index in the triangle plot for δ = 0.5 also suggests IBD but the pattern is fading for higher δ values, suggesting an older event.

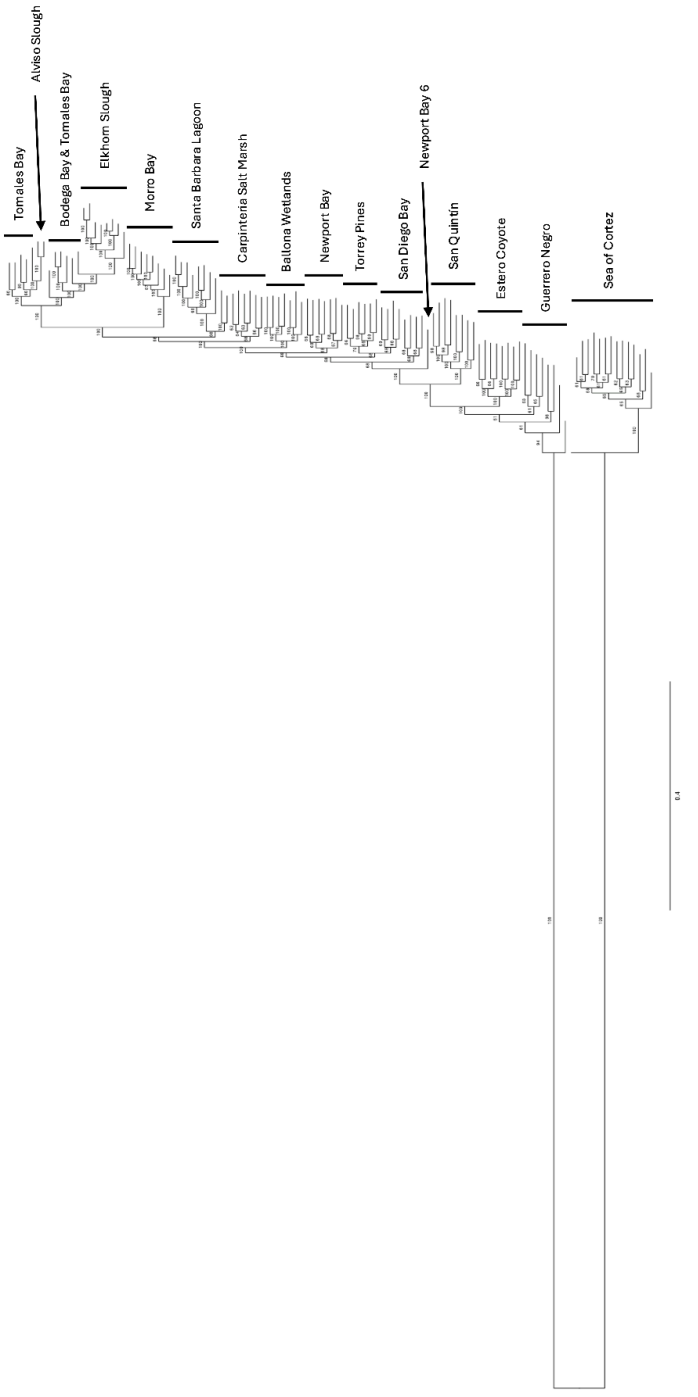

Figure 2.3 - Maximum‑likelihood tree constructed using the pruned SNP set. This unrooted tree is shown midpoint‑oriented. Branch labels are UFBoot support values (1000 iterations).

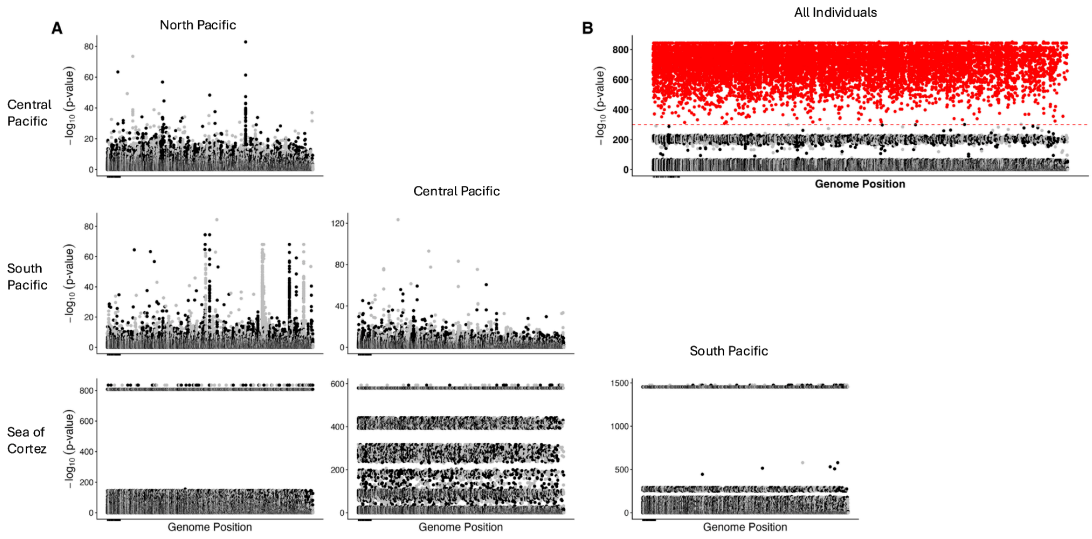

Figure 2.4 - Panel A) Manhattan plots of pcadapt ⁻log10(p values) based on pairwise comparisons of longjaw mudsucker populations. Panel B) Manhattan plot of all Pacific and Sea of Cortez populations. Outlier loci are shown in red.

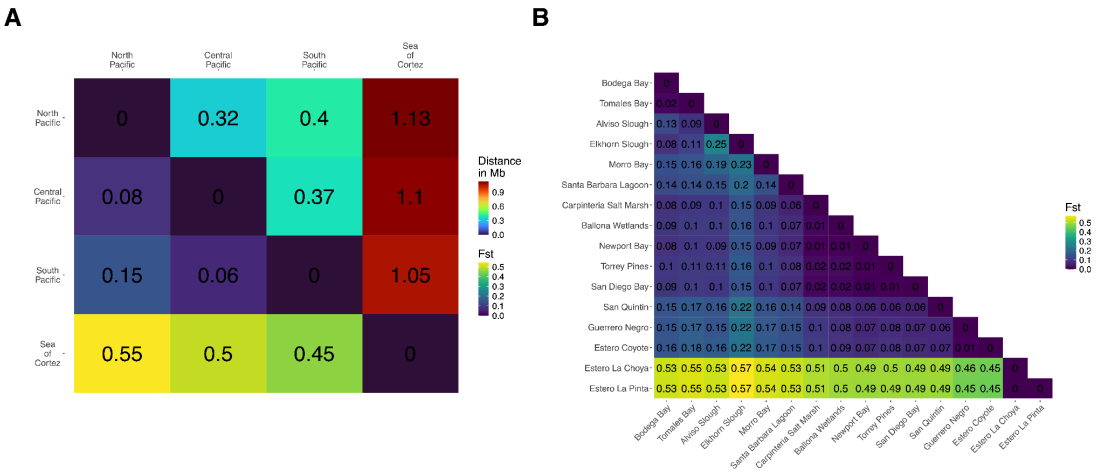

Figure 2.5 - Gene flow among populations of Longjaw mudsucker. Panel A) Pairwise F_{ST} and Distance (Mb) among the four major populations. Panel B) Pairwise F_{ST} values among sampling localities.

TABLES

Table 2.1 - Sampling locations and sampling numbers of longjaw mudsucker *(G. mirabilis)*.

Sample Code	Locality	Number	Latitude, longitude
BOD	Bodega Bay	4	38.314665, -123.032296
TOM	Tomales Bay	7	38.211092, -122.924553
ALV	Alviso Slough	2	37.438783, -121.961516
ELK	Elkhorn Slough	8	36.855232, -121.752256
MRR	Morro Bay	8	35.335103, -120.821086
SBL	Santa Barbara Lagoon	8	34.410845, -119.850049
CSM	Carpinteria Salt Marsh	8	34.397514, -119.526709
BAL	Ballona Wetlands	7	33.962682, -118.445631
NEW	Newport Lagoon	8	33.633381, -117.884788
TOR	Torrey Pines	6	32.931994, -117.251956
SDB	San Diego Bay	8	32.586942, -117.105197
SQU	San Quintín	8	30.463467, -115.975372
GNE	Guerrero Negro	8	27.982412, -114.083009
ECO	Estero Coyote	8	26.808196, -113.478885
ELC	Estero la Choya	8	31.340528, -113.613102
ELP	Estero la Pinta	7	31.248660, -113.260285

Chapter 3: Selection on reproductive compatibility drives incipient speciation across a habitat break in the barred surfperch (*Amphistichus argenteus*)

ABSTRACT

Dispersal limitation and environmental heterogeneity can interact to generate pronounced population structure and promote divergence in coastal marine fishes. The barred surfperch (*Amphistichus argenteus*), a viviparous species lacking a pelagic larval stage, offers an opportunity to examine how life history, habitat context, and selection shape population genomic patterns along a dynamic coastline. Using low-coverage whole-genome resequencing of 253 individuals spanning the complete range of the species from northern California to Baja California, Mexico, we identify four genetic groups based on PCA, ancestry (sNMF), and phylogenetic analyses: three mainland clades separated by breaks at Big Sur and the Palos Verdes Peninsula, and a highly differentiated Santa Rosa Island lineage of likely central-coast origin. Pairwise mainland clade F_{ST} values range from 0.03–0.12; island–mainland divergence ranges from 0.10–0.17. Triangle plots reveal a pronounced deficit of genomic intermediates across the Palos Verdes break, a signature inconsistent with isolation-by-distance and indicative of two deeply diverged lineages. Near-fixed allele-frequency differences at multiple unlinked loci functionally enriched for spermatogenesis and fertilization suggest that this break is not merely a historical artifact but is being actively maintained by selection on reproductive compatibility. This focused selection signal at loci governing spermatogenesis and sperm motility mirrors findings in the confamilial black surfperch along the same coastline. Combined with a long–documented difference in breeding phenology near this break, this strongly suggests the northern–central and southern lineages are reproductively

isolated and may represent incipient species. This work establishes the first population genomic baseline for a recreationally important apelagic marine fish and illustrates how historical vicariance, habitat discontinuity, and selection on reproductive compatibility together initiate speciation in vertebrates with limited dispersal.

INTRODUCTION

Species diversity within sympatric systems has been a longstanding focus in evolutionary biology, particularly in marine environments where clear physical barriers to dispersal are often absent (Faria et al., 2021; Lessios & Robertson, 2006; Palumbi, 1994; Rocha & Bowen, 2008). Many marine organisms display a bipartite life history strategy, in which pelagic larvae disperse and are transported over potentially vast distances by ocean currents. Under such conditions, high gene flow and weak population structure is expected (Cowen & Sponaugle, 2009; Palumbi, 1992; Pinsky et al., 2017). Yet, gene flow is affected by a variety of factors beyond pelagic larval duration (PLD), as studies have shown a weak correlation of PLD with gene flow (Riginos et al., 2014; Weersing & Toonen, 2009). Present-day population structure often reflects a complex interplay among historical vicariance, dispersal, demographic processes, and selection, necessitating system-specific context (Bowen et al., 2020; Hedgecock et al., 2007; Hellberg et al., 2002; Nosil, 2012; Selkoe & Toonen, 2011).

While PLD and dispersal potential are often emphasized in marine population genomic studies, patterns of gene flow in sympatric marine populations can be shaped by numerous factors including adult movement (Green et al., 2015), oceanographic patterns (e.g., currents, eddies, storm-driven events) (Cowen & Sponaugle, 2009), habitat availability (Selkoe et al., 2016), effective population size (N_e) (Hedgecock & Pudovkin, 2011), and local selective pressures (Bernardi et al., 2016; Sanford & Kelly, 2011). Genetic differentiation among populations cannot be interpreted as a simple proxy for dispersal, nor does genetic homogeneity necessarily indicate high gene flow (Hedgecock et al., 2007); rather, robust inference requires integrating genetic patterns with ecological and demographic context. These interacting processes determine not only whether individuals move, but whether migrants survive and successfully reproduce, ultimately governing effective gene flow (Hellberg et al., 2002; Slatkin, 1987). Understanding how such factors generate and maintain population structure is a central goal of molecular ecology (Nosil, 2012; Seehausen et al., 2014).

Marine fishes provide a wide variety of systems for investigating the role of these processes, and insights from population genomics carry important applied implications (Carbonneau et al., 2024; Perrot et al., 2026). Patterns of genetic structure can directly affect how populations are defined, conserved, and managed. For example, genomic analyses have informed conservation strategies for threatened marine species (Gruenthal et al., 2014; Prince et al., 2017) and have revealed unexpected population structure in commercially important fishes (Bradbury et al.,

2010; G. C. Longo et al., 2025), underscoring the need to manage biologically distinct units informed by population genomics. Population genomics provides a powerful framework for linking evolutionary processes to conservation and fisheries management in marine systems.

Applying population genomics frameworks to taxa with unusual life histories can yield valuable insight into how dispersal, demography, and selection interact to shape population structure. Surfperches (family *Embiotocidae*), are an informative group for examining the relationship between life history and population structure. The surfperches comprise an adaptive radiation in which niche partitioning (e.g., kelp–, seagrass–, rocky reef–, sand–associated species) has been implicated in ecological speciation within the family (DeMartini, 1969; G. Longo & Bernardi, 2015). The family includes 23 species found in temperate coastal waters of the North Pacific from Mexico to Japan, but are absent from northern latitudes along the Aleutian Islands. All surfperches are viviparous, with internal fertilization and direct development, giving birth to fully-formed juveniles. Consequently, they differ from most marine fishes in lacking a pelagic larval stage, also known as “apelagic” species. This reproductive mode is expected to reduce dispersal and promote population structure (Bernardi, 2000, 2005). Previous genomic work on embiotocids has revealed pronounced phylogeographic subdivision consistent with these expectations (Toy, 2022; Toy et al., 2025). However, linking patterns of divergence to ecological specialization represents an important next step, particularly through the first

population genomic investigation within the sand-dwelling surfperches (genus *Amphistichus*).

The barred surfperch (*Amphistichus argenteus*) is a sand–specialist that occupies surf-zone habitats along much of the California coast and into the northern Baja California peninsula, spanning approximately ten degrees of latitude (28°–38° N) (Figure 1). Sandy beach habitat is broadly but discontinuously distributed across this range, interrupted at multiple points by rocky headlands and complex shorelines (Allen, 2006). The species' apelagic life history further restricts long-distance dispersal to juvenile and adult movement (Bernardi, 2000, 2005). Gene flow is therefore expected to depend on the location and continuity of sandy habitat along the coast. Barred surfperch are unique among surfperches because they support a popular recreational fishery (Carlisle et al., 1960; Oda et al., 2019). Recent evaluations of marine protected areas (MPAs) in California reported significantly higher abundances of juvenile and large adult barred surfperch within protected areas (Marraffini et al., 2024), suggesting that fishing pressure affects population size and age structure in this species. Such exploitation may reduce effective population size by disproportionately removing large reproductive individuals, potentially increasing the influence of genetic drift and promoting population divergence even in the presence of occasional dispersal (Hauser et al., 2002; Hauser & Carvalho, 2008). This combination of ecological, life-history, and anthropogenic factors presents a rare opportunity to examine how multiple evolutionary drivers operate simultaneously to shape population genomic patterns in the barred surfperch.

In the present study, we address: (1) to what extent does gene flow occur among coastal populations?; (2) what is the genetic status of Santa Rosa island individuals?; (3) is there evidence for local adaptation, particularly genomic regions under strong selection? Here, we address these questions using genome-wide nuclear markers generated through low-coverage whole-genome resequencing. By integrating population structure analyses, estimates of gene flow, and genome scans for selection, we evaluate whether divergence in barred surfperch reflects genome-wide isolation, heterogeneous genomic differentiation maintained by selection, or a combination of both. This work provides new insight into how low dispersal, demography, and selection interact to shape genomic variation in nearshore marine fishes and establishes a baseline for assessing future genetic change under ongoing environmental and anthropogenic pressures.

METHODS

Reference Genome

We collected an adult male specimen by hook and line from Carmel, CA, USA (36.5546°N, 121.93103°W) for reference genome sequencing and assembly (CDFW permit S-221030003-23277-001). We visually sexed the specimen and sampled fin tissue for DNA extraction, after which the specimen was preserved in 95% ethanol and stored at -20°C. Genomic DNA was extracted from fin tissue using the Qiagen DNeasy Blood & Tissue Kit and high molecular weight was confirmed on a 0.7% agarose gel. Genomic DNA was then used to prepare libraries for Nanopore

long-read sequencing. We prepared two libraries for two separate Nanopore sequencing runs using the SQK-LSK109 chemistry kit and protocol. Libraries were sequenced on a Nanopore MinION device using two R9.4.1 flow cells. In total, we obtained 9.69 Gb of raw sequence data (1,598,015 raw reads), with an N50 of 7.2kb. Raw reads were base-called with Guppy v6.4.6 (Oxford Nanopore Technologies, Oxford, UK), retaining only reads that passed the default quality filters. Outputs from the two runs were concatenated into a single dataset, and read length and quality distributions were characterized with NanoStat (De Coster et al., 2018) and the FastQ utilities (Cock et al., 2010). Adapter and primer sequences were trimmed with Porechop (https://github.com/rrwick/Porechop), and post-trimming quality was re-verified with FastQC (Andrews, 2010). To balance read depth against sequence quality, we used NanoFilt (De Coster et al., 2018) to generate two filtered datasets: one comprising reads ≥1000 bp with Q ≥3, and a second comprising reads ≥500 bp with Q ≥5. Each dataset was independently assembled de novo using wtdbg2 (Ruan & Li, 2020), and consensus sequences were produced with wtpoa-cns (Ruan & Li, 2020). Assembly contiguity was summarized with assembly-stats (https://github.com/sanger-pathogens/assembly-stats), and gene-content completeness was evaluated with BUSCO v5.2.2 against the actinopterygii_odb10 lineage database (Manni et al., 2021). The assemblies were then polished by realigning the input reads with minimap2 (Li, 2018) and applying consensus correction in Racon (Vaser et al., 2017). Finally, we screened the polished assemblies for contamination with Kraken2 (Wood et al., 2019) and removed any contigs classified as bacterial.

To improve genome completeness, the A. argenteus genome assembly was scaffolded using RagTag v2.1.0 (Alonge et al., 2019). The reference genome of Embiotoca jacksoni (GenBank accession no. GCA_022577435), a closely related species, was used as a guide (Bernardi et al., 2022). Pairwise alignments between the draft assembly and the reference genome were generated using minimap2 v2.30 (Li, 2018) with the asm5 preset. RagTag was then used to infer contig order and orientation based on conserved synteny. Assembly quality and completeness were assessed before and after scaffolding using BUSCO v5 (Simão et al., 2015) with the Actinopterygii_odb10 lineage dataset. Based on improvements in scaffold contiguity and gene completeness, the RagTag-scaffolded assembly was used as the reference genome for all downstream analyses.

Population Genomics Sampling

We collected 253 fish by hook and line at 21 sites, covering the range of the species (Table 1) (CDFW permit S-221030003-23277-001; RPRS permit CHIS-2024-SCI-0005). Specimens were sampled for fin clips, sexed, photographed, and returned to the environment. Fin clips were immediately placed in 95% ethanol and kept as cool as possible in the field (ambient temperature, or on ice) until reaching the laboratory where they were stored at −20°C.

DNA Extraction and Library Preparation

Genomic DNA was isolated from all samples using the Promega Wizard Genomic DNA Purification Kit, and DNA quality was assessed on a 0.7% agarose gel. Sequencing libraries were prepared from genomic DNA using half-reaction

volumes of the NEBNext UltraExpress FS DNA Library Prep Kit (New England Biolabs), together with unique dual indices supplied by the sequencing facility. The enzymatic fragmentation step was carried out with a 20-minute incubation, and libraries were amplified for five PCR cycles using dual unique index primers. Final libraries were eluted in 15 μL of 0.1× TE buffer. Library concentrations were measured using a Qubit fluorometer (Invitrogen), and fragment size distribution was evaluated on an Agilent system prior to pooling. Pooled libraries were sequenced as 150 bp paired-end reads on an Illumina NovaSeq 6000 (S4) platform at the Vincent J. Coates Genomics Sequencing Laboratory at the University of California, Berkeley, targeting approximately 8× genome-wide coverage.

The analyses described in Sections 2.4–2.5 and 2.7–2.11 were implemented using a published workflow from Toy et al., available on GitHub (https://github.com/Arthur-Oules/EJA_lcwgs_GenPop).

Map Generation

To contextualize the observed genetic variation within its geographical and environmental setting, sampling sites were overlaid onto an oceanographic map constructed from E.U. Copernicus Marine Service Information products. Sea surface temperature data were drawn from 'SST_GLO_SST_L4_NRT_OBSERVATIONS_010_001' (1/20th degree grid) (European Union-Copernicus Marine Service, 2015), while surface current data were obtained from 'GLOBAL_ANALYSISFORECAST_PHY_001_024' (1/12th degree horizontal resolution) (European Union-Copernicus Marine Service, 2016) (Figure 1B). Daily values through 2023 were averaged across

both datasets. High-resolution coastlines were sourced from the Natural Earth dataset and imported into RStudio via the ne_states function from the rnaturalearth 1.0.1 package (Massicotte & South, 2025).

Mapping and Variant Calling

Read mapping and variant calling were carried out against the reference genome described in Section 2.1, using the snpArcher workflow (Mirchandani et al., 2024) to produce high-quality variant calls suitable for downstream analysis. Raw reads underwent quality filtering and adapter trimming with fastp (Chen et al., 2018) prior to alignment against the concatenated reference genome with bwa mem (Li, 2013) using recommended parameters. Per-individual variant calling was then performed with Sentieon Haplotyper (Kendig et al., 2019), followed by joint genotyping via Sentieon Genotyper to produce a combined VCF (variant call format) file. Sites that failed GATK hard filtering thresholds defined in the snpArcher workflow were removed. We generated a second SNP dataset by applying minor allele frequency (MAF) filtering and removing multiallelic sites and indels. To reduce noise from rare variants, we used bcftools (Danecek et al., 2021) to retain only variants with MAF ≥ 0.01. With the exception of Manhattan plots and genes under selection, the following methods and results are described using the MAF-filtered, biallelic SNP data set.

Sex-associated regions

Prior to population genomic analyses, sex-associated genomic regions were identified and excluded from the dataset to prevent confounding signals from

distorting inference of population structure. An initial PCADAPT run on the full SNP dataset revealed that PC2 (5.59% variance explained) separated males and females into two discrete groups within each geographic clade (Figure S6A), a pattern inconsistent with broad-scale population structure. Inspection of the corresponding Manhattan plot confirmed that this signal was driven by a small number of contigs exhibiting extraordinarily elevated $-\log_{10}$(p-values) far exceeding the Bonferroni significance threshold relative to the broader genomic background (Figure S6B). Five contigs displaying this extreme differentiation (9.1, 10.1, 45.1, 75.1, and 82.1) were flagged as putatively sex-linked. For each contig, the genomic interval spanning all SNPs with $-\log_{10}$(p-value) > 50 was extracted and written to a BED file. These regions were then masked from the VCF using bcftools view (Danecek et al., 2021), and all subsequent analyses were conducted on the resulting autosomal SNP dataset.

Population F_{ST} and Hamming Distances

We used the --fst option in PLINK v2.00a6 (Chang et al., 2015) to estimate Wright's fixation index (F_{ST}) for each pair of populations, applying the Weir and Cockerham (1984) estimator. In addition, we calculated pairwise Hamming distances among individuals using the --distance function in PLINK v1.9.0 beta 7 (Chang et al., 2015), which reports differences in allele counts. Population-level Hamming distances were then obtained by averaging distances between individuals from different populations, with a correction applied based on within-population distances.

Analysis of Population Structure and Ancestry

To assess population structure, we conducted a genome-wide principal component analysis (PCA) using PCADAPT v4.4.0 (Privé et al., 2020). Raw data were first converted to PED format using the --export option in PLINK v2.00a6 (Chang et al., 2015), and subsequently transformed into GENO format for ancestry inference with sNMF using the ped2geno function in LEA v3.16.0 (Frichot & François, 2015). Individual ancestry proportions and allele frequencies were estimated using the snmf function, which implements a non-negative matrix factorization framework with least-squares optimization. This approach yields results comparable to STRUCTURE or ADMIXTURE analyses under similar assumptions (e.g., absence of genetic drift and linkage equilibrium in ancestral populations), while not requiring Hardy–Weinberg equilibrium and offering improved computational efficiency for large SNP datasets (>10,000 markers) (Frichot et al., 2014).

The snmf procedure also computes a cross-entropy criterion based on masked genotype prediction, which is used to evaluate model fit across different numbers of ancestral populations (K) and to select among replicate runs for a given K (Frichot et al., 2014). To identify the most appropriate value of K, we performed 30 independent runs of sNMF for K values ranging from 2 to 8 and compared cross-entropy scores (Figure S2). The α-regularization parameter was kept at its default value of 10, as its influence is primarily relevant for smaller datasets (<10,000 SNPs). The optimal run and corresponding K value, defined by the lowest cross-entropy, are presented in Figure 1C.

Isolation-by-Distance (IBD)

We evaluated the relationship between genetic differentiation and geographic separation (isolation by distance, IBD). Geographic distances among sampling sites were calculated as great-circle distances using the distm() function with distVincentyEllipsoid from the R package geosphere (Hijmans, 2010). Genetic distances were estimated from a pruned SNP dataset in 10 kb windows using Hudson's F_{ST}, implemented with the allel.windowed_hudson_fst() function in scikit-allel v1.3.6 (Miles, 2015). We also used the PCA results, treating sample coordinates in principal component space as an additional representation of genetic structure for distance-based comparisons. In addition, pairwise Cavalli-Sforza and Edwards chord distances were calculated using dist.genpop() (method = 2) in the R package adegenet v2.1.10 (Jombart, 2008). To further explore the relationship between genetic and geographic distance, we fit linear models using the base R function stats::lm() (e.g., $F_{ST} \sim m * \text{Physical Distance} + b$). The significance of isolation by distance patterns was assessed using Mantel permutation tests implemented with mantel.randtest() in adegenet v2.1.10 (Jombart, 2008).

Triangle Plots

To distinguish among alternative explanations for spatial genetic structure, we used triangle plots (Wiens & Colella, 2025), which depict the relationship between hybrid index and interclass heterozygosity. In a two-population framework, the hybrid index represents the proportion of ancestry derived from each parental population, while interclass heterozygosity quantifies the fraction of loci carrying alleles from both parental groups. Under this model, parental individuals are expected to occupy

opposite vertices of the triangle (hybrid indices of 0 and 1 with zero interclass heterozygosity), whereas F1 hybrids and later-generation backcrosses should fall above the Hardy–Weinberg expectation boundary. Individuals deviating from Hardy–Weinberg equilibrium may instead plot below this boundary (Wiens & Colella, 2024).

Simulation work indicates that this approach can help discriminate isolation by distance (IBD) from neutral diffusion (ND) when introgression is relatively recent (Wiens & Colella, 2025). Specifically, IBD is expected to produce a negative linear relationship between hybrid index and interclass heterozygosity, whereas ND generates a curved distribution within the triangle space.

Triangle plots were generated using the triangulaR package (Wiens & Colella, 2024). We first applied the alleleFreqDiff() function to filter loci based on allele frequency differences (δ) of 0.5, 0.75, and 1 between parental populations (n = 11 (STI) and n = 15 (PLM)), thereby identifying ancestry-informative markers (i.e., loci fixed or nearly fixed between parental groups). This yielded 41,957, 8,116, and 721 retained SNPs, respectively. Hybrid index and interclass heterozygosity were then calculated for each filtered dataset using the hybridIndex() function.

Outlier Loci Analysis

To identify candidate loci under selection, the VCF file was converted to BED format using the --make-bed function in PLINK v2.00a6 (Chang et al., 2015). We then applied the PCADAPT 4.4.0 package (Privé et al., 2020) to perform a PCA on the centered and scaled genotype matrix. For each SNP, a Mahalanobis distance —

representing each point's deviation from the main distribution — was computed across the K retained principal components, then squared and divided by the genomic inflation factor to yield a test statistic approximating a chi-squared distribution with K degrees of freedom, from which p-values were derived. The number of components retained (K) was determined using Cattell's rule, with biplots examined to confirm that retained PCs contributed meaningfully to data structure. Negative log-transformed p-values were plotted along the genome to produce Manhattan plots, and Bonferroni correction was applied for multiple testing with a significance cutoff of $\alpha = 0.01$. Outlier SNP positions were subsequently mapped onto an annotated reference genome of the closely related species *Embiotoca jacksoni* (GenBank Accession: GCF_022577435.1-RS_2024_1) (Bernardi et al., 2022) to extract associated gene annotations.

Phylogenetic Analysis

We created a maximum-likelihood (ML) tree by converting the VCF into a single sequence alignment using vcf2phylip.py (Ortiz, 2019). The resulting FASTA alignment was used to infer an ML tree in IQ-TREE v2.2.2.6 (Minh et al., 2020) under the GTR+G4 substitution model, with branch support assessed using 1000 replicates of the ultrafast bootstrap approximation (UFBoot; Hoang et al., 2018). The ML consensus tree was visualized and annotated in FigTree (https://github.com/rambaut/figtree).

RESULTS

Reference Genome of A. argenteus

The final reference genome assembly spanned 571 Mb across 354 scaffolds, with a mean scaffold length of 16 Mb, a maximum scaffold length of 25.55 Mb, an N50 of 16.12 Mb, and an overall GC content of 41.6%. Scaffolding with RagTag substantially reduced assembly fragmentation, decreasing the number of scaffolds from 3,150 to 354, and improved BUSCO completeness from 92.6% to 94.1%, driven primarily by a reduction in fragmented BUSCOs. Assembly completeness was assessed using BUSCO v5.2.2 (Simão et al., 2015) with the actinopterygii_odb10 dataset (n = 3,640 genes) in genome mode using MetaEuk as the gene predictor, yielding 94.1% complete BUSCOs (93.4% single-copy, 0.7% duplicated), 1.0% fragmented, and 4.9% missing.

Sampling, Loci, and Polymorphism Statistics

We sampled 253 individuals of *Amphistichus argenteus* from Stinson Beach (California, USA) in the north to the Playa la Misión (Baja California, Mexico) in the south, covering nearly the entire known latitudinal range of the species (Figure 1B). Our sampling included one of the northern California Channel Islands, Santa Rosa Island (California, USA) (Figure 1B, Table 1). Low coverage genomic sequencing of our samples resulted in an average 8.5X coverage per individual. In total, 15,805,202 SNPs were discovered before any filters were applied. The GATK best practices filters (Van Der Auwera et al., 2013) were then applied to this dataset; 2,389,190 were removed, leaving 13,416,012 SNPs. We further filtered the dataset as described in the

methods section, resulting in 1,432,059 biallelic SNPs. Population genomic analyses were performed using the filtered data set, unless stated otherwise.

Population Structure

Population structure was first assessed using a Principal Components Analysis (PCA). PC1 (7.0% of variance) separated Santa Rosa Island individuals from all coastal populations, while PC2 (3.1%) resolved three distinct coastal clusters corresponding to northern (Stinson Beach to Carmel), central (Gamboa to Venice Beach), and southern (Newport Beach to Playa La Misión) populations (Figure 1A). The sharp discontinuities between clusters, rather than a gradual spread of points along the axes, are consistent with discrete population breaks rather than continuous isolation-by-distance, a pattern that persists when the PCA is restricted to only mainland individuals (Figure S1).

These patterns were mirrored by the ancestry analysis, in which cross-entropy minimization identified K=4 as the optimal number of ancestral genetic populations (Figures 1C, S2, S3). The three coastal clades each correspond to a distinct genetic cluster, with abrupt transitions in ancestry proportions between Carmel and Gamboa in the north and between Venice Beach and Newport Beach in the south (Figure 1C). The central clade shows modest admixture with both the northern and southern clusters, which may reflect ongoing gene flow across these boundaries or retention of shared ancestral polymorphisms. Santa Rosa Island individuals comprised the fourth cluster, showing largely distinct ancestry with minimal signal of shared ancestry with any mainland population.

Isolation by Distance, Neutral Diffusion, and Admixture

The genetic structuring of *A. argenteus* can be described as a separation of the coastal populations from the island population, with the coastal individuals showing a distinct clustering of three unique ancestral populations. Pairwise F_{ST} estimates between the Santa Rosa Island population and each of the coastal groups are 0.13 (Northern Clade), 0.10 (Central Clade), and 0.17 (Southern Clade); pairwise F_{ST} estimates between coastal populations was 0.03 (Northern vs. Central), 0.06 (Central vs. Southern), and 0.12 (Northern vs. Southern) (Figure 5A). Comparisons among coastal sampling sites do not exceed 0.17 (Figure 5B). A phylogenetic analysis based on F_{ST} distances again shows the same pattern, with the coastal populations grouping together as separate clusters separate from the Santa Rosa Island population (Figure S4).

Our analysis of isolation by distance was conducted for the coastal populations, excluding the Santa Rosa Island population. Isolation-by-distance (IBD) was found to be present and significant (Mantel test $p = 0.001$) (Figure 2A), indicating that geographic distance contributes to genetic differentiation among coastal populations. However, a significant IBD signal alone does not distinguish between continuous gene flow across a single panmictic population and secondary contact between historically isolated groups followed by admixture (Wiens & Colella, 2025). To further investigate this, we examined triangle plots of hybrid index against interclass heterozygosity across three allele frequency thresholds (δ; Figure 2B). At $\delta = 0.5$, individuals from each sampling site group tightly, with moderate separation

between northern/central and southern trajectories, consistent with two partially differentiated groups sharing a substantial proportion of alleles. At $\delta = 0.75$, the gap between the northern/central and southern clusters widened, and interclass heterozygosity increased, suggesting that loci of moderate-to-high differentiation disproportionately mark the boundary between these groups. The pattern was most striking at $\delta = 1$, where individuals collapsed to opposite corners of the triangle with a large gap between groups, a signature of fixed differences between two deeply diverged groups with no shared alleles at the most differentiated loci. Some individuals retain intermediate hybrid index values across thresholds, suggesting residual shared ancestry or limited admixture across the break.

This progression stands in contrast to what would be expected under a simple IBD model, in which points would form a continuous spread across the hybrid index axis with uniformly elevated interclass heterozygosity, rather than discrete clusters at opposing poles. Taken together, the triangle plot results suggest that the coastal population structure of *A. argenteus* reflects a deep divergence between northern–central and southern lineages rather than clinal differentiation driven by geographic distance. The sharpest break in this system falls between Venice Beach (Los Angeles County) and Newport Beach (Orange County), across the Palos Verdes Peninsula. The peninsula itself represents a substantial physical barrier to coastal movement as a rocky headland. Palos Verdes interrupts the otherwise continuous nearshore habitat that *A. argenteus* occupies.

Evolutionary Relationships of Populations

A maximum likelihood tree of individuals (Figure 3) and a neighbor-joining tree of populations constructed from pairwise F_{ST} values (Figure S4) both recovered three main clades corresponding to the northern, central, and southern populations identified in the PCA and ADMIXTURE analyses. In both trees, the southern clade (Newport Beach–Playa la Misión) is monophyletic, while the central clade (Gamboa–Venice Beach) is paraphyletic with respect to the northern clade. In the F_{ST} tree, Carmel is subtended by a notably longer branch than the other northern sites, suggesting it is more genetically distinct within the northern clade. The Santa Rosa Island lineage nests within the central clade in the ML tree, with two individuals from central clade sites (Moonstone Beach and Morro Beach) grouping in close proximity to the island lineage, a pattern consistent with ongoing gene flow between the island and adjacent central mainland populations or retention of shared ancestral variation. This placement is consistent with the pairwise F_{ST} data, in which Santa Rosa Island shows the lowest differentiation from the central clade ($F_{ST} = 0.10$) relative to the northern ($F_{ST} = 0.13$) and southern ($F_{ST} = 0.17$) clades (Figure 5). Together, these lines of evidence are most consistent with a central mainland origin for the Santa Rosa Island population.

Outlier Loci and Signals of Selection

We identified outlier loci using PCADAPT, a non-F_{ST} based method well suited for detecting selection in structured populations (Toy et al., 2025). Principal components (PCA), ADMIXTURE, and phylogenetic analyses (Figures 1A, 1C, 3) all show that the species forms four genetically distinct groups: a northern clade, central

clade, southern clade, and island clade. This pattern is further supported by genome-wide Manhattan plots. Pairwise plots comparing each of these 4 groups reveal different levels of divergent alleles (figure 4A). Depending on the comparison, the number of outlier loci (SNPs) varied between 31 and 440 (out of 1,432,059), which mapped to 19 to 119 genes, respectively. When all mainland locations were considered (Figure 4B), we identified 316 outlier loci and 69 genes as outliers.

Genes Under Directional Selection – male reproductive biology

The single strongest signal in the scan was at SDK2B (sidekick cell adhesion molecule 2b; $-\log_{10} p \approx 413$; Figure S5A), where Carmel (CAR) was fixed for the reference allele while all other mainland localities were near-fixed for the alternate (Figure S5B). Among the 69 annotated PCADAPT outliers detected across mainland localities, five additional loci stood out for their combination of statistical strength and biological coherence: MFN2, FAT3A, IGF3, TRDMT1, and CEP290 (Figure S5A). All five have well-supported roles in male reproductive biology. CEP290 encodes a transition-zone ciliary protein required for sperm flagellum assembly (Wu et al., 2020). MFN2 mediates mitochondrial fusion in the spermatid midpiece and is required for sperm motility and energetics (Varuzhanyan et al., 2019). IGF3 is a teleost-specific insulin-like growth factor expressed predominantly in the testis and acts as a central regulator of spermatogenesis (Nóbrega et al., 2015). TRDMT1/DNMT2 is a tRNA cytosine methyltransferase that packages paternal small-RNA cargo for intergenerational transmission (Zhang et al., 2018). FAT3A, an atypical cadherin, has been identified as a target of fisheries-induced evolution

affecting age and size at maturity (Sun et al., 2025). Per-population genotype proportions at each of the five reproductive–biology loci (Figure S5B) revealed a consistent pattern: northern and central populations were nearly fixed for the reference homozygote, while southern populations were nearly fixed for the alternate homozygote. Heterozygotes were rare and were largely restricted to a handful of transitional populations.

DISCUSSION

Dispersal

The genomic patterns we recovered for *Amphistichus argenteus* are consistent with the limited dispersal capacity that is characteristic of surfperches (Baltz, 1984; Wourms, 1981) whose newborn juveniles are released directly into the nearshore surf zone at relatively large body sizes (42–53 mm) (Carlisle et al., 1960). Marine fishes lacking a pelagic larval stage have been known to exhibit limited gene flow (Bernardi, 2005; Toy et al., 2025; Weersing & Toonen, 2009), as opportunities for long-distance dispersal are restricted to juvenile or adult movement. In a species with restricted movement, the effective migration rate between geographically separated localities is low, and genetic drift becomes a comparatively more potent force in driving genomic divergence between sites (Hellberg et al., 2002; Slatkin, 1987). Data presented here extend this expectation to the sand-dwelling surfperches (*Amphistichus*), which have not previously been examined at the population genomic level.

Genetic structure among coastal *A. argenteus* populations was lower than that reported for other surfperches (Toy et al., 2025), with pairwise F_{ST} values as low as 0.01 for localities up to 400 km apart (e.g., Moonstone Beach–Venice Beach; Figure 5). Within clades, adjacent coastal populations generally show low differentiation (F_{ST} = 0.01—0.02), a pattern suggesting substantial regional adult movement. Notable exceptions include sites flanking the boundaries of the major clades (e.g., Carmel–Gamboa F_{ST} = 0.10; Venice Beach–Newport Beach F_{ST} = 0.05). These breaks coincide with complex regions. Although kelp cover off Palos Verdes has declined substantially since the 1950s (Foster & Schiel, 1985), these breaks coincide with areas of complex rocky headlands and historically extensive kelp habitat (Jacobs et al., 2004), suggesting that habitat structure may have contributed to reduced alongshore dispersal. This pattern supports our prediction that habitat discontinuities may contribute to reduced gene flow.

The phylogeographic break we observe near Palos Verdes is particularly notable in light of mark-recapture data. Carlisle et al. (1960) documented an individual tagged in Long Beach (south of Palos Verdes), that was recovered in Santa Monica (north of Palos Verdes), corresponding to a movement of roughly 50 km across the peninsula. This observation contrasts with our genetic results and underscores an important distinction between dispersal and effective gene flow. Individuals may disperse, yet if migrants or their offspring experience reduced survival or reproductive success, genetic structure may persist (Lowe & Allendorf, 2010). We speculate whether post-dispersal processes, such as selection against

migrants or reduced habitat-specific performance, could contribute to maintaining divergence across this region. Notably, Carlisle et al. (1960) also reported mixing among adjacent beaches, a pattern that aligns with our finding of low differentiation among neighboring sandy sites.

Depth distribution adds a further constraint. Barred surfperch occupy the breaking surf along sandy beaches and are rarely observed deeper than 27 m, although captures as deep as 73 m have been reported (Carlisle et al., 1960). This shallow-water ecology makes the deep water separating coastal California from the Channel Islands a likely dispersal barrier, although rare dispersal events to offshore islands may occur (e.g., in heavy storms). Consistent with this expectation, island and mainland populations are highly differentiated (F_{ST} = 0.12–0.19), indicating that successful cross-channel dispersal events are rare. The historical and biogeographic origin of this island lineage is addressed in detail in Section 4.2.

The significant Mantel correlation between geographic and genetic distance (p = 0.001; Figure 2A) is consistent with an IBD signal. However, it more likely reflects the cumulative effect of crossing discrete habitat discontinuities (Carmel–Big Sur; Los Angeles–Orange County) than a smooth, distance-dependent decay in gene flow along the entire coastline, as is observed in the Black surfperch (Toy et al., 2025). This interpretation is supported by discrete clades visualized by PCA (Figure 1A), ADMIXTURE (Figure 1C), pairwise F_{ST} (Figure 5) and long branches subtending certain localities (e.g., Carmel; Gamboa) in the F_{ST} tree (Figure S4). Such locality-specific divergence is expected when semi-isolated populations accumulate

drift independently of neighboring sites. Carmel is the most distinct of these. It exhibits the highest pairwise F_{ST} in the matrix (0.19; Carmel–Santa Rosa Island) and is consistently elevated against every other site (Figure 5). Carmel Beach is a small pocket of sandy habitat (~1 km) flanked on both sides by a rocky coastline. To the north are the headlands of Pebble Beach and Pacific Grove, and to the south are the cliffs and granite outcroppings of Point Lobos and Big Sur. These features interrupt the continuous sandy-beach habitat *A. argenteus* requires for alongshore movement. Functioning as a long-standing habitat "island", the Carmel population may receive very few immigrants from neighboring sandy beaches and may have accumulated drift largely independently.

Limited dispersal and genetic drift alone do not fully account for the broad coastal structure we observe. Triangle plots show discrete clusters at opposing poles of the hybrid index at high δ rather than a continuous spread (Figure 2B). This type of pattern is difficult to generate under a pure stepping-stone model of drift and migration along a linear coastline. A simple IBD scenario predicts a smooth gradient of intermediate genotypes, not the bimodal genotype distributions we observe at the most differentiated loci (Wiens & Colella, 2025). The bimodal pattern instead implies that the northern–central and southern groups carry many fixed allele differences, a signature of historically isolated lineages. Equally informative is the gap in the center of the triangle plot: individuals with intermediate hybrid index are absent (Figure 2B), indicating that contemporary hybridization between the two groups is rare, even at sites immediately adjacent to the Palos Verdes break. We therefore interpret the

coastal structure as the joint outcome of (i) chronically low dispersal, which makes drift effective at regional scales, and (ii) a historical reduction in gene flow across the Palos Verdes Peninsula, which permitted the northern–central and southern groups to drift apart as semi-independent evolutionary units before any secondary contact.

Historical context: Pleistocene sea-level oscillations and the Santarosae Island complex

The population structure recovered for *A. argenteus* must be interpreted in the context of the dynamic Quaternary history of the California coastline. The southern California Bight is one of the best-studied regions of coastal phylogeographic discontinuity in the northeastern Pacific. Phylogeographic breaks at the Los Angeles region, Palos Verdes Peninsula, and Point Conception have been documented in multiple dispersal-limited fishes and invertebrates along the southern California coast (Bernardi, 2000; Dawson, 2001; Dawson et al., 2002; Jacobs et al., 2004), although species with broader dispersal capacity often show no breaks (Burton, 1998; Palumbi, 1994). The deep break we observe across the Palos Verdes Peninsula in *A. argenteus* falls within the dispersal-limited pattern, and its concordance with previously described breaks in the confamilial surfperches *Embiotoca jacksoni* and *E. lateralis* (Bernardi, 2005) suggests a potential shared historical driver rather than a unique feature of *A. argenteus*.

Repeated Pleistocene glacial–interglacial cycles produced sea-level fluctuations on the order of 120m (Siddall et al., 2003), reorganizing the position and connectivity of sandy-beach habitats along the coast. These events likely fragmented

continuously distributed nearshore species into regional refugia separated by stretches of unsuitable rocky or steeply incised shoreline (Marko et al., 2010). The Illinoian glaciation (~400 ka) has been invoked to account for the timing of the *E. jacksoni* north–south split (Bernardi, 2005). A similar mechanism resulting in the divergence of northern and southern lineages during a glacial lowstand, followed by post-glacial range expansion and secondary contact at Palos Verdes, provides a parsimonious explanation for the bimodal triangle-plot configuration (Figure 2B) and the fixed-difference outliers (Figure S5) reported here. Under this scenario, the contemporary genetic break is in part a legacy of historical allopatry, with dispersal limitation along the present-day coastline maintaining but not generating the divergence.

Origin of the Santa Rosa Island population. The geographic origin of island populations is of particular interest in marine phylogeography. The deep 30 km Santa Barbara Channel is a major barrier for a species lacking a pelagic dispersive stage, and successful crossings are likely restricted to rare events such as offshore transport during major storms. Such crossings were more likely during Pleistocene lowstands, when the channel was far narrower. Though the northern Channel Islands (Anacapa, Santa Cruz, Santa Rosa, and San Miguel) are distinct islands today, they previously existed as a single island, Santarosae (Orr, 1968), during the Last Glacial Maximum when sea level was ~120 m lower (Siddall et al., 2003), and the distance from the mainland was just ~7 km (Kennett et al., 2008). It is therefore possible that island populations were established on Santarosae prior to the isolation of Santa Rosa Island

as sea levels rose. This hypothesis is supported by mitochondrial data from the black surfperch (*Embiotoca jacksoni*), which exhibits shared maternal haplotypes among the four northern Channel Islands, suggesting historical connectivity or a common source population (Bernardi, 2000).

Several lines of evidence point to a central-coast founder source for the Santa Rosa Island lineage. The lineage forms a discrete ancestry component (Figure 1C, S2), separates cleanly from the mainland on PC1 (Figure 1A), and exhibits high F_{ST} (0.10–0.17; Figure 5) against every mainland clade. Its placement within the central clade in the ML tree (Figure 3) and its minimum F_{ST} against central sites (Figure 5) point to what is today the central coast as the likely source population. The magnitude of differentiation is consistent with a founder event followed by sustained drift in a small, isolated population. Two central-coast individuals (Moonstone Beach 14, Morro Beach 1; Figure 3) cluster near the island lineage which could reflect occasional contemporary exchange or incomplete lineage sorting. Future work sampling additional Channel Islands would help resolve whether the Santarosae founding signal is shared across the system.

Taken together, the population structure of *A. argenteus* is best understood as the joint product of two interacting forces operating on different timescales: deep historical events that established the major lineages we observe today, and the restricted contemporary dispersal regime that prevents those lineages from being homogenized by gene flow in the present. This combination of historical vicariance and ongoing dispersal limitation is similar in form to what has been proposed for *E.*

jacksoni (Bernardi, 2000, 2005) and the tidewater goby *Eucyclogobius newberryi* (Dawson et al., 2001, 2002), and reinforces the California coast as a region in which Pleistocene history is preserved in extant phylogeographic structure with particular fidelity in low-dispersal, habitat-restricted marine taxa.

Natural Selection

Set against the drift-dominated demographic backdrop of this system, the PCADAPT scans reveal a focused signal of locus-specific selection that drift alone cannot explain. Across all mainland individuals, an outlier scan identified 316 outlier loci in 69 genes (Figure 4B), with the strongest outliers reaching extreme values and configured in a way that is qualitatively dissimilar to the rest of the genome. At the five top loci (MFN2, FAT3A, IGF3, TRDMT1, CEP290; Figure S5B), northern and central populations are nearly fixed for the reference homozygote while southern populations are nearly fixed for the alternate homozygote, with heterozygotes appearing at only a handful of transitional sites. This bimodal, near-private configuration at a non-random subset of loci is the type of pattern that requires divergent selection of sufficient strength to overcome ongoing gene flow, or historical allopatry maintained by current dispersal limitation, or a combination of the two.

The functional composition of these outliers maps remarkably closely to outliers recovered for the black surfperch (Toy et al., 2025). In their five-cluster system, the top divergent loci among groups mapped to SPAG1 (sperm-associated antigen 1-like), IZUMO1 (the canonical sperm–egg fusion factor), SMOX (spermine oxidase), and SPATA18 (spermatogenesis associated 18). Among our top *A. argenteus*

outliers are: CEP290 (transition-zone ciliary protein required for sperm flagellum assembly; (Z. Wu et al., 2020)), MFN2 (mitochondrial fusion in the spermatid midpiece, sperm motility and energetics; (Varuzhanyan et al., 2019)), IGF3 (teleost-specific testis-restricted regulator of spermatogenesis; (Nóbrega et al., 2015)), and TRDMT1/DNMT2 (tRNA cytosine methyltransferase, paternal small-RNA cargo and epigenetic transmission; (Zhang et al., 2018)). While these represent a different set of genes, they converge on the same functional axis: spermatogenesis, sperm motility and structural integrity, fertilization, and paternal-effect transmission to offspring. The probability that two independent population-genomic scans, on two confamilial species sampled along the same coastline but occupying very different ecological guilds (Amphistichinae: sandy-beach surf zone; Embiotocinae: rocky reef and kelp forest), would each recover their strongest selection signal in male-gamete biology is low. The most parsimonious interpretation is that the shared reproductive biology of Embiotocidae (i.e., internal fertilization, viviparity, sperm storage of up to several months, delayed fertilization, and multiple paternity (Bond & Forsgren, 2022; Froeschke et al., 2007; Izumiyama et al., 2020; Reisser et al., 2009)) generates persistent sexual and post-copulatory selection on sperm performance and gamete recognition. Such selection can drive rapid divergence at reproductive loci even in the face of high dispersal, as documented in *Echinometra* sea urchins, where it produces species-level fertilization barriers between broadly distributed congeners (Palumbi & Metz, 1991). In low-dispersal embiotocids, the same kind of selection appears to drive analogous divergence at the population level within a species, likely because

restricted gene flow allows locus-specific differences to accumulate among demographically isolated populations. The near-fixed, functionally clustered hits we recover here, and that are seen in *E. jacksoni* (Toy et al., 2025), are consistent with this scenario (Bernardi, 2013; Palumbi, 2009).

This convergent functional signature strengthens the inference that the coastal genetic break across the Palos Verdes Peninsula is not merely a historical artifact but is being actively maintained, and perhaps deepened, by selection on traits directly relevant to reproductive compatibility. This process is often referred to as reinforcement (Servedio & Noor, 2003). The combination of near-fixed allele-frequency differences at multiple unlinked loci that are functionally enriched for fertilization and spermatogenesis is the same configuration interpreted as supporting evidence of incipient or cryptic speciation between coastal *E. jacksoni* and their offshore island lineages (Toy et al., 2025). Our system shows a similar signature, ultimately suggesting that the northern–central and southern *A. argenteus* lineages are best viewed as partially reproductively isolated rather than as endpoints of a smooth cline.

The outlier set is not exclusively reproductive, however. For example, FAT3A, an atypical cadherin identified as a GWAS hit for body weight under experimental size-selective harvest in marine medaka (Sun et al., 2025), emerged as one of our strongest candidates. Given the intense size-selective fishing pressure experienced by *A. argenteus*, this locus may reflect selection on growth- or size-related traits, although this possibility remains speculative. The species' broad latitudinal range also

raises the possibility that thermal and other abiotic clines impose additional selection along the coast, a hypothesis that environmental association analyses could test directly.

Incipient Speciation – Reproductive Isolation across the Palos Verdes peninsula

The central evolutionary finding of this study is that barred surfperch collected on either side of the Palos Verdes Peninsula carry fixed or near-fixed allele-frequency differences at a small set of loci governing spermatogenesis, sperm motility, and paternal epigenetic transmission (MFN2, IGF3, TRDMT1, CEP290; Figure S5B). Genome-wide pairwise F_{ST} across this break is 0.06 (Venice Beach–Newport Beach; Figure 5) signaling that populations remain broadly similar across most of the genome. Two non-mutually-exclusive explanations can produce this disconnect: divergence is evolutionarily recent and the broader genome has not yet differentiated, or persistent gene flow homogenizes the neutral background while divergent selection sustains differentiation at a focused set of loci. Both converge on the same conclusion. The northern–central and southern coastal lineages of A. argenteus are not the ends of a continuous cline; they are two semi-independent reproductive units, and their divergence is concentrated in genes that govern reproductive compatibility.

This molecular signature aligns with the earliest reported biological discontinuity in A. argenteus. Carlisle et al. (1960) found that the reproductive season is earlier in southern populations than in northern ones (north of Oxnard): gravid females sampled in late February already carried advanced embryos in the south,

while females further north still carried unfertilized eggs. They attributed this latitudinal phenology gradient to water temperature, but in light of our data it raises a second possibility: allochronic isolation, in which divergence is reinforced by asynchronous breeding seasons (Taylor & Friesen, 2017). If southern females reach advanced gravidity weeks before northern ones, gametic mismatch reduces opportunities for gene exchange even where the lineages contact, and the reproductive-gene differences we recover may both reflect and reinforce this temporal separation.

Considering (i) fixed and near-fixed differences at multiple unlinked reproductive-biology loci (Figure S5B), (ii) the bimodal hybrid-index distribution with an empty central region indicating an absence of contemporary hybrids (Figure 2B), (iii) the documented north–south gradient in breeding phenology consistent with allochronic isolation (Carlisle et al., 1960; Taylor & Friesen, 2017), and (iv) the sharp ancestry transition between Venice Beach and Newport Beach (Figure 1A, 1C), a cogent argument can be made that the northern–central and southern *A. argenteus* lineages represent incipient species. Under the Biological Species Concept (Mayr, 1963), species are defined by reproductive isolation, and our data are consistent with substantial restrictions on contemporary gene flow between the two lineages. Under the Genotypic Cluster Species Concept (Mallet, 1995), species are recognized as discrete genetic clusters separated by a deficit of intermediates, and the pronounced bimodality in ancestry and hybrid-index values is consistent with this expectation. The modest genome-wide F_{ST} is not a counterargument: reproductive isolation

routinely persists at small subsets of loci against a backdrop of low overall differentiation (e.g., "islands of speciation" pattern; Nosil et al., 2009; C.-I. Wu, 2001), and the profile we recover fits that model directly. We therefore interpret the northern–central and southern mainland *A. argenteus* lineages as effectively reproductively isolated under modern operational species concepts. These lineages may constitute incipient species and are the most informative system in our dataset for understanding how speciation begins in apelagic marine vertebrates.

CONCLUSION

Barred surfperch (*Amphistichus argenteus*) are sand-associated nearshore fishes that share the defining reproductive traits of the surfperch family. Internal fertilization, sperm storage, delayed fertilization, and multiple paternity are characteristics that create persistent post-copulatory selection on sperm performance and gamete recognition. Our genomic analyses reveal a hierarchical pattern of population structure along the California coastline shaped by the interplay of historical vicariance, habitat discontinuity, and dispersal limitation, with the deepest break occurring across the Palos Verdes Peninsula. A combination of discrete ancestry clusters, pronounced bimodality in hybrid-index distributions with an absence of contemporary intermediates, documented north–south differences in breeding phenology consistent with allochronic isolation, and near-fixed allele-frequency differences at loci governing spermatogenesis, sperm motility, and paternal epigenetic transmission provides a compelling case that the northern–central

and southern lineages of *A. argenteus* represent incipient species. This conclusion is reinforced by the striking convergence between our top selection candidates and the reproductive-biology loci independently identified under selection in the confamilial black surfperch, *Embiotoca jacksoni*, surveyed along the same coastline. The surfperch family comprises 23 recognized species that share these reproductive life-history characteristics while differing markedly in ecology. This makes them an unusually tractable comparative framework for understanding how dispersal regime and habitat context modulate the pace and genomic architecture of speciation. Future work can build on results presented here, taking advantage of this unique family to more deeply understand how differences in ecology shape genomic divergence, and how selection on reproductive compatibility can drive populations toward reproductive isolation even across short geographic distances. Lastly, this work establishes a genomic baseline for a recreationally important species, underscoring the broader value of population genomics for informing the management of biologically distinct units under ongoing environmental and anthropogenic pressures.

FIGURES

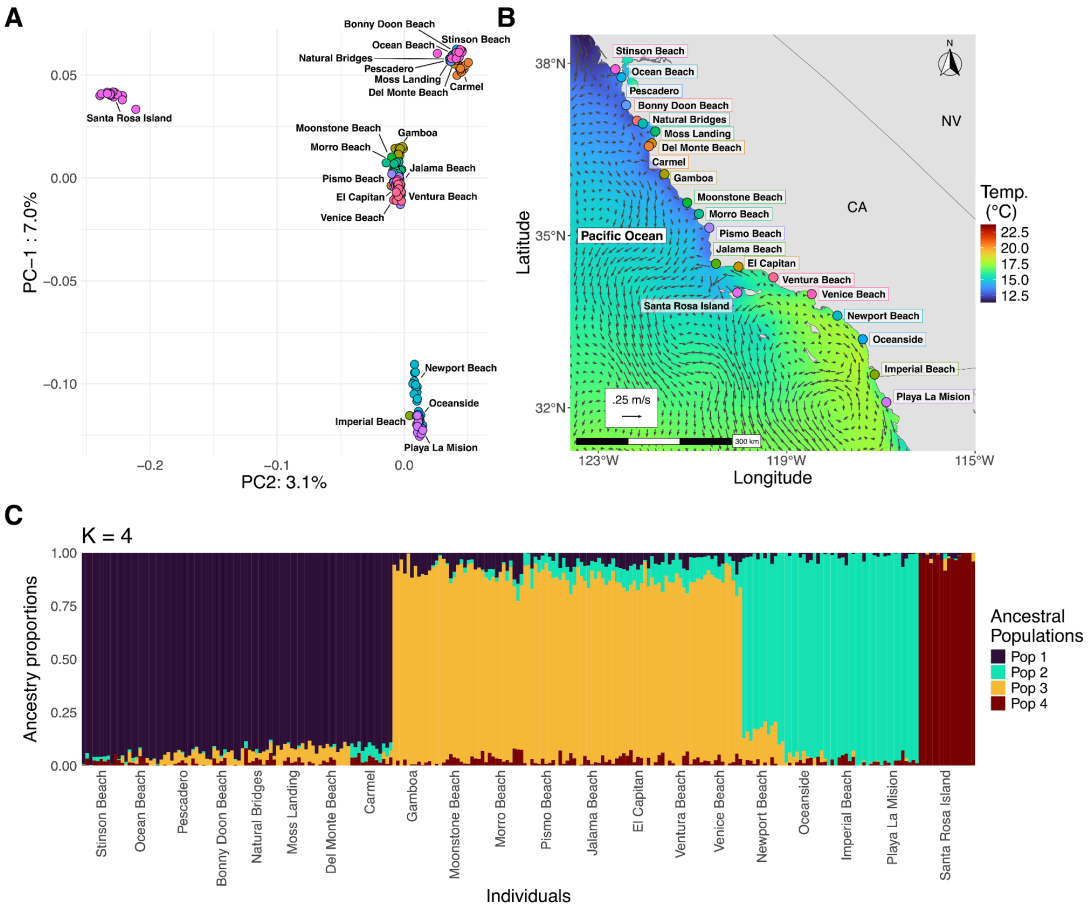

Figure 3.1 - Genetic population structure of the barred surfperch, *Amphistichus argenteus*. Panel A) Principle Components Analysis (PCA) of 253 individuals' whole genomes of *A. argenteus*, colored by sampling location. Panel B) Map of the study region showing sampling locations used in this study and yearly average current and sea surface temperatures. Panel C) Ancestry plot divided into four genetic clusters. Each vertical bar represents an individual, populations are indicated on the x-axis and are placed from north to south along the coastline. Santa Rosa Island individuals are placed on the far right side of the plot.

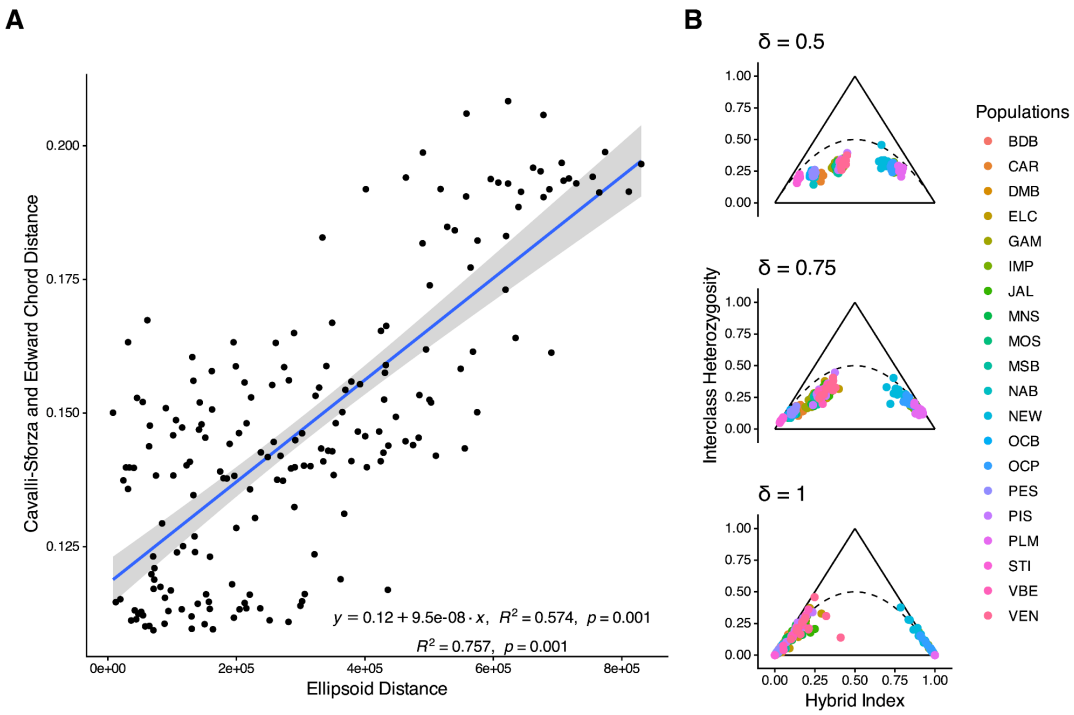

Figure 3.2. - Isolation by Distance in coastal populations of the barred surfperch, *Amphistichus argenteus*. Panel A). Mantel test of the correlation between the geographical distance matrix and the genetic distance matrix between mainland populations. First line at the bottom right shows linear regression fit and associated R^2 and p‑value, second line shows Mantel test R^2 and p‑value. Panel B). Triangle plots using the same populations and colour codes as in Figure 1. Triangle plots with Stinson Beach and Playa la Misión set as parental populations, allele frequency thresholds (δ) of 0.5, 0.75 and 1 (top to bottom) allow the selection of Ancestry‑Informative Markers. Upper part of the triangle over the dotted lines represents the expected position of hybrids and backcrossed individuals under Hardy–Weinberg equilibrium (Wiens and Colella 2024a, 2024b). The mantel test displays statistically significant correlation between the geographical and genetic distance between populations which is characteristic of an isolation by distance (IBD) scenario. Greater separation between North–Central (left) and Southern (right) trajectories at higher δ values suggests a deep coastal divergence driven by highly differentiated and fixed loci.

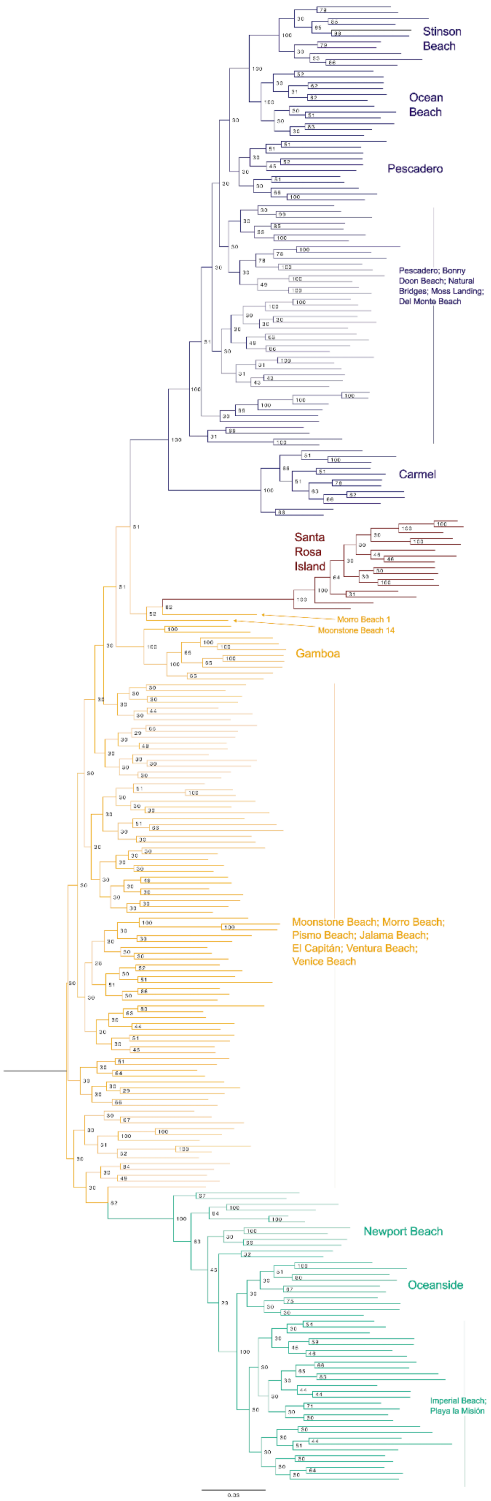

Figure 3.3 - Midpoint rooted maximum‑likelihood (ML) tree constructed using the full SNP set. Clades coloured according to the 4 groups discussed in this paper. Branch labels are UFBoot support values (1000 iterations).

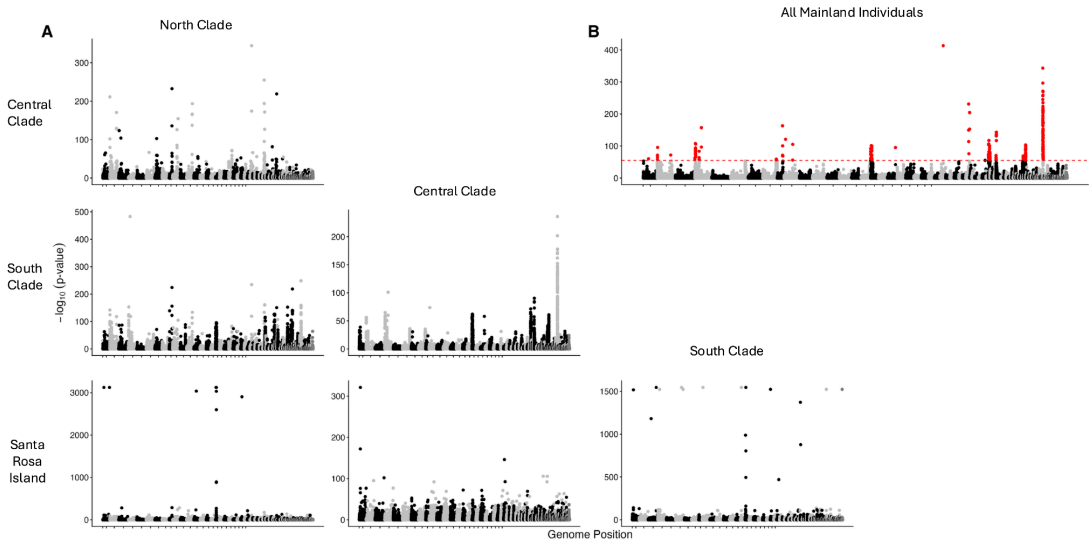

Figure 3.4 - Panel A) Manhattan plots of pcadapt ⁻log10(p values) based on pairwise comparisons of barred surfperch populations. Panel B) Manhattan plot of all individuals. Outlier loci are shown in red.

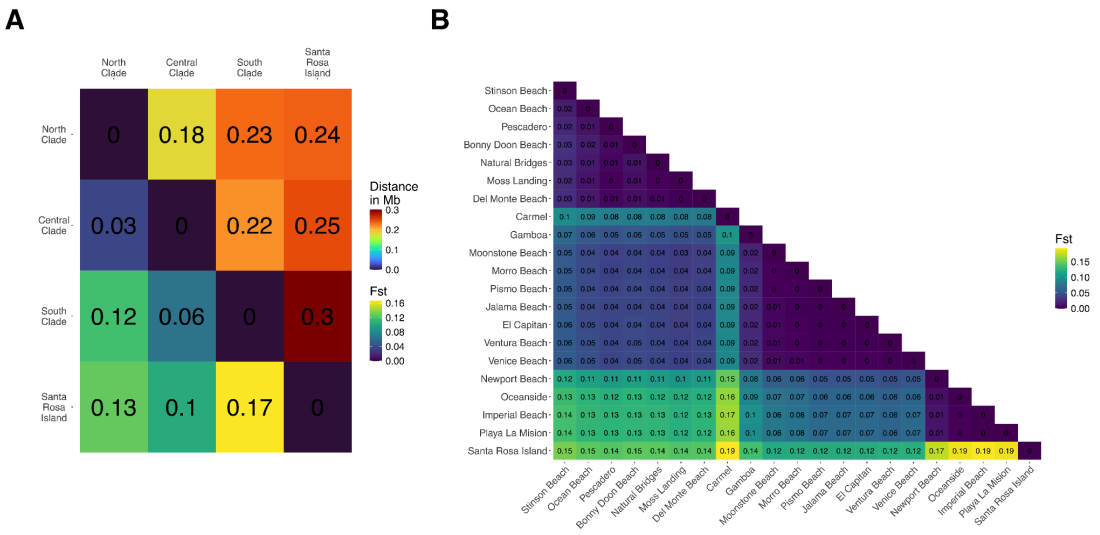

Figure 3.5 - Gene flow among populations of barred surfperch. Panel A) Pairwise F_{ST} and mean pairwise genetic distance among the four major clades. Panel B) Pairwise F_{ST} values among sampling localities.

TABLES

Table 3.1 - Sampling locations and sampling numbers of barred surfperch *(A. argenteus)*.

Sample Code	Locality	Number	Latitude, longitude
STI	Stinson Beach	11	37.89693, -122.64186
OCB	Ocean Beach	12	37.75644, -122.51085
PES	Pescadero	11	37.27028, -122.41141
BDB	Bonny Doon Beach	11	37.00027, -122.18192
NAB	Natural Bridges, Santa Cruz	10	36.95055, -122.05795
MOS	Moss Landing	10	36.81449, -121.79142
DMB	Del Monte Beach, Monterey	11	36.60573, -121.86814
CAR	Carmel	12	36.55460, -121.93103
GAM	Gamboa, Big Sur	11	36.06963, -121.60053
MNS	Moonstone Beach	14	35.57372, -121.11332
MSB	Morro Beach	12	35.37790, -120.86465
PIS	Pismo Beach	13	35.13758, -120.64323
JAL	Jalama Beach	14	34.50959, -120.50182
ELC	El Capitán	12	34.45885, -120.02544
VEN	Ventura Beach	11	34.27198, -119.28497
SRI	Santa Rosa Island	16	34.00760, -120.04822
VBE	Venice Beach	12	33.98036, -118.46986
NEW	Newport Beach	12	33.60553, -117.92509
OCP	Oceanside	12	33.19462, -117.38526
IMP	Imperial Beach	11	32.57960, -117.13296
PLM	Playa la Misión	15	32.09928, -116.88482

Synthesis

Understanding how marine populations diverge and persist in a rapidly changing ocean requires integrating ecological, physiological, and evolutionary perspectives. Patterns of population structure observed today reflect not only contemporary dispersal and environmental conditions, but also the cumulative effects of historical isolation, habitat continuity, and selection acting over evolutionary time. I designed this dissertation as a comparative investigation of how population structure is generated and maintained across marine fish systems that differ in biogeographic context, life history, and environmental regime. By combining physiological experiments with whole-genome population genomic analyses, this work provides new insight into how historical processes, environmental heterogeneity, and selection interact to shape population divergence in marine fishes.

In Chapter 1, I demonstrated that disjunct Pacific and Sea of Cortez lineages of the longjaw mudsucker (*Gillichthys mirabilis*) exhibit subtle but biologically meaningful differences in thermal physiology. Broad thermal performance curves were largely similar between populations, though Pacific individuals reached a higher maximum metabolic rate, while Sea of Cortez individuals showed lower thermal sensitivity and higher acute thermal tolerance following warm pre-conditioning. This contrast, higher peak performance in Pacific fish versus greater thermal tolerance and plasticity in Sea of Cortez fish, mirrors the demands of each population's thermal environment, though a two-lineage comparison cannot establish adaptation. These patterns align with environmental temperature data showing that Sea of Cortez

habitats experience both higher maximum temperatures and faster short-term thermal fluctuations. Together, these results indicate that population-level physiological differentiation can be revealed under ecologically relevant thermal contexts and suggest that both plasticity and local adaptation may contribute to shaping thermal tolerance in this species.

In Chapter 2, I used low-coverage whole-genome resequencing to examine genomic divergence between Pacific and Sea of Cortez populations of *G. mirabilis*. I found exceptionally high levels of genomic differentiation, extensive fixed differences, and reciprocal phylogenetic monophyly, patterns consistent with long-term isolation. In addition to genome-wide divergence, I identified numerous candidate loci under selection, including genes involved in protein homeostasis and the endoplasmic reticulum stress response, as well as ion and solute transporters implicated in osmoregulation, consistent with selection acting on genes tied to the thermal and osmotic contrasts between regions. These results indicate that divergence between Pacific and Sea of Cortez populations reflects not only genetic drift under prolonged isolation, but potentially adaptive differentiation associated with contrasting environmental conditions. Collectively, the physiological and genomic results support the interpretation that these populations represent deeply diverged evolutionary lineages that may warrant recognition as distinct species.

In Chapter 3, I extended this population genomic framework to the barred surfperch (*Amphistichus argenteus*), a nearshore species lacking a pelagic larval stage. In contrast to the mudsucker, barred surfperch exhibited modest genome-wide

differentiation among most coastal localities, yet resolved into four genetically distinct groups, three coastal clades plus a highly divergent Santa Rosa Island lineage, with strong divergence concentrated at a subset of loci. Patterns of structure were closely associated with habitat continuity, with low differentiation along continuous sandy coastline and pronounced breaks at rocky headlands and the deep water separating the mainland from offshore islands. Although tagging data indicate that individuals can move across some of these boundaries, genomic patterns suggest that successful gene flow is limited, highlighting the importance of post-dispersal processes in maintaining structure. At the sharpest coastal break, across the Palos Verdes Peninsula, genome scans identified candidate genes under selection related to spermatogenesis, sperm motility, and fertilization, a signal that parallels findings in a confamilial surfperch along the same coastline and, together with a documented difference in breeding phenology across this break, suggests that selection on reproductive compatibility is actively maintaining, and may be deepening, this divergence.

Taken together, these chapters demonstrate that contemporary population structure in marine fishes is best understood as a mosaic shaped by interacting historical, ecological, and evolutionary processes. Deep historical isolation can produce genome-wide divergence approaching species-level differentiation, as observed in the longjaw mudsucker, whereas partial connectivity combined with selection can generate heterogeneous genomic divergence and persistent population structure, as observed in barred surfperch. Across both systems, environmental

heterogeneity emerges as a consistent driver of physiological and genomic differentiation, reinforcing the role of local selective pressures in shaping evolutionary trajectories.

This work has important implications for both evolutionary theory and applied management. Results presented here show that simple predictors such as dispersal potential or geographic distance alone are insufficient to explain population structure, as much of the genome may appear connected while key regions linked to environmental performance and physiological tolerance remain strongly differentiated, consistent with heterogeneous genomic divergence and potential adaptive structure. These findings highlight the value of considering population structure, ecological context, and evidence of adaptive differentiation when evaluating population boundaries. Managing genetically and ecologically distinct populations as a single stock risks eroding locally adapted variation critical for long-term persistence, and, as demonstrated in other harvested marine species, the consequences of collapsing discrete evolutionary units into one stock can be more severe than those of managing a largely panmictic population as multiple units (Kerr et al., 2017; Spies & Punt, 2015).

Moreover, the observation that populations differ in physiological tolerance and in genes associated with environmental performance suggests that responses to future climate change will be uneven across the species' range. Populations adapted to extreme or variable environments may exhibit greater resilience to certain stressors, whereas others may be more vulnerable or dependent on local recruitment.

Recognizing and preserving this adaptive heterogeneity will likely be critical for maintaining productivity and evolutionary potential under ongoing environmental change.

Overall, this dissertation advances understanding of how population divergence arises and persists in marine fishes by demonstrating that population structure reflects layered histories and ongoing ecological and evolutionary processes. By integrating physiological and genomic perspectives across contrasting systems, this work provides a more comprehensive framework for interpreting marine population structure and for anticipating how marine species may respond to continued environmental change.

Appendix 1: Supplementary material for Chapter 1

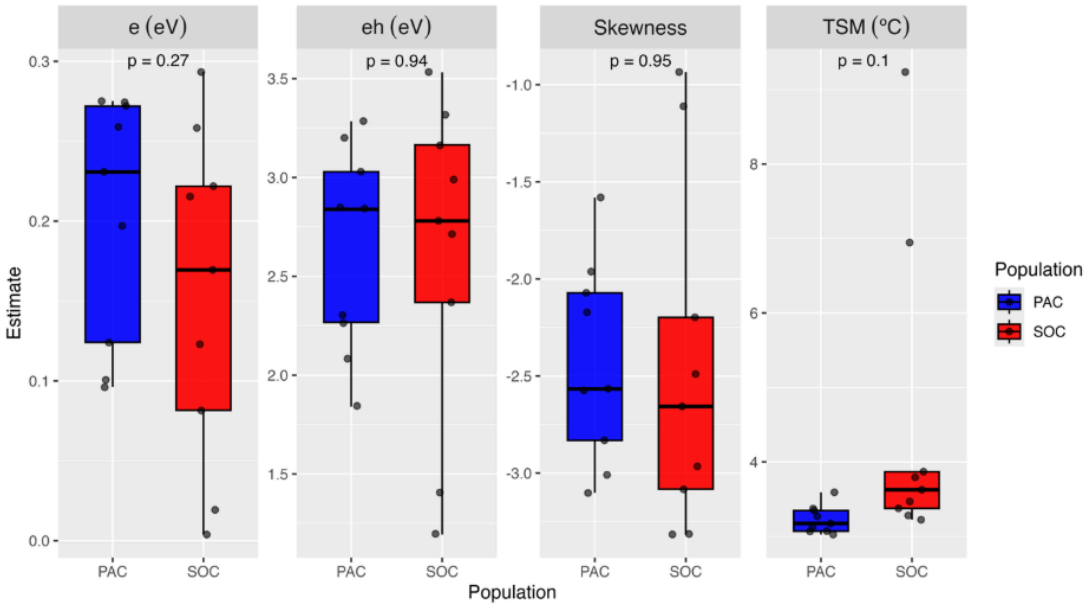

Figure S1.1 - Thermal performance curve (TPC) parameters. Parameter estimates of e, eh, skewness, and thermal safety margin (TSM) derived from the Sharpe–Schoolfield high-temperature model (sharpeschoolhigh_1981) are shown for each lineage following acclimation to common garden ambient temperature (15.2 ± 1.3°C) for > 4 wks. The PAC lineage in blue is shown and the SOC population in red. Welch's two-sample t-tests were used to assess differences in each parameter between lineages; corresponding p-values are reported above each panel. Definitions of all parameters are provided in Table 1.2 and values are reported in Table 1.3.

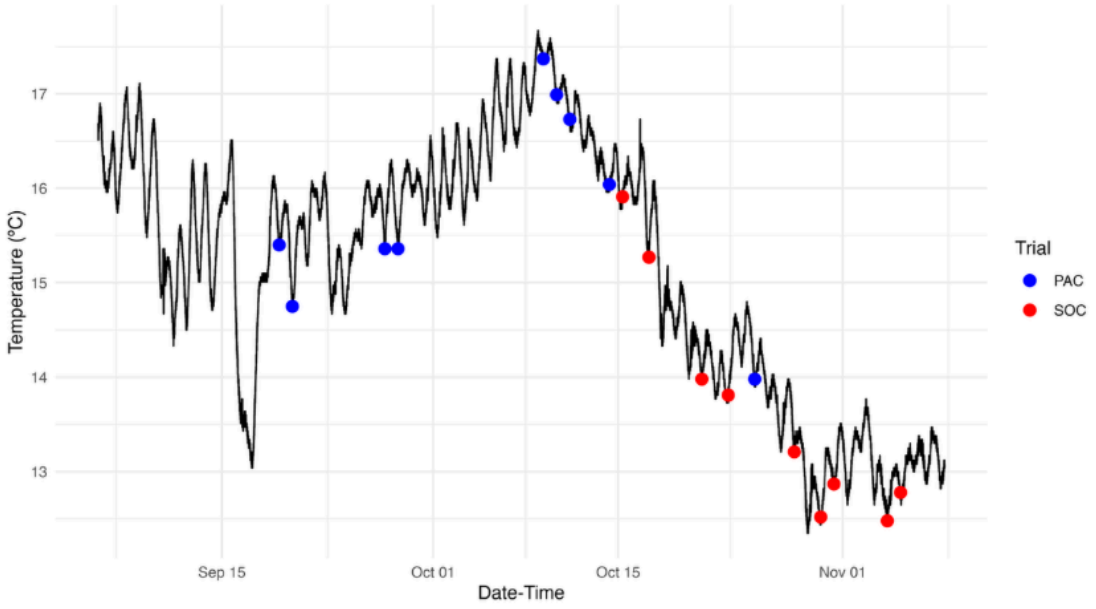

Figure S1.2 - Common garden ambient temperature during experimental acclimation. Ambient temperature in the wet laboratory recorded continuously over time during the common garden acclimation period. The solid line shows ambient tank temperature, while points indicate the dates on which experimental TPC trials were conducted. Trials from the PAC lineage are shown in blue and those from the SOC lineage in red. Temperatures the morning of each trial ranged from 12.48–17.37°C.

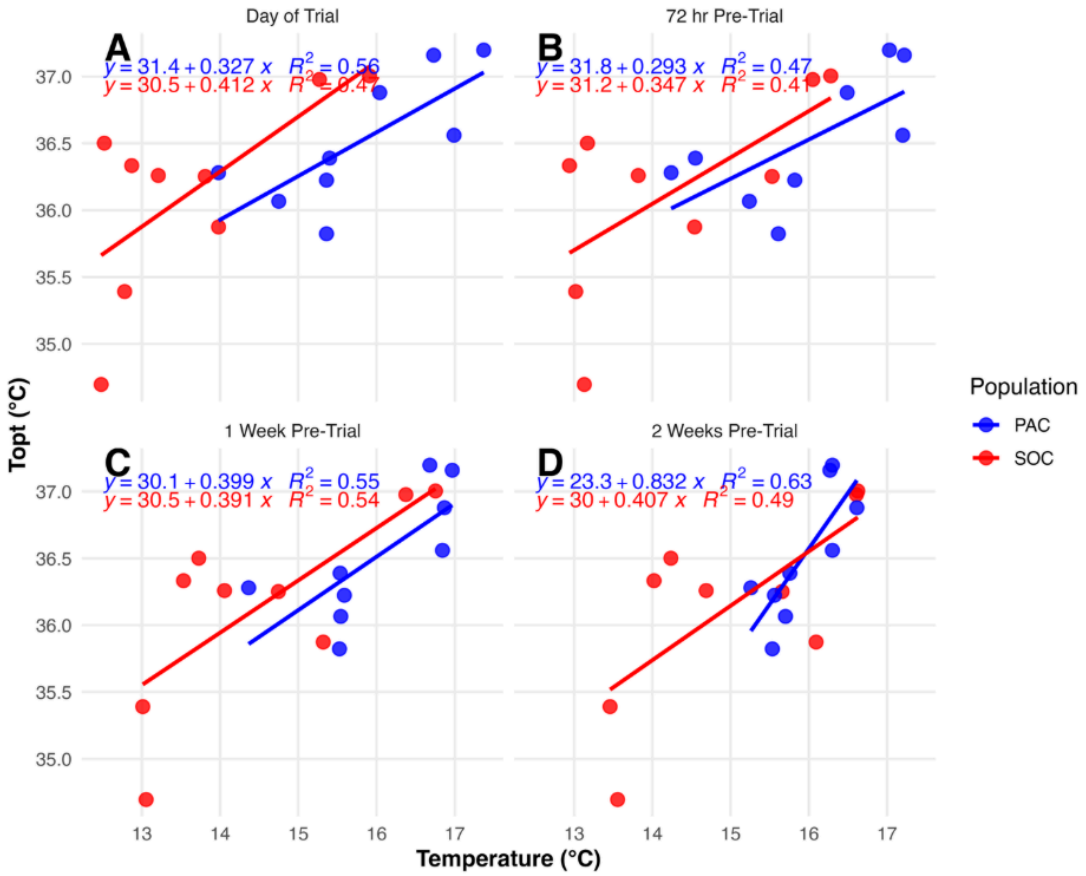

Figure S1.3 - Short-term thermal history and Topt. Relationship between ambient seawater temperature on the day of respirometry trials and estimated thermal optima (T_{opt}) based on individual respirometry data fit to Sharpe–Schoolfield high‑temperature model (Schoolfield et al., 1981). Panels show thermal history at different timescales prior to each trial: A) temperature on the day of respirometry; B) average temperature 72 h prior; C) average temperature 1 week prior; D) average temperature 2 weeks prior. Points represent individual fits, colored by lineage (PAC = blue, SOC = red). Solid lines show lineage-specific linear regressions; equations and R^2 values are shown for each lineage.

Table S1.1 - CT_{max}: Inter-Lineage Comparison. CT_{max} values from Figure 1.6 are reported, including mean, standard deviation, and p-value comparing the two lineages (SOC and PAC) pre-conditioned to ambient common garden temperature (13.5 ± 1.8°C) or 25ºC for 12 hours using two-tailed Welch's t-test.

Acclimation or Pre-Conditioning Temp. (ºC)	**mean (SOC)**	**mean (PAC)**	**sd (SOC)**	**sd (PAC)**	***p***
Ambient	36.29	36.44	0.67	0.41	0.5987
25	37.44	36.71	0.43	0.21	**0.0015****

Table S1.2 - CT_{max}: Acclimatory Response. CT_{max} values from Figure 1.7 are reported, including mean, standard deviation, and p-value comparing the CT_{max} at each acclimation or pre-conditioning temperature (Common Garden Ambient (13.5 ± 1.8°C) and 25ºC) for each lineage using two-tailed Welch's t-test.

Lineage	**mean (ambient)**	**mean (25ºC)**	**sd (ambient)**	**sd (25ºC)**	***p***
SOC	36.29	37.44	0.67	0.43	**0.0016****
PAC	36.44	36.71	0.41	0.21	0.1297

Appendix 2: Supplementary material for Chapter 2

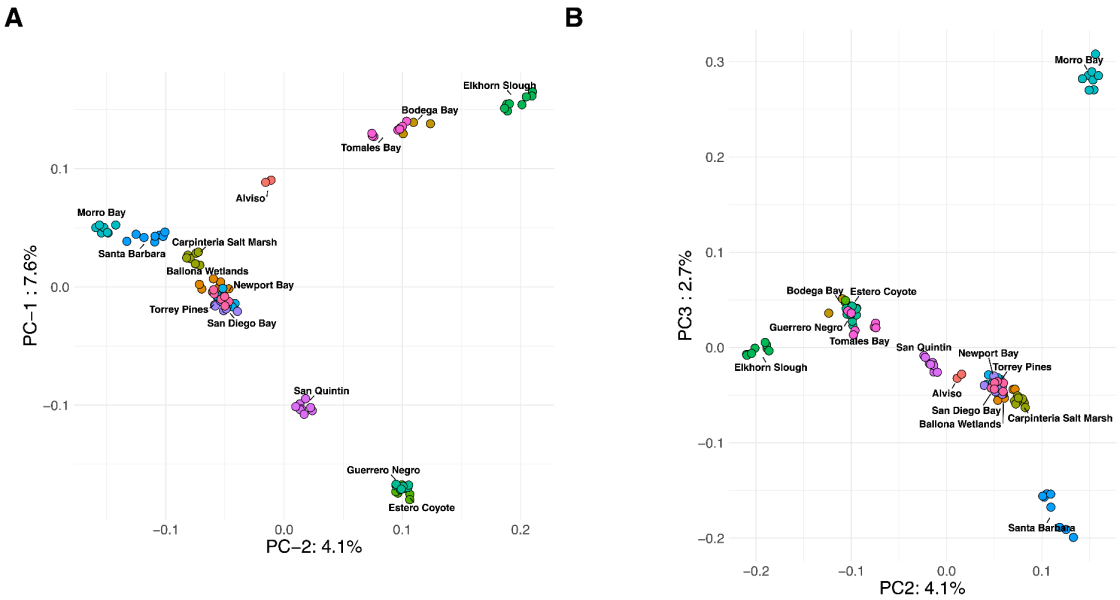

Figure S2.1 - Principal Component Analysis (PCA) of 98 individuals' whole genomes of Longjaw mudsuckers, Gillichthys mirabilis, from Pacific sampling localities. Individuals are coloured by sampling locations showing distribution of samples for the first and second principal components (Panel A) and the second and third principal components (Panel B).

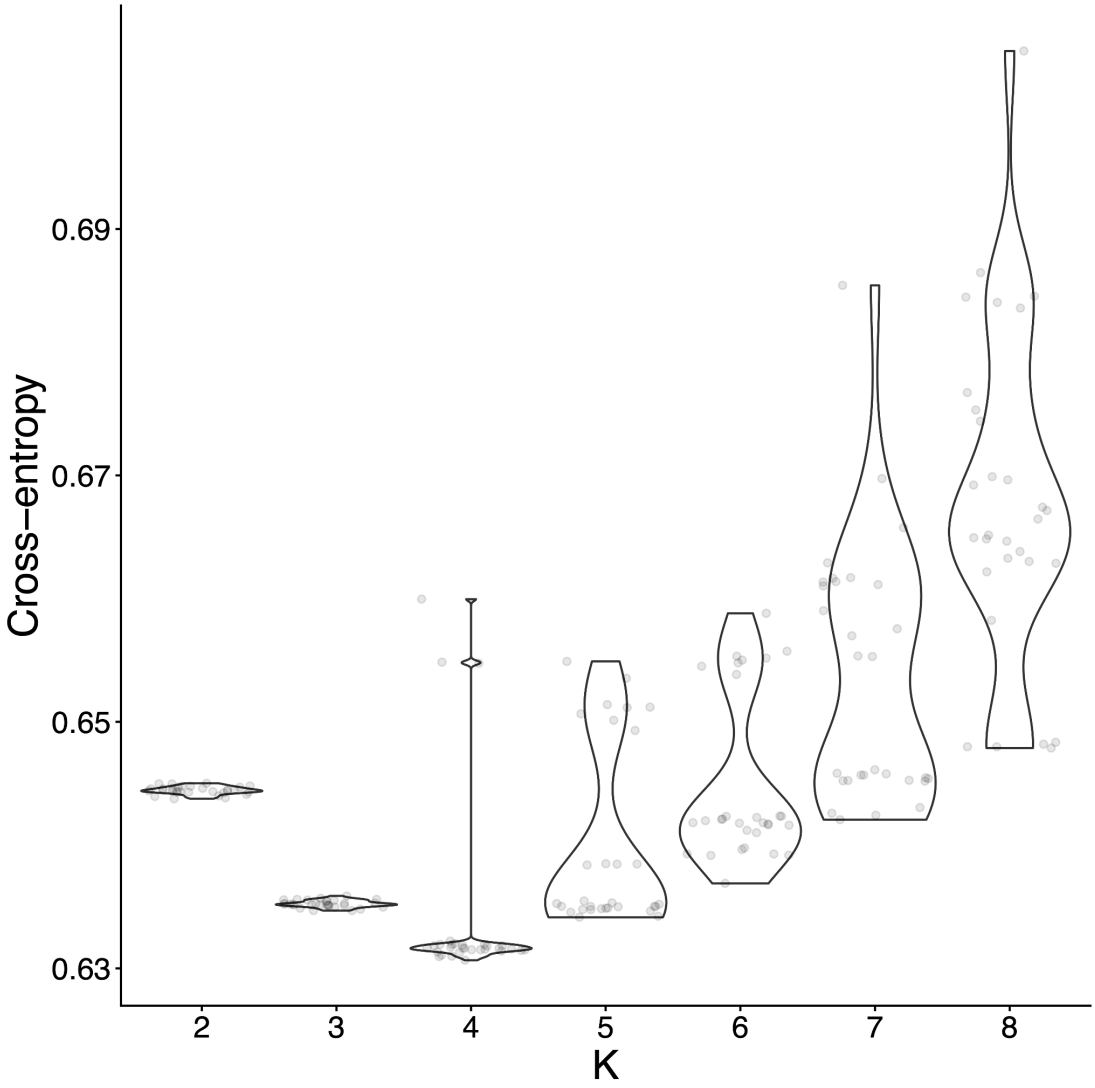

Figure S2.2 - Cross entropy plot, indicating the most likely value of K (K = 4)—i.e the lowest cross-entropy—in ancestry(sNMF) analysis.

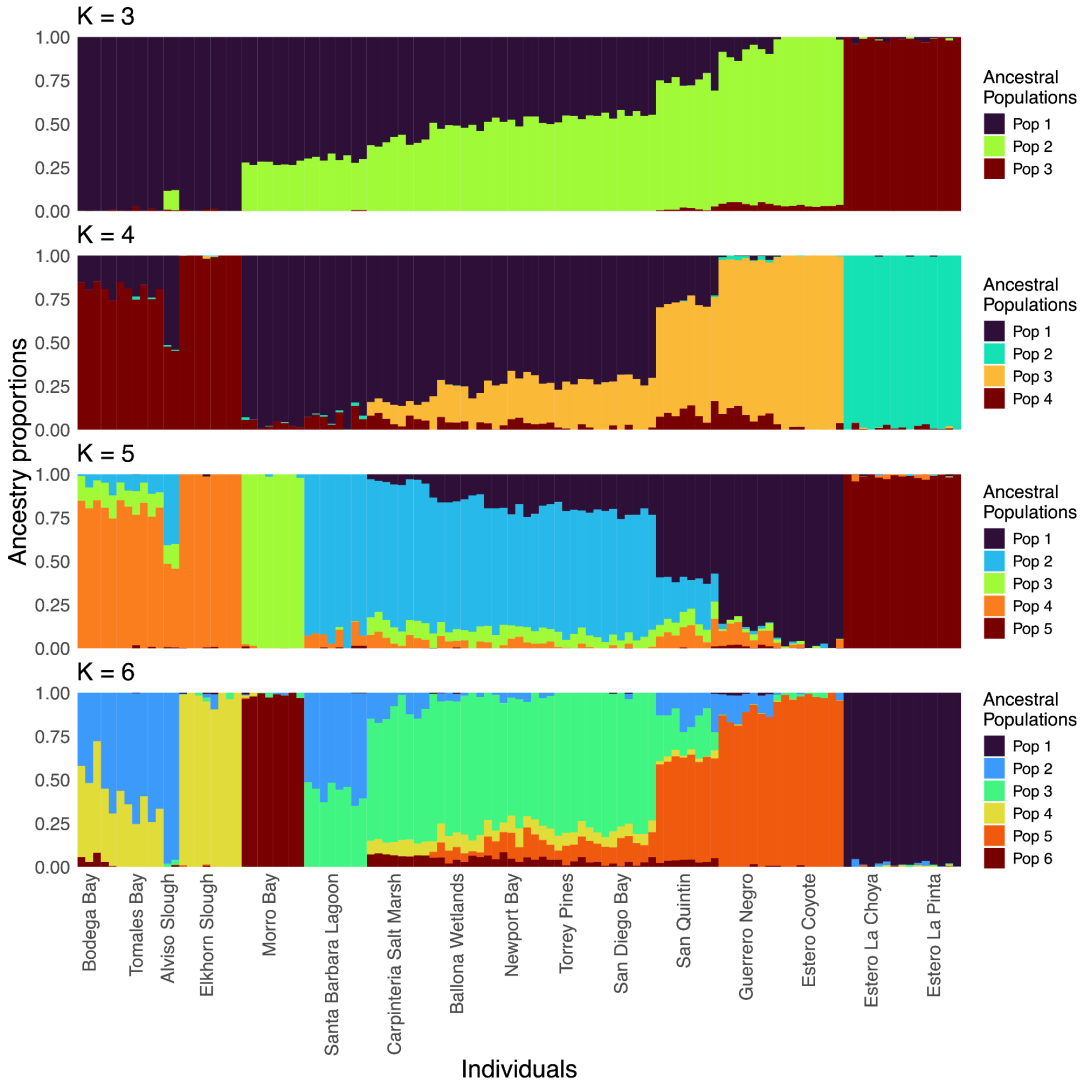

Figure S2.3 - Ancestry (admixture) plots generated using sNMF with alternative values of K (genetic clusters).

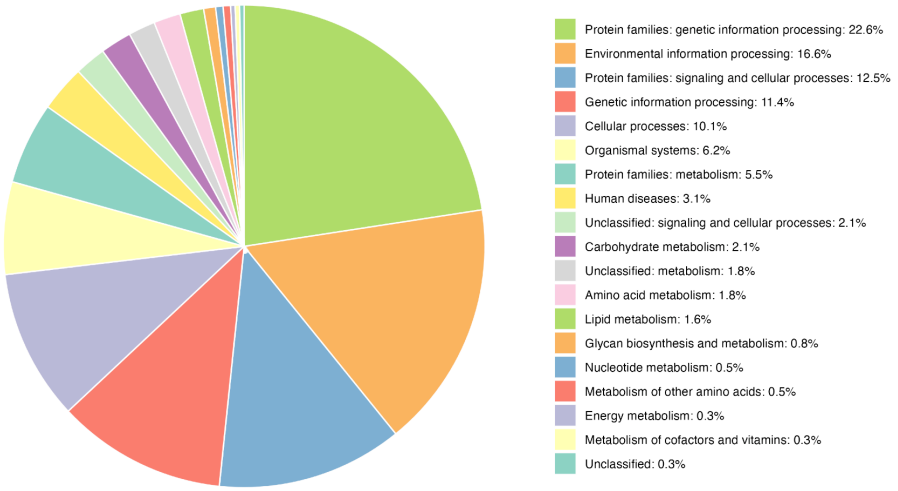

Figure S2.4 - KEGG analysis of outlier protein coding genes found in Longjaw mudsuckers.

Appendix 3: Supplementary material for Chapter 3

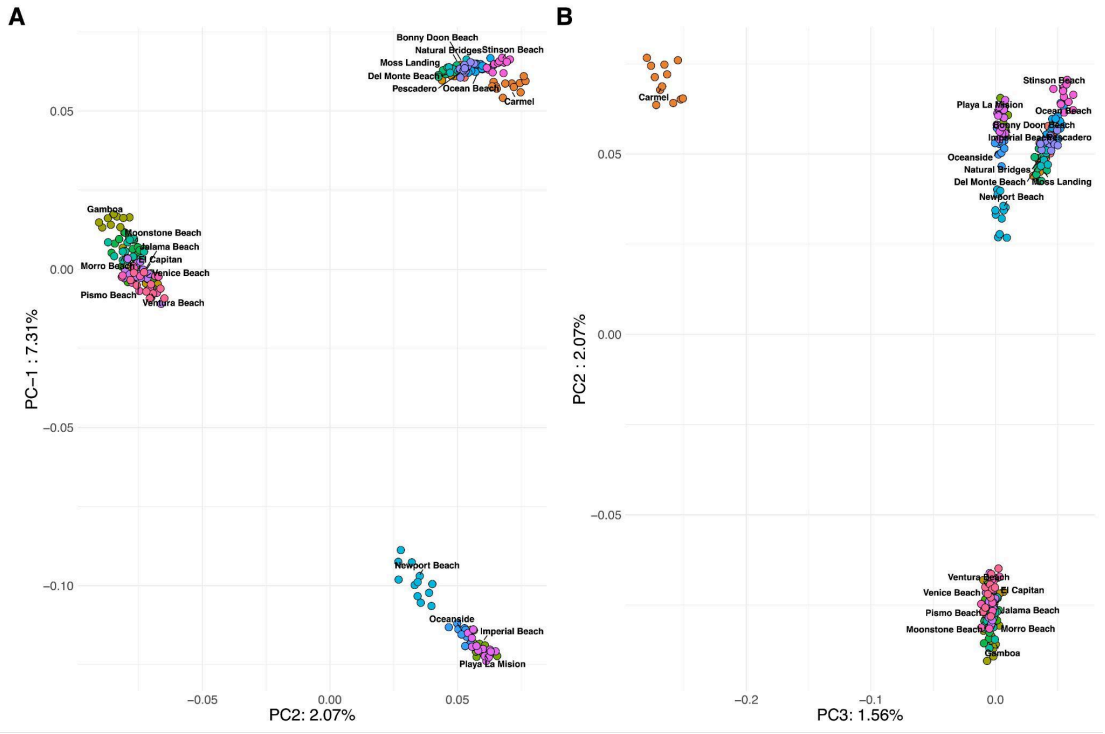

Figure S3.1 - Principal Component Analysis (PCA) of 253 individuals' whole genomes of barred surfperch, *Amphistichus argenteus*, from mainland sampling localities. Individuals are coloured by sampling locations showing distribution of samples for the first and second principal components (Panel A) and the second and third principal components (Panel B).

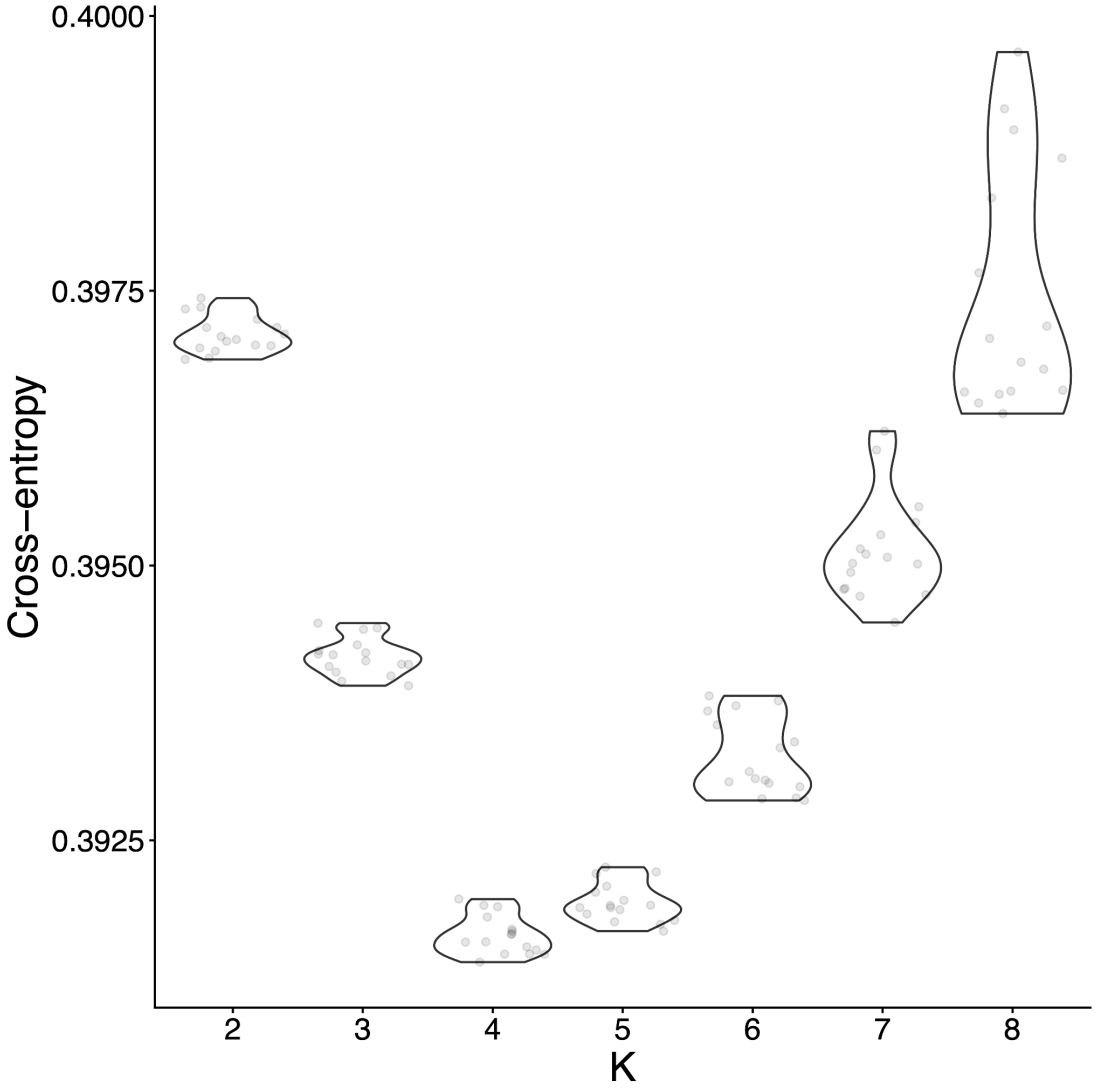

Figure S3.2 - Cross entropy plot, indicating the most likely value of *K* (*K* = 4)—*i.e* the lowest cross-entropy—in ancestry(sNMF) analysis.

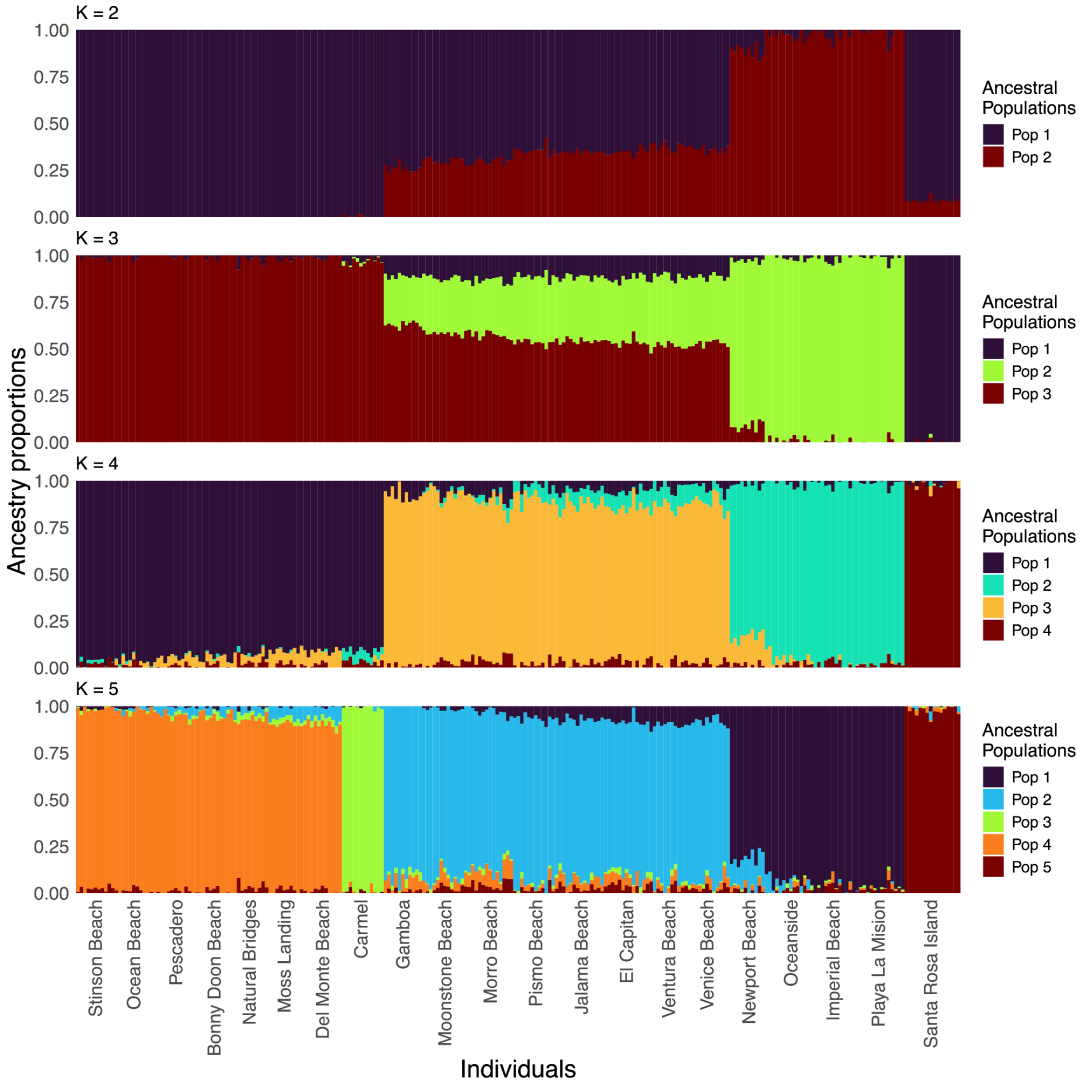

Figure S3.3 - Ancestry (admixture) plots generated using sNMF with alternative values of K (genetic clusters).

Figure S3.4 - A midpoint-rooted population-level tree generated using the ape package, using the matrix of pairwise population F_{ST} values as a distance matrix (Figure 5).

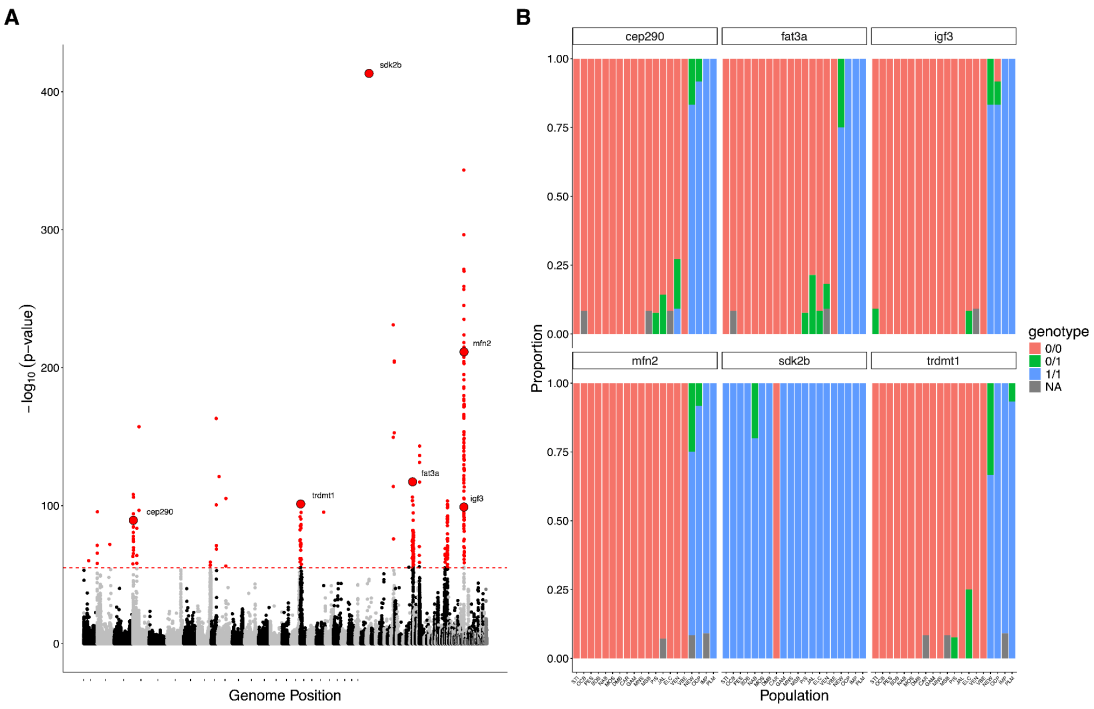

Figure S3.5 - Outlier loci identified by PCADAPT in mainland *Amphistichus argenteus*. Panel A) Manhattan plot of PCADAPT $-\log_{10}$(*p*-values) across the genome for the all-mainland scan, with alternating black and gray shading distinguishing scaffolds. Outlier SNPs above the Bonferroni-corrected significance threshold (dashed red line) are shown in red. Six loci of biological interest are labeled: SDK2B (the strongest signal in the scan), and five additional outliers — CEP290, MFN2, IGF3, TRDMT1, and FAT3A. Panel B) Per-population genotype proportions at each of the six labeled loci. Sampling localities are ordered from north (left) to south (right) on the x-axis. Bars show the proportion of individuals carrying the reference homozygote (0/0, red), heterozygote (0/1, green), alternate homozygote (1/1, blue), or missing data (NA).

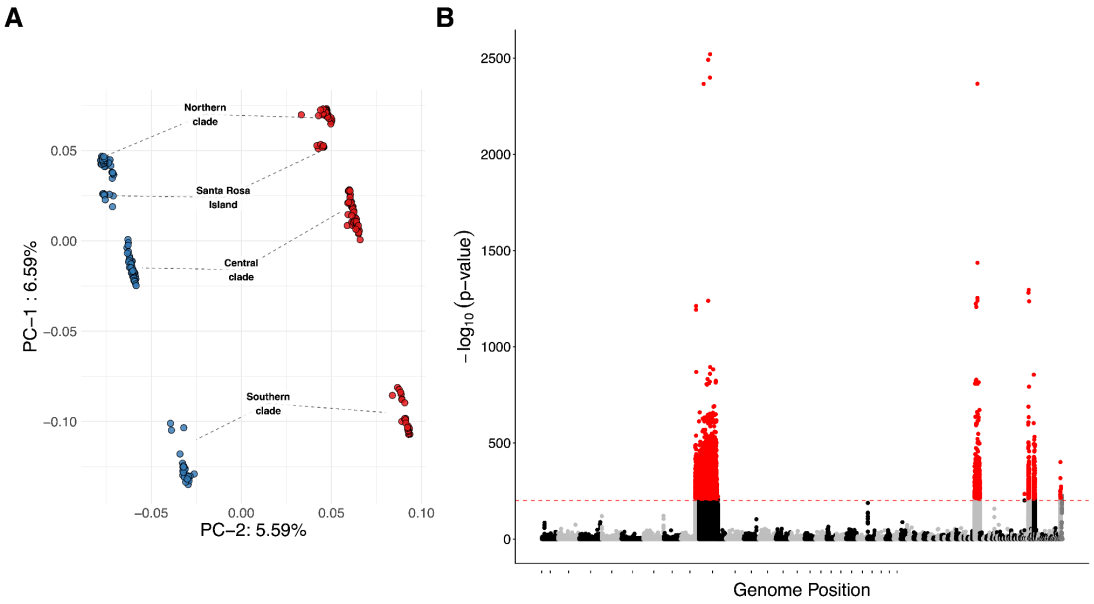

Figure S3.6 - Identification of sex-associated genomic regions in *Amphistichus argenteus* prior to population genomic analyses. Panel A) Principal Component Analysis (PCA) of 253 individuals based on the full SNP dataset. PC1 (6.59%) recovers the four geographic groups, while PC2 (5.59%) separates males (blue) from females (red) within each clade. Panel B) Manhattan plot of PCADAPT $-\log_{10}(p\text{-values})$ from the same K = 3 scan, with alternating black and gray shading distinguishing scaffolds. SNPs above the Bonferroni-corrected significance threshold (dashed red line) are shown in red. Five contigs (9.1, 10.1, 45.1, 75.1, 82.1) carry large regions containing SNPs with extreme $-\log_{10}(p\text{-values})$, consistent with sex linkage, and were masked from downstream analyses.

Literature Cited

Albalat, R., & Cañestro, C. (2016). Evolution by gene loss. Nature Reviews Genetics, 17(7), 379–391. https://doi.org/10.1038/nrg.2016.39

Allen, L. (2006). The Ecology of Marine Fishes California and Adjacent Waters. University of California Press. https://doi.org/10.1525/california/9780520246539.001.0001

Alonge, M., Soyk, S., Ramakrishnan, S., Wang, X., Goodwin, S., Sedlazeck, F. J., Lippman, Z. B., & Schatz, M. C. (2019). RaGOO: Fast and accurate reference-guided scaffolding of draft genomes. Genome Biology, 20(1), 224. https://doi.org/10.1186/s13059-019-1829-6

Anastasiadi, D., Díaz, N., & Piferrer, F. (2017). Small ocean temperature increases elicit stage-dependent changes in DNA methylation and gene expression in a fish, the European sea bass. Scientific Reports, 7(1), 12401. https://doi.org/10.1038/s41598-017-10861-6

Andrews, S. (2010). FastQC: a quality control tool for high throughput sequence data. Available online at: http://www.bioinformatics.babraham.ac.uk/projects/fastqc

Angilletta, M. J. (2009). Thermal Adaptation: A Theoretical and Empirical Synthesis. OUP Oxford. https://books.google.com/books?id=5RRREAAAQBAJ

Baltz, D. M. (1984). Life history variation among female surfperches (Perciformes: Embiotocidae). Environmental Biology of Fishes, 10(3), 159–171. https://doi.org/10.1007/BF00001123

Barlow, G. (1963). Species structure of the gobiid fish Gillichthys mirabilis from coastal sloughs of the eastern Pacific. 17(1): 47-72.

Barlow, G. W. (1961). Intra- and interspecific differences in rate of oxygen consumption in Gobiid fishes of the genus Gillichthys. The Biological Bulletin. (world). https://doi.org/10.2307/1539427

Beldade, R., Pedro, T., & Gonçalves, E. J. (2007). Pelagic larval duration of 10 temperate

cryptobenthic fishes. Journal of Fish Biology, 71(2), 376–382. https://doi.org/10.1111/j.1095-8649.2007.01491.x

Berg, P. R., Jentoft, S., Star, B., Ring, K. H., Knutsen, H., Lien, S., Jakobsen, K. S., & André, C. (2015). Adaptation to Low Salinity Promotes Genomic Divergence in Atlantic Cod (Gadus morhua L.). Genome Biology and Evolution, 7(6), 1644–1663. https://doi.org/10.1093/gbe/evv093

Bernardi, G. (2000). Barriers to Gene Flow in Embiotoca Jacksoni, a Marine Fish Lacking a Pelagic Larval Stage. Evolution, 54(1), 226–237. https://doi.org/10.1111/j.0014-3820.2000.tb00023.x

Bernardi, G. (2005). PHYLOGEOGRAPHY AND DEMOGRAPHY OF SYMPATRIC SISTER SURFPERCH SPECIES, EMBIOTOCA JACKSONI AND E. LATERALIS ALONG THE CALIFORNIA COAST: HISTORICAL VERSUS ECOLOGICAL FACTORS. Evolution, 59(2), 386–394. https://doi.org/10.1111/j.0014-3820.2005.tb00997.x

Bernardi, G. (2013). Speciation in fishes. Molecular Ecology, 22(22), 5487–5502. https://doi.org/10.1111/mec.12494

Bernardi, G., Azzurro, E., Golani, D., & Miller, M. R. (2016). Genomic signatures of rapid adaptive evolution in the bluespotted cornetfish, a Mediterranean Lessepsian invader. Molecular Ecology, 25(14), 3384–3396. https://doi.org/10.1111/mec.13682

Bernardi, G., Findley, L., & Rocha-Olivares, A. (2003). VICARIANCE AND DISPERSAL ACROSS BAJA CALIFORNIA IN DISJUNCT MARINE FISH POPULATIONS. Evolution, 57(7), 1599–1609. https://doi.org/10.1111/j.0014-3820.2003.tb00367.x

Bernardi, G., Toy, J. A., Escalona, M., Marimuthu, M. P. A., Sahasrabudhe, R., Nguyen, O., Sacco, S., Beraut, E., Toffelmier, E., Miller, C., & Shaffer, H. B. (2022). Reference Genome of the Black Surfperch, Embiotoca jacksoni (Embiotocidae, Perciformes), a California Kelp Forest Fish That Lacks a Pelagic Larval Stage. Journal of Heredity, 113(6), 657–664. https://doi.org/10.1093/jhered/esac034

Beron-Vera, F. J., & Ripa, P. (2002). Seasonal salinity balance in the Gulf of California. Journal of Geophysical Research: Oceans, 107(C8), 15-1-15–15. https://doi.org/10.1029/2000JC000769

Bond, E. R., & Forsgren, K. L. (2022). External reproductive structures of male Surfperches (Embiotocidae): Urogenital papilla. The Anatomical Record, 305(7), 1712–1723. https://doi.org/10.1002/ar.24836

Bowen, B. W., Forsman, Z. H., Whitney, J. L., Faucci, A., Hoban, M., Canfield, S. J., Johnston, E. C., Coleman, R. R., Copus, J. M., Vicente, J., & Toonen, R. J. (2020). Species Radiations in the Sea: What the Flock? Journal of Heredity, 111(1), 70–83. https://doi.org/10.1093/jhered/esz075

Bradbury, I. R., Hubert, S., Higgins, B., Borza, T., Bowman, S., Paterson, I. G., Snelgrove, P. V. R., Morris, C. J., Gregory, R. S., Hardie, D. C., Hutchings, J. A., Ruzzante, D. E., Taggart, C. T., & Bentzen, P. (2010). Parallel adaptive evolution of Atlantic cod on both sides of the Atlantic Ocean in response to temperature. Proceedings of the Royal Society B: Biological Sciences, 277(1701), 3725–3734. https://doi.org/10.1098/rspb.2010.0985

Brennan, R. S., Galvez, F., & Whitehead, A. (2015). Reciprocal osmotic challenges reveal mechanisms of divergence in phenotypic plasticity in the killifish Fundulus heteroclitus. Journal of Experimental Biology, 218(8), 1212–1222. https://doi.org/10.1242/jeb.110445

Bugg, W. S., Yoon, G. R., Schoen, A. N., Laluk, A., Brandt, C., Anderson, W. G., & Jeffries, K. M. (2020). Effects of acclimation temperature on the thermal physiology in two geographically distinct populations of lake sturgeon (Acipenser fulvescens). Conservation Physiology, 8(1), coaa087. https://doi.org/10.1093/conphys/coaa087

Burton, R. S. (1998). INTRASPECIFIC PHYLOGEOGRAPHY ACROSS THE POINT CONCEPTION BIOGEOGRAPHIC BOUNDARY. Evolution, 52(3), 734–745.

https://doi.org/10.1111/j.1558-5646.1998.tb03698.x

Butlin, R. K., & Faria, R. (2024). Local adaptation and reproductive isolation: When does speciation start? Evolutionary Journal of the Linnean Society, 3(1), kzae003. https://doi.org/10.1093/evolinnean/kzae003

Carbonneau, A., April, J., Normandeau, E., Ferchaud, A., Nadeau, V., & Bernatchez, L. (2024). Population genomics, life‐history tactics, and mixed‐stock subsistence fisheries in the northernmost American Atlantic salmon populations. Evolutionary Applications, 17(2), e13654. https://doi.org/10.1111/eva.13654

Carlisle, J. G., Schott, J. W., & Abramson, N. J. (1960). Fish Bulletin No. 109. The Barred Surfperch (Amphistichus argenteus Agassiz) in Southern California. https://escholarship.org/uc/item/9fh0623k

Chang, C. C., Chow, C. C., Tellier, L. C., Vattikuti, S., Purcell, S. M., & Lee, J. J. (2015). Second-generation PLINK: Rising to the challenge of larger and richer datasets. Gigascience, 4(1), s13742-015-0047–0048. https://doi.org/10.1186/s13742-015-0047-8

Chen, S., Zhou, Y., Chen, Y., & Gu, J. (2018). fastp: An ultra-fast all-in-one FASTQ preprocessor. Bioinformatics, 34(17), i884–i890. https://doi.org/10.1093/bioinformatics/bty560

Cock, P. J. A., Fields, C. J., Goto, N., Heuer, M. L., & Rice, P. M. (2010). The Sanger FASTQ file format for sequences with quality scores, and the Solexa/Illumina FASTQ variants. Nucleic Acids Research, 38(6), 1767–1771. https://doi.org/10.1093/nar/gkp1137

Cowen, R. K., & Sponaugle, S. (2009). Larval Dispersal and Marine Population Connectivity. Annual Review of Marine Science, 1(1), 443–466. https://doi.org/10.1146/annurev.marine.010908.163757

Crawford, D. L., Schulte, P. M., Whitehead, A., & Oleksiak, M. F. (2020). Evolutionary

physiology and genomics in the highly adaptable killifish (Fundulus heteroclitus). Comprehensive Physiology, 10(2), 637-671.

Currie, R. J., Bennett, W. A., & Beitinger, T. L. (1998). Critical thermal minima and maxima of three freshwater game-fish species acclimated to constant temperatures. Environmental Biology of Fishes, 51(2), 187–200. https://doi.org/10.1023/A:1007447417546

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

Dawson, M. N. (2001). Phylogeography in coastal marine animals: A solution from California? Journal of Biogeography, 28(6), 723–736. https://doi.org/10.1046/j.1365-2699.2001.00572.x

Dawson, M. N., Louie, K. D., Barlow, M., Jacobs, D. K., & Swift, C. C. (2002). Comparative phylogeography of sympatric sister species, Clevelandia ios and Eucyclogobius newberryi (Teleostei, Gobiidae), across the California Transition Zone. Molecular Ecology, 11(6), 1065–1075. https://doi.org/10.1046/j.1365-294X.2002.01503.x

Dawson, M. N., Staton, J. L., & Jacobs, D. K. (2001). PHYLOGEOGRAPHY OF THE TIDEWATER GOBY, EUCYCLOGOBIUS NEWBERRYI (TELEOSTEI, GOBIIDAE), IN COASTAL CALIFORNIA. Evolution, 55(6), 1167–1179. https://doi.org/10.1111/j.0014-3820.2001.tb00636.x

De Coster, W., D'Hert, S., Schultz, D. T., Cruts, M., & Van Broeckhoven, C. (2018). NanoPack: Visualizing and processing long-read sequencing data. Bioinformatics, 34(15), 2666–2669. https://doi.org/10.1093/bioinformatics/bty149

DeLiberto, A. N., Drown, M. K., Ehrlich, M. A., Oleksiak, M. F., & Crawford, D. L. (2022). Feeling the heat: Variation in thermal sensitivity within and among populations.

Journal of Experimental Biology, 225(21), jeb244831. https://doi.org/10.1242/jeb.244831

DeMartini, E. E. (1969). A correlative study of the ecology and comparative feeding mechanism morphology of the Embiotocidae (surf-fishes) as evidence of the family's adaptive radiation into available ecological niches. Wasmann J. Biol, 27(2), 177–247.

Dietz, T. J. (1994). Acclimation of the threshold induction temperatures for 70-kDa and 90-kDa heat shock proteins in the fish Gillichthys mirabilis. Journal of Experimental Biology, 188(1), 333–338. https://doi.org/10.1242/jeb.188.1.333

Dietz, T. J., & Somero, G. N. (1992). The threshold induction temperature of the 90-kDa heat shock protein is subject to acclimatization in eurythermal goby fishes (genus Gillichthys). Proceedings of the National Academy of Sciences, 89(8), 3389–3393. https://doi.org/10.1073/pnas.89.8.3389

European Union-Copernicus Marine Service. (2015). Global Ocean OSTIA Sea Surface Temperature and Sea Ice Analysis [Dataset]. Mercator Ocean International. https://doi.org/10.48670/MOI-00165

European Union-Copernicus Marine Service. (2016). Global Ocean 1/12° Physics Analysis and Forecast updated Daily [Dataset]. Mercator Ocean International. https://doi.org/10.48670/MOI-00016

Faria, R., Johannesson, K., & Stankowski, S. (2021). Speciation in marine environments: Diving under the surface. Journal of Evolutionary Biology, 34(1), 4–15. https://doi.org/10.1111/jeb.13756

Feder, J. L., Egan, S. P., & Nosil, P. (2012). The genomics of speciation-with-gene-flow. Trends in Genetics, 28(7), 342–350. https://doi.org/10.1016/j.tig.2012.03.009

Foster, M. S., & Schiel, D. R. (1985). The Ecology of Giant Kelp Forests in California: A Community Profile. U.S. Fish and Wildlife Service.

Fox, J., & Weisberg, S. (2019). An R Companion to Applied Regression (Third Edition).

Frichot, E., & François, O. (2015). LEA: An R package for landscape and ecological association studies. Methods in Ecology and Evolution, 6(8), 925–929. https://doi.org/10.1111/2041-210X.12382

Frichot, E., Mathieu, F., Trouillon, T., Bouchard, G., & François, O. (2014). Fast and Efficient Estimation of Individual Ancestry Coefficients. Genetics, 196(4), 973–983. https://doi.org/10.1534/genetics.113.160572

Froeschke, B., Allen, L. G., & Pondella, D. J. (2007). Life History and Courtship Behavior of Black Perch, Embiotoca jacksoni (Teleostomi: Embiotocidae), from Southern California. Pacific Science, 61(4), 521–531. https://doi.org/10.2984/1534-6188(2007)61[521:LHACBO]2.0.CO;2

Frölicher, T. L., & Laufkötter, C. (2018). Emerging risks from marine heat waves. Nature Communications, 9(1), 650. https://doi.org/10.1038/s41467-018-03163-6

Futuyma, D. J., & Kirkpatrick, M. (2017). Evolution (Fourth edition). Sinauer Associates, Inc., Publishers.

Gandra, M., Assis, J., Martins, M. R., & Abecasis, D. (2021). Reduced Global Genetic Differentiation of Exploited Marine Fish Species. Molecular Biology and Evolution, 38(4), 1402–1412. https://doi.org/10.1093/molbev/msaa299

Garcia, E., Simison, W. B., & Bernardi, G. (2020). Patterns of Genomic Divergence and Signals of Selection in Sympatric and Allopatric Northeastern Pacific and Sea of Cortez Populations of the Sargo (Anisotremus davidsonii) and Longjaw Mudsucker (Gillichthys mirabilis). Journal of Heredity, 111(1), 57–69. https://doi.org/10.1093/jhered/esz071

Garland Jr, T., & Adolph, S. C. (1991). Physiological differentiation of vertebrate populations. Annual Review of Ecology and Systematics, 193-228.

Garland Jr, T., & Adolph, S. C. (1994). Why not to do two-species comparative studies: limitations on inferring adaptation. Physiological Zoology, 67(4), 797-828.

Green, A. L., Maypa, A. P., Almany, G. R., Rhodes, K. L., Weeks, R., Abesamis, R. A., Gleason, M. G., Mumby, P. J., & White, A. T. (2015). Larval dispersal and movement patterns of coral reef fishes, and implications for marine reserve network design. Biological Reviews, 90(4), 1215–1247. https://doi.org/10.1111/brv.12155

Gruenthal, K. M., Witting, D. A., Ford, T., Neuman, M. J., Williams, J. P., Pondella, D. J., Bird, A., Caruso, N., Hyde, J. R., Seeb, L. W., & Larson, W. A. (2014). Development and application of genomic tools to the restoration of green abalone in southern California. Conservation Genetics, 15(1), 109–121. https://doi.org/10.1007/s10592-013-0524-5

Hauser, L., Adcock, G. J., Smith, P. J., Bernal Ramírez, J. H., & Carvalho, G. R. (2002). Loss of microsatellite diversity and low effective population size in an overexploited population of New Zealand snapper (Pagrus auratus). Proceedings of the National Academy of Sciences, 99(18), 11742–11747. https://doi.org/10.1073/pnas.172242899

Hauser, L., & Carvalho, G. R. (2008). Paradigm shifts in marine fisheries genetics: Ugly hypotheses slain by beautiful facts. Fish and Fisheries, 9(4), 333–362. https://doi.org/10.1111/j.1467-2979.2008.00299.x

Hedgecock, D., Barber, P. H., & Edmands, S. (2007). Genetic Approaches to Measuring Connectivity. Oceanography, 20(3), 70–79.

Hedgecock, D., & Pudovkin, A. I. (2011). Sweepstakes Reproductive Success in Highly Fecund Marine Fish and Shellfish: A Review and Commentary. Bulletin of Marine Science, 87(4), 971–1002. https://doi.org/10.5343/bms.2010.1051

Hellberg, M. E., Burton, R. S., Neigel, J. E., & Palumbi, S. R. (2002). Genetic assessment of connectivity among marine populations. Bulletin of Marine Science, 70(1), 273–290.

Hijmans, R. J. (2010). geosphere: Spherical Trigonometry (p. 1.5-20) [Dataset]. https://doi.org/10.32614/CRAN.package.geosphere

Hoang, D. T., Chernomor, O., Von Haeseler, A., Minh, B. Q., & Vinh, L. S. (2018). UFBoot2: Improving the Ultrafast Bootstrap Approximation. Molecular Biology and Evolution, 35(2), 518–522. https://doi.org/10.1093/molbev/msx281

Holsinger, K. E., & Weir, B. S. (2009). Genetics in geographically structured populations: Defining, estimating and interpreting FST. Nature Reviews Genetics, 10(9), 639–650. https://doi.org/10.1038/nrg2611

Huang, D., & Bernardi, G. (2001). Disjunct Sea of Cortez–Pacific Ocean Gillichthys mirabilis populations and the evolutionary origin of their Sea of Cortez endemic relative, Gillichthys seta. Marine Biology, 138(2), 421–428. https://doi.org/10.1007/s002270000454

Huang, J., Li, Y., Liu, Z., Kang, Y., & Wang, J. (2018). Transcriptomic responses to heat stress in rainbow trout Oncorhynchus mykiss head kidney. Fish & Shellfish Immunology, 82, 32–40. https://doi.org/10.1016/j.fsi.2018.08.002

Huey, R. B., & Kingsolver, J. G. (1993). Evolution of Resistance to High Temperature in Ectotherms. The American Naturalist, 142, S21–S46. https://doi.org/10.1086/285521

Izumiyama, M., Westphal, M. F., & Crow, K. D. (2020). In the surf zone: Reproductive strategy of the calico surfperch (Amphistichus koelzi) in a comparative context. Journal of Fish Biology, 96(4), 939–949. https://doi.org/10.1111/jfb.14283

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

Jombart, T. (2008). adegenet: A R package for the multivariate analysis of genetic markers. Bioinformatics, 24(11), 1403–1405. https://doi.org/10.1093/bioinformatics/btn129

Kanehisa, M., Araki, M., Goto, S., Hattori, M., Hirakawa, M., Itoh, M., Katayama, T., Kawashima, S., Okuda, S., Tokimatsu, T., & Yamanishi, Y. (2007). KEGG for linking genomes to life and the environment. Nucleic Acids Research, 36(Database),

D480–D484. https://doi.org/10.1093/nar/gkm882

Kendig, K. I., Baheti, S., Bockol, M. A., Drucker, T. M., Hart, S. N., Heldenbrand, J. R., Hernaez, M., Hudson, M. E., Kalmbach, M. T., Klee, E. W., Mattson, N. R., Ross, C. A., Taschuk, M., Wieben, E. D., Wiepert, M., Wildman, D. E., & Mainzer, L. S. (2019). Sentieon DNASeq Variant Calling Workflow Demonstrates Strong Computational Performance and Accuracy. Frontiers in Genetics, 10, 736. https://doi.org/10.3389/fgene.2019.00736

Kennett, D., Kennett, J., West, G., Erlandson, J., Johnson, J., Hendy, I., West, A., Culleton, B., Jones, T., & Staffordjr, T. (2008). Wildfire and abrupt ecosystem disruption on California's Northern Channel Islands at the Ållerød–Younger Dryas boundary (13.0–12.9ka). Quaternary Science Reviews, 27(27–28), 2530–2545. https://doi.org/10.1016/j.quascirev.2008.09.006

Kern, P., Cramp, R. L., & Franklin, C. E. (2015). Physiological responses of ectotherms to daily temperature variation. Journal of Experimental Biology, jeb.123166. https://doi.org/10.1242/jeb.123166

Kontopoulos, D., Smith, T. P., Barraclough, T. G., & Pawar, S. (2020). Adaptive evolution shapes the present-day distribution of the thermal sensitivity of population growth rate. PLOS Biology, 18(10), e3000894. https://doi.org/10.1371/journal.pbio.3000894

Kültz, D. (2005). MOLECULAR AND EVOLUTIONARY BASIS OF THE CELLULAR STRESS RESPONSE. Annual Review of Physiology, 67(1), 225–257. https://doi.org/10.1146/annurev.physiol.67.040403.103635

Lavín, M. F., & Marinone, S. G. (2003). An Overview of the Physical Oceanography of the Gulf of California. In O. U. Velasco Fuentes, J. Sheinbaum, & J. Ochoa (Eds.), Nonlinear Processes in Geophysical Fluid Dynamics: A tribute to the scientific work of Pedro Ripa (pp. 173–204). Springer Netherlands. https://doi.org/10.1007/978-94-010-0074-1_11

Lessios, H. A., & Robertson, D. R. (2006). Crossing the impassable: Genetic connections in 20 reef fishes across the eastern Pacific barrier. Proceedings of the Royal Society B: Biological Sciences, 273(1598), 2201–2208. https://doi.org/10.1098/rspb.2006.3543

Lewis, B. P., Green, R. E., & Brenner, S. E. (2003). Evidence for the widespread coupling of alternative splicing and nonsense-mediated mRNA decay in humans. Proceedings of the National Academy of Sciences, 100(1), 189–192. https://doi.org/10.1073/pnas.0136770100

Li, H. (2013). Aligning sequence reads, clone sequences and assembly contigs with BWA-MEM (Version 2). arXiv. https://doi.org/10.48550/ARXIV.1303.3997

Li, H. (2018). Minimap2: Pairwise alignment for nucleotide sequences. Bioinformatics, 34(18), 3094–3100. https://doi.org/10.1093/bioinformatics/bty191

Logan, C. A., & Buckley, B. A. (2015). Transcriptomic responses to environmental temperature in eurythermal and stenothermal fishes. Journal of Experimental Biology, 218(12), 1915–1924. https://doi.org/10.1242/jeb.114397

Logan, C. A., & Somero, G. N. (2010). Transcriptional responses to thermal acclimation in the eurythermal fish Gillichthys mirabilis (Cooper 1864). American Journal of Physiology-Regulatory, Integrative and Comparative Physiology, 299(3), R843–R852. https://doi.org/10.1152/ajpregu.00306.2010

Logan, C. A., & Somero, G. N. (2011). Effects of thermal acclimation on transcriptional responses to acute heat stress in the eurythermal fish Gillichthys mirabilis (Cooper). American Journal of Physiology-Regulatory, Integrative and Comparative Physiology, 300(6), R1373–R1383. https://doi.org/10.1152/ajpregu.00689.2010

Longo, G. C., D'Amelio, K., Larson, W., Enciso, C. E., Torre, J., Minich, J. J., Michael, T. P., & Craig, M. T. (2025). Population Genomics Reveals Panmixia in Pacific Sardine (Sardinops sagax) of the North Pacific. Evolutionary Applications, 18(9), e70154. https://doi.org/10.1111/eva.70154

Longo, G., & Bernardi, G. (2015). The evolutionary history of the embiotocid surfperch radiation based on genome-wide RAD sequence data. Molecular Phylogenetics and Evolution, 88, 55–63. https://doi.org/10.1016/j.ympev.2015.03.027

Losos, J. B., & Ricklefs, R. E. (2009). Adaptation and diversification on islands. Nature, 457(7231), 830–836. https://doi.org/10.1038/nature07893

Lowe, W. H., & Allendorf, F. W. (2010). What can genetics tell us about population connectivity? Molecular Ecology, 19(15), 3038–3051. https://doi.org/10.1111/j.1365-294X.2010.04688.x

Lynn, J. (1967). SEASONAL VARIATION OF TEMPERATURE AND SALINITY AT 10 METERS IN THE CALIFORNIA CURRENT.

Madeira, D., Narciso, L., Cabral, H. N., & Vinagre, C. (2012). Thermal tolerance and potential impacts of climate change on coastal and estuarine organisms. Journal of Sea Research, 70, 32–41. https://doi.org/10.1016/j.seares.2012.03.002

Maeda, K., Hanahara, N., Uehara, M., & Tachihara, K. (2022). Larval study revealed diversity and life‐history traits of crypto‐benthic eel gobies. Journal of Fish Biology, 101(6), 1411–1427. https://doi.org/10.1111/jfb.15209

Malhotra, J. D., & Kaufman, R. J. (2007). The endoplasmic reticulum and the unfolded protein response. Seminars in Cell & Developmental Biology, Degradation of Misfolded Glycoproteins, 18(6), 716–731. https://doi.org/10.1016/j.semcdb.2007.09.003

Mallet, J. (1995). A species definition for the modern synthesis. Trends in Ecology & Evolution, 10(7), 294–299.

Manni, M., Berkeley, M. R., Seppey, M., Simão, F. A., & Zdobnov, E. M. (2021). BUSCO Update: Novel and Streamlined Workflows along with Broader and Deeper Phylogenetic Coverage for Scoring of Eukaryotic, Prokaryotic, and Viral Genomes. Molecular Biology and Evolution, 38(10), 4647–4654. https://doi.org/10.1093/molbev/msab199

Marinone, S. G., Ulloa, M. J., Parés-Sierra, A., Lavín, M. F., & Cudney-Bueno, R. (2008). Connectivity in the northern Gulf of California from particle tracking in a three-dimensional numerical model. Journal of Marine Systems, 71(1), 149–158. https://doi.org/10.1016/j.jmarsys.2007.06.005

Marko, P. B., Hoffman, J. M., Emme, S. A., Mcgovern, T. M., Keever, C. C., & Nicole Cox, L. (2010). The 'Expansion-Contraction' model of Pleistocene biogeography: Rocky shores suffer a sea change? Molecular Ecology, 19(1), 146–169. https://doi.org/10.1111/j.1365-294X.2009.04417.x

Marraffini, M. L., Hamilton, S. L., Marin Jarrin, J. R., Ladd, M., Koval, G., Madden, J. R., Mangino, I., Parker, L. M., Emery, K. A., Terhaar, K., Hubbard, D. M., Miller, R. J., & Dugan, J. E. (2024). Evaluating the influence of marine protected areas on surf zone fish. Conservation Biology, 38(6), e14296. https://doi.org/10.1111/cobi.14296

Massicotte P, & South A. (2025). Rnaturalearth: World Map Data from Natural Earth. R package version 1.1.0.9000. https://docs.ropensci.org/rnaturalearth/

Mayr, E. (1963). Animal Species and. Evolution, 521.

McGourty, C. R., Hobbs, J. A., Bennett, W. A., Green, P. G., Hwang, H.-M., Ikemiyagi, N., Lewis, L., & Cope, J. M. (2009). Likely Population-Level Effects of Contaminants on a Resident Estuarine Fish Species: Comparing Gillichthys mirabilis Population Static Measurements and Vital Rates in San Francisco and Tomales Bays. Estuaries and Coasts, 32(6), 1111–1120. https://doi.org/10.1007/s12237-009-9177-5

Miles, A. (2015). scikit-allel: A Python package for exploring and analysing genetic variation data (Version 1.3.6). GitHub. https://github.com/cggh/scikit-allel

Minh, B. Q., Schmidt, H. A., Chernomor, O., Schrempf, D., Woodhams, M. D., Von Haeseler, A., & Lanfear, R. (2020). IQ-TREE 2: New Models and Efficient Methods for Phylogenetic Inference in the Genomic Era. Molecular Biology and Evolution, 37(5), 1530–1534. https://doi.org/10.1093/molbev/msaa015

Mirchandani, C. D., Shultz, A. J., Thomas, G. W. C., Smith, S. J., Baylis, M., Arnold, B., Corbett-Detig, R., Enbody, E., & Sackton, T. B. (2024). A Fast, Reproducible, High-throughput Variant Calling Workflow for Population Genomics. Molecular Biology and Evolution, 41(1), msad270. https://doi.org/10.1093/molbev/msad270

Momigliano, P., Jokinen, H., Fraimout, A., Florin, A.-B., Norkko, A., & Merilä, J. (2017). Extraordinarily rapid speciation in a marine fish. Proceedings of the National Academy of Sciences, 114(23), 6074–6079. https://doi.org/10.1073/pnas.1615109114

Nosil, P. (2012). Ecological speciation. Oxford University Press.

Nosil, P., & Feder, J. L. (2012a). Genomic divergence during speciation: Causes and consequences. Philosophical Transactions of the Royal Society B: Biological Sciences, 367(1587), 332–342. https://doi.org/10.1098/rstb.2011.0263

Nosil, P., & Feder, J. L. (2012b). Widespread yet heterogeneous genomic divergence. Molecular Ecology, 21(12), 2829–2832. https://doi.org/10.1111/j.1365-294X.2012.05580.x

Nosil, P., Funk, D. J., & Ortiz‐Barrientos, D. (2009). Divergent selection and heterogeneous genomic divergence. Molecular Ecology, 18(3), 375–402. https://doi.org/10.1111/j.1365-294X.2008.03946.x

Nóbrega, R. H., Morais, R. D. V. D. S., Crespo, D., De Waal, P. P., De França, L. R., Schulz, R. W., & Bogerd, J. (2015). Fsh Stimulates Spermatogonial Proliferation and Differentiation in Zebrafish via Igf3. Endocrinology, 156(10), 3804–3817. https://doi.org/10.1210/en.2015-1157

Oda, K. T., Meyer, K. C., & Reilly, P. N. (2019). Barred Surfperch, Amphistichus argenteus, and Redtail Surfperch, Amphistichus rhodoterus, Enhanced Status Report. 95.

Orr, P. C. (1968). Prehistory of Santa Rosa Island. Santa Barbara Museum of Natural History.

Ortiz, E. M. (2019). vcf2phylip v2.0: Convert a VCF matrix into several matrix formats for

phylogenetic analysis. https://doi.org/DOI:10.5281/zenodo.254086

Padfield, D., O'Sullivan, H., & Pawar, S. (2021). rTPC and nls.multstart: A new pipeline to fit thermal performance curves in r. Methods in Ecology and Evolution, 12(6), 1138–1143. https://doi.org/10.1111/2041-210X.13585

Palumbi, S. R. (1992). Marine speciation on a small planet. Trends in Ecology & Evolution, 7(4), 114–118. https://doi.org/10.1016/0169-5347(92)90144-Z

Palumbi, S. R. (1994). GENETIC DIVERGENCE, REPRODUCTIVE ISOLATION, AND MARINE SPECIATION. Annual Review of Ecology and Systematics, 25(1), 547–572. https://doi.org/10.1146/annurev.es.25.110194.002555

Palumbi, S. R. (2009). Speciation and the evolution of gamete recognition genes: Pattern and process. Heredity, 102(1), 66–76. https://doi.org/10.1038/hdy.2008.104

Palumbi, S. R., & Metz, E. C. (1991). Strong reproductive isolation between closely related tropical sea urchins (genus Echinometra). Molecular Biology and Evolution, 8(2), 227–239. https://doi.org/10.1093/oxfordjournals.molbev.a040642

Perrot, O., Harris, L. N., EHTO, Legagneux, P., Moore, J.-S., & Bernatchez, L. (2026). Application of a GT-seq SNP panel to quantify interannual variation in mixed-stock composition in the largest commercial fishery for Arctic Charr (Salvelinus alpinus) in Canada. Arctic Science. https://doi.org/10.1139/as-2025-0084

Pinsky, M. L., Saenz-Agudelo, P., Salles, O. C., Almany, G. R., Bode, M., Berumen, M. L., Andréfouët, S., Thorrold, S. R., Jones, G. P., & Planes, S. (2017). Marine Dispersal Scales Are Congruent over Evolutionary and Ecological Time. Current Biology, 27(1), 149–154. https://doi.org/10.1016/j.cub.2016.10.053

Place, S. P., & Hofmann, G. E. (2001). Temperature interactions of the molecular chaperone Hsc70 from the eurythermal marine goby Gillichthys mirabilis. Journal of Experimental Biology, 204(15), 2675–2682. https://doi.org/10.1242/jeb.204.15.2675

Prince, D. J., O'Rourke, S. M., Thompson, T. Q., Ali, O. A., Lyman, H. S., Saglam, I. K.,

Hotaling, T. J., Spidle, A. P., & Miller, M. R. (2017). The evolutionary basis of premature migration in Pacific salmon highlights the utility of genomics for informing conservation. Science Advances, 3(8), e1603198. https://doi.org/10.1126/sciadv.1603198

Privé, F., Luu, K., Vilhjálmsson, B. J., & Blum, M. G. B. (2020). Performing Highly Efficient Genome Scans for Local Adaptation with R Package pcadapt Version 4. Molecular Biology and Evolution, 37(7), 2153–2154. https://doi.org/10.1093/molbev/msaa053

Rabosky, D. L., Chang, J., Title, P. O., Cowman, P. F., Sallan, L., Friedman, M., Kaschner, K., Garilao, C., Near, T. J., Coll, M., & Alfaro, M. E. (2019). Data from: An inverse latitudinal gradient in speciation rate for marine fishes (Version 1, p. 116034679 bytes) [Dataset]. Dryad. https://doi.org/10.5061/DRYAD.FC71CP4

Raimondi, P. (2011). Variation in entrainment impact estimation based on different measures of acceptable uncertainty. California Energy Commission, PIER Energy Related Environmental Research Program. CEC, 500(2011), 020.

Rautsaw, R. M., Schramer, T. D., Acuña, R., Arick, L. N., DiMeo, M., Mercier, K. P., Schrum, M., Mason, A. J., Margres, M. J., Strickland, J. L., & Parkinson, C. L. (2021). Genomic Adaptations to Salinity Resist Gene Flow in the Evolution of Floridian Watersnakes. Molecular Biology and Evolution, 38(3), 745–760. https://doi.org/10.1093/molbev/msaa266

Reisser, C. M. O., Beldade, R., & Bernardi, G. (2009). Multiple paternity and competition in sympatric congeneric reef fishes, Embiotoca jacksoni and E. lateralis. Molecular Ecology, 18(7), 1504–1510. https://doi.org/10.1111/j.1365-294X.2009.04123.x

Riddle, B. R., Hafner, D. J., Alexander, L. F., & Jaeger, J. R. (2000). Cryptic vicariance in the historical assembly of a Baja California Peninsular Desert biota. Proceedings of the National Academy of Sciences, 97(26), 14438–14443. https://doi.org/10.1073/pnas.250413397

Riginos, C., Buckley, Y. M., Blomberg, S. P., Treml, E. A. (2014). Dispersal Capacity Predicts Both Population Genetic Structure and Species Richness in Reef Fishes. The American Naturalist, 184(1), 52–64. https://doi.org/10.1086/676505

Rocha, L. A., & Bowen, B. W. (2008). Speciation in coral‐reef fishes. Journal of Fish Biology, 72(5), 1101–1121. https://doi.org/10.1111/j.1095-8649.2007.01770.x

Roden, G. I., & Groves, G. W. (1959). Recent oceanographic investigations in the Gulf of California.

Ruan, J., & Li, H. (2020). Fast and accurate long-read assembly with wtdbg2. Nature Methods, 17(2), 155–158. https://doi.org/10.1038/s41592-019-0669-3

Saha, A., Jorde, P. E., Sodeland, M., Korslund, L. M., & Knutsen, H. (2025). Pelagic Larval Duration and Isolation by Distance in Coastal Species. Evolutionary Applications, 18(12), e70187. https://doi.org/10.1111/eva.70187

Salazar Sawkins, A., & Bernardi, G. (2026). Population genomics and speciation of a Baja California disjunct species, the longjaw mudsucker (Gillichthys mirabilis). Ecology and Evolution, in review. https://doi.org/10.22541/authorea.15003180/v1

Sanchez-Cabeza, J.-A., Herrera-Becerril, C. A., Carballo, J. L., Yáñez, B., Álvarez-Sánchez, L. F., Cardoso-Mohedano, J.-G., & Ruiz-Fernández, A. C. (2022). Rapid surface water warming and impact of the recent (2013–2016) temperature anomaly in shallow coastal waters at the eastern entrance of the Gulf of California. Progress in Oceanography, 202, 102746. https://doi.org/10.1016/j.pocean.2022.102746

Sanford, E., & Kelly, M. W. (2011). Local Adaptation in Marine Invertebrates. Annual Review of Marine Science, 3(1), 509–535. https://doi.org/10.1146/annurev-marine-120709-142756

Schluter, D. (2001). Ecology and the origin of species. Trends in Ecology & Evolution, 16(7), 372–380. https://doi.org/10.1016/S0169-5347(01)02198-X

Schoolfield, R. M., Sharpe, P. J. H., & Magnuson, C. E. (1981). Non-linear regression of

biological temperature-dependent rate models based on absolute reaction-rate theory. Journal of Theoretical Biology, 88(4), 719–731. https://doi.org/10.1016/0022-5193(81)90246-0

Schulte, P. M. (2015). The effects of temperature on aerobic metabolism: Towards a mechanistic understanding of the responses of ectotherms to a changing environment. Journal of Experimental Biology, 218(12), 1856–1866. https://doi.org/10.1242/jeb.118851

Schulte, P. M., Healy, T. M., & Fangue, N. A. (2011). Thermal Performance Curves, Phenotypic Plasticity, and the Time Scales of Temperature Exposure. Integrative and Comparative Biology, 51(5), 691–702. https://doi.org/10.1093/icb/icr097

Scott, G. R., & Johnston, I. A. (2012). Temperature during embryonic development has persistent effects on thermal acclimation capacity in zebrafish. Proceedings of the National Academy of Sciences, 109(35), 14247–14252. https://doi.org/10.1073/pnas.1205012109

Seehausen, O., Butlin, R. K., Keller, I., Wagner, C. E., Boughman, J. W., Hohenlohe, P. A., Peichel, C. L., Saetre, G.-P., Bank, C., Brännström, Å., Brelsford, A., Clarkson, C. S., Eroukhmanoff, F., Feder, J. L., Fischer, M. C., Foote, A. D., Franchini, P., Jiggins, C. D., Jones, F. C., … Widmer, A. (2014). Genomics and the origin of species. Nature Reviews Genetics, 15(3), 176–192. https://doi.org/10.1038/nrg3644

Selkoe, K. A., D'Aloia, C. C., Crandall, E. D., Iacchei, M., Liggins, L., Puritz, J. B., Heyden, S. von der, & Toonen, R. J. (2016). A decade of seascape genetics: Contributions to basic and applied marine connectivity. Marine Ecology Progress Series, 554, 1–19. https://doi.org/10.3354/meps11792

Selkoe, K. A., & Toonen, R. J. (2011). Marine connectivity: A new look at pelagic larval duration and genetic metrics of dispersal. Marine Ecology Progress Series, 436, 291–305. https://doi.org/10.3354/meps09238

Servedio, M. R., & Noor, M. A. F. (2003). The Role of Reinforcement in Speciation: Theory and Data. Annual Review of Ecology, Evolution, and Systematics, 34(Volume 34, 2003), 339–364. https://doi.org/10.1146/annurev.ecolsys.34.011802.132412

Shumate, A., & Salzberg, S. L. (2021). Liftoff: Accurate mapping of gene annotations. Bioinformatics, 37(12), 1639–1643. https://doi.org/10.1093/bioinformatics/btaa1016

Siddall, M., Rohling, E. J., Almogi-Labin, A., Hemleben, Ch., Meischner, D., Schmelzer, I., & Smeed, D. A. (2003). Sea-level fluctuations during the last glacial cycle. Nature, 423(6942), 853–858. https://doi.org/10.1038/nature01690

Sidell, B. D., Johnston, I. A., Moerland, T. S., & Goldspink, G. (1983). The eurythermal myofibrillar protein complex of the mummichog (Fundulus heteroclitus): Adaptation to a fluctuating thermal environment. Journal of Comparative Physiology, 153(2), 167–173. https://doi.org/10.1007/BF00689620

Simão, F. A., Waterhouse, R. M., Ioannidis, P., Kriventseva, E. V., & Zdobnov, E. M. (2015). BUSCO: Assessing genome assembly and annotation completeness with single-copy orthologs. Bioinformatics, 31(19), 3210–3212. https://doi.org/10.1093/bioinformatics/btv351

Slatkin, M. (1987). Gene Flow and the Geographic Structure of Natural Populations. Science, 236(4803), 787–792. https://doi.org/10.1126/science.3576198

Solovyev, M. M., Kashinskaya, E. N., Rogozhin, E. A., & Moyano, F. J. (2021). Seasonal changes in kinetic parameters of trypsin in gastric and agastric fish. Fish Physiology and Biochemistry, 47(2), 381–391. https://doi.org/10.1007/s10695-020-00919-0

Somero, G. N. (2020). The cellular stress response and temperature: Function, regulation, and evolution. Journal of Experimental Zoology Part A: Ecological and Integrative Physiology, 333(6), 379–397. https://doi.org/10.1002/jez.2344

Sorte, C., & Stachowicz, J. (2011). Patterns and processes of compositional change in a California epibenthic community. Marine Ecology Progress Series, 435, 63–74.

https://doi.org/10.3354/meps09234

Sun, P., Gan, G., Tian, Y., Liang, Z., Yang, Y., & Heino, M. (2025). Genomic Adaptation of Marine Medaka (Oryzias melastigma) Under Artificial Life-History Selection. Molecular Ecology, 34(14), e17833. https://doi.org/10.1111/mec.17833

Taylor, R. S., & Friesen, V. L. (2017). The role of allochrony in speciation. Molecular Ecology, 26(13), 3330–3342.

Thornton, J. E., Du, P., Jing, L., Sjekloca, L., Lin, S., Grossi, E., Sliz, P., Zon, L. I., & Gregory, R. I. (2014). Selective microRNA uridylation by Zcchc6 (TUT7) and Zcchc11 (TUT4). Nucleic Acids Research, 42(18), 11777–11791. https://doi.org/10.1093/nar/gku805

Toy, J. A. (2022). Adaptation Genomics of Surfperch Populations in the Context of Rapid Environmental Change [Ph.D., University of California, Santa Cruz]. In ProQuest Dissertations and Theses. Dissertations & Theses @ University of California; ProQuest Dissertations & Theses A&I (2774212933). https://www.proquest.com/dissertations-theses/adaptation-genomics-surfperch-populations-context/docview/2774212933/se-2?accountid=14523

Toy, J. A., Oulès, A., Longo, G., Shaffer, H. B., & Bernardi, G. (2025). Population Genomics, Local Adaptation and Cryptic Speciation in a Temperate Reef Fish, the Black Surfperch, Embiotoca jacksoni, Using Genome‐Wide Resequencing. Molecular Ecology, 34(22), e70073. https://doi.org/10.1111/mec.70073

Van Der Auwera, G. A., Carneiro, M. O., Hartl, C., Poplin, R., Del Angel, G., Levy‐Moonshine, A., Jordan, T., Shakir, K., Roazen, D., Thibault, J., Banks, E., Garimella, K. V., Altshuler, D., Gabriel, S., & DePristo, M. A. (2013). From FastQ Data to High‐Confidence Variant Calls: The Genome Analysis Toolkit Best Practices Pipeline. Current Protocols in Bioinformatics, 43(1). https://doi.org/10.1002/0471250953.bi1110s43

Varuzhanyan, G., Rojansky, R., Sweredoski, M. J., Graham, R. L., Hess, S., Ladinsky, M. S.,

& Chan, D. C. (2019). Mitochondrial fusion is required for spermatogonial differentiation and meiosis. eLife, 8, e51601. https://doi.org/10.7554/eLife.51601

Vaser, R., Sović, I., Nagarajan, N., & Šikić, M. (2017). Fast and accurate de novo genome assembly from long uncorrected reads. Genome Research, 27(5), 737–746. https://doi.org/10.1101/gr.214270.116

Velotta, J. P., McCormick, S. D., Whitehead, A., Durso, C. S., & Schultz, E. T. (2022). Repeated Genetic Targets of Natural Selection Underlying Adaptation of Fishes to Changing Salinity. Integrative and Comparative Biology, 62(2), 357–375. https://doi.org/10.1093/icb/icac072

Verberk, W. C. E. P., Henry, E., Leiva, F. P., Barbarossa, V., & Schipper, A. M. (2026). Heat Tolerance and Its Plasticity in Freshwater and Marine Fishes Reflect Exposure to Extremes and Seasonal Variation in Habitat Temperatures. Ecology Letters, 29(3), e70341. https://doi.org/10.1111/ele.70341

Verri, T., Terova, G., Romano, A., Barca, A., Pisani, P., Storelli, C., & Saroglia, M. (2012). The SoLute Carrier (SLC) family series in teleost fish. Functional Genomics in Aquaculture, 24, 219–320.

Vogel, C., Bashton, M., Kerrison, N. D., Chothia, C., & Teichmann, S. A. (2004). Structure, function and evolution of multidomain proteins. Current Opinion in Structural Biology, 14(2), 208–216. https://doi.org/10.1016/j.sbi.2004.03.011

Waterbury, C. R., Sutton, T. M., Kelley, A. L., & López, J. A. (2024). Effects of temperature acclimation on the upper thermal tolerance of two Arctic fishes. Conservation Physiology, 12(1), coae001. https://doi.org/10.1093/conphys/coae001

Weersing, K., & Toonen, R. (2009). Population genetics, larval dispersal, and connectivity in marine systems. Marine Ecology Progress Series, 393, 1–12. https://doi.org/10.3354/meps08287

Weir, B. S., & Cockerham, C. C. (1984). Estimating F-Statistics for the Analysis of Population

Structure. Evolution, 38(6), 1358. https://doi.org/10.2307/2408641

Weisel, G. F. (1947). Breeding Behavior and Early Development of the Mudsucker, a Gobiid Fish of California. Copeia, 1947(2), 77. https://doi.org/10.2307/1438636

White, C., Selkoe, K. A., Watson, J., Siegel, D. A., Zacherl, D. C., & Toonen, R. J. (2010). Ocean currents help explain population genetic structure. Proceedings of the Royal Society B: Biological Sciences, 277(1688), 1685–1694. https://doi.org/10.1098/rspb.2009.2214

Wiens, B. J., & Colella, J. P. (2024). triangulaR: an R package for identifying AIMs and building triangle plots using SNP data from hybrid zones. Evolutionary Biology. https://doi.org/10.1101/2024.03.28.587167

Wiens, B. J., & Colella, J. P. (2025). That's Not a Hybrid: How to Distinguish Patterns of Admixture and Isolation By Distance. Molecular Ecology Resources, 25(3), e14039. https://doi.org/10.1111/1755-0998.14039

Wolf, J. B. W., & Ellegren, H. (2017). Making sense of genomic islands of differentiation in light of speciation. Nature Reviews Genetics, 18(2), 87–100. https://doi.org/10.1038/nrg.2016.133

Wood, D. E., Lu, J., & Langmead, B. (2019). Improved metagenomic analysis with Kraken 2. Genome Biology, 20(1), 257. https://doi.org/10.1186/s13059-019-1891-0

Wourms, J. P. (1981). Viviparity: The Maternal-Fetal Relationship in Fishes. American Zoologist, 21(2), 473–515.

Wu, C.-I. (2001). The genic view of the process of speciation. Journal of Evolutionary Biology, 14(6), 851–865.

Wu, Z., Pang, N., Zhang, Y., Chen, H., Peng, Y., Fu, J., & Wei, Q. (2020). CEP290 is essential for the initiation of ciliary transition zone assembly. PLOS Biology, 18(12), e3001034. https://doi.org/10.1371/journal.pbio.3001034

Xiao, J., Zhong, H., & Feng, H. (2023). Post-translational modifications and regulations of

RLR signaling molecules in cytokines-mediated response in fish. Developmental & Comparative Immunology, 141, 104631. https://doi.org/10.1016/j.dci.2023.104631

Xue, Y., Kuok, C., Xiao, A., Zhu, Z., Lin, S., & Zhang, B. (2010). Identification and expression analysis of mical family genes in zebrafish. Journal of Genetics and Genomics, 37(10), 685–693. https://doi.org/10.1016/S1673-8527(09)60086-2

Yoklavich, M. M., Stevenson, M., & Cailliet, G. M. (1992). Seasonal and spatial patterns of ichthyoplankton abundance in Elkhorn Slough, California. Estuarine, Coastal and Shelf Science, 34(2), 109–126. https://doi.org/10.1016/S0272-7714(05)80099-X

Zhao, X., Li, L., Li, C., Liu, E., Zhu, H., & Ling, Q. (2022). Heat stress-induced endoplasmic reticulum stress promotes liver apoptosis in largemouth bass (Micropterus salmoides). Aquaculture, 546, 737401. https://doi.org/10.1016/j.aquaculture.2021.737401