UNIVERSITY OF CALIFORNIA
SANTA CRUZ

GENOMIC ANALYSIS OF DISJUNCT MARINE FISH POPULATIONS OF THE NORTHEASTERN PACIFIC AND SEA OF CORTEZ

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

DOCTOR OF PHILOSOPHY

in

ECOLOGY AND EVOLUTIONARY BIOLOGY

by

Eric Garcia

December 2018

The Dissertation of Eric Garcia is approved:

Professor Giacomo Bernardi, Chair

Professor Pete Raimondi

Brian Simison, Ph.D.

Professor Octavio Aburto

Lori Kletzer
Vice Provost and Dean of Graduate Studies

Table of Contents

List of Tables

General Introduction

Tempo and mode of divergence in Baja California disjunct fishes based on a genome-wide calibrated substitution rate.

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*)

Genomic divergence and signals of convergent selection between northeastern Pacific and Sea of Cortez disjunct populations of four marine fishes

List of Figures

General Introduction

Tempo and mode of divergence in Baja California disjunct fishes based on a genome-wide calibrated substitution rate.

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*)

Genomic divergence and signals of convergent selection between northeastern Pacific and Sea of Cortez disjunct populations of four marine fishes

Abstract

Genomic analysis of disjunct marine fish populations of the northeastern Pacific and Sea of Cortez

Eric Garcia

The formation of the Baja California peninsula separated the distributions of 19 marine fishes into disjunct Pacific and Sea of Cortez populations. Similarly, their Pacific distributions cross phylogeographic points that diminish the genetic connectivity of their populations. This resulted in multiple species experiencing a gradient of gene flow and an extraordinary framework to study mechanisms of divergence and signals of selection under different scenarios of isolation. Genetic isolation in these species has previously been studied using only a handful of markers. In this dissertation, Restriction Site-Associated DNA (RADseq) is used to genotype thousands of genome-wide makers to study the evolutionary history, characterized genomic patterns of divergence among populations, and search for signals of drift and selection, in four of these species: the sargo, *Anisotremus davidsonii* (ADA)*;* the longjaw mudsucker, *Gillichthys mirabilis* (GMI), the California sheephead, *Semicossyphus pulcher* (SPU); and the zebraperch, *Kyphosus azureus* (KAZ). While, no evidence of isolation was found for SPU, estimated dates the disjunction of ADA, GMI, and KAZ, were 60, 284, and 23 thousand years, respectively. A dispersal episode where species migrated northwards and then became isolated by the warming of the seawater temperate in the south of the Baja California peninsula at the end of

the last glaciation period, and a vicariant isolation resulting from the closure of a mid-peninsular seaway, are the plausible historical events that produced the current distributions of these species. Lower than expected levels of genomic gene flow (significant p-values) were seen across Point Conception in GMI (F_{ST}=0.15), Punta Eugenia in ADA (F_{ST}=0.02), and across the peninsula in KAZ, GMI and ADA (F_{ST}=0.03, 0.11, and 0.23, respectively). Furthermore, when comparing disjunct populations, 19 to 46 % of outlier loci matched coding genes in all species and analyses identified 15 genomic regions, potentially involved in processing environmental information, metabolism, immune response, and possibly reproduction, diverging in more than one of these species. Results are interpreted as supporting (1) the idea that ADA and GMI disjunct populations are in the initial phases of allopatric speciation and (2) the presence of convergent selection in these species.

Dedication

To my family, friends, and mentors.

Acknowledgements

I would like to express my deep gratitude to my partner and family for supporting me throughout my Ph.D. adventure. Thank you for making my life so much better. I am forever in doubt to my committee members, Drs. Giacomo Bernardi, Pete Raimondi, Brian Simison, and Octavio Aburto, for the guidance and support at all times during this dissertation. Thank you to all EEB professors and fellow Ph.D. students for your lessons and discussions, which improved my understanding of evolution and academic life. Finally, I would like to thank all my friends for making Santa Cruz an incredible place to call home for all these years.

This work was supported financially by the Consejo Nacional de Ciencia y Tecnología (CONACYT) in Mexico, UC Mexus, The Myers Trust, STARTS scholarship, Friends of the Long Marine Lab, and by the Graduate Student Association and the department of Ecology and Evolutionary Biology at UCSC. I would also like to thank Dr. Larry G. Allen for providing fish illustrations.

GENERAL INTRODUCTION

Studying Evolution in the Genomic Era

Genetic differences are the building blocks of biodiversity. How this variation distinguishes populations and eventually creates new species has been a central theme of investigation among evolutionary biologists. The way scientists have perceived the speciation process and the methodologies they used to answer questions about evolution have been evolving through time. Speciation was first proposed by Darwin to be the result of natural selection but with the rise of the modern synthesis, isolation, where populations can divergence without any selective pressures, was instead considered to be the main driver of differentiation (Coyne & Orr 2004; Darwin 1859; Gaither et al. 2015). Much of the research in the field had thus focused in populations with some geographic separation until recently, when many authors have documented an active role of natural selection in several sympatric systems (Schluter 2009; Crow et al 2010; Nosil 2012; Bowen et al 2013). (Nosil 2012; Crow et al. 2010)As a result, views of speciation are currently more holistic and include spatial and ecological components that often operate in conjunction during speciation events (Diekmann 1999; Bernardi 2013).

Regardless of the forces driving differentiation, a population must first have genetic variation in order to diverge from others. Variation is acquired by a population through random mutations, recombination or the gene shuffling in sexual organisms, or the introduction of new alleles by individuals migrating from a

different population. Each population has then a unique gene pool that is the outcome of complex historical and contemporary evolutionary trajectories. Existing variation within a population can be maintained by micro-evolutionary processes such as balancing selection or diminished by drift and other types of selection.

Structure between populations, whether or not they are isolated, can be produced randomly through genetic drift (acting mostly upon neutral loci) or selectively by natural selection (by definition acting only upon non-neutral loci). When populations interbreed, structure will be shaped by the interplay of gene flow, drift and selection. On one hand, populations might become more similar to each other by gene flow resulting in an overall homogenization of variation throughout the genome. On the other hand, natural selection is considered to be the major force responsible for any existing differentiation between populations with high gene flow (Nosil et al. 2009; Gaither et al. 2015), except for small effective populations where drift can also substantially change the frequencies of random alleles.

Therefore, it is important to include neutral and non-neutral loci to obtain a more complete picture of the evolutionary histories of populations and species alike. Due to technological limitations, geneticists have analyzed the speciation process for a long time with only a handful of loci (Hohenlohe et al. 2010), which in some studies were a mix of neutral and selected loci but in most were only of one type. Yet, a new suite of genetic tools often referred as Next Generation Sequencing (NGS) have revolutionized the study of genomics and increased the ability of researchers to investigate components of evolution in both, individual and holistic approaches.

Genetic variation can now be studied at the genomic level with technologies such as Restriction site-Associated DNA (RADseq) that can produce amounts of loci many orders of magnitude greater than First Generation Sequencing methodologies (Baird et al. 2008; Hohenlohe et al. 2010). Traditional population parameters can be measured at the genomic level and analyses yield higher confidence levels and statistical power.

Furthermore, discrepancies in the allele frequencies between populations allow calculating statistics such as F_{ST} that estimate differentiation. These statistics can be measured throughout the entire genomic data and models can be run to predict the range of differentiation that is expected to occur randomly through genetic drift. Loci embedded within this range represent neutral loci that evolve freely (Figure 1; Hohenlohe et al. 2010; Nosil et al. 2009). In contrast, loci with lower differentiation (lower F_{ST} values) than neutral loci are candidate regions to be under balancing selection. This maintains a high diversity of alleles which can be useful in genes involved, for instance, in pathogen defenses, immune and inflammation responses (Hohenlohe et al. 2010). Similarly, loci with higher differentiation than neutral loci are candidate regions to be under divergent selection which are the loci experiencing different selective pressures and driving local adaptation in populations (Hohenlohe et al. 2010; Nosil et al. 2009). However, outlier loci are initially only considered "candidates" since some of these loci might illustrate higher levels of differentiation than neutral loci only because they are present in adjacent, non-adaptive areas, linked to the actual selected loci (Hohenlohe et al. 2010; Nosil et al. 2009; Bernardi 2013).

Outlier loci can then either be the actual genes under selection or simply linked loci that are "hitchhiking" the effects of selection and creating islands of speciation, or areas in the genome with substantially higher differentiation (Figure 1; Nosil et al. 2009; Bernardi 2013).

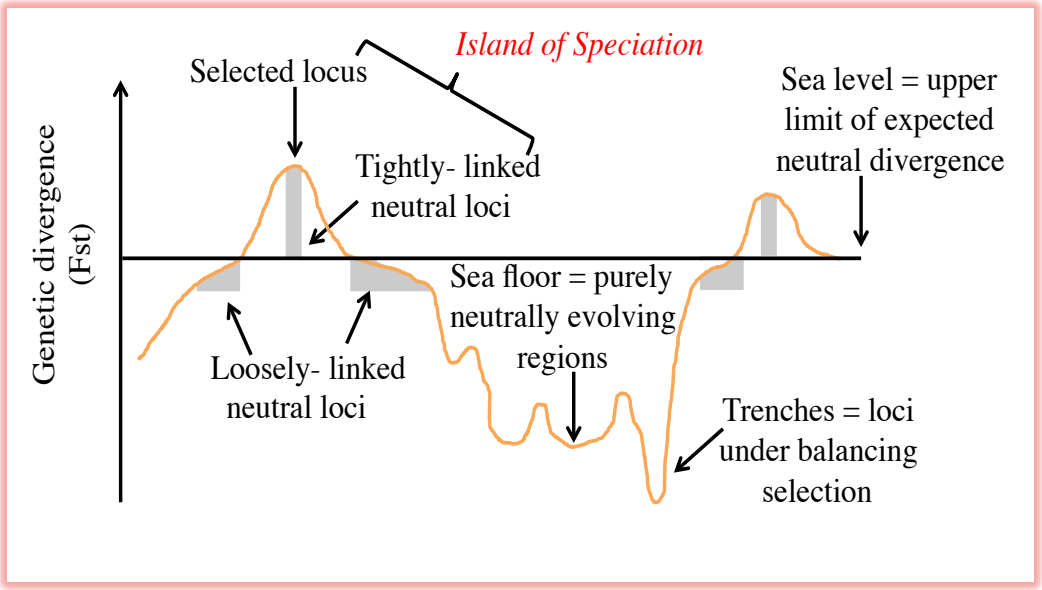


Figure 1. Visual representation of expected differentiation along the genome (orange line). Loci with significantly higher differentiation create islands of speciation uplifted by loci under selection. Areas with significantly lower differentiation represent loci under balancing selection. Loci between these two extremes are neutral loci (modified from Nosil et al. 2009).

RADseq uses enzymes that cut DNA at particular restriction sites producing thousands of sequences which are distributed throughout the genome (Baird et al. 2008). Once outlier loci are filtered, they can be compared to the online database Basic Local Alignment Search Tool (BLAST) to reveal if any of them match known gene sequences (Altschul et al. 1997). Subsequently, with knowledge of gene

function, we can then infer the possible ecological or environmental mechanisms that might be responsible for the differential selective pressure between populations (Gaither et al. 2015). Recent studies discovered that on average, 5 to 10% of the total number of loci obtained by RADseq showed signs of divergent selection (i.e. higher differentiation than neutral loci) and some of these loci where successfully correlated to an ecological process in the studied system (Nosil et al. 2009).

Any organism and even any population can have unique evolutionary trajectories shaped by many factors such as behavior, historical events affecting population size, and the combined forces of micro-evolutionary processes acting in the past and present time. Thus, much of our ability to investigate how the different drivers of evolution (gene flow, drift, and selection) produce divergence in allopatric and sympatric populations, depends as well on finding good systems where to apply these modern genetic tools. Marine fishes offer the opportunity to study these components in both individual and holistic approaches (Bernardi 2013). For instance, while some fishes exhibit large distributions with little or no genetic structure (Bowen et al. 2006; Schultz et al. 2007), the formation of the Isthmus of Panama approximately 2.8 to 3.5 million years ago (Coates and Obando, 1996; Collins, 1996; Craig et al., 2004; Knowlton, 1993; Lessios, 2008; Marko, 2002; Thomson et al., 2000), acted as an absolute barrier to gene flow for several others. In-between these two extremes, the Baja California disjunct fishes form an extraordinary system of marine fishes with shared evolutionary histories and a mix of allopatric and sympatric populations where to study genomic divergence at different levels of isolation.

The Baja California Disjunct Fishes

The Baja California disjunct fishes are a group of 19 temperate species that shared similar distributions from central California to southern Baja California, in the northern Eastern Pacific, and northern and central zones of the Sea of Cortez (here also referred as the "Gulf", Figure 2, Table 1, Bernardi et al. 2003).

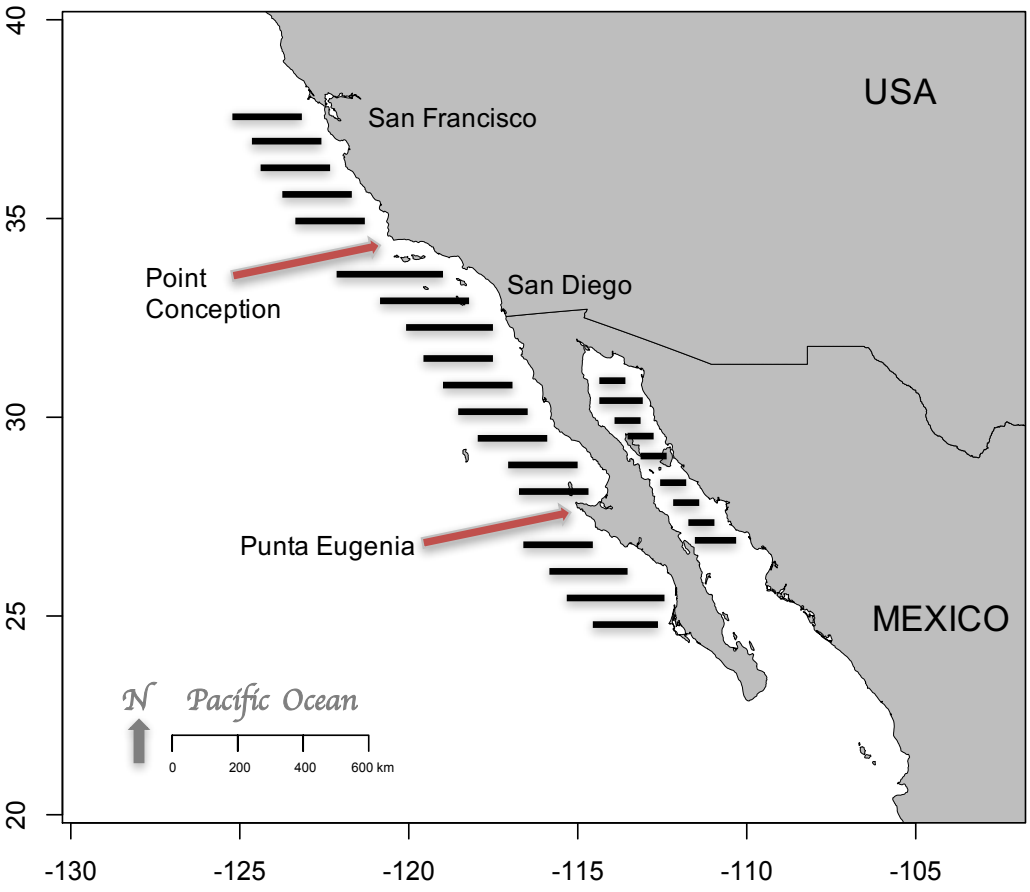


Figure 2. Typical Baja California disjucnt distribution.

Table 1. List and distribution of the 19 Baja California disjunct fish species (Froese & Pauly; Thomson et al., 2000; Bernardi et al., 2003). Pacific distributions are more specific than Gulf. Notes in the Gulf distribution are given when available.

Family	Species	Common name	Range
Atherinidae	*Leuresthes tenuis/sardina*	grunion	San Francisco to Magdalena Bay (southern Pacific Baja)
Girellidae	*Girella nigricans /simplicidens*	opaleye	San Francisco to southern Baja (common in upper gulf)
Haemulidae	*Anisotremus davidsonii*	sargo	Monterey to Magdalena Bay (rare north of Point Conception (Miller and lea 1972)
Blenniidae	*Hypsoblennius jenkinsi*	mussel blenny	Southern California to Baja
Chaenopsidae	*Chaenopsis alepidota*	oragethroat pikeblenny	Southern California to Baja. Upper gulf.
Serranidae	*Paralabrax maculatofasciatus*	spotted sand bass	Monterey to Baja
Gobiidae	*Gillichthys mirabilis*	longjaw mudsucker	Tomales bay to Bahia Magdanela. Mulege and Bahia Santa Maria, Culiacan, Sinaloa to the north of the gulf.
Gobiidae	*Lythrytpnus dalli*	blue banded boby	Southern California to central Baja
Blenniidae	*Hypsoblennius gentilis*	bay blenny	Monterey to Bahia Magdalena
Kyphosidae	*Kyphosus azureus*	zebraperch	Monterey to Baja
Labridae	*Halichoeres semicinctus*	rock wrasse	Point Conception to Bahia Magdalena (Isla Guadalupe too). Bahia de la Paz - Mazatlan and north in the gulf
Labridae	*Semicossyphus pulcher*	california sheephead	Monterey to Baja (uncommon south of Isla Angel de la Guardia and Guaymas)
Scorpaenidae	*Sebastes macdonaldi*	Mexican rockfish	California to southern Baja
Scorpaenidae	*Scorpaena guttata*	scorpionfish	Santa Cruz to Punta Abreojos (southern Baja, isla Guadalupe too)
Embiotocidae	*Zalembius rosaceus*	pink surfperch	Point Delgada, northern California to Bahia San Cristobal, southern Baja
Polyprionidae	*Stereolepis gigas*	giant seabass	Humboldt Bay to southern Baja. Upper Gulf
Agonidae	*Xenetremus ritteri*	flagfin poacher	Malibu to central Baja. Upper Gulf
Pleuronectidae	*Hypsopsetta guttulata*	diamond turbot	Cape Mendocino, northern California to southern Baja
Pleuronectidae	*Pleuronichthys verticalis*	hornyhead turbot	Point Reyes (San Francisco) to southern Baja

Populations along the Pacific coast occur without geographic barriers but gene flow might be limited by temperature and oceanographic discontinuities at Point Conception and Punta Eugenia (Allen et al. 2006). In contrast, there are no obvious barriers within the Gulf but these populations are isolated from the Pacific by the actual Baja California peninsula and the warmer seawater temperate at southernmost regions (Bernardi et al. 2003; Medina & Walsh 2000; Thomson et al. 2000; Present 1987; Tranah & Allen 1999). The end result is 19 natural cases of allopatric and sympatric populations with different levels of gene flow.

The taxonomy, ecology and life histories of these species do not illustrate patterns that can help explain the similar distributions. These species belong to 14 different families and have diverse ecologies from inhibiting estuaries to deep reefs, or having small to large body sizes, different diets, or short to long pelagic larval durations (i.e. 30 days in some and 120 in others). Yet, two genetic patterns based on few mitochondrial and nuclear markers were revealed among these species (Bernardi et al. 2003). In the first, 8 of 12 studied species showed deep genetic differences between the Pacific and Gulf populations as well as shallower differences across Point Conception and Punta Eugenia. The other 4 species did not show noticeable divergence between any of their populations. Here, we employed NGS techniques to the two species showing the highest structure: the sargo, *Anisotremus davidsonii* (grunts, Haemulidae) and the longjaw mudsucker, *Gillichthys mirabilis* (gobies, Gobiidae), as well as two of the species showing no divergence: the California sheephead, *Semicossyphus pulcher* (wrasses, Labridae) and the zebraperch,

Kyphosus azureus (sea chubs, Kyphosidae), to examine how micro-evolutionary processes affect populations in a gradient of isolation (Figure 3).

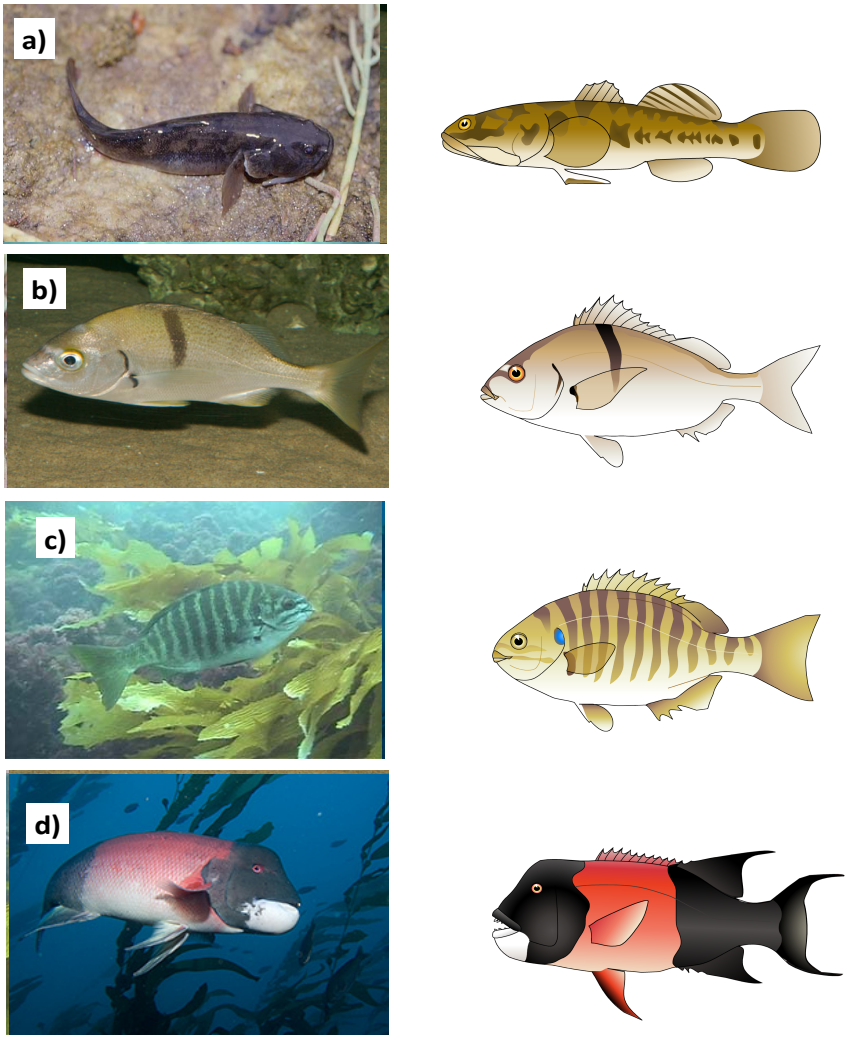


Figure 3. The studied species:
a) longjaw mudsucker, *Gillichthys mirabilis;*
b) sargo, *Anisotremus davidsonii;*
c) zebraperch, *Hermosilla azurea;*
d) California sheephead, *Semicossyphus pulcher.*

Permission for illustrations provided by Dr. Larry .G Allen

In this dissertation, two pairs of congeneric geminate grunt species are sequenced using RADseq to internally calibrate the genomic mutation rate within the *Anisotremus* genus. This substitution rate is then utilized to estimate time and predict most likely hypothesis of divergence that produced the disjunct population in the focal species. Subsequently, thousands of loci are used to characterize genome-wide gene flow and to compare patterns of genomic structure and signals of selection between sympatric and allopatric Baja California populations of the sargo and longjaw mudsucker. Finally, this study compares patterns of genomic differentiation between the allopatric populations and identifies coding genes diverging convergently in more than one of these temperate marine fishes with diverse ecological backgrounds. While each chapter is written and should be treated as an independent, stand-alone study, all three combined conform a useful framework to continue exploring the mechanisms in which isolation and selection shape the genome and drive adaptation in organisms.

References

Allen, L.G., Pondella, D.J. & Horn, M. h., 2006. *The Ecology of Marine Fishes: California and Adjacent Waters*, UC Press.

Altschul, S.F. et al., 1997. Gapped BLAST and PSI-BLAST: a new generation of protein database search programs. *Nucleic acids research*, 25(17), pp.3389–402.

Baird, N.A. et al., 2008. Rapid SNP discovery and genetic mapping using sequenced RAD markers. *PLoS ONE*, 3(10), pp.1–7.

Bernardi, G., 2013. Speciation in fishes. *Molecular Ecology*, 22(22), pp.5487–5502.

Bernardi, G., Findley, L. & Rocha-Olivares, A., 2003. Vicariance and dispersal across Baja California in disjunct marine fish populations. *Evolution; international journal of organic evolution*, 57(7), pp.1599–1609.

Coyne, J.A. & Orr, H.A., 2004. *Speciation*, Sunderland, Massachusetts: Sinauer Associates.

Crow, K.D., Munehara, H. & Bernardi, G., 2010. Sympatric speciation in a genus of marine reef fishes. *Molecular Ecology*, 19(10), pp.2089–2105.

Darwin, C., 1859. *On the Origin of Species by Means of Natural Selection, Or, the Preservation of Favoured Races in the Struggle for Life*, London.

Gaither, M.R. et al., 2015. Genomic signatures of geographic isolation and natural selection in coral reef fishes. *Molecular Ecology*, 24(7), pp.1543–1557.

Hohenlohe, P.A. et al., 2010. Population genomics of parallel adaptation in threespine

stickleback using sequenced RAD tags. *PLoS Genetics*, 6(2).

Medina, M. & Walsh, P.J., 2000. Comparison of Four Mendelian Loci of the California Sea Hare (Aplysia californica) from Populations of the Coast of California and the Sea of Cortez. *Marine Biotechnology*, 2, pp.449–455.

Nosil, P., 2012. *Ecological speciation*, New York: Oxford University Press Inc.

Nosil, P., Funk, D.J. & Ortiz-Barrientos, D., 2009. Divergent selection and heterogeneous genomic divergence. *Molecular Ecology*, 18(3), pp.375–402.

Present, T.M.C., 1987. Genetic Differentiation of Disjunct Gulf of California and Pacific Outer Coast Populations of Hypsoblennius jenkinsi Genetic Differentiation of Disjunct Gulf of California and Pacific Outer Coast Populations of Hypsoblennius jenkinsi. *Copeia*, 1987(4), pp.1010–1024.

Thomson, D.A., Findley, L.T. & Kersitch, A.N., 2000. *Reef fishes of the Sea of Cortez: the rocky shore fishes of the Gulf of California.*, Austin, TX: The University of Texas Press.

Tranah, G.J. & Allen, L.G., 1999. *Morphologic and genetic variation among six populations of the spotted sand bass, Paralabrax maculatofasciatus, from southern California to the upper Sea of Cortez*, Los Angeles, CA.

CHAPTER 1

Tempo and Mode of Divergence in Baja California Disjunct Fishes Based on a Genome-Wide Calibrated Substitution Rate.

Abstract

The distribution of the Baja California disjunct species, where populations in the Sea of Cortez occurred isolated from the Pacific, was a product of multiple vicariant and dispersal events. This study first produces a time calibrated genome-wide substitution rate from thousands of RADseq markers by genotyping two trans-Isthmian *Anisotremus* geminate fish clades. Then, this rate is utilized to estimate divergence time of disjunct populations of the congeneric sargo, *Anisotremus davidsonii,* and three other Baja California disjunct fishes: the longjaw mudsucker, *Gillichthys mirabilis*, the California sheephead, *Semicossyphus pulcher,* and the zebraperch, *Kyphosus azureus*. Finally, produced dates of divergence are compared to previous estimates and used to predict the most likely hypothesis of isolation in each species. Observed genomic substitution rate was faster, and suggested narrower and younger ranges of divergence time than previous estimates based on individual markers. No evidence of isolation was found for the sheephead. Estimated dates of divergence were 60, 284, and 23 thousand years, for the sargo, mudsucker, and zebraperch, respectively. A dispersal episode where species migrated northwards and then became isolated by the warming of the seawater temperate in the south of the Baja California peninsula at the end of the last glaciation period, and a vicariant

isolation resulting from the closure of a mid-peninsular seaway, are the plausible historical events that produced the current disjunction in the distributions of these species.

Keywords: allopatric populations, dispersal, isolation, speciation, vicariance.

Introduction

Disjunct populations are formed after a physical or other type of barrier separates the populations of a species (Palumbi 1992), a process that might lead to allopatric speciation events and can be used to study early drivers of evolution (Mayr 1942; Endler 1977; Coyne & Orr 2004; Bernardi 2013). Compared to terrestrial or fresh water systems, barriers to gene flow in marine systems are less obvious but do form in the way of geographical separation as well as differences in oceanographic features such as current regimes, temperature, and salinity (Helfman et al. 2009; Bernardi 2013; Rocha et al. 2002). The Baja California Peninsula in Northeastern Mexico, offers a remarkable opportunity to examine natural cases of population disjunctions including several terrestrial organisms (Upton & Murphy 1997; B. R. Riddle et al. 2000; Brett R. Riddle et al. 2000; B R Riddle et al. 2000) as well as marine invertebrates (Medina & Walsh 2000), mammals (Maldonado et al. 1995) and fishes (Miller & Lea 1972; Thomson et al. 2000; Bernardi & Lape 2005; Bernardi et al. 2003; Present 1987; Huang & Bernardi 2001; Stepien et al. 2001). Among these, fishes present diverse mechanisms of isolation discussed next.

There are 19 temperate marine fishes that inhabit the Pacific coast of California and the Baja California and also have populations in the northern and central portions of the Sea of Cortez (also called Gulf of California and here also referred simply as the "Gulf"), where populations are thought to be isolated in part by the higher seawater temperature at the south of the Baja California peninsula (where these species are absent, Bernardi et al., 2003; Bernardi and Lape, 2005; Huang and

Bernardi, 2001; Miller and Lea, 1972; Present, 1987; Stepien et al., 2001; Thomson et al., 2000, Supplementary Table 1, Supplementary Figure 1).

Three hypotheses lead the efforts to explain the vicariant and dispersal events that produced these disjunct distributions. Approximately 5 millions years ago (Mya), the San Andres fault began separating the Baja California peninsula from the mainland until creating the Sea of Cortez an island within (Grismer 2000; Bernardi 2014; B. R. Riddle et al. 2000). During this process, a seaway in the north of the peninsula and another one in the south (near the current location of La Paz), connected the Pacific Ocean and the Gulf but were uplifted and closed approximately 3 Mya (Bernardi and Lape, 2005; Riddle et al., 2000). This event is here referred as the "ancient seaway divergence hypothesis." An additional seaway was later created and uplifted approximately 1 to 1.6 Mya in the middle of the peninsula (Riddle et al. 2000). This closure represents a second possible divergent scenario referred by this study as the "Mid-peninsular seaway divergence hypothesis." Finally, a third possibly explanation, here refereed as the "Post-glacial divergence hypothesis," predicts that after the closure of these seaways, it remained possible for temperate fishes to migrate in cold water around the peninsula until the last glaciation period ended approximately 12,000 years ago (ya) and seawater temperature was increased in the southern portion of the peninsula (Bernardi et al., 2003; Brusca, 1973; Riddle et al., 2000).

Genetic differences in a handful of markers between Pacific and Gulf populations of several disjunct species have been linked to these hypotheses before

(Bernardi & Lape 2005; Huang & Bernardi 2001; Bernardi et al. 2003; Stepien et al. 2001; Thomson et al. 2000; Miller & Lea 1972; Bernardi 2014; Terry et al. 2000; Brusca 1973; Present 1987; Poortvliet et al. 2013). The most comprehensive study analyzed the genetic patterns of 12 disjunct species and found deep divergence across the Pacific and Gulf populations of 8 species but high gene flow levels for the other 4 (Bernardi et al. 2003). Yet, these patterns might change if populations are examined with an increased number of loci. Here, thousands of markers are used to assess divergence time between disjunct populations of two species with some of the highest observed divergence, the sargo, *Anisotremus davidsonii* (ADA) and the longjaw mudsucker, *Gillichthys mirabilis* (GMI), and two species that previously showed little or no divergence, the California sheephead, *Semicossyphus pulcher* (SPU) and the zebraperch, *Kyphosus azureus* (KAZ).

The Pacific and Gulf populations of the sargo and longjaw mudsucker were proposed to have diverged from 0.16 to 0.64 Mya (based on divergence and coalescence analysis using mtDNA cyt-b and nDNA S7) by the post-glacial warming of the seawater temperature and from 0.76 to 2.3 Mya (based on mtDNA cyt-b) by the closure of the mid-peninsular seaway, respectively (Huang & Bernardi 2001; Bernardi et al. 2003; Bernardi & Lape 2005). However, given the large range of dates for separation, other hypotheses of isolation were not ruled out but considered less parsimonious. In the case of the California sheephead and the zebraperch, previous studies found little genetic evidence of isolation and estimated high gene flow across the peninsula (Bernardi et al. 2003; Poortvliet et al. 2013). Yet again, as the utilized

markers in these studies might lack sufficient resolution to detect a very recent cessation to gene flow and as mutation rates vary among different species and loci, different patterns of divergence are likely to emerged by examining many more loci (Avise 2000; Bernardi et al. 2003; Coyne & Orr 2004).

This study estimates internally-calibrated times for the divergence in the focal species by using two pairs of *Anisotremus* geminates to obtained an average rate of substitution from thousands of genome-wide RADseq loci. Geminate taxa refer to organisms that had some populations separated by the uplifting of the Isthmus of Panama and speciated into sister species in the Pacific and Atlantic Oceans (Jordan 1908). As a closure of the Isthmus, at approximately 3 Mya, has been heavily supported (O'Dea et al. 2016; Coates & Obando 1996; Collins 1996; Craig et al. 2004; Lessios 2008; Marko 2002; Thomson et al. 2000; Knowlton et al. 2003), trans-Isthmian geminate lineages have been used regularly to calibrate the time of population divergence, speciation events, and phylogeographic patterns in fishes (Bernardi & Lape 2005; Tavera et al. 2012; Tariel et al. 2016; Campbell et al. 2018; Thacker 2017; Bowen et al. 2013; Helfman et al. 2009). While examining the utility of RADseq to estimate dates of separation, we test if analyzing disjunct populations with large number of loci might uncover unseen divergence in sheephead and zebraperch, and estimate different or smaller ranges for the time of disjunction in sargo and mudsucker. Finally, we investigate the phylogeographic history of these species by matching produced estimates of divergence time to the mentioned hypotheses of isolation: post-glacial, mid-peninsular, or ancient seaways divergence.

Materials and Methods

Sampling

Samples were collected from Northeastern Pacific (NEP) and Gulf populations of the four focal Baja California disjunct species and from the Tropical Eastern Pacific (TEP) and the Western Atlantic (WA) for two pairs of geminate species (Table 1, see supplementary Table 2 and 3 for specific locations). Minnow traps were used to collect mudsuckers. Zebraperch and all *Anisotremus* samples were collected by spearing or by hand netting juveniles in tide pools. California sheephead individuals were speared or obtained from local fisherman.

Table 1. Number of samples collected per species and region.

Baja California disjunct species	Northern Eastern Pacific (NEP)	Sea of Cortez (Gulf)
Anisotremus davidsonii (ADA)	15	22
Gillichthys mirabilis (GMI)	34	21
Semicossyphus pulcher (SPU)	16	10
Kyphosus azureus (KAZ)	10	22
Geminate species	Tropical Eastern Pacific (TEP)	Western Atlantic (WA)
Anisotremus interruptus (AIN)	3	
Anisotremus taeniatus (ATA)	3	
Anisotremus surinamensis (ASU)		3
Anisotremus virginicus (AVI)		3

DNA Extraction and Library Preparation

Genomic DNA was isolated from fin, muscle, or gill, following the Qiagen DNeasy 96 Tissue Kit for purification of DNA recommendations (QIAGEN,

Valencia, California, USA). A single RAD library was constructed with the original protocol and using SBf1 for digestion, sonication for fragmentation, and magnetic beads for size selection (Miller et al. 2007; Baird et al. 2008). Individual barcodes were ligated to samples which were then sequenced at UC Berkeley (Vincent J. Coates Genomics Sequencing Laboratory) on a single illumina HiSeq 2000 lane.

Filtering and Discovering Markers

Sample de-multiplexing, and loci and Single Nucleotide Polymorphisms (SNPs) filtering and discovering, were performed using modified Perl scripts (Miller et al. 2012) and STACKS version 2 (Catchen et al. 2011; Catchen et al. 2013). Only reads with quality scores of 90% or higher (Phred score=33) and containing exact matches to the SBf1 restriction site were maintained for the analysis. Final sequence length was 80-bp after quality filters and the removal of restriction site and barcodes.

In a single *de novo* map in STACKS 2 for all five species of *Anisotremus,* stacks (suspected orthologous loci) were created if they showed a minimum depth of three (-m) and a maximum of three mismatches per loci in each sample (-M). Given that this analysis contained five different (albeit closely related) species, we allowed a maximum of 7 mismatches between catalog loci (-n). Separate analyses were performed for the other three species with an increased depth of 8x (-m), the same number of mismatches between loci in the individuals (-M) but allowing only two differences when building catalogs (-n). Subsequent *Anisotremus* data consisted of only loci that were present in all geminate individuals and at the same time, in least

80 percent of the individuals from the Pacific and Gulf sargo populations (*populations* –p 6 –r 0.8). Similarly, loci in the other species must have been present in at least 80 percent of the Pacific and Gulf samples to be considered for downstream analyses (*populations* –p 2 –r 0.8).

Estimating Time of Disjunction

The first analysis consisted of calculating the calibrated rate of substitution in *Anisotremus* RADseq loci. We accomplished this by using the closure of the Isthmus of Panama as the divergence time for the geminates *A. taeniatus / A. virginicus* and for *A. interruptus / A. surinamensis.* Including all five *Anisotremus* species in the same analysis permitted calculating this rate simultaneously with estimating the time of the divergence between sargo disjunct populations. For this, sequence alignments were constructed to build a Bayesian phylogeny with BEAST v. 2 (Bouckaert et al. 2014) while using a General Time Reversible (GTR) substitution model in combination with a lognormal relaxed clock-model and a birth-death model for priors. To allow for direct comparison to previous divergence estimates (Bernardi & Lape 2005), a prior divergence range of 3.1 to 3.5 My was specified using a normal distribution in both pairs of geminates. 10 billion generations were run with tree sampling every million.

Alignment were produced by converting a phylip output file from STACKS maintaining all the individuals into nexus format in Mesquite (Maddison & Maddison 2018). This nexus file was then converted into a BEAST .xml input file using

mentioned parameters in BEAUti 2. Output from the Markov chain Monte Carlo and ESS values were monitored using Tracer v.1.7 and a maximum clade credibility time tree with median node heights was selected using TreeAnnotator v.2.5 (Bouckaert et al. 2014). This tree was then visualized and modified in FigTree .1.4.2 (Bouckaert et al. 2014).

Subsequently, the *Anisotremus* RAD loci substitution rate was used to estimate the divergence between the disjunct populations of the other species. In order to do this, we first calculated the average number of differences between Pacific and Gulf population of all disjunct species. The genepop output file from STACKS was converted into Arlequin format using PGDSpider (Lischer & Excoffier 2012) and the corrected average pairwise difference between populations (which also accounts for differences within populations) was calculated in Arlequin v.3.5.1.2 (Excoffier & Lischer 2010). The *Anisotremus* RAD loci rate, as well as previously published Cyt-b mutation rates for the same geminate pairs (Bernardi & Lape 2005), were then applied to back calculate time estimates of divergence on the other species lacking external sources of calibration.

Results and discussion

Markers and Single Nucleotide Polymorphisms

The number of orthologous loci passing all quality and population filters range from 4,232 to 9,340 per species (Figure 2). Single nucleotide polymorphisms (SNPs) from these loci were used to estimate the time in which disjunct populations

became isolated. The alignment used to estimate calibrated Anisotremus substitution rates and the divergence in sargo populations had a length of 14,309 polymorphic characters. Two *A. virginicus* individuals were dropped from the analysis as they showed very low coverage in early *de novo* analyses.

Genomic Rate of Evolution

Misleading relationships can be observed between species if taxon sampling is low when analyzing a group of organisms (Lessios 2008; Hamilton et al. 2017). In our case, a phylogeny of the entire *Anisotremus* genus confirms that the species within our geminate clades are indeed sister taxa (Bernardi et al. 2008). Furthermore, given the cooling conditions already happening at the northern and southern latitudes during and after the closure of the Isthmus, it is unlikely that these species maintained gene flow by migrating around the continents (Nof & Van Gorder 2003; Hodell & Warnke 1991; Lessios 2008). In this regard, our estimated rate of evolution is obtained with a degree of confidence. However, specifying times of divergence to geminate clades is not a trivial process that can affect estimated rates. We encountered a considerable range of dates for the closure of the Isthmus (approximately 2.8 to 3.5 Mya) even when we considered only the traditionally accepted dates (O'Dea et al. 2016; Coates & Obando 1996; Collins 1996; Craig et al. 2004; Lessios 2008; Marko 2002; Thomson et al. 2000; Knowlton et al. 2003). As the principal goal of this study was to compare the produced RADseq loci substitution rate to previous estimates based on individuals markers (Bernardi & Lape 2005), our

BEAST analysis was performed indicating both geminate clades as priors with a possible divergence from 3.1 to 3.5 My. The *Anisotremus* average genomic substitution rate from was 0.0088 mutations per million years. In turn, the Bayesian phylogeny predicted a divergence of 60 thousand years (ky) for the sargo disjunct populations (ESS values were high and stable. 95% Highest Posterior Density, HPD, 0–220 ky) (Table 2, Figure 1).

Table 2. Divergence time estimates based on internally-calibrated *Anisotremus* RAD loci substitution rate (3.1-3.5 My geminate priors) and observed average distance between disjunct populations (% Diff.). A substitution rate of 8.8 μ (10^{-3}) was estimated in this study and the median (Est. Div.) and range of divergence (Est. Range of Div.) were calculated for sargo. Bernardi and Lape (2005) estimated the range of divergence in Cyt-b for the sargo by applying a rate of 1.7-1.5 μ (10^{-8}) to *A. virginicus / A. taeniatus* and a rate of 0.9-1.0 μ (10^{-8}) to *A. surinamensis / A. interruptus*. These rates are applied to the observed number of differences in RAD loci to back calculate divergence in the other species. The lower end of divergence for sargo was zero (low end of HPD) thus preventing this calculation in the other species.

	RAD loci rate				Cyt- b rate
	Loci	% Diff.	Est. Div.	Est. Range of Div.	Est. Range of Div.
Sargo	5915	19.54	60 ky	up to 220 ky	330 - 641 ky
Longjaw mudsucker	8468	92.56	284 ky	up to 1 My	1.56 - 3.04 My
Zebraperch	4232	7.58	23 ky	up to 90 ky	130 - 250 ky
California sheephead	9340	n/a	n/a	n/a	n/a

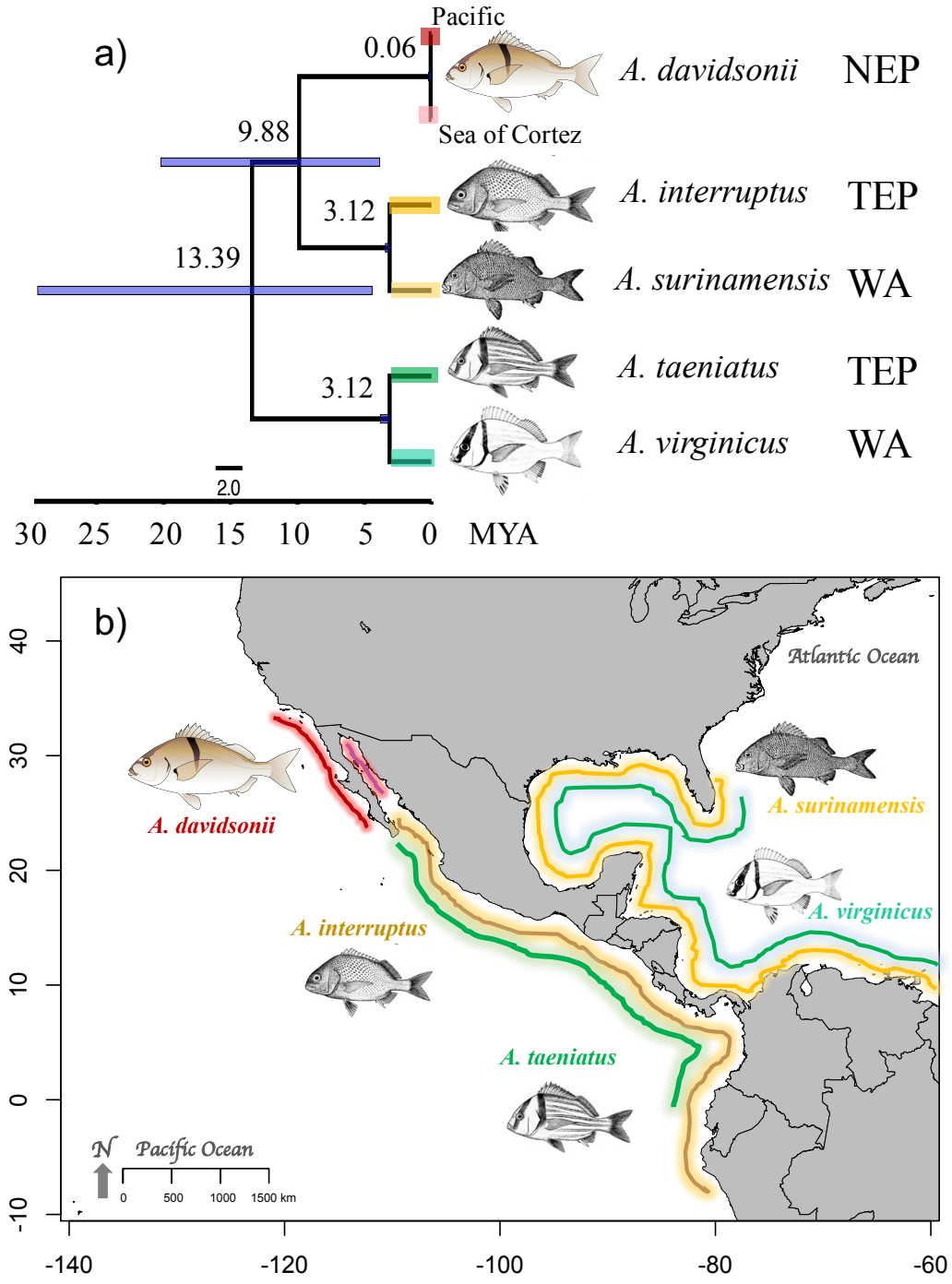


Figure 1. a) Time-calibrated phylogeny of studied *Anisotremus* species using both geminate clades as priors with pre-set divergence time from 3.1 to 3.5 My. Numbers specify the estimated date of node divergence (in million years). Blue bars represent the 95% Highest Posterior Densities. NEP, Northeastern Pacific; TEP, tropical Eastern Pacific; WA, Western Atlantic. b) Species distributions.

Time and Mode of Divergence in Disjunct Populations

Considering the dynamic evolutionary histories of these species, our genomic substitution rate produced dates of divergence somewhat comparable with previous estimates (Table 2). Overall, our time estimates are younger as a result of the surprisingly faster average rate of evolution observed in produced RAD loci than in mitochondrial Cyt-b (Bernardi & Lape 2005). Our estimates also have smaller ranges of variation.

As in other studies (Poortvliet et al. 2013; Bernardi et al. 2003), evidence of isolation was not found between disjunct populations of sheephead. The variation seen between populations within the Gulf and Pacific was higher than between the two regions. This produced a negative corrected distance between sheephead disjunct populations which prevented further calculations and rose questions about its classification as a disjunct species. In contrast, this study revealed previously unseen divergence between disjunct populations of zebraperch. As expected, zebraperch divergence was young and shallow using any of the three substitution rates (Table 2). This divergence time was considerable far from the one million mark leaving the post-glacial divergence hypothesis has the only plausible explanation for the observed divergence.

Results for the longjaw mudsucker are also very interesting as well. To recall, divergence was previously estimated to range from 0.76 to 2.3 My (Huang & Bernardi 2001). Authors attributed this divergence to a possible vicariant event via the closure of the mid-peninsular seaway but given the large variation in these

estimates, a post-glacial separation was not ruled out but considered less parsimonious. When applying the mutation rates from Bernardi and Lape (2005) to the average distance in RAD loci between disjunct populations, produced estimates ranged from approximately 1.5 to 3 My. While, this range opens the door to a possible ancient divergence, the RAD loci estimates are more in tune with either an early dispersal and isolation event or vicariance produced by the mid-peninsular seaway hypothesis. The probability of the latter appears higher as that it occurs in the middle of combined time estimates.

In the case of the sargo, previous estimates of divergence ranged from 330 to 641 ky (or from 160 to 641 ky if coalescence analyses are included). This was considered to be the result of dispersal around the peninsula or the closure of an additional seaway that has not been discovered yet (Bernardi & Lape 2005). Our analysis produced considerably younger estimates (up to 220 ky, Table 2) cementing the idea of sargo populations migrating north and later becoming isolated by the warming of the sea water temperature in the southern tip of the peninsula. Although it is outside the focus of this study, it is worth noting that the produced phylogeny suggests that the separation of sargo and the geminate clade of *A. surinamensis* and *A. interruptus* occurred before the closure of the Isthmus.

Potential Sources of Discordance

In order to allow direct comparisons with previous estimates, this study assumed a divergence time of 3.1 to 3.8 My for both *Anisotremus* geminate clades.

However, previous studies have highlighted the possibility of at least one of these trans-Isthmian pairs to have diverged at a different time than at the closure of the Isthmus, as sequence divergence in the clade conformed by *A. taeniatus and A. virginicus,* was found to be consistently higher than that between *A. surinamensis* and *A. interruptus* in multiple markers (Bernardi & Lape 2005; Lessios 2008; Bernardi et al. 2008; Bermingham et al. 1997). This can be the result of either different rates of evolution or different divergence time in the two clades, or a combination of both. Evidence of differential substitution rates (such as very different branch lengths) in the two clades is not present in previous phylogenies (Bernardi & Lape 2005; Bernardi et al. 2008). Alternately, the different dates of divergence might be a result of differences in ecology such as the usage of mangrove and reef habitat. For example, *Anisotremus taeniatus* and *A. virginicus* could have separated before the closure of the Isthmus if they were more dependent on reef habitat that might have disappeared in the Isthmus area as this became shallower (Thomson et al. 2000; Lessios 2008; Froese & Pauly n.d.; Knowlton et al. 2003). Finally, a vicariance date of 2 my has also been considered for *A. surinamensis* and *A. interruptus,* based on a proposed Pleistocene breach of the Isthmus (Bernardi & Lape 2005). When using these two pairs of geminates and traditional dates for the closure of the Isthmus of Panama, indicating older dates of divergence for *A. taeniatus* / *A. virginicus* and younger for *A. surinamensis* / *A. interruptus,* might help finding the upper and lower limits of possible divergence times in disjunct populations.

An additionally potential issue when estimating divergence dates using RADseq is that produced loci, and associated rates of substitution, are likely not the same in different species. However, while different rates of change among loci are unequivocal, a consensus rate might be reachable purely giving the large numbers of loci produced by this technology. If other analyses of geminate lineages produce similar RAD loci substitution rates, estimates like ours might be widely applicable to date divergence, and other evolutionary events, in systems without external sources of temporal calibration. Time calibration analyses include many variables, which require iterative processing and adjusting, our analysis was a comparison of estimates given fixed variables. Futures analyses will include diverse dates for geminates species, different models of evolution, and estimating dates of divergence using only suspected neutral RAD loci.

Conclusion

Unlike the sheephead, which along with its congenerics show anti-tropical distributions, ancestral lineages of the sargo, zebraperch, and mudsucker, were likely present in tropical waters in the Pacific and Atlantic oceans. Genetic evidence from this is other studies, as well as the current distributions of related species in the same genera and families, are concordant with a phylogeographic scenario in which these species migrated northwards from lower latitudes (Bernardi et al. 2003; Bernardi & Lape 2005; Bernardi et al. 2008; Tavera et al. 2012; Thomson et al. 2000; Froese & Pauly n.d.; Helfman et al. 2009; Huang & Bernardi 2001). These species might have

been peripheral populations at the edge as their physiological and adaptive capabilities that accumulated genetic differences from divergent selective pressures until eventually speciated as a result of the invasion to new environments (Rocha et al. 2005; Coyne & Orr 2004). Disjunct pupations in these species were likely isolated by either the closure of a mid-peninsular seaway approximately 1 Mya, or by the retreat of the ocean thermocline into higher latitudes by the end of the last glaciation, in combination with an earlier migration of these species into the Pacific coast of Baja California. While this scenario is probably true for other disjunct species as well, it is definitively not universal within Baja California disjuncts. For instance, the Mexican rockfish, *Sebastes macdonaldi,* and the pink surfperch, *Zalembius rosaceus,* evolved from two spectacular Northeastern Pacific species radiations. The Baja California Disjunct species continues to offer a remarkable framework where to test hypothesis of species biogeography and study the early mechanisms of speciation.

References

Avise, J.C., 2000. Phylogeography: The History and Formation of Species. Harvard University Press, Cambridge, MA.

Baird, N.A., Etter, P.D., Atwood, T.S., Currey, M.C., Shiver, A.L., Lewis, Z.A., Selker, E.U., Cresko, W.A., Johnson, E.A., 2008. Rapid SNP discovery and genetic mapping using sequenced RAD markers. PLoS One 3, 1–7. https://doi.org/10.1371/journal.pone.0003376

Bermingham, E., McCafferty, S.S., Martin, A.P., 1997. Fish biogeography and molecular clocks: perspectives from the Panamian Isthmus, in: T.D., K., Stepien, C.A. (Eds.), Molecular Systematics of Fishes. Academic Press Inc., San Diego, CA, pp. 113–128.

Bernardi, G., 2014. Baja California disjunctions and phylogeographic patterns in sympatric California blennies. Front. Ecol. Evol. 2, 1–9. https://doi.org/10.3389/fevo.2014.00053

Bernardi, G., 2013. Speciation in fishes. Mol. Ecol. 22, 5487–5502. https://doi.org/10.1111/mec.12494

Bernardi, G., Alva-Campbell, Y.R., Gasparini, J.L., Floeter, S.R., 2008. Molecular ecology, speciation, and evolution of the reef fish genus Anisotremus. Mol. Phylogenet. Evol. 48, 929–35. https://doi.org/10.1016/j.ympev.2008.05.011

Bernardi, G., Findley, L., Rocha-Olivares, A., 2003. Vicariance and dispersal across Baja California in disjunct marine fish populations. Evolution 57, 1599–1609. https://doi.org/10.1554/02-669

Bernardi, G., Lape, J., 2005. Tempo and mode of speciation in the Baja California disjunct fish species Anisotremus davidsonii. Mol. Ecol. 14, 4085–4096. https://doi.org/10.1111/j.1365-294X.2005.02729.x

Bouckaert, R., Heled, J., Kühnert, D., Vaughan, T., Wu, C.H., Xie, D., Suchard, M.A., Rambaut, A., Drummond, A.J., 2014. BEAST 2: A Software Platform for Bayesian Evolutionary Analysis. PLoS Comput. Biol. 10, 1–6. https://doi.org/10.1371/journal.pcbi.1003537

Bowen, B.W., Rocha, L.A., Toonen, R.J., Karl, S.A., 2013. The origins of tropical marine biodiversity. Trends Ecol. Evol. 28, 317–376.

Brusca, R.C., 1973. A handbook to the common intertidal invertebrates of the Gulf of California. University of Arizona Press, Tucson, AZ.

Campbell, M.A., Robertson, D.R., Vargas, M.I., Allen, G.R., McMillan, W.O., 2018. Multilocus molecular systematics of the circumtropical reef-fish genus *Abudefduf* (Pomacentridae): history, geography and ecology of speciation. PeerJ 6, e5357. https://doi.org/10.7717/peerj.5357

Catchen, J., Hohenlohe, P.A., Bassham, S., Amores, A., Cresko, W.A., 2013. Stacks: An analysis tool set for population genomics. Mol. Ecol. 22, 3124–3140. https://doi.org/10.1111/mec.12354

Catchen, J.M., Amores, A., Hohenlohe, P., Cresko, W., Postlethwait, J.H., 2011. *Stacks* : Building and Genotyping Loci *De Novo* From Short-Read Sequences. G3: Genes|Genomes|Genetics 1, 171–182.

https://doi.org/10.1534/g3.111.000240

Coates, A., Obando, J., 1996. The geologic evolution of the Central American isthmus, in: Jackson, J., Budd, A., Coates, A. (Eds.), Evolution and Environment in Tropical America. University of Chicago Press, Chicago, pp. 21–56.

Collins, T., 1996. Molecular comparisons of transisthmian species pairs: rates and patterns of evolution., in: Jackson, J.B.C., Budd, A.F., Coates, A. (Eds.), Evolution and Environment in Tropical America. University of Chicago Press, Chicago, pp. 303–334.

Coyne, J.A., Orr, H.A., 2004. Speciation. Sinauer Associates, Sunderland, Massachusetts.

Craig, M.T., Hastings, P.A., Pondella, D.J., 2004. Speciation in the Central American Seaway: The importance of taxon sampling in the identification of trans-isthmian geminate pairs. J. Biogeogr. 31, 1085–1091. https://doi.org/10.1111/j.1365-2699.2004.01035.x

Endler, J.A., 1977. Geographic Variation, Speciation and Clines. Princeton University Press.

Excoffier, L., Lischer, H.E.L., 2010. Arlequin suite ver 3.5: A new series of programs to perform population genetics analyses under Linux and Windows. Mol. Ecol. Resour. 10, 564–567. https://doi.org/10.1111/j.1755-0998.2010.02847.x

Froese, R., Pauly, D., n.d. Fishbase [WWW Document]. URL

http://www.fishbase.org (accessed 5.5.16).

Grismer, L.L., 2000. Evolutionary biogeography on Mexico's Baja California peninsula: A synthesis of molecules and historical geology. Proc. Natl. Acad. Sci. 97, 14017–14018. https://doi.org/10.1073/pnas.260509697

Hamilton, H., Saarman, N., Short, G., Sellas, A.B., Moore, B., Hoang, T., Grace, C.L., Gomon, M., Crow, K., Brian Simison, W., 2017. Molecular phylogeny and patterns of diversification in syngnathid fishes. Mol. Phylogenet. Evol. 107, 388–403. https://doi.org/10.1016/j.ympev.2016.10.003

Helfman, G.S., Collette, B.B., Facey, D.E., Bowen, B.W., 2009. The diversity of fishes, Second. ed. Wiley-Blackwell.

Hodell, D.A., Warnke, D.A., 1991. Climatic evolution of the Southern Ocean during the Pliocene epoch from 4.8 to 2.6 million years ago. Quat. Sci. Rev. 10, 205–214. https://doi.org/10.1016/0277-3791(91)90019-Q

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. Mar. Biol. 138, 421–428.

Jordan, D.S., 1908. The law of geminate species. Am. Nat. 42, 73–80.

Knowlton, N., Weight, L. a, Solorzano, L.A., Mills, D.K., Bermingham, E., 1993. Divergence in Proteins, Mitochondria! DNA, and Reproductive Compatibility Across the Isthmus of Panama. Science (80-.). 260, 1629–1632.

Lessios, H.A., 2008. The Great American Schism: Divergence of Marine Organisms After the Rise of the Central American Isthmus. Annu. Rev. Ecol. Evol. Syst. 39, 63–91. https://doi.org/10.1146/annurev.ecolsys.38.091206.095815

Lischer, H.E.L., Excoffier, L., 2012. PGDSpider: An automated data conversion tool for connecting population genetics and genomics programs. Bioinformatics 28, 298–299. https://doi.org/10.1093/bioinformatics/btr642

Maddison, W.P., Maddison, D.R., 2018. Mesquite: a modular system for evolutionary analysis.

Maldonado, J.E., Davila, F.O., Stewart, B.S., Geffen, E., 1995. Intraspecific genetic differentiation in California sea lions (Zalophus californianus) from Southern California and the Gulf of California. Mar. mammal Sci. 11, 46–58.

Marko, P.B., 2002. Fossil calibration of molecular clocks and the divergence times of geminate species pairs separated by the Isthmus of Panama. Mol. Biol. Evol. 19, 2005–2021. https://doi.org/10.1093/oxfordjournals.molbev.a004024

Mayr, E., 1942. Systematics and the Origin of Species from the View- point of a Zoologist. Columbbia University Press, New York, New York.

Medina, M., Walsh, P.J., 2000. Comparison of Four Mendelian Loci of the California Sea Hare (Aplysia californica) from Populations of the Coast of California and the Sea of Cortez. Mar. Biotechnol. 2, 449–455. https://doi.org/10.1007/s101260000020

Miller, D.J., Lea, R.N., 1972. Guide to the coastal marine fishes of California. Berkeley, CA.

Miller, M., Dunham, J., Amores, a, Cresko, W., Johnson, E., 2007. genotyping using restriction site associated DNA (RAD) markers. Genome Res. 17, 240–248. https://doi.org/10.1101/gr.5681207

Miller, M.R., Brunelli, J.P., Wheeler, P.A., Liu, S., Rexroad, C.E., Palti, Y., Doe, C.Q., Thorgaard, G.H., 2012. A conserved haplotype controls parallel adaptation in geographically distant salmonid populations. Mol. Ecol. 21, 237–249. https://doi.org/10.1111/j.1365-294X.2011.05305.x

Nof, D., Van Gorder, S., 2003. Did an open Panama Isthmus correspond to an invasion of Pacific water into the Atlantic? J. Phys. Ocean 33, 1324–1336. https://doi.org/10.1175/1520-0485(2003)033<1324:DAOPIC>2.0.CO;2

O'Dea, A., Lessios, H.A., Coates, A.G., Eytan, R.I., Restrepo-Moreno, S.A., Cione, A.L., Collins, L.S., De Queiroz, A., Farris, D.W., Norris, R.D., Stallard, R.F., Woodburne, M.O., Aguilera, O., Aubry, M.P., Berggren, W.A., Budd, A.F., Cozzuol, M.A., Coppard, S.E., Duque-Caro, H., Finnegan, S., Gasparini, G.M., Grossman, E.L., Johnson, K.G., Keigwin, L.D., Knowlton, N., Leigh, E.G., Leonard-Pingel, J.S., Marko, P.B., Pyenson, N.D., Rachello-Dolmen, P.G., Soibelzon, E., Soibelzon, L., Todd, J.A., Vermeij, G.J., Jackson, J.B.C., 2016. Formation of the Isthmus of Panama. Sci. Adv. 2, 1–12. https://doi.org/10.1126/sciadv.1600883

Palumbi, S.R., 1992. Marine speciation on a small planet. Trends Ecol. Evol. 7, 114–118.

Poortvliet, M., Longo, G.C., Selkoe, K., Barber, P.H., White, C., Caselle, J.E., Perez-Matus, A., Gaines, S.D., Bernardi, G., 2013. Phylogeography of the California sheephead, Semicossyphus pulcher: The role of deep reefs as stepping stones and pathways to antitropicality. Ecol. Evol. 3, 4558–4571. https://doi.org/10.1002/ece3.840

Present, T.M.C., 1987. Genetic Differentiation of Disjunct Gulf of California and Pacific Outer Coast Populations of Hypsoblennius jenkinsi Genetic Differentiation of Disjunct Gulf of California and Pacific Outer Coast Populations of Hypsoblennius jenkinsi. Copeia 1987, 1010–1024.

Riddle, B.R., Hafner, D.J., Alexander, L.F., 2000. Comparative phylogeography of Baileys' pocket mouse (Chaetodipus baileyi) and the Peromyscus eremicus species group: historical vicariance of the Baja California Peninsular Desert. Mol. Phylogenet. Evol. 17, 161–72. https://doi.org/10.1006/mpev.2000.0842

Riddle, B.R., Hafner, D.J., Alexander, L.F., 2000. Phylogeography and systematics of the Peromyscus eremicus species group and the historical biogeography of North American warm regional deserts. Mol. Phylogenet. Evol. 17, 145–160. https://doi.org/10.1006/mpev.2000.0841

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. Proc. Natl.

Acad. Sci. 97, 14438–14443. https://doi.org/10.1073/pnas.250413397

Rocha, L.A., Bass, A.L., Robertson, D.R., Bowen, B.W., 2002. Adult habitat preferences, larval dispersal, and the comparative phylogeography of three Atlantic surgeonfishes (Teleostei: Acanthuridae). Mol. Ecol. 11, 243–252. https://doi.org/10.1046/j.0962-1083.2001.01431.x

Rocha, L.A., Robertson, D.R., Roman, J., Bowen, B.W., 2005. Ecological speciation in tropical reef fishes. Proc. R. Soc. B 272, 573–579. https://doi.org/10.1098/2004.3005

Stepien, C.A., Rosenblatt, R.H., Bargmeyer, B.A., 2001. Phylogeography of the Spotted Sand Bass , Paralabrax maculatofasciatus : Divergence of Gulf of California and Pacific Coast Populations. Evolution (N. Y). 55, 1852–1862.

Tariel, J., Longo, G.C., Bernardi, G., 2016. Tempo and mode of speciation in Holacanthus angelfishes based on RADseq markers. Mol. Phylogenet. Evol. 98, 84–88. https://doi.org/10.1016/j.ympev.2016.01.010

Tavera, J.J., Acero, A.P., Balart, E.F., Bernardi, G., 2012. Molecular phylogeny of grunts (Teleostei, Haemulidae), with an emphasis on the ecology, evolution, and speciation history of new world species. BMC Evol. Biol. 12, 57. https://doi.org/10.1186/1471-2148-12-57

Terry, A., Bucciarelli, G., Bernardi, G., 2000. Restricted Gene Flow and Incipient Speciation in Disjunct Pacific Ocean and Sea of Cortez Populations of a Reef Fish Species , Girella nigricans. Evolution (N. Y). 54, 652–659.

Thacker, C.E., 2017. Patterns of divergence in fish species separated by the Isthmus of Panama. BMC Evol. Biol. 17, 1–14. https://doi.org/10.1186/s12862-017-0957-4

Thomson, D.A., Findley, L.T., Kersitch, A.N., 2000. Reef fishes of the Sea of Cortez: the rocky shore fishes of the Gulf of California. The University of Texas Press, Austin, TX.

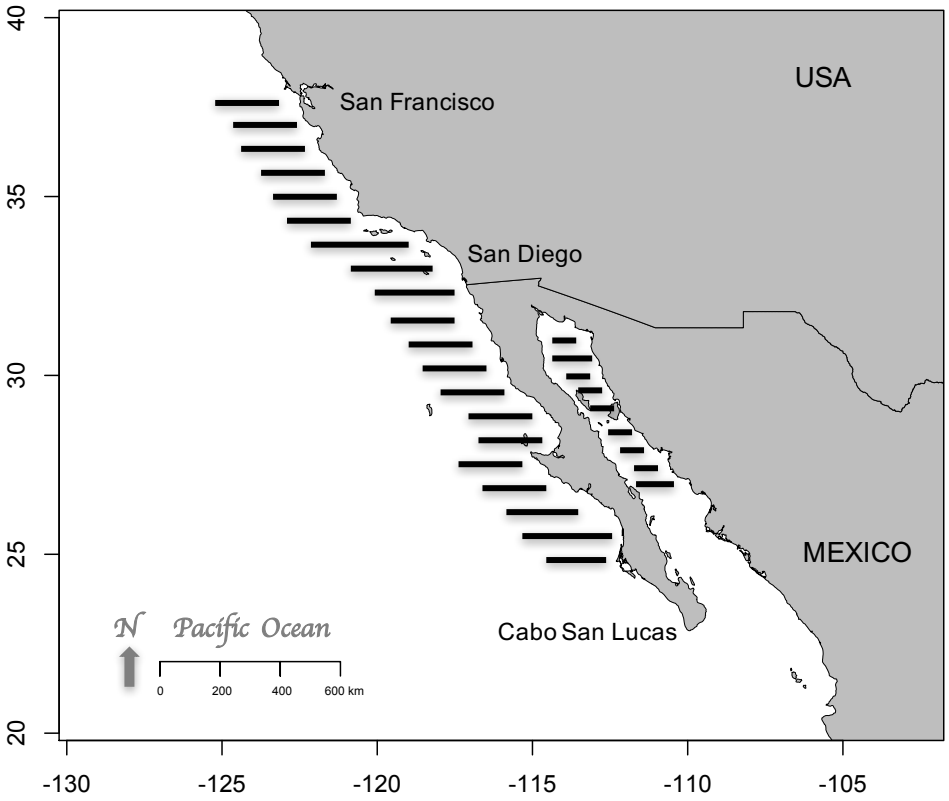
Upton, D.E., Murphy, R.W., 1997. Phylogeny of the Side-Blotched Lizards (Phrynosomatidae:Uta) Based on mtDNA Sequences: Support for a Midpeninsular Seaway in Baja California. Mol. Phylogenet. Evol. 8, 104–113. https://doi.org/10.1006/mpev.1996.0392

Supplementary Information:

Supplementary Table 1**.** List of the 19 Baja California disjunct fish species

(Bernardi et al. 2003).

Family	Species	Common name
Atherinidae	*Leuresthes tenuis/ L. sardina*	grunion
Girellidae	*Girella nigricans / G. simplicidens*	opaleye
Haemulidae	*Anisotremus davidsonii*	sargo
Blenniidae	*Hypsoblennius jenkinsi*	mussel blenny
Blenniidae	*Hypsoblennius gentilis*	bay blenny
Chaenopsidae	*Chaenopsis alepidota*	oragethroat pikeblenny
Serranidae	*Paralabrax maculatofasciatus*	spotted sand bass
Gobiidae	*Gillichthys mirabilis*	longjaw mudsucker
Gobiidae	*Lythrytpnus dalli*	blue banded boby
Kyphosidae	*Kyphosus azureus*	zebraperch
Labridae	*Halichoeres semicinctus*	rock wrasse
Labridae	*Semicossyphus pulcher*	California sheephead
Scorpaenidae	*Sebastes macdonaldi*	Mexican rockfish
Embiotocidae	*Zalembius rosaceus*	pink surfperch
Scorpaenidae	*Scorpaena guttata*	scorpionfish
Polyprionidae	*Stereolepis gigas*	giant seabass
Agonidae	*Xenetremus ritteri*	flagfin poacher
Pleuronectidae	*Hypsopsetta guttulata*	diamond turbot
Pleuronectidae	*Pleuronichthys verticalis*	hornyhead turbot



Supplementary Figure 1. Typical distribution of Baja California disjucnt fishes.

Supplementary Table 2. Sampling sites for the *Anisotremus* geminate species. *A. interruptus,* AIN; *A. taeniatus,* ATA; *A. surinamensis,* ASU; *A. virginicus,* AVI.

Site	AIN	ATA	ASU	AVI
Tropical Eastern Pacific				
Playa Troncoles, Costa Rica	3			
Puntarenas, Costa Rica		3		
Western Atlantic				
Florida, USA			3	
Guarapari Islands, Brazil				3

Supplementary Table 3. Pacific and Sea of Cortez (Gulf) sampling sites (ordered by latitude) and number of samples per species. SPU, *Semicossyphus pulcher*; KAZ, *Kyphosus azureus*; GMI, *Gillichthys mirabilis*; ADA, *Anisotremus davidsonii*.

Region	**Site**	SPU	KAZ	GMI	ADA
Pacific		**16**	**10**	**34**	**15**
	Monterey Bay (MB)			7	
	Goleta Slough (GOS)			10	
	Carpinteria Slough(CAS)			7	
	Catalina Island (CI)	5			2
	San Diego (SD)		10		7
	Guadalupe Island (GUA)	4			
	Guerrero Negro (GNE)			4	
	Punta Eugenia (PEU)				1
	Bahia Tortugas (TOR)	4			
	Punta San Roque (PSR)				5
	Estero El Coyote (ECO)			6	
	Puerto San Carlos (PSC)	3			
Gulf		**10**	**12**	**21**	**22**
	Estero La Choya (ELC)			10	
	Punta Choya (PC)	3			5
	Estero Morua (EMO)			5	
	Estero La Pinta (ELP)			6	
	Bahia Kino (BK)				12
	Bahia de los Angeles	5	12		5
	(BLA)	2			
	Los Frailes (LFR)				

CHAPTER 2

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*).

Abstract

The Pacific and Sea of Cortez allopatric populations of the sargo, *Anisotremus davidsonii,* (Haemulidae), and the longjaw mudsucker, *Gillichthys mirabilis,* (Gobiidae), are separated by the Baja California peninsula, and sympatric populations along the Northeastern Pacific cross established phylogeographic breaks that impact gene flow. Thus, these taxa offer an excellent multi-species framework where to study the mechanisms of divergence and signatures of selection under a gradient of isolation. Here, thousands of loci are genotyped from 48 sargos and 73 mudsuckers using RADseq, to characterize population connectivity, and investigate if sympatric and allopatric populations show unique genomic signatures of divergence, drift, or selection. Statistically significant divergence was seen across Point Conception in mudsucker (F_{ST}=0.15), Punta Eugenia in sargo (F_{ST}=0.02), and across the peninsula in both species (F_{ST}=0.11, and 0.23, in sargo and mudsucker, respectively). Structure across the peninsula in presumed neutral loci was highest in the mudsucker but both species showed evidence of strong selection. Larger number of presumed loci under selection were seen between allopatric populations (586-721 vs 13-224 between sympatric) and these were more differentiated (F_{ST}=0.36-1 vs 0.19-0.94 between

sympatric). Furthermore, the majority of outliers between sympatric populations accumulated at the lower end of the range of differentiation in contrast to a more even distribution between allopatric populations. This might be an indication of a prevalence of soft selective sweeps when sympatric populations are adapting to a mosaic of adaptive peaks and a more even contribution of soft and hard sweeps during adaptation to very distinct peaks in allopatry.

Keywords: adaptive evolution, Baja California, incipient speciation, differential selection, ecological divergence, selective sweeps, sympatric populations.

Introduction

Under the biological species concept, speciation is the process by which two or more populations accumulate sufficient genetic differences to reach reproductive isolation (Mayr 1942). In allopatry, divergence can be initiated by the appearance of a barrier, physically separating populations (vicariance), or by occasional dispersal events where individuals are able to cross an existing barrier which normally impedes open migration between sites (dispersal). In the absence of gene flow populations may diverge randomly through genetic drift acting on neutral loci or selectively by local adaptation changing the allele frequencies of the coding portions of the genome.

Alternatively, populations might also differentiate in sympatry (without a physical barrier), or close geographic proximity (Rocha et al. 2005), if they experience distinct selective pressures from different environments and local adaptation overwhelms the effects on gene flow (Rundle & Nosil 2005; Bernardi 2013; Coyne & Orr 2004). In contrast, high levels of gene flow may prevent the accumulation of advantageous alleles in populations when gene flow is higher than selective pressure (Coyne & Orr 2004; Sexton et al. 2014). Yet, the effects of gene flow are more complex than a simple unequivocal homogenization of genetic variation. In fact low levels of gene flow might be beneficial to local adaptation by providing new alleles from other populations that might impart higher local fitness (Sexton et al. 2014). Therefore, divergence is the outcome of the interplay between drift, selection, and the level of gene flow between populations. Comprehensive studies examining each of these processes can shed light on the evolutionary histories

of species and the similarities and differences between the way different modes of speciation operate in populations and allow them to persist and adapt to new environments.

Sympatric and Allopatric Populations of the Baja California Disjunct Fishes

Marine fishes with allopatric populations are scarce due to the paucity of absolute barriers to gene flow in the ocean (such as the isthmus of Panama) and the great dispersal potential of the pelagic larval stage of most fishes (Helfman et al. 2009; Bernardi 2013; Rocha et al. 2002). The Baja California disjunct species are a rare example of temperate marine fishes (19 species) in which populations have been isolated by specific vicariant and dispersal events (Bernardi et al. 2003; Brusca 1973; B. R. Riddle et al. 2000) (see Supplementary Table for the full list of species). These species count with populations that have no geographic barriers to dispersal throughout their Pacific distribution from central California to approximately Bahia Magdalena, in southern Baja California. Likewise, populations occur without obvious barriers in the northern and central zones of the Sea of Cortez (also known as Gulf of California and here simply referred as the Gulf). However, Gulf populations are separated from the Pacific by the Baja California peninsula and the warmer seawater temperatures at the southern point of the peninsula, where these species are either rare or absent (see Supplementary Figure S1 for a representation of the typical Baja California disjunct distribution).

Based on the respective presence or absence of geographic barriers, this study considers Gulf populations to be allopatric to the Pacific and populations within the Pacific and Gulf distributions to be sympatric. While disjunct populations are considered an early step toward allopatric speciation (Endler 1977), the absence of geographic barriers in the Northeastern Pacific does not intrinsically translate into panmixia as gene flow in marine fishes in this area is affected by environmental factors (Allen et al. 2006; Bernardi & Lape 2005; Huang & Bernardi 2001; Bernardi et al. 2003) . Point Conception (near Santa Barbara, California) is an established biogeographic boundary where fish communities north and south of this point change drastically due to temperature and oceanographic discontinuities (Briggs 1974; Dawson et al. 2006). Similarly, Punta Eugenia (near the middle of the Baja California peninsula) has been shown to be a phylogeographic barrier that lowers genetic connectivity of some fish populations (Bernardi 2000; Bernardi and Talley 2000; Terry el al. 2000; Huang and Bernardi 2001; Stepien et al. 2001; Schinske et al. 2010). Bernardi et al. (2013) found that disjunct species with low gene flow across the Baja California peninsula (8 out of 12 studied species), also presented decreased gene flow across Punta Eugenia and Point Conception. The end result is multiple species with populations experiencing a gradient of gene flow and an extraordinary system to study the mechanisms of divergence and signals of selection under different scenarios of isolation.

Despite significant differences in their ecologies, the sargo, *Anisotremus davidsonii* (grunts, Haemulidae), and the mudsucker, *Gillichthys mirabilis* (gobies,

Gobiidae), are among the species that showed most divergence (based on mitochondrial cytochrome b, mtCYTB) across Point Conception and Punta Eugenia as well as high differentiation between disjunct populations (Table 1). In fact, their disjunct populations have been proposed to be in the process of incipient speciation as they formed well-supported clades that are sister to each other and reciprocally monophyletic (Bernardi et al. 2003; Bernardi & Lape 2005). Given that divergence is not homogeneous along the genome, however, these patterns might differ when assessing divergence with an increased number of markers.

Here, we scanned thousands of loci throughout the genome using Restriction site-Associated DNA sequencing (RADseq) to characterize the gene flow between sympatric and allopatric Northeastern and Sea of Cortez populations of sargo and mudsucker. Subsequently, by analyzing differentiation in neutral and outlier loci in combination and separately, we investigated the role of drift and selection in creating population structure. We finally searched for commonalities in the patterns of genomic divergence and signals of selection between sympatric and allopatric populations of the focal species. We specifically ask (1) if the genomic patterns of divergence across the peninsula and mentioned Pacific phylogeographic breaks are concordant with the previously observed mtCTYB divergence? (2) Do genomic patterns of differentiation change when analyzing neutral and outlier loci separately? and finally (3), can we identify patterns of outlier loci divergence that are specific to sympatric and allopatric populations?

Table 1. General characteristics and previously available genetic information (based on mt CYTB) of studied species.

Family	Species	Distribution Pacific/Gulf	Habitat and depth	PLD (days)	Gene flow across Point Conception	Gene flow across Punta Eugenia	Gene flow across peninsula	Proposed Gulf/Pacific divergence time & mode	References
Gobiidae	*Gillichthys mirabilis* (longjaw mudsucker)	Tomales bay to Bahia Magdanela /upper & central gulf	Intertidal shallow soft bottoms, estuaries and sloughs.	unknown, but larvae settle at 8-12mm	Moderate to high	Extremely low	Extremely low	0.76 to 2.3 mya.	Huang & Bernardi (2001) Bernardi et al. (2003) FishBase
Haemulidae	*Anisotremus davidsonii* (sargo)	Monterey to central Baja / upper & central gulf	Rocky reef. Occasionally sand bottoms. Up to 60m	40 to 50	n/a	Low	Extremely low	0.16 to 0.64 mya.	Bernardi et al. (2003) Bernardi & Lape (2005) FishBase

Methods

Collections and DNA extractions

Sargo specimens were collected by polespear on SCUBA or free diving. Mudsucker individuals were collected using minnow traps at multiple locations throughout their Pacific and Gulf distributions (Figure 1). See Table 2 for numbers of samples per population and region.

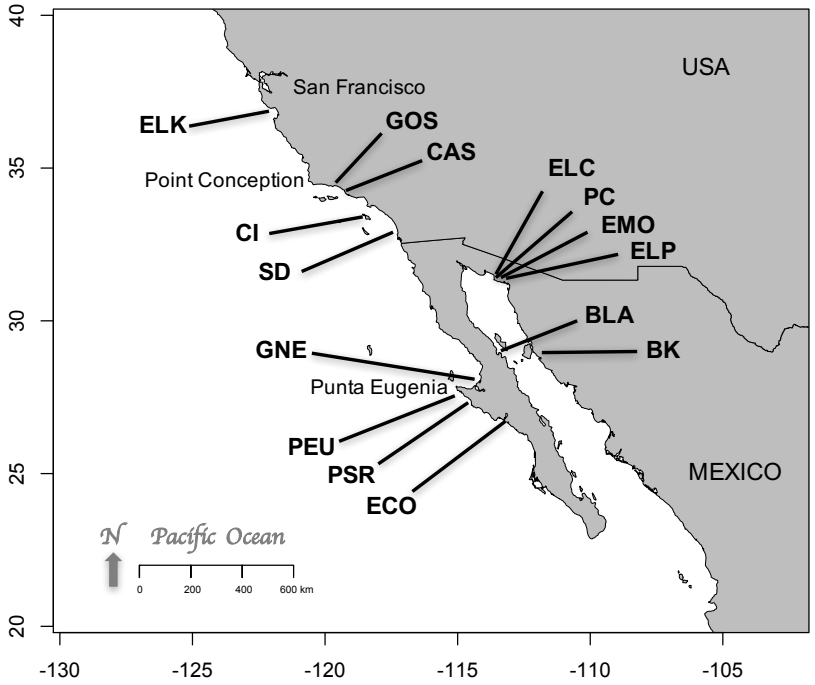


Figure 1. Pacific and Sea of Cortez sampling localities and well-established phylogeographic breaks, Point Conception and Punta Eugenia. See Table 2 for the number of samples per species. ELK, Elkhorn Slough; GOS, Goleta Slough; CAS, Carpinteria Sough; CI, Catalina Island; SD, San Diego; GNE, Guerrero Negro; PEU, Punta Eugenia; PSR, Punta San Roque; ECO, Estero El Coyote; ELC, Estero La Choya; PC, Punta Choya; EMO, Estero Morua; ELP, Estero La Pinta; BK, Bahia Kino; BLA, Bahia de Los Angeles.

Table 2. Location and number of samples per species.

Region	Location	*Gillichthys mirabilis*	*Anisotremus davidsonii*
Pacific	(Pac)	**45**	**18**
	(CAL)	30	10
California	Elkhorn Slough (ELK)	11	
	Goleta Slough (GOL)	10	
	Carpinteria Slough(CAS)	9	
	Catalina Island (CI)		3
	San Diego (SD)		7
Baja			
	(BCS)	15	8
California	Gerrero Negro (GNE)	8	
	Punta Eugenia (PEU)		2
	Punta San Roque (PSR)		6
	Estero El Coyote (ECO)	7	
Sea of Cortez	(Gulf)	**28**	**30**
	Estero La Choya (ELC)	10	
	Punta Choya (PC)		10
	Estero Morua (EMO)	9	
	Estero La Pinta (ELP)	9	
	Bahia Kino (BK)		12
	Bahia de los Angeles (BLA)		8

Only one of our sampling sites for the mudsucker (Elkhorn Slough) occurs north of Point Conception. While the sargo might be found at these latitudes during warm water events such as El Nino, it is generally rare and does not have established populations north of Point Conception. Between Point Conception and Punta Eugenia, samples were collected in three sites for the mudsucker and two for the sargo (Figure 1, Table 2). However, since only three sargos were obtained from Catalina Island (CI), samples from CI and San Diego (SD) were pooled into a single

Californian population (CAL) for the subsequent analyses. South of Punta Eugenia, mudsuckers were collected from one site and sargos from two. Yet, these two sargo sites (Punta Eugenia, PEU and Punta San Roque, PSR) were also pooled together to form a single Baja California sargo population (BCS). We sampled three distinct locations for each species within the northern and central zones of the Sea of Cortez. Fin clips, gill, or muscle tissues, were extracted from specimens and stored in 95% ethanol at room temperature in the field and at -80 in the lab. Qiagen DNeasy 96 Tissue Kits for purification of DNA from animal tissues (QIAGEN, Valenica, California, USA) were utilized following the manufacturer's protocol to extract genomic DNA from subsampled tissue.

Marker Genotyping, Discovery, and Validation.

We followed the original protocol utilizing the Sbf1 restriction enzyme to digest DNA (Miller et al. 2007; Baird et al. 2008) and produced two RAD libraries. Each genomic DNA sample contained a concentration of 400ng which was then physically sheared using a Covaris S2 sonicator using an intensity of 5 for 30 seconds, 10% duty cycle, and cycles/burst of 200. Final amplification PCR was performed with 50µl reaction volumes and 18 amplification cycles. Subsequent purification and size selection steps were accomplished using Ampure XP beads (Agencourt). Unique barcodes were ligated to all samples which were later sequenced using two illumina HiSeq 2000 lanes at UC Berkeley (Vincent J. Coates Genomics Sequencing Laboratory).

Discovery and genotyping of single nucleotide polymorphism (SNP) was performed using the STACKS software version 1.29 (Catchen et al. 2011; Catchen et al. 2013) and modified Perl scripts (Miller et al. 2012). We excluded any reads without the 6-bp barcode or an exact match to the SBf1 restriction site sequence, or with a probability of sequencing error higher than 10% (Phred score=33). Final quality-filtered reads consisted of 80-bp after removing barcodes and restriction sites. Population scripts in STACKS used these reads as the input for the population genomic analysis. The analysis of allopatric populations of both species started by running population scripts with all the Pacific individuals pooled as a single population and all the Gulf individuals as another. For the analysis of sympatric populations of *G. mirabilis,* one population script was run considering each site as a separate population. For sympatric populations of *A. davidsonii,* Pacific individuals were pooled into two populations, California (CAL) and Baja California (BCS), and Gulf sites were kept as separate populations.

After substantial exploration of the different datasets and to enable using the same parameters in all population analyses, only reads with a minimum depth coverage of 6x (m=6) that were present in at least 60% (r=0.60) of the corresponding population were included in the analyses. We further minimized the probability of linkage in our data by using the write_single_SNP option to select only the first SNP in reads that had more than one SNP. Reads passing the quality and population filters are presented as our total loci in Table 3.

Analysis of Population Genomic Divergence

After the various population analyses in STACKS, the genepop output file from the population program of STACKS was converted into Arlequin format using PGDSpider (Lischer & Excoffier 2012). Arlequin version 3.5.1.2 (Excoffier & Lischer 2010) was then used to compute genomic indexes of diversity for each population and determine genetic distances between populations. Fixation index (F_{ST}) was calculated using the total number of loci and computing a distance matrix under 10000 permutations, a significance level of 0.05, and allowing a maximum of 0.1 missing level per site. Usable loci, polymorphic loci, percent polymorphism (polymorphic loci divided by usable loci), and gene diversity or theta Θ, are also reported from the output of Arlequin (Table 3). Z-tests of proportions (with the significance level set to 0.01) were conducted in the Ausvet EpiTools calculator (http://epitools.ausvet.com.au/content.php?page=z-test-2) to determine if the average amount of polymorphism observed in sympatric Pacific and Gulf population, as well as between allopatric populations, was statistically different.

We used different approaches to explore and visualize genomic structure using all, neutral, and outlier loci. First, we performed a Discriminant Analysis of Principal Components (DAPC) (Jombart et al. 2010) on our dataset containing all loci. DAPC combines the benefits of discriminant and principal component analyses and is particularly useful to study differences between clusters (or populations) as it utilizes a multivariate approach to explore the entire variation in the data and minimizes that within clusters. This analysis was performed using the ADEGENET

package (Jombart 2008) in R (R Core Team 2013) with the Structure file produced by the populations program in STACKS as input. The algorithm *find.clusters* identified the plausible number of clusters by comparing Bayesian Information Criterion (BIC) values and the cross-validation tool *xvalDapc* determined the number of principal components that were retained.

We subsequently examined genetic structure in neutral and outlier loci, separately, to investigate if genetic drift and selection have shaped population divergence in unique patterns. Putative neural loci were obtained by excluding the outlier loci from the total number of loci. The cataloging of outlier loci, or putative loci under selection, is explained below. We ran population scripts employing the blacklist and whitelist options, for neutral and outlier loci respectively. Structure output files were analyzed using a Bayesian approach in STRUCTURE version 2.3.4 (Pritchard et al. 2000) to determine if the assignment of samples into genetic clusters differed based on neutral and outlier loci. For the neutral loci data set, a range of *K* from one to seven for *A. davidsonii* and one to ten for *G. mirabilis* (corresponding to the number of sites for each species plus two more possible hypothetical populations) were performed with 10 000 as the burn-in parameter and 100 000 replicates under the admixture model. STRUCTURE runs with the outlier loci followed the same parameters but with a range of *K* from one to five for both species. Only the *K* with the highest likelihood according to the Evanno method (Evanno et al. 2005) implemented in Structure Harvester (Earl & VonHoldt 2012) are illustrated.

Cataloging and Analyzing F_{ST} Outliers

Although working with F_{ST} outliers might incorporate a series of shortfalls (Bierne et al. 2011, 2013; Lotterhos & Whitlock 2015), they are commonly used to show evidence of the diverging effects of selection between populations (Bernardi et al. 2016; Gaither et al. 2015; Longo & Bernardi 2015; Stockwell et al. 2016). Outlier loci were identified directly from the *phistats* output file of STACKS by selecting loci above a threshold value equal to three standard deviations above the mean AMOVA F_{ST} (or loci with the top 0.03% divergence values).

We then investigated the patterns of loci abundance and divergence between populations. To begin the outlier analysis, a population script was run using a whitelist containing the outlier identifications and the resulting structure files were used to create STRUCTURE plots following the same parameters as with the neutral loci (see above). Subsequently, violin plots (density and boxplot hybrid graphics) were produced using the package ggplot2 and geom_violin( ) (Wickham 2016) in R to observe the F_{ST} range of outliers and their abundance along this range. We also performed Z-tests to determine whether the relative proportion of outliers (i.e. the number of outlier loci according to the corresponding total number of loci) was statistically different between intraspecific and interspecific populations. In both species, we tested if the proportion corresponding to the average number of outliers found in the sympatric populations was different from that of the allopatric populations and whether differences in Pacific and Gulf outlier proportions were significant.

Outlier sequences were then uploaded into the GenBank database specifying an Expect threshold (E-value) of 10^{-6}, which returns only matches with a one in a million chance to be paired with a record by chance alone. We finally documented matches to protein coding genes as well as matches to sequences without annotation and present the percentages of each category in horizontal stacked bar plots built in R. Similarly, Z-tests were conducted to determine if the percent of outliers matching to coding genes was statically different between populations.

Results

Loci and Polymorphism Statistics

Genomic DNA was sequenced from a total of 121 samples (73 *G .mirabilis* and 48 *A. davidsonii)* in two illumina lanes producing approximately 200 millions reads. The *de novo* program from STACKS created a total of 3,994,606 unique stacks and identified a total of 708,055 SNPs (averaging to 33,013 stacks and 5,852 SNPs per individual). The number of loci passing all filters in the population scripts with Pacific and Gulf pooled populations of *G. mirabilis* and *A. davidsonii* were 15,058 and 4, 379, respectively (Table 3). Total loci resulting from the scripts with multiple populations were 4, 316 for *G. mirabilis* and 15,338 for *A. davidsonii.* For both species, percent polymorphism was statistically higher in the pooled Gulf population than in the pooled Pacific population (72% to 25% in *G. mirabilis* and 72% to 49% *A. davidsonii;* z- value=10.1, p-value=0, and z-value=13.5, p-value=0). Percent

Table 3. Locus, polymorphism, and genetic diversity statistics of the Pacific and Sea of Cortez (Gulf) populations per species (after quality and population filters). Percent polymorphism was calculated by dividing the number of polymorphic loci by the usable loci. ELK, Elkhorn Slough; GOL, Goleta Slough; CAS, Carpinteria Slough; GNE, Gerrero Negro; ECO, Estero El Coyote; Pac, Pacific; Gulf, Sea of Cortez; ELC, Estero La Choya; EMO, Estero Morua; ELP, Estero La Pinta; CAL, California; BCS, Baja California; PC, Punta Choya; BKI, Bahia Kino; BLA, Bahia de los Angeles.

Species	Site	Samples	Total loci	Usable loci	Polym. loci	% Polym.	Tetha Θ
G.mirabilis							
	ELK	11	4316	3002	270	9	0.01845
	GOL	10	4316	3900	304	8	0.01991
	CAS	9	4316	159	13	8	0.02902
	GNE	8	4316	24	7	29	0.08680
	ECO	7	4316	124	19	15	0.05113
	Pacific	45	15058	352	89	25	0.02770
	ELC	10	4316	3344	1315	39	0.07139
	EMO	9	4316	129	38	29	0.05618
	ELP	9	4316	44	17	39	0.06654
	Gulf	28	15058	168	121	72	0.05166
A. davidsonii							
	CAL	10	15338	8182	2609	32	0.06383
	BCS	8	15338	5780	2046	35	0.07524
	Pacific	18	4379	2107	1031	49	0.06214
	PC	10	15338	3108	1111	36	0.05951
	BK	12	15338	8464	3693	44	0.06592
	BLA	8	15338	3191	1141	36	0.07564
	Gulf	30	4379	1391	1004	72	0.05963

polymorphism ranged from 8% to 39% and from 32% to 49%, in discrete mudsucker and sargo populations, respectively. The average polymorphism in sympatric Gulf

populations was also statistically higher than in Pacific populations for either specie (z-value=13.4 and p-value=0 for mudsucker populations; z-value=5.8 and p-value=0.0001 for sargo populations). See Table 3 for full loci and polymorphism statistics.

Table 4. Distance matrix reporting F_{ST} values (below diagonal) and significance (above diagonal) between sympatric populations of *Anisotremus davidsonii* and *Gillichthys mirabilis.* Results from pooled allopatric Pacific and Gulf populations are given below matrices. ELK, Elkhorn Slough; GOS, Goleta Slough; CAS, Carpinteria Sough; CI, Catalina Island; SD, San Diego; GNE, Guerrero Negro; PEU, Punta Eugenia; PSR, Punta San Roque; ECO, Estero El Coyote; ELC, Estero La Choya; PC, Punta Choya; EMO, Estero Morua; ELP, Estero La Pinta; BK, Bahia Kino; BLA, Bahia de Los Angeles.

Gillichthys mirabilis

	ELK	GOL	CAS	GNE	ECO	ELC	EMO	ELP
ELK		+	+	+	+	+	+	+
GOL	0.15		-	+	-	+	+	+
CAS	0.12	0		+	-	+	+	+
GNE	0.18	0.05	0.04		-	+	+	+
ECO	0.15	0	0	0		+	+	+
ELC	0.30	0.25	0.20	0.13	0.16		-	-
EMO	0.31	0.26	0.21	0.14	0.19	0		-
ELP	0.28	0.23	0.19	0.10	0.14	0	0	

Pacific vs Gulf Fst= 0.23 +

Anisotremus davidsonii

	CAL	BCS	PC	BK	BLA
CAL		+	+	+	+
BCS	0.02		+	+	+
PC	0.12	0.11		-	-
BK	0.12	0.11	0		-
BLA	0.12	0.11	0.01	0	

Pacific vs Gulf Fst = 0.11 +

Genomic Divergence and Structure between Sympatric and Allopatric Populations

The first step in our population comparison was to characterize the genomic F_{ST} in Arlequin using the total number of loci. The allopatric sargo and mudsucker populations (Pacific and Gulf pooled populations) respectively presented a genomic F_{ST} of 0.11 and 0.23, and both were significant (using a significance level of 0.05). Pacific populations of the sargo (CAS and BCS) diverged by a F_{ST}=0.02278 and this was significant as well. In contrast, differentiation between any pair of the Gulf populations only reached F_{ST} values lower than 0.01. Latitudinal pairwise F_{ST} between Pacific mudsucker populations ranged from 0 to 0.14575 and all Gulf comparisons resulted in F_{ST}=0 (See Table 4 for the complete distance matrix).

Subsequently, in order to determine if drift and/or selection have produced unique patterns of population structure, we dissected the observed genomic divergence by performing a DAPC analysis using all loci together and STRUCTURE plots with neutral and outlier loci individually. Differences in the genetic composition of allopatric populations for both species resulted in distinct clusters in the DAPC graphs (Figure 2). Interestingly, Pacific sargo population and Gulf mudsucker populations are also visibly well-separated. STRUCTURE plots showed different patterns of genomic structure when we analyzed either neutral or outlier loci (Figure 3). When using neutral loci in both species, Pacific individuals largely appeared to belong to a common cluster with Gulf individuals, but not without showing noticeable differences. In contrast, when outlier loci are utilized, plots showed an extremely high probability that every Pacific individual belong to a

separate population to that of all Gulf individuals. Structure Harvester selected a K of 4 for the analysis of neutral loci in *A. davidsonii* and a K of 2 for every other treatment.

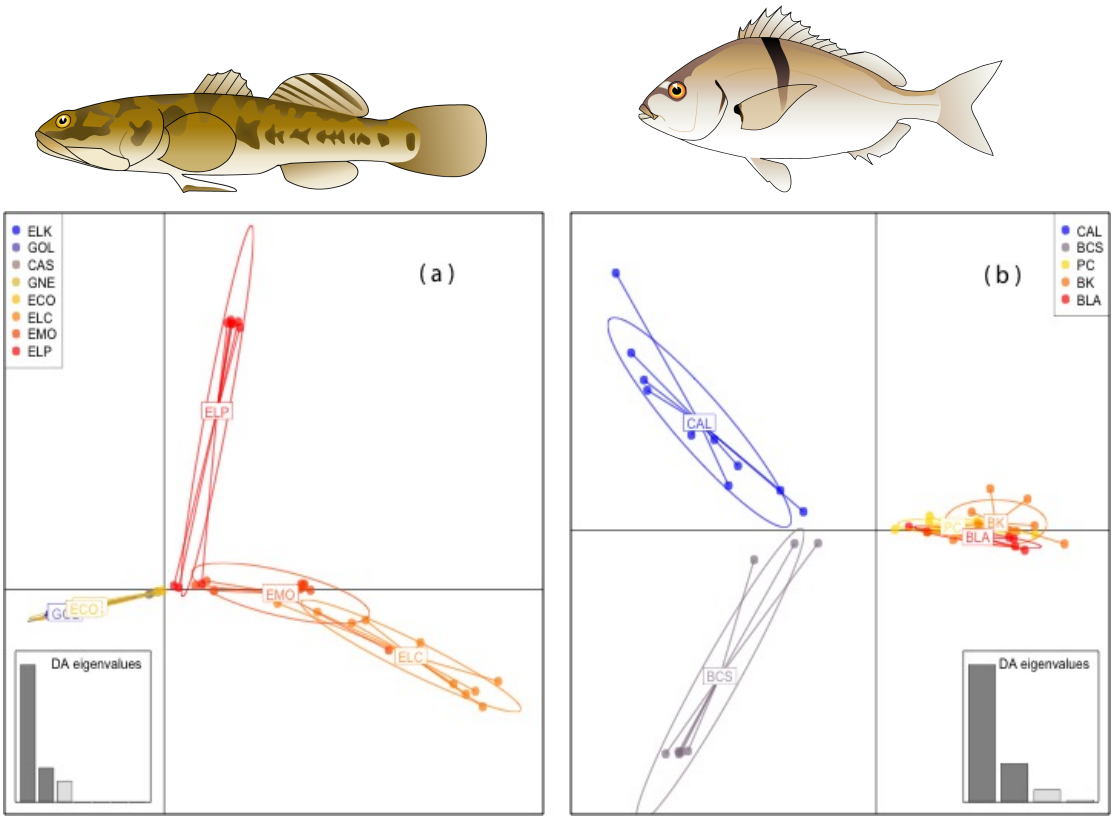


Figure 2. DAPC cluster plots of *Gillichthys mirabilis* (a) and *Anisotremus davidsonii* (b) populations. For both species, Pacific populations cluster to the left of vertical line (ELK, GOL, CAS, GNE and ECO for *G. mirabilis*; CAL and BCS for *A. davisonii*) and Sea of Cortez populations cluster to the right (ELC, EMO and ELP for *G. mirabilis*; PC, BK and BLA for *A. davisonii*). Plots were created in R using the adegenet package with 3 and 4 discriminant functions and retaining 14 and 12 principal components for *G. mirabilis* and *A. davisonii,* respectively. Number of principal components was selected by the validation tool xvalDapc.

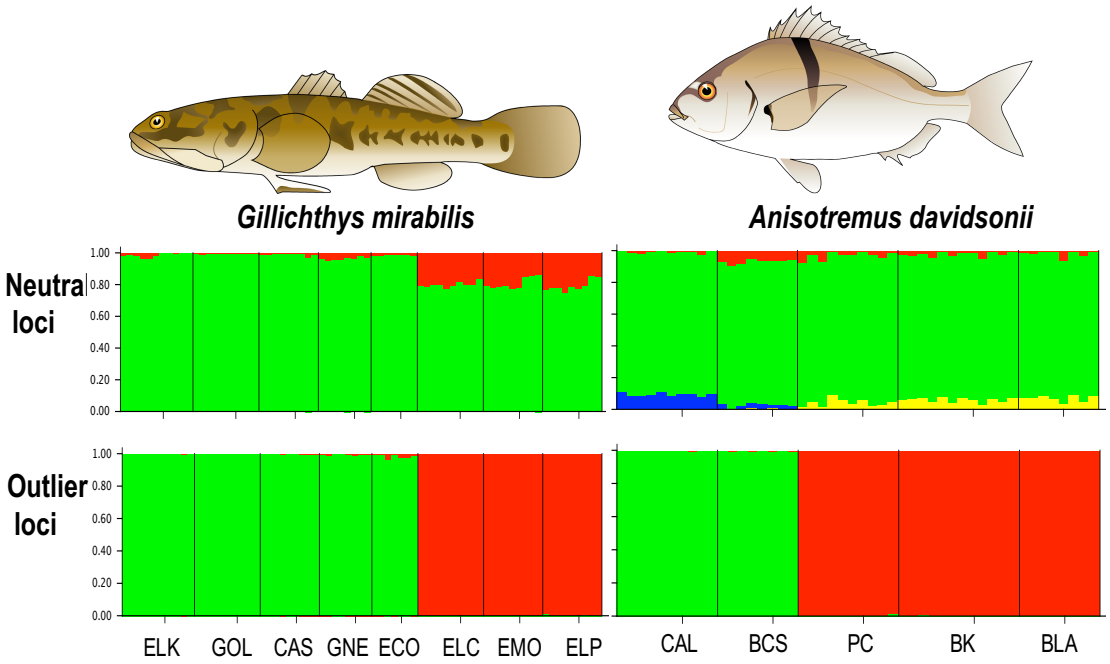


Figure 3. STRUCTURE plots with Bayesian assignment of individual into distinct genetic clusters or populations (green, red, blue, or yellow) based on presumed neutral loci (top panels) and outlier loci or loci suspected to be under selection (bottom panels). Structure Harvester selected a k of 4 for neutral loci in *A. davidsonii* and a K of 2 for every other analysis. Structure Harvester results are presented in Supplementary Figure 2. For *G. mirabilis*, neutral loci= 4169 and outlier loci= 147. For *A. davidsonii*, neutral loci= 15058 and outlier loci= 275. ELK, Elkhorn Slough; GOS, Goleta Slough; CAS, Carpinteria Sough; CI, Catalina Island; SD, San Diego; GNE, Guerrero Negro; PEU, Punta Eugenia; PSR, Punta San Roque; ECO, Estero El Coyote; ELC, Estero La Choya; PC, Punta Choya; EMO, Estero Morua; ELP, Estero La Pinta; BK, Bahia Kino; BLA, Bahia de Los Angeles.

Outlier Loci and Evidence of Selection

Outliers were identified for latitudinal pairwise Pacific populations and for all pairwise comparisons within the Gulf by selecting loci with a differentiation higher than three standard deviation from the mean AMOVA F_{ST} in the *phistats* file

produced by the STACKS population scripts. The number of outliers between allopatric population of the sargo and mudsucker were 721 loci with a F_{ST} range from 0.36 to 1 and 586 loci with a F_{ST} range from 0.72 to 1, respectively (Figure 4). Among these, allopatric populations of sargo and mudsucker had 40 and 49 fixed loci, respectively. When comparing CAL and BCS sargo populations, 202 outlier loci were identified with F_{ST} values ranging from 0.24 to 0.77. In the Gulf, population pairwise comparisons resulted in 192 to 224 outlier loci and F_{ST} values ranging from 0.19 to 0.81. Analyses of mudsucker sympatric populations yielded a range of outlier loci from 13 to 22 with an F_{ST} range of 0.19 to 0.94 in the Pacific and 57 to 76 outlier loci, with an F_{ST} range of 0.18 to 0.55 in the Gulf (Figure 4).

Given that the population program did not produce equal total number of loci in the sympatric and allopatric analyses, we compared the proportions of the total loci that were classified as outliers. We found that the average proportion of outliers observed in sympatric sargo populations was not different than that in the sympatric mudsucker populations (z-value=2.3, p-value=0.02). However, the proportions of outliers between allopatric populations of the two species were significantly different (z-value=29.2, p-value=0). In fact, with the exception of finding no statistically differences between the average outlier proportion in the Pacific and Gulf sargo populations, all other comparison (Pacific sargo *v.s.* Pacific mudsucker populations; sympatric *v.s.* allopatric populations within each species; and Pacific *v.s.* Gulf mudsucker populations) were statistically different.

We further classified outliers based on whether they matched known coding regions, sequences without annotation, or produced no match at all in GenBank. Out of the 721 and 586 outlier loci identified between allopatric sargo and mudsucker populations, 25% and 17% were paired to coding genes by the Blastn tool, respectively (Figure 4). This percent difference was statistically significant (z-value=3.6, p-value=0.0004). On average, 38% and 27% of the outliers between sympatric Pacific sargo and mudsucker populations produced matches to coding genes and 35% and 16% did the same in Gulf sargo and mudsucker populations, respectively.

While the percent difference between Pacific sargo and mudsucker populations was not statistically different (z-value=0.9, p-value=0.3424), the gene matches among Gulf populations was significantly different (z-value=2.8, p-value=0.0046). Sargo presented a statistically higher average percentage of outliers in sympatric *v.s.* allopatric populations (z-value=4.2, p-value=0.0001). Percentages and number of loci falling into each category for every population analysis are given in Figure 4. Supplementary File 1 provides the list of all the loci (symparic and allopatric) matching a coding gene with their respective F_{ST} values and GenBank gene descriptions.

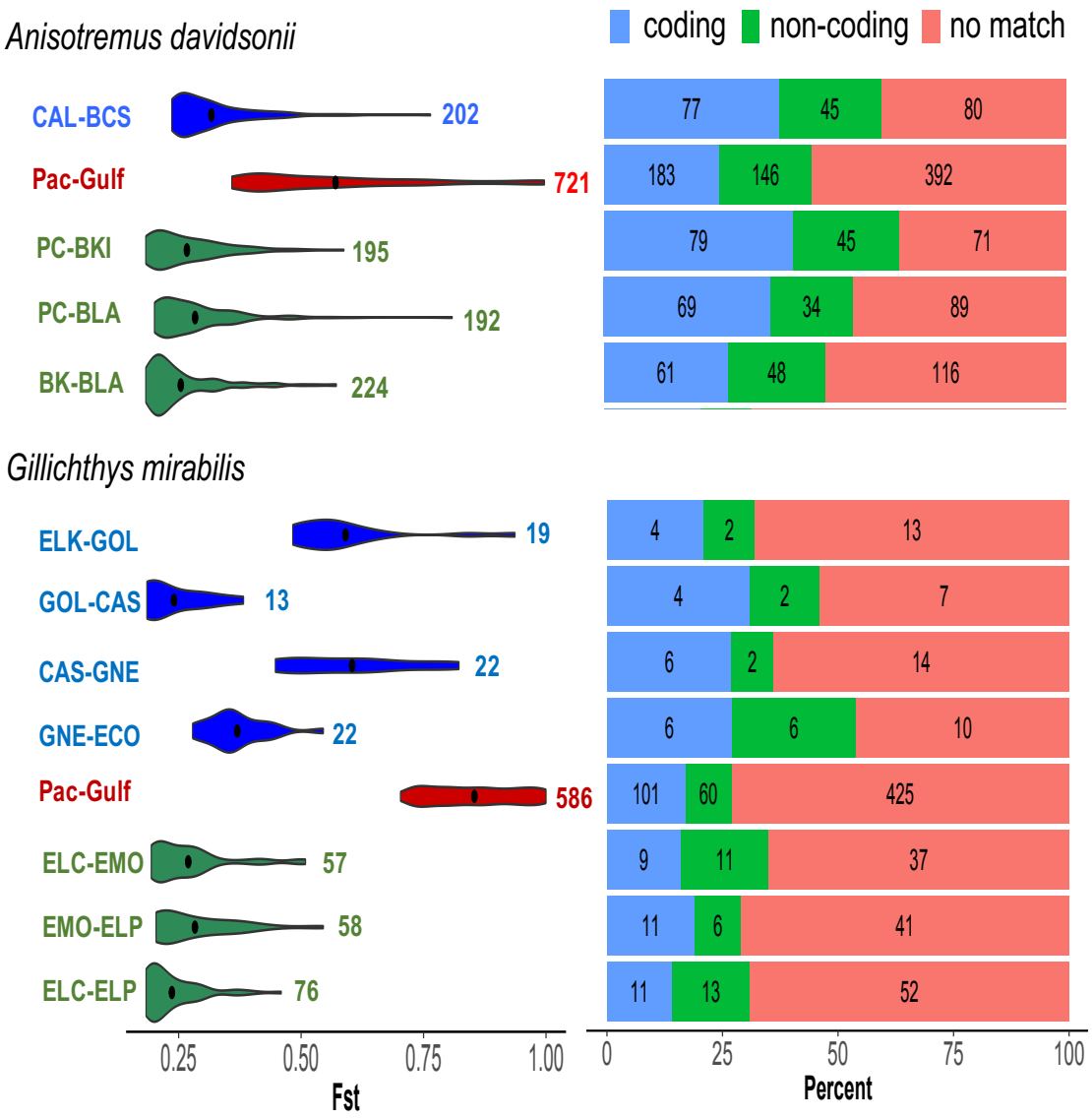


Figure 4. Outlier loci statistics in pairwise comparisons of populations of *Anisotremus davidsonii* and *Gillichthys mirabilis.* Violin plots (left panels) depict the relative density of outliers over the overall outlier F_{ST} distribution. Numbers at the end of shapes give the total number of outlier loci. Horizontal stacked bar plots (right panels) illustrate the percent of loci matching to a protein-coding gene (blue stacks), or a sequence with no annotation (green stacks), or returning no matches using the BLAST tool in GenBank (salmon colored stacks). Numbers inside stacked bars give the actual number of outliers in each category. ELK, Elkhorn Slough; GOL, Goleta Slough; CAS, Carpinteria Slough; GNE, Gerrero Negro; ECO, Estero El Coyote; Pac, Pacific; Gulf, Sea of Cortez; ELC, Estero La Choya; EMO, Estero Morua; ELP, Estero La Pinta; CAL, California; BCS, Baja California; PC, Punta Choya; BKI, Bahia Kino; BLA, Bahia de los Angeles.

Discussion

The Baja California disjunct species are an exceptional group of species with a shared evolutionary history where genomic divergence, local adaptation, and modes of selections can be examined. This study analyzed two disjunct species that previously showed population structure based on mtCYTB and searched for common patterns of genomic divergence and adaptation in sympatric and allopatric populations. Overall, the output from illumina sequencing yielded less reads in mudsucker than sargo individuals, most likely due to poorer condition and older tissue samples. This was also reflected in the lower number of loci observed in mudsucker populations. The lower number of loci in allopatric than in sympatric analysis of sargo populations is likely a product of pooling populations that already contained intra-population variation and thus impeding a large number of loci from passing established filters. Regardless of these limitations, a sufficient number of loci passed all filters to allow us to reach some comparative and meaningful results (Table 3).

Divergence and Structure between Allopatric Populations

According to every evaluation in this study (fixation indexes, DAPCs, and STRUCTURE analyses) which consistently revealed strong structure between the Pacific and Gulf populations, gene flow across the peninsula appears to be extremely low. Genomic distance based on F_{ST} values was large and significant, and DAPCs graphs clearly separated the two regions in a single plot axis. The different genetic

patterns observed in STRUCTURE plots when analyzing putative neutral and loci under selection provide potential evidence that this divergence has been produced in combination by drift and selection. When analyzing putative neutral loci, Pacific and Gulf sargo individuals showed distinctive clusters (blue and yellow) and mudsuckers from the Gulf possessed a visibly higher probability of belonging to an alternate genetic pool (red). When analyzing the outlier loci, the probability of any Pacific individual belonging to the same genetic pool than any Gulf individual, for both species, is extremely low or null. The low levels of gene flow could have allowed drift to make such noticeable impacts on disjunct populations producing the observed unique genetic patters. Similarly, isolation might have also aided selection in creating the drastically different genetic compositions seemed in the outlier loci. The observed higher levels of polymorphism in Gulf than in Pacific populations, might be a reflection of the environmental diversity in the Gulf (Thomson et al. 2000), or for instance, the proposed older age and bigger effective sizes of the sargo population in this region (Bernardi & Lape 2005).

Divergence and Structure between Sympatric Populations

Point Conception and Punta Eugenia appear to be an important discontinuity for gene flow between Pacific populations of both species as suggested by the significant F_{ST} values between ELK and rest of the mudsucker populations as well as between the CAL and BCS sargo populations. Yet, GNE and ECO mudsucker populations did not show differentiation across Punta Eugenia preventing the

genomic divergence from entirely agreeing with previously observed mitochondrial F_{ST} patterns. Besides the comparisons of populations north and south of these phylogeographic points, results also reported a low but significant F_{ST} value between the mudsucker populations from the Carpinteria Slough (CAS) and Gerrero Negro (GNE). The large geographic distance as well as strong upwelling regimes (at the San Quintin region) between them, might be responsible for such divergence.

Interestingly, Arlequin detected very little or no divergence between Gulf populations for both species but the DAPC showed some displacement in the sargo populations and substantial differentiation between the three studied mudsucker locations. The fact that F_{ST} values could not always predict how distantly populations would be placed in DAPC plots, which in turn revealed structure not visible by F_{ST} values alone, highlights how complex structure between these populations can be and the importance of analyzing it with different methods. On one hand, although the three *Esteros* where we collected the gobies are close to each other, isolation might exist as these are very large estuaries and mudsuckers prefer habitats occuring in outcrops or small channels remarkably high in the intertidal zone. Moreover, this area of the Sea of Cortez experiences very large tidal fluxes that often disconnect mudsucker habitat from the main flow in the estuary which might then prevent the exportation of mudsucker larvae. On the other hand, sargo populations are relatively far from each other but the series of islands between the peninsula and mainland might be acting as stepping stones allowing larvae to cross the Gulf.

Outlier loci and Signals of Selection

Our results strongly suggest the presence of differential selection in the Pacific and Gulf populations of each species. The two regions contain very different outlier gene pools as analyses identified large number of substantially diverged loci, including many with fixed differences (40 for sargo and 49 for mudsucker allopatric populations). This divergence is strikingly illustrated in the STRUCTURE plots where almost every individual is homogeneous for the gene pool of its region. The relative proportion of outliers (considering the initial total loci in each analysis) was higher in allopatric than in sympatric comparisons. Moreover, our exploratory examination revealed that outliers from allopatric populations enclosed substantially higher ranges of differentiation than in sympatry and these were distributed more evenly throughout such range. In general, the great majority of outliers diverging between sympatric populations lied close to the lower end of the Fst range instead. This "outlier evenness" in allopatry might be a byproduct distribution of a group of outliers with higher differentiation but it might also be an exclusive signature of allopatric speciation. For instance, while selection can be favoring different alleles in the loci that accumulate in the lower end of the F_{ST} range in sympatric populations, existing gene flow could be holding back the fixation process of these potentially adaptive alleles. Local adaptation in sympatry might therefore be reached predominantly by soft selective sweeps. In contrast, at some level of isolation presumably only attainable in allopatry, some of these alleles can avoid the effects of gene flow and diverge more easily obtaining the mentioned "outlier evenness." This

in turn, might indicate a more balanced operation of soft and hard selective sweeps in allopatric adaptation. While this study did not test this hypothesis given that for example, the outliers found in sympatry and allopatry are not the same loci, this phenomenon merits further investigation in other systems.

Overall, the relative proportion of outliers between latitudinal pairwise sympatric populations was concordant with the observed levels of genomic distance. For instance, this proportion was statistically higher in Pacific populations of the sargo than of the mudsucker but the average proportion between intra- and interspecific Gulf populations did not differ. As mentioned above, the range of F_{ST} values of sympatric outliers was not as high as in allopatry but it was still considerably high and outliers were more densely distributed toward the lower end of this range. The two Pacific mudsucker population pairs with significant F_{ST} values (ELK-GOL and CAS-GNE) accordingly illustrated the two highest ranges of differentiation of all sympatric comparisons of mudsucker populations.

Subsequent evidence of selection includes having large proportions of outliers (up to 25% in allopatry and 38% in sympatry) matching a known coding gene in GenBank. While sympatric comparisons produced overall less numbers of outliers than in allopatry, there was no statistical difference in the average proportion of loci matching genes. Mentioned phylogenetic breaks, upwelling events, habitat discontinuities, as well as the natural latitudinal temperature gradient, might be part of the selective pressures producing the observed divergence in Pacific populations. Similarly, environmental differences between central and northern zones of the Sea of

Cortez and the tidal regime in combination with habitat choice might do the same for Gulf sargo and mudsucker populations, respectively. This study was successful in identifying patterns of outlier differentiation between the studied sympatric and allopatric populations. A more in-depth revision of the identified outliers (as well as those loci matching a sequence without annotation but that could be adaptive) might shine light on specific details about how sympatric an allopatric operate in these and possibly other populations as well.

Conclusion

The Baja California Peninsula acts as an effective barrier to gene flow between Pacific and Sea of Cortez populations of both species, which show signatures of drift and adaptation to their specific environments. Point Conception and Punta Eugenia moderate the migration of individuals in the Pacific but might create different patterns of divergence depending on the utilized markers. Divergence patterns in outlier loci might be directly correlated with the level of isolation between corresponding populations. Comparisons of allopatric populations produced substantially larger number of outlier loci than sympatric analysis and these were also more evenly distributed along higher ranges of differentiation. Yet, for every population analysis (sympatric and allopatric) at least 16% of the outlier loci matched a known coding gene. These results pinpoint signatures of adaptation in every population and potentially indicate the prevalence of soft sweeps between sympatric populations and a more balanced contribution of soft and hard sweeps in allopatry.

This study illustrates how RADseq data can be used to explore signatures of sympatric and allopatric speciation as well as the utility of examining genomic divergence using diverse methodologies and while separating neutral and non-neutral loci in analyses. Future population genomic studies are urged to employ multiple techniques and scrutinize the type of loci to use the most appropriate approach to answer formulated questions.

References

Allen, L. G., Pondella, D. J. & Horn, M. h. (2006) *The Ecology of Marine Fishes: California and Adjacent Waters*. UC Press.

Baird, N. A., Etter, P. D., Atwood, T. S., Currey, M. C., Shiver, A. L., Lewis, Z. A., Selker, E. U., Cresko, W. A. & Johnson, E. A. (2008) Rapid SNP Discovery and Genetic Mapping Using Sequenced RAD Markers. *PLoS ONE* **3**, 1–7.

Bernardi, G. & Lape, J. (2005) Tempo and Mode of Speciation in the Baja California Disjunct Fish Species Anisotremus Davidsonii. *Molecular Ecology* **14**, 4085–4096.

Bernardi, G., Findley, L. & Rocha-Olivares, A. (2003) Vicariance and Dispersal across Baja California in Disjunct Marine Fish Populations. *Evolution; international journal of organic evolution* **57**, 1599–1609.

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**, 3384–3396.

Bierne, N., Welch, J., Loire, E., Bonhomme, F. & David, P. (2011) The Coupling Hypothesis: Why Genome Scans May Fail to Map Local Adaptation Genes. *Molecular Ecology* **20**, 2044–2072.

Bierne, N., Roze, D. & Welch, J. J. (2013) Pervasive Selection or Is It.? Why Are FSToutliers Sometimes so Frequent? *Molecular Ecology* **22**, 2061–2064.

Briggs, J. C. (1974) *Marine Zoogeography*. New York: McGraw-Hill.

Brusca, R. C. (1973) *A Handbook to the Common Intertidal Invertebrates of the Gulf of California.* Tucson, AZ: University of Arizona Press.

Catchen, J., Hohenlohe, P. A., Bassham, S., Amores, A. & Cresko, W. A. (2013) Stacks: An Analysis Tool Set for Population Genomics. *Molecular Ecology* **22**, 3124–3140.

Catchen, J. M., Amores, A., Hohenlohe, P., Cresko, W. & Postlethwait, J. H. (2011) *Stacks* : Building and Genotyping Loci *De Novo* From Short-Read Sequences. *G3: Genes|Genomes|Genetics* **1**, 171–182.

Dawson, M. N., Waples, R. S. & Bernardi, G. (2006) Phylogeography. In *The Ecology of Marine Fishes: California and Adjacent Waters* p. UC Press Larry G. Allen, Daniel J. Pondella, II, and Michael H. Horn Eds.

Earl, D. & VonHoldt, B. (2012) STRUCTURE HARVESTER: A Website and Program for Visualizing STRUCTURE Output and Implementing the Evanno Method. *Conservation Genetics Resources* **4**, 359–361.

Endler, J. A. (1977) *Geographic Variation, Speciation and Clines*. Princeton University Press.

Evanno, G., Regnaut, S. & Goudet, J. (2005) Detecting the Number of Clusters of Individuals Using the Software STRUCTURE: A Simulation Study. *Molecular Ecology* **14**, 2611–2620.

Excoffier, L. & Lischer, H. E. L. (2010) Arlequin Suite Ver 3.5: A New Series of Programs to Perform Population Genetics Analyses under Linux and Windows. *Molecular Ecology Resources* **10**, 564–567.

Gaither, M. R., Bernal, M. A., Coleman, R. R., Bowen, B. W., Jones, S. A., Simison, W. B. & Rocha, L. A. (2015) Genomic Signatures of Geographic Isolation and Natural Selection in Coral Reef Fishes. *Molecular Ecology* **24**, 1543–1557.

Jombart, T. (2008) Adegenet: A R Package for the Multivariate Analysis of Genetic Markers. *Bioinformatics* **24**, 1403–1405.

Jombart, T., Devillard, S., Balloux, F., Falush, D., Stephens, M., Pritchard, J., Pritchard, J., Stephens, M., Donnelly, P., Corander, J., et al. (2010) Discriminant Analysis of Principal Components: A New Method for the Analysis of Genetically Structured Populations. *BMC Genetics* **11**, 94.

Lischer, H. E. L. & Excoffier, L. (2012) PGDSpider: An Automated Data Conversion Tool for Connecting Population Genetics and Genomics Programs. *Bioinformatics* **28**, 298–299.

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.

Lotterhos, K. E. & Whitlock, M. C. (2015) The Relative Power of Genome Scans to Detect Local Adaptation Depends on Sampling Design and Statistical Method.

Molecular Ecology **24**, 1031–1046.

Mayr, E. (1942) *Systematics and the Origin of Species from the View- Point of a Zoologist*. New York, New York: Columbbia University Press.

Miller, M., Dunham, J., Amores, a, Cresko, W. & Johnson, E. (2007) Genotyping Using Restriction Site Associated DNA (RAD) Markers. *Genome Research* **17**, 240–248.

Miller, M. R., Brunelli, J. P., Wheeler, P. A., Liu, S., Rexroad, C. E., Palti, Y., Doe, C. Q. & Thorgaard, G. H. (2012) A Conserved Haplotype Controls Parallel Adaptation in Geographically Distant Salmonid Populations. *Molecular Ecology* **21**, 237–249.

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

R Core Team. (2013) R: A Language and Environment for Statistical Computing. Vienna, Austria: R Foundation for Statistical Computing 2013, doi:ISBN 3-900051-07-0.

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**, 14438–14443.

Rocha, L. A., Robertson, D. R., Roman, J. & Bowen, B. W. (2005) Ecological Speciation in Tropical Reef Fishes. *Proceedings of the Royal Society B* **272**,

573–579.

Sexton, J. P., Hangartner, S. B. & Hoffmann, A. A. (2014) Genetic Isolation by Environment or Distance: Which Pattern of Gene Flow Is Most Common? *Evolution* **68**, 1–15.

Stockwell, B. L., Larson, W. A., Waples, R. K., Abesamis, R. A., Seeb, L. W. & Carpenter, K. E. (2016) The Application of Genomics to Inform Conservation of a Functionally Important Reef Fish (Scarus Niger) in the Philippines. *Conservation Genetics* **17**, 239–249.

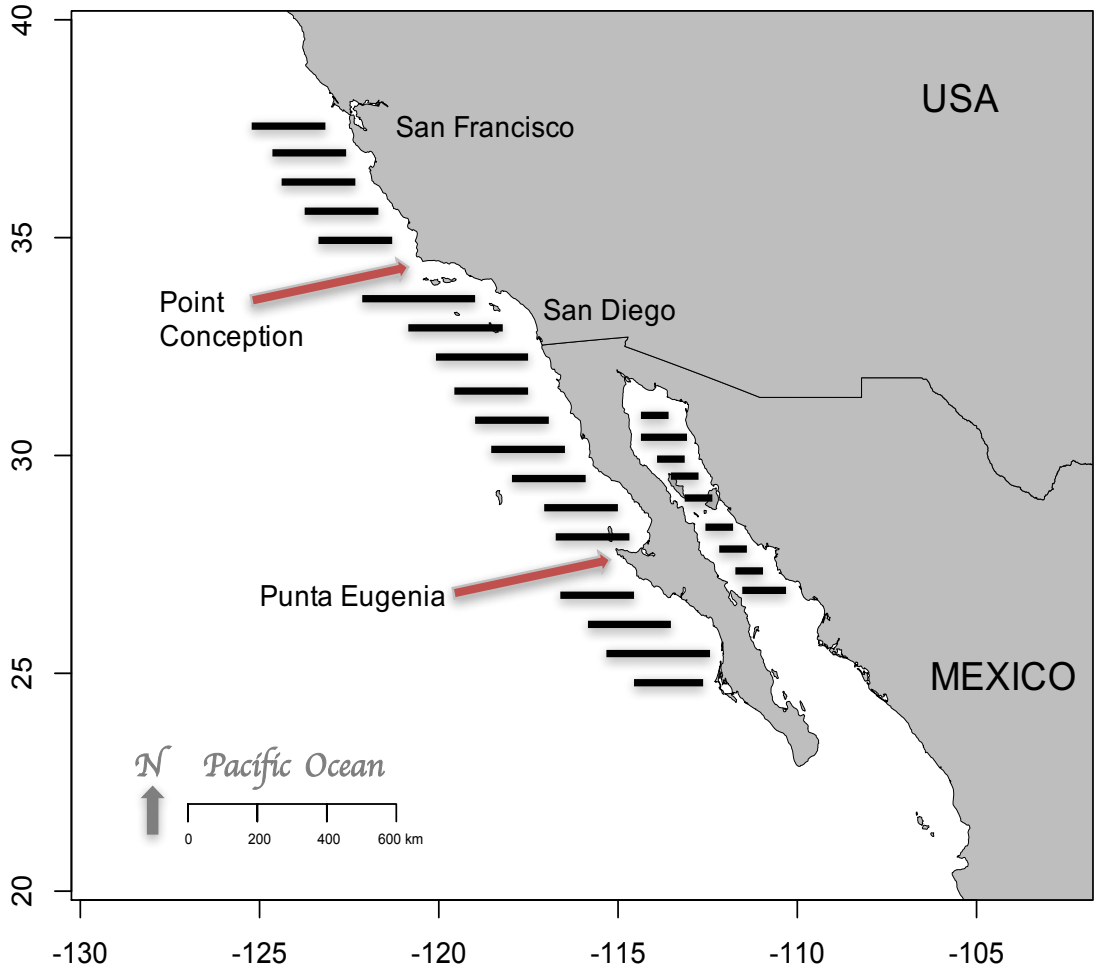
Thomson, D. A., Findley, L. T. & Kersitch, A. N. (2000) *Reef Fishes of the Sea of Cortez: The Rocky Shore Fishes of the Gulf of California.* Austin, TX: The University of Texas Press.

Wickham, H. (2016) *Ggplot2: Elegant Graphics for Data Analysis*. New York.

Supplementary Information

Supplementary Table 1**.** List of the 19 Baja California disjunct fish species (Bernardi et al. 2003).

Family	Species	Common name
Atherinidae	*Leuresthes tenuis/ L. sardina*	grunion
Girellidae	*Girella nigricans / G. simplicidens*	opaleye
Haemulidae	*Anisotremus davidsonii*	sargo
Blenniidae	*Hypsoblennius jenkinsi*	mussel blenny
Chaenopsidae	*Chaenopsis alepidota*	oragethroat pikeblenny
Serranidae	*Paralabrax maculatofasciatus*	spotted sand bass
Gobiidae	*Gillichthys mirabilis*	longjaw mudsucker
Gobiidae	*Lythrytpnus dalli*	blue banded boby
Blenniidae	*Hypsoblennius gentilis*	bay blenny
Kyphosidae	*Kyphosus azureus*	zebraperch
Labridae	*Halichoeres semicinctus*	rock wrasse
Labridae	*Semicossyphus pulcher*	California sheephead
Scorpaenidae	*Sebastes macdonaldi*	Mexican rockfish
Embiotocidae	*Zalembius rosaceus*	pink surfperch
Scorpaenidae	*Scorpaena guttata*	scorpionfish
Polyprionidae	*Stereolepis gigas*	giant seabass
Agonidae	*Xenetremus ritteri*	flagfin poacher
Pleuronectidae	*Hypsopsetta guttulata*	diamond turbot
Pleuronectidae	*Pleuronichthys verticalis*	hornyhead turbot



Supplementary Figure 1. Typical distribution of the Baja California disjunct species and location of the well-established biogeographic breaks in the Pacific.

Supplementary Figure 2. Structure Harvester Output:

Anisotremus davidsonii

15058 neutral loci (15338 total loci -280 outlier loci). K=4 was selected.

K	Reps	Mean LnP(K)	Stdev LnP(K)	Ln'(K)	|Ln''(K)|	Delta K
1	10	-205806.92000	14.135440	—	—	—
2	10	-200435.53000	224.562300	5371.390000	8836.390000	39.349392
3	10	-203900.53000	564.896762	-3465.000000	3439.920000	6.089467
4	10	-203925.61000	622.339092	-25.080000	194718.330000	312.881406

K	Reps	Mean LnP(K)	Stdev LnP(K)	Ln'(K)	|Ln''(K)|	Delta K
5	10	-398669.02000	328814.801505	-194743.410000	362571.210000	1.102661
6	10	-230841.22000	85264.319795	167827.800000	288243.680000	3.380590
7	10	-351257.10000	235171.117020	-120415.880000	—	—

280 outlier loci. K=2 was selected.

K	Reps	Mean LnP(K)	Stdev LnP(K)	Ln'(K)	|Ln''(K)|	Delta K
1	10	-10718.04000	0.365756	—	—	—
2	10	-3313.470000	0.266875	7404.570000	7136.180000	26739.792691
3	10	-3045.080000	0.561348	268.390000	805.260000	1434.512273
4	10	-3581.950000	1579.984620	-536.870000	952.640000	0.602943
5	10	-3166.180000	398.719154	415.770000	—	—

Gillichthys mirabilis

4169 neutral loci (4316 total loci -147 outlier loci). K=2 was selected.

K	Reps	Mean LnP(K)	Stdev LnP(K)	Ln'(K)	|Ln''(K)|	Delta K
1	10	-59229.560000	6.831496	—	—	—
2	10	-53640.810000	105.544392	5588.750000	189325.41000	1793.79882
3	10	-237377.47000	239035.29124	-183736.66000	161469.280000	0.675504
4	10	-259644.85000	341260.719055	-22267.38000	100243.130000	0.293744
5	10	-382155.36000	319468.970318	-122510.51000	101526.080000	0.317796
6	10	-403139.79000	334488.287904	-20984.430000	113259.570000	0.338605
7	10	-537383.79000	258593.19520	-134244.00000	180627.570000	0.698501
8	10	-491000.22000	212888.002932	46383.570000	316609.720000	1.487213
9	10	-761226.370000	295510.172897	-270226.15000	359754.900000	1.217403
10	10	-671697.620000	348340.494014	89528.750000	—	—

147 outlier loci. K=2 was selected.

K	Reps	Mean LnP(K)	Stdev LnP(K)	Ln'(K)	|Ln''(K)|	Delta K
1	10	-10718.040000	0.365756	—	—	—
2	10	-3313.470000	0.266875	7404.570000	7136.180000	26739.792691
3	10	-3045.080000	0.561348	268.390000	805.260000	1434.512273
4	10	-3581.950000	1579.984620	-536.870000	952.640000	0.602943
5	10	-3166.180000	398.719154	415.770000	—	—

CHAPTER 3

Genomic Divergence and Signals of Convergent Selection between Northeastern Pacific and Sea of Cortez Disjunct Populations of Four Marine Fishes

Abstract

Disjunct populations offer the opportunity to study adaptation as well as the early mechanisms of allopatric speciation and the genomic effects of isolation. Unlike terrestrial organisms, or even freshwater fishes, disjunct distributions are uncommon in marine fishes. Yet, the formation of the Baja California peninsula has produced 19 natural cases of marine fishes with disjunct populations in the Pacific and Sea of Cortez. Genetic isolation in some of these species has repeatedly been studied based on few markers. Here, we explored the Baja California disjunction further, using genomic markers. This study, based on RADseq, genotyped thousands of loci from genome-wide scans to document patterns of genomic divergence in disjunct population of the sargo, *Anisotremus davidsonii* (ADA)*;* the longjaw mudsucker, *Gillichthys mirabilis* (GMI)*;* the California sheephead, *Semicossyphus pulcher* (SPU); and the zebraperch, *Kyphosus azureus* (KAZ). Furthermore, we examine the unique patterns of genomic structure drafted by drift and selection, and search for coding regions showing higher than expected differentiation between disjunct populations of multiple species. Depending on the species, approximately 5 to 10 thousand loci were genotyped. Disjunct populations of the sheephead showed little genetic evidence of isolation (F_{ST}=0.004). While populations at the south end of the peninsula have not

been detected near the surface, migrant sheephead might be maintaining gene flow between Pacific and Sea of Cortez populations using deep sea reefs as stepping-stones. Population structure between zebraperch populations was detected for the first time ever. This was shallow (F_{ST}=0.03) but statistically significant (p-value=0.0004), and consistent with an early time of isolation. In contrast, divergence and structure for sargo (F_{ST}=0.1) and mudsucker (F_{ST}=0.31), were considerably higher, thus supporting the idea that these species are in the initial stages of allopatric speciation. We observed different patterns of divergence in presumed neutral and non-neutral loci. All species showed evidence of drift (except for the sheephead) and selection (including sheephead) in STRUCTURE plots, strongly suggesting the presence of unique selective pressures in the two regions. Large numbers of outlier loci (presumed loci under selection) were found (from 120 to 330) and a high proportion of these were also matched to protein coding genes (from 19% to 46%). We identified 15 genomic regions of convergence in more than one of these species (40 outlier loci from all four species were matched to these regions). These regions are suspected to be involved in processing environmental information, metabolism, development, immune response, and possible reproduction. Yet, validation of these loci across larger population samples and analyses of the specific mechanisms in which these loci might impart adaptive advantages in these populations are still needed.

Keywords: adaptive evolution, allopatric populations, Baja California, incipient speciation, population divergence, temperature isolation.

Introduction

Species with disjunct distributions have the potential to remain the same species or diverge depending on how effectively the barrier separating their populations impedes gene flow (Palumbi, 1992). Disjunct populations provide an excellent opportunity to study the evolutionary processes of one of the main forces of speciation, allopatry (Endler 1977; Coyne & Orr 2004).

Given the long pelagic larval durations (PLD) in most marine fishes (averaging 30 days but may be as long as 2 years) and the scarcity of obvious physical barriers preventing gene flow in the oceans, marine fishes were traditionally considered to exhibit open populations with larvae that disperse long distances and undermine the genetic structure of the species (Helfman et al. 2009) However, the lack of support for a correlation between PLD and genetic structure in a wide array of marine fishes (Bowen et al. 2006; Shulman & Bermingham 1995; Bay et al. 2006), the development of the idea of self-recruitment (Jones et al. 1999; Swearer et al. 2002), and the documentation of the ability of larvae to navigate and make active choices about where to settle (Leis et al. 2003; Leis & Lockett 2005; Leis & Carson-Ewart 2002; Leis & Carson-Ewart 2000; Planes et al. 2002; Selkoe et al. 2006; Bernardi et al. 2012; Pujolar et al. 2006), have highlighted the need to investigate the role of adults and the overall ecology of the species in producing genetic variation between fish populations (Rocha et al. 2002; Poortvliet et al. 2013)

Additionally, although hard physical barriers in the oceans are few (e.g. the closure of the isthmus of Panama), other types of barriers such as seawater

temperature, salinity and freshwater discharge can also isolate fish populations and provide systems where to study the speciation process in diverse scenarios of isolation (Rocha et al. 2002; Bernardi & Lape 2005).

The Baja California Disjunct Fishes

The Baja California disjunct fishes are a group of 19 temperate species inhabiting the Pacific coast from central California to central Baja California Sur (the southern state in the peninsula) and northern and central zones of the Sea of Cortez (also called Gulf of California and here simply referred as the Gulf), where populations are isolated by the warmer seawater temperature at the south of the peninsula (and where these species are absent; supplementary Figure 1 and supplementary Table 1). These taxa offer an extraordinary system of independent natural cases of allopatry with hypothesized isolation dates ranging from at least 2 million to 12,000 years (Miller & Lea 1972; Medina & Walsh 2000; Thomson et al. 2000; Bernardi et al. 2003). The majority of previous genetic analyses studying allopatry in fishes of the Baja California peninsula have focused on one or two disjunct species (Present 1987; Tranah & Allen 1999; Terry et al. 2000; Huang & Bernardi 2001; Stepien et al. 2001). Yet, Bernardi et al. (2003) examined few mitochondrial and nuclear markers in 12 of these taxa and found 8 species with fixed differences (and low gene flow) and 4 with no notable divergence (and high gene flow) between Pacific and Gulf populations. However, the results of this and the other previous studies were limited in that (1) they utilized only a handful of markers and

(2) these do not possess enough resolution to detect a very recent cessation of gene flow. Given that differentiation is not uniform throughout the genome, utilizing modern genetic tools (Next Generation Sequencing, NGS) that allow for massive parallel sequencing of neutral and selected loci, might reveal undetected divergence and distinct patterns of differentiation across the peninsula. Similarly, by examining separately each type of loci (neutral vs selected), the effects of drift and selection on allopatric populations and their patterns among species can be investigated.

This study performs genome-wide scans to search for signals of selection and determine the levels of genomic differentiation and gene flow between Gulf and Pacific populations of a subset of the Baja California disjunct species: the sargo, *Anisotremus davidsonii* (ADA)*;* the longjaw mudsucker, *Gillichthys mirabilis* (GMI)*;* the California sheephead, *Semicossyphus pulcher* (SPU); and the zebraperch, *Kyphosus azureus* (KAZ).These four species were selected in order to provide a range of population divergence time and ecological characteristics where to compare genomic variation (Table 1).

On one hand, with low gene flow and high differentiation based on mtDNA, sargo and longjaw mudsucker Pacific and Gulf populations are proposed to have diverged 0.160 to 0.64 mya and 0.76 to 2.3 mya, respectively (Huang & Bernardi 2001; Bernardi et al. 2003; Bernardi & Lape 2005). Moreover, Bernardi et al. (2003) reported signals of incipient speciation in these species as their Pacific and Gulf populations illustrated reciprocal monophylies, i.e. they formed separate sister clades

Table 1. Species characteristic and previously available genetic information.

Family	Species	Pacific distribution / Gulf distribution	Habitat and depth	PLD (days)	Markers	Gene flow across peninsula	Proposed Gulf/Pacific divergence time	References
Labridae	*Semicossyphus pulcher* (california sheephead)	Monterey to Cape Region in Baja/ Northern Gulf	Rocky reefs and kelp beads. Up to 55m.	30	mt CYB mt CR microsat	High	n/a	Bernardi et al. (2003), Poortvliet et al. (2013), Miller & Lea (1972)
Kyphosidae	*Kyphosus azureus* (zebraperch)	Monterey to central Baja/ Northern, central Gulf	Shallow rocky reefs. Up to 27m.	unknown	mt CR	High	n/a	Bernardi et al. (2003), Thomson et al. (2000)
Gobiidae	*Gillichthys mirabilis* (longjaw mudsucker)	Tomales bay to Bahia Magdalena/ Northern, central Gulf	Intertidal soft bottoms, estuaries and sloughs. Shallow only.	unknown, but larvae settle till 8-12mm	mt CYB	Extremely low	0.76 to 2.3 mya.	Huang & Bernardi (2001), Bernardi et al. (2003), Thomson et al. (2000)
Haemulidae	*Anisotremus davidsonii* (sargo)	Point Conception to Bahia Magdalena/ Northern, central Gulf	Rocky reef. Occasionally sand bottoms. Up to 60m	40 to 50	mt CYB	Extremely low	0.16 to 0.64 mya.	Bernardi et al. (2003), Bernardi & Lape (2005), Thomson et al. (2000)

(supplementary Figure 2). On the other hand, California sheephead and zebraperch showed high levels of gene flow across the peninsula (Bernardi et al. 2003).

Ecologically, zebraperch and sargo are most similar as they tend to prefer shallow rocky reefs, although sargo is also found on sandy bottoms and deeper reefs. Juveniles of both species are also common in tide pools. In contrast, California sheephead can be found anywhere from shallow to deep reefs and longjaw mudsucker is restricted to upper intertidal habitat (Thomson et al. 2000; Huang & Bernardi 2001; Bernardi et al. 2003; Poortvliet et al. 2013).

Using Restriction site-Associated DNA sequencing (RADseq), we generated thousands of genome-wide markers to explore the patterns of genomic divergence among these four species and searched for coding genes that have diverged in disjunct populations. Our main goal was to determine if the genomic divergence mirrors the previously observed mtDNA patterns. In addition, we assessed divergence differences using neutral and outlier loci (or loci suspected to be under selection). Finally, search for common genomic regions that might have diverged simultaneously in disjunct populations of multiple species, as evidence of convergent selection.

Materials and Methods

Sample Collection

Specimens of the sargo, zebraperch, and California sheephead were collected from multiple locations throughout the species distribution (Figure 1), while tide pooling, spearfishing on SCUBA or free diving, or tissue was sampled from local

fishermen. Longjaw mudsuckers were obtained from several sites using minnow traps in their habitats. For this population analysis, however, we only focused on the differences between the two main regions. Sample sizes for the Pacific and Sea of Cortez varied from 10 to 34 and from 10 to 22 individuals per species, respectively (Table 2). Fin or gill tissue was sampled from specimens and placed in 95% ethanol at room temperature while in the field but at -80ºC in the lab.

Genotyping, Discovery and Validation of Single Nucleotide Polymorphisms.

Genomic DNA was extracted from fin or gill tissue from collected specimens using the Qiagen DNeasy 96 Tissue Kit for purification of DNA from animal tissues (QIAGEN, Valencia, California, USA). The construction of two RAD libraries followed the original protocol utilizing the Sbf1 restriction enzyme to digest DNA (Miller et al. 2007; Baird et al. 2008). Samples were individually barcoded and sequenced in two lanes on an illumina HiSeq 2000 at UC Berkeley (Vincent J. Coates Genomics Sequencing Laboratory).

We used the STACKS software version 1.29 (Catchen et al. 2011; Catchen et al. 2013) and modified Perl scripts (Miller et al. 2012) for single nucleotide polymorphism (SNP) discovery and genotyping. Reads with more than 10% probability of sequencing error (Phred score=33) and without the exact sequence of the SBf1 restriction site or the 6-bp barcode were excluded from the analysis. Barcodes and restriction sites were also removed from fragments resulting in 80-bp quality filtered reads.

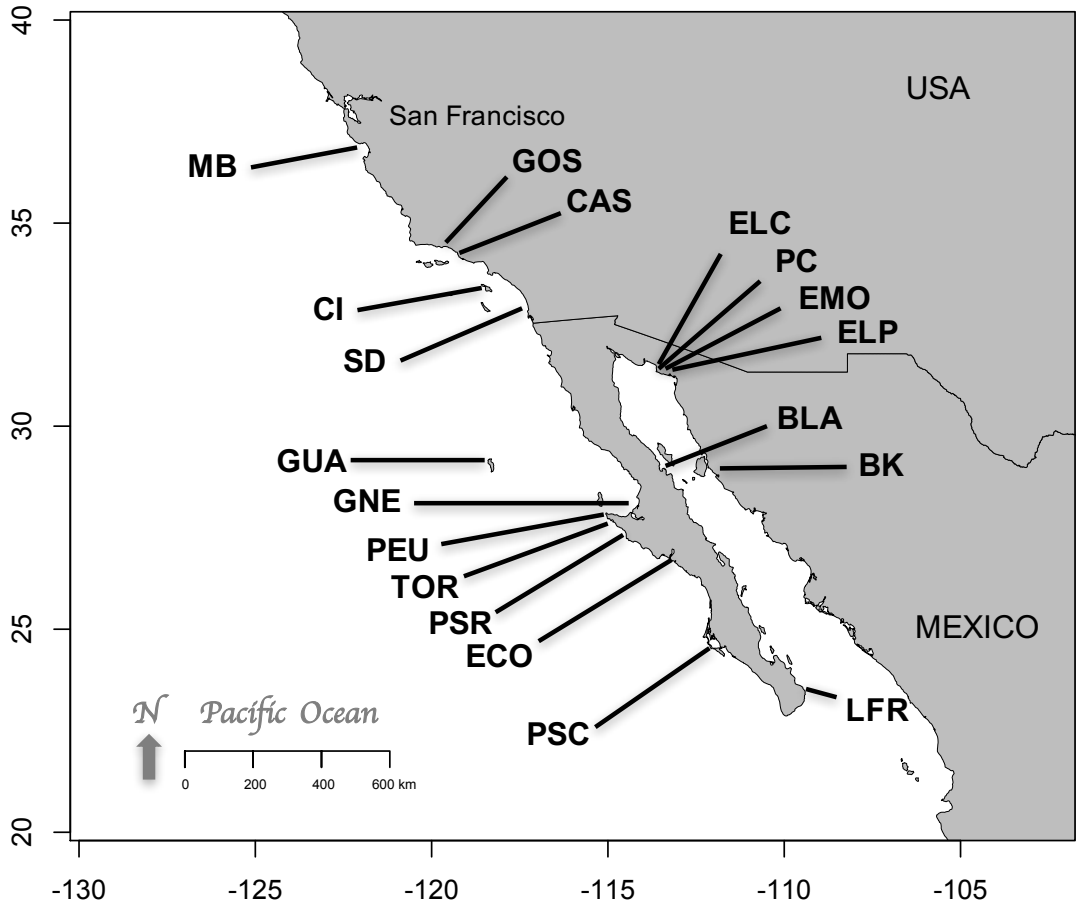


Figure 1. Pacific and Sea of Cortez sampling localities. See Table 2 for the number of samples per species. MB, Monterey Bay; GOS, Goleta Slough; CAS, Carpinteria Sough; CI, Catalina Island; SD, San Diego; GUA, Guadalupe Island; GNE, Guerrero Negro; PEU, Punta Eugenia; TOR, Bahia Tortugas; PSR, Punta San Roque; ECO, Estero El Coyote; PSC, Puerto San Carlos; ELC, Estero La Choya; PC, Punta Choya; EMO, Estero Morua; ELP, Estero La Pinta; BK, Bahia Kino; BLA, Bahia de Los Angeles; LFR, Los Frailes.

These reads were utilized as the input for the population genomic analyses performed using the population scripts in STACKS. We selected only putative SNPs that exhibited a minimum depth coverage of 8x (m=8) and were present in at least 80% (r=0.80) of pooled individuals from the Pacific and Sea of Cortez. If any read had more than one SNP, only the first SNP in that read was considered in the analysis

to avoid linkage disequilibrium. Reads passing the quality and population filters represent our total loci in Table 3.

Table 2. Pacific and Sea of Cortez (Gulf) collecting sites (ordered by latitude) and samples per species. SPU, *Semicossyphus pulcher*; KAZ, *Kyphosus azureus*; GMI, *Gillichthys mirabilis*; ADA, *Anisotremus davidsonii*.

Region	**Site**	SPU	KAZ	GMI	ADA
Pacific		**16**	**10**	**34**	**15**
	Monterey Bay (MB)			7	
	Goleta Slough (GOS)			10	
	Carpinteria Slough(CAS)			7	
	Catalina Island (CI)	5			2
	San Diego (SD)		10		7
	Guadalupe Island (GUA)	4			
	Guerrero Negro (GNE)			4	
	Punta Eugenia (PEU)				1
	Bahia Tortugas (TOR)	4			
	Punta San Roque (PSR)				5
	Estero El Coyote (ECO)			6	
	Puerto San Carlos (PSC)	3			
Gulf		**10**	**12**	**21**	**22**
	Estero La Choya (ELC)			10	
	Punta Choya (PC)	3			5
	Estero Morua (EMO)			5	
	Estero La Pinta (ELP)			6	
	Bahia Kino (BK)				12
	Bahia de los Angeles	5	12		5
	(BLA)	2			
	Los Frailes (LFR)				

Analysis of Population Genomic Diversity and Differentiation

To determine population genomic diversity indexes and distance, we converted the genepop output file from the population script of STACKS into Arlequin format using PDGSpider (Lischer & Excoffier 2012). Fixation index (F_{ST}) between populations, based on the total loci, was calculated in Arlequin version 3.5.1.2 (Excoffier & Lischer 2010) with 10000 permutations, a significance level of 0.05 and computing a distance matrix. From the output of Arlequin, we also report the total number of loci, usable loci, polymorphic loci, percent polymorphism (polymorphic loci divided by total loci), and genetic diversity or theta (Table 3).

Genomic population structure was explored and visualized using three different approaches. First, we manually modified the STRUCTURE output file from the population script of STACKS to perform a Discriminant Analysis of Principal Components (DAPC) (Jombart et al. 2010) using the total number of loci. DAPC is a multivariable approach that is particularly useful to analyze the differences between clusters (or populations) because, while combining the benefits of discriminant and principal component analyses, it explores the entire variation in the data and minimizes that within clusters. DAPC plots are used to illustrate the relationship between Pacific and Gulf genetic pools. DAPC was performed in R (R Core Team 2013) using the ADEGENET package (Jombart 2008). The plausible number of clusters was identified comparing Bayesian Information Criterion (BIC) values in the algorithm *find.clusters* and the number of principal components retained was obtained using the cross-validation tool *xvalDapc.*

Subsequently, by analyzing genetic structure using neutral and outlier loci, we investigated the possibility that genetic drift and selection created different patterns of divergence between the two regions (our two other structure approaches). The identification of outlier loci, or candidate loci to be under selection, is described below. Neutral loci consist of the total number of loci minus the outlier loci. A population script was run for each type of loci using the whitelist and blacklist flags and the output genepop file was converted into STRUCTURE format in PDGSpider. These files, which included population identification, were utilized to analyze genetic clusters within neutral and outlier loci using a Bayesian approach in STRUCTURE version 2.3.4 (Pritchard et al. 2000). We performed 10 replicates for each *K* (from one to five) with a burn-in parameter of 10 000 and running 100 000 replicates under the admixture model. We present only the *K* with the highest likelihood according to the Evanno method (Evanno et al. 2005) in STRUCTURE HARVESTER (Earl & VonHoldt 2012).

F_{ST} Outliers

We used the loci with the highest differentiation to highlight the potential diverging effects of selection between the Gulf and the Pacific. While caveats were identified when using F_{ST} outliers as a proxy for regions under selection (Bierne et al. 2011; Bierne et al. 2013; Lotterhos & Whitlock 2015), they remain a common approach to show evidence of selection in populations (Bernardi et al. 2016; Stockwell et al. 2016; Longo & Bernardi 2015; Gaither et al. 2015). Loci were

considered as outliers when falling within the top 0.03% (or three standard deviations above the mean) of F_{ST} values from the output of the population scripts in STACKS.

Outlier loci were used in the STRUCTURE plots and subsequent violin plot which shows the F_{ST} range of outliers and their relative density along their differentiation distribution. Violin plots correspond to the density and boxplot hybrid graphic created in R using the package ggplot2 and geom_violin( ) (Wickham 2016). Outliers identified by both methods were compared against the GenBank database using an Expect threshold (E-value) of 0.000001 corresponding to a one in a million chance to get a match by chance alone. We recorded matches to protein coding genes, as well as chromosomal regions. Finally, we classified protein coding genes diverging simultaneously in more than one species into KEGG assignments (Kanehisa et al. 2008; Ogata et al. 1999) and searched for their functions or pathways in the GeneCard online database (www.genecards.org), since it also provides information from the Gene Ontology, UniProt, GTEx, Entrez Gene, and other databases.

Results

Single Nucleotide Polymorphism Discovery

A total of 140 individuals (75 from the Pacific and 65 from the Gulf) among the four disjunct species (Table 2) were sequenced in two illumina lanes producing approximately 230 million reads. After quality filters, our database consisted of approximately 6.5 million reads and one million SNPs, which averaged to 45997 reads and 7081 SNPs per individual. After population filters selecting for reads with

a minimum of an 8x coverage and present in at least 80% of the corresponding population (Pacific or Gulf), the total number of loci (RADtags) per species ranged from 4858 to 10532 (mean = 8345). Species showed consistently higher numbers of usable and polymorphic loci in the Gulf than in Pacific populations. Locus, polymorphism, and genetic diversity statistics per species and region are given in Table 3.

Genomic structure between Pacific and Sea of Cortez populations

Percent polymorphism was very similar between Pacific and Gulf populations in the zebraperch but considerably higher in the Gulf for sargo and mudsucker populations. Only the California sheephead clearly showed greater polymorphism in the Pacific but the difference was smaller than in the sargo and especially for the mudsucker (Table 3). All species showed higher genetic diversity, theta, in Gulf populations. The genomic F_{ST} (using the total number or loci) between Pacific and Gulf populations ranged from 0.004 in the California sheephead to 0.31 in the longjaw mudsucker (Table 3). Genomic divergence was low and non-significant for the sheephead, low but significant for the zebraperch, and large and significant for the last two species, sargo and mudsucker (Table 3).

Table 3. Locus, polymorphism, and genetic statistics of the Pacific and Sea of Cortez (Gulf) populations per species (after quality and population filters). Percent polymorphism was calculated by dividing the number of polymorphic loci by the total loci. Neutral loci represent the total loci minus the outlier loci from STACKS.

Species	*Semiccosyphus pulcher*		*Kyphosus azureus*		*Gillichthys mirabilis*		*Anisotremus davisonii*	
Region	Pacific	Gulf	Pacific	Gulf	Pacific	Gulf	Pacific	Gulf
Number of individuals	16	10	10	12	34	21	15	22
Total loci	10532	10532	4858	4858	8700	8700	9288	9288
Usable loci	1716	2066	694	4631	1233	2143	1825	2046
Polymorphic loci	1486	1509	500	3297	281	1717	907	1437
Percent polymorphism	14%	14%	10%	68%	3%	20%	10%	15%
Genetic diversity	0.1914	0.19905	0.17603	0.19438	0.02436	0.08436	0.0773	0.07973
F_{ST}		0.004 (-)		0.03 (+)		0.31 (+)		0.1 (+)
Outlier loci		259		120		330		207
Neutral loci		10273		4738		8370		8031

DAPC analyses

Two patterns emerged from the DAPC plots using one discriminant function and all loci. Pacific and Gulf clusters of the California sheephead and zebraperch overlapped while they did not for the sargo and mudsucker (Figure 2). In each species (except for the zebraperch), multiple peaks within each region might indicate further structure within our pooled populations.

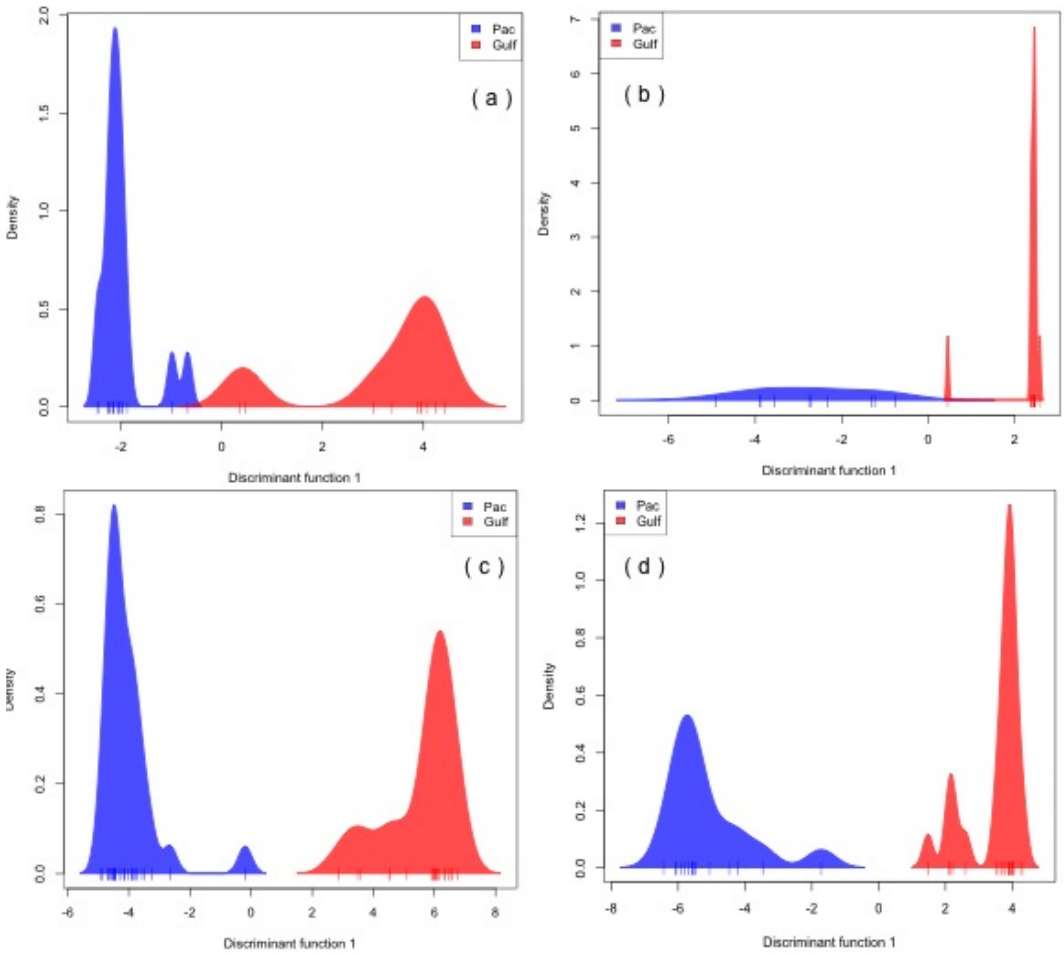


Figure 2. DAPC cluster plots of Pacific (Pac; left – blue peaks) and Sea of Cortez (Gulf; right- red peaks) per species based on all loci. Plots were created in R using the adegenet package with one discriminant and retaining the number of principal components selected by xvalDapc; 18 for *Semicossyphus pulcher* (a); 8 for *Kyphosus azureus* (b); 10 for *Gillichthys mirabilis* (c); and 20 for *Anisotremus davidsonii* (d).

STRUCTURE analyses

Structure Harvester selected a k of 2 for all STRUCTURE plots with the exception of the plots using neutral loci for GMI and ADA, where a k of 3 was

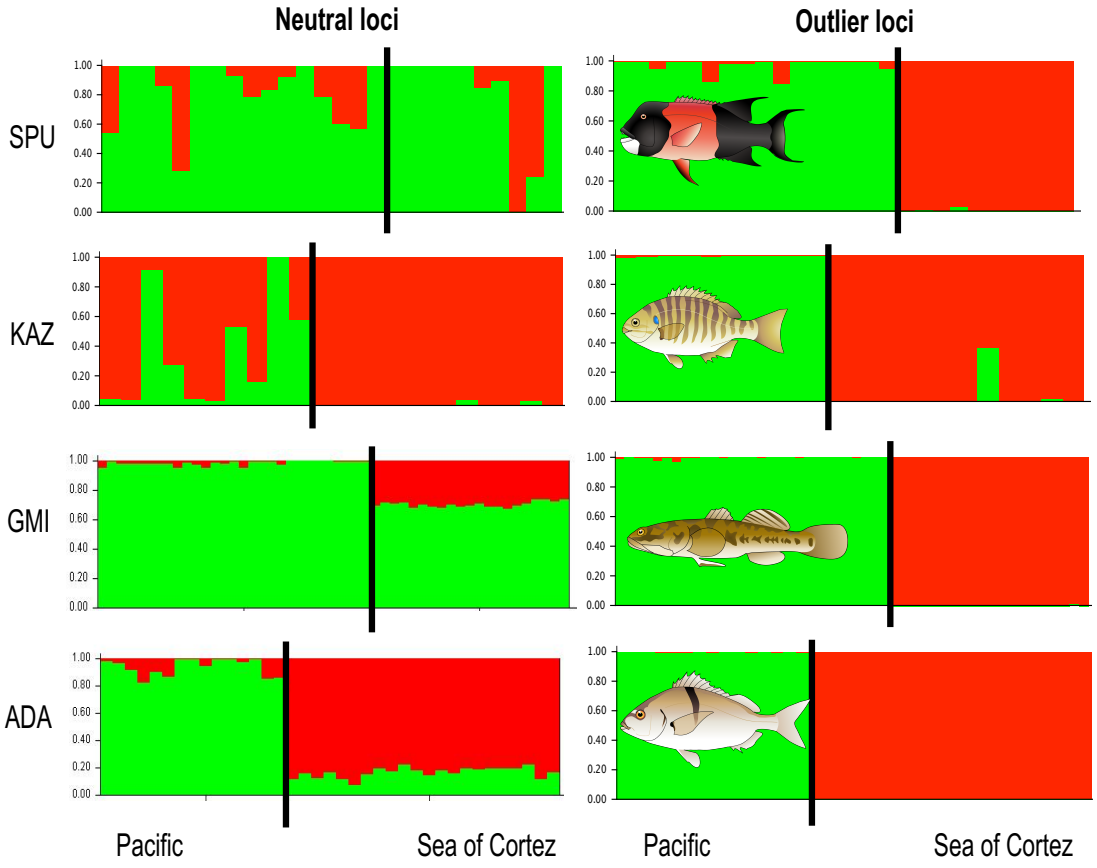


Figure 3. STRUCTURE plots based on presumed neutral loci (left panels) and outlier loci or loci suspected to be under selection (right panels) and individual Bayesian assignment into distinct genetic clusters or populations (Green, Red or blue). For each panel, individuals to the left of the black bar were collected in the Pacific and the others in the Sea of Cortez. Structure Harvester selected a k of 2 (from 1 to 5) for both types of loci for SPU (*Semicossyphus pulcher*: neutral loci= 10,273, outlier loci= 259) and KAZ (*Kyphosus azureus*: neutral loci= 4,738, outlier loci=120 loci). A k of 3 was selected for neutral loci in GMI (*Gillichthys mirabilis*: neutral loci= 8,370) and ADA (*Anisotremus davidsonii*: neutral loci= 8,031) but a k of 2 for outlier loci (GMI: outlier loci=300; ADA: outlier loci= 207). Full Structure Harvester results are presented in Supplementary Figure 3.

selected (Supplementary Figure 3). Significantly higher structure was seen in plots using outlier loci (except for SPU) demonstrating different patterns of divergence when we compare Pacific and Gulf populations using neutral and outlier loci (Figure 3).

Outlier loci and evidence of selection

The number of outliers ranged from 120 to 303 as identified keeping loci three standard deviations above the mean F_{ST} (Table 3). The F_{ST} values for loci ranged from 0.15 to 0.48 for SPU, from 0.23 to 0.52 for KAZ, from 0.33 to 1 for ADA, and from 0.66 to 1 for GMI (Figure 4). Overall, all loci were significantly more differentiated in GMI and ADA including 20 fixed loci between Pacific and Gulf populations of GMI and 15 for ADA. The density of outliers along their F_{ST} distribution is shown in Figure 4. STRUCTURE plots based on STACKS outlier loci show well differentiated genetic clusters between the two regions, except for SPU. Several individuals in both regions were found to be genetically homogeneous for their respective clusters (Figure 3).

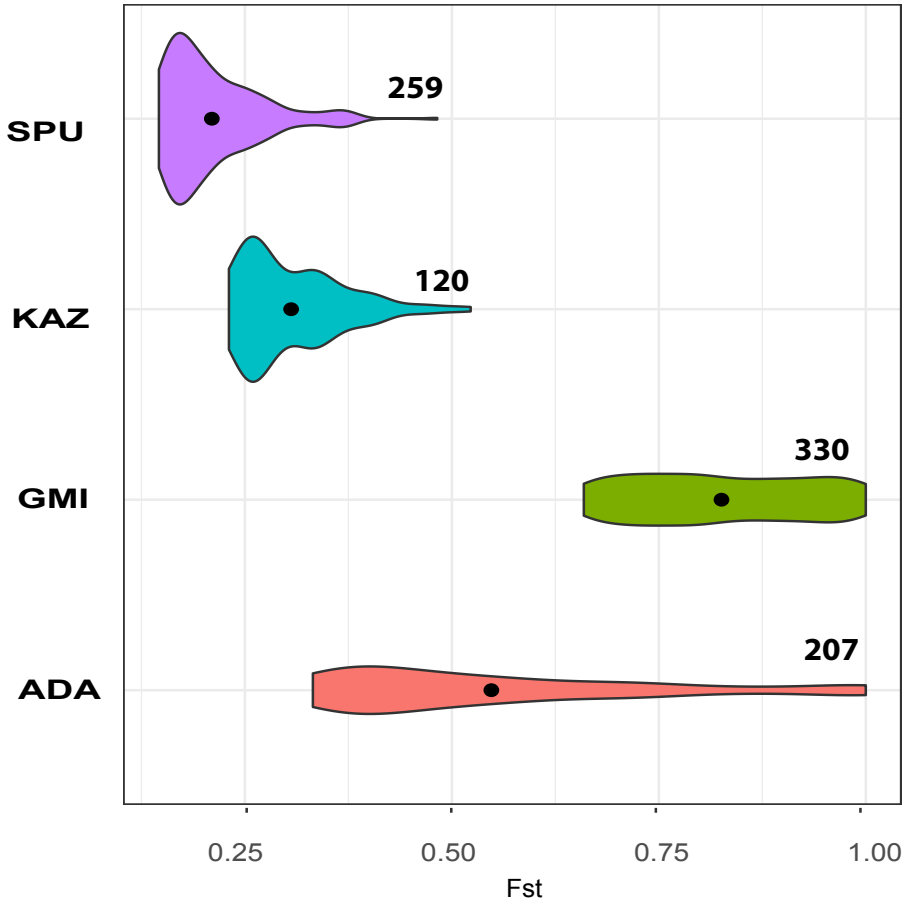


Figure 4. F_{ST} distribution and corresponding relative density of outlier loci (from STACKS) per species. Numbers above shapes depict the total number of outlier loci per species. SPU, *Semicossyphus pulcher*; KAZ, *Kyphosus azureus*; GMI, *Gillichthys mirabilis*; ADA, *Anisotremus davidsonii*.

The percent of outliers with a GenBank match to a protein coding gene ranged from 19% to 46%. The percent of outliers with a GenBank match to a sequence or a location in a chromosome, but with not gene information, ranged from 13% to 18%. Loci that did not produce a match in GenBank consisted of 42% to 65% of the total outliers. Figure 5 shows the total number of loci per species matching a gene or sequence and those returning not match at all.

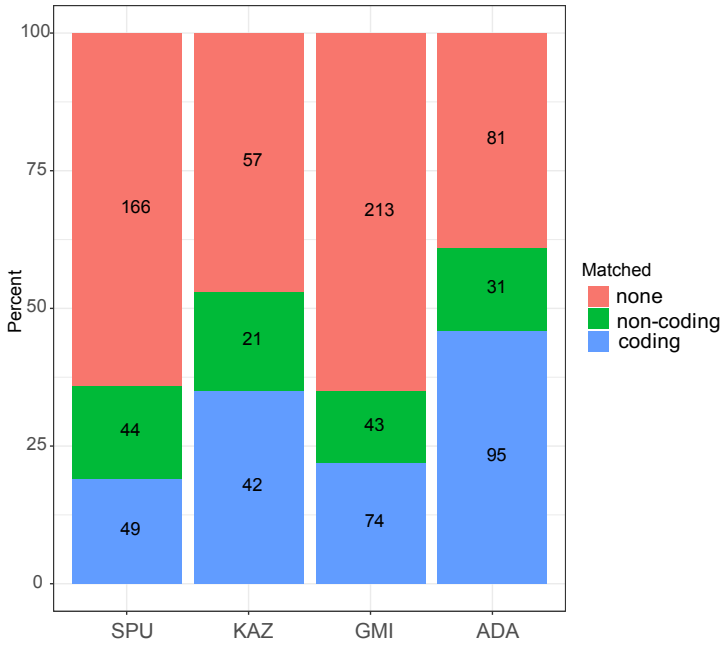


Figure 5. Stacked barplot showing the percentages of outliers giving a GenBank match to a gene (protein or mRNA coding) (blue bottom stack), a sequence or chromosome location without gene information (green middle stack) or not match found (red top stack). Actual numbers of outlier loci are given inside each stack.

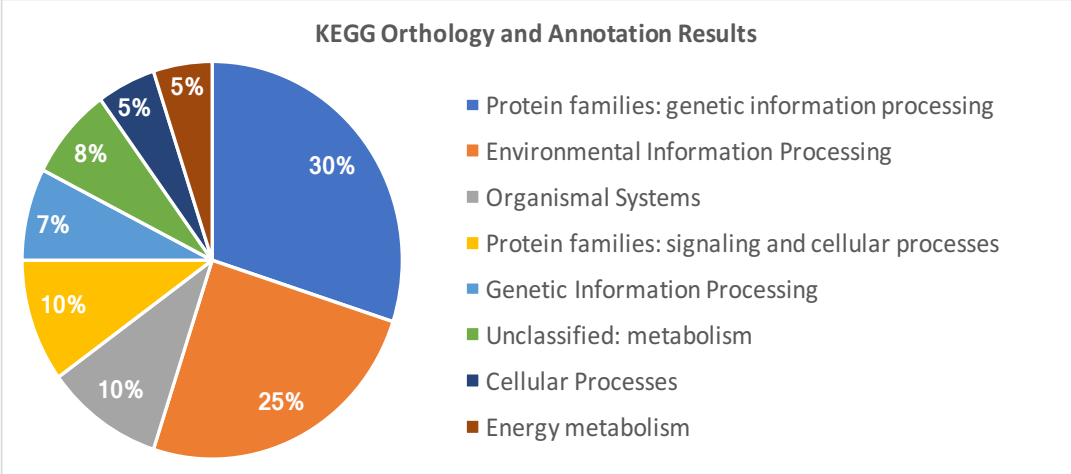


Figure 6. KEGG assignments of outlier loci (matching to the same gene in GenBank in more than one species) using the KEGG's orthology and link annotation tool (KOALA). Pie depicts the relative percentage of each category.

Table 4. List and function of genes (from outlier loci) presumably under selection and that are diverging between Pacific and Gulf populations of more than one species. Numbers give F_{ST} values (or range of values if multiple loci within a species matched the same gene family or gene type) between disjunct populations in each species. SPU, *Semichossyphus pulcher;* KAZ, *Kyphosus azureus;* GMI, *Gillichthys mirabilis;* ADA, *Anisotremus davidsonii.* Gene names were obtained from Blastn in GenBank. Gene functions were obtained from KEGG, GeneCards, Gene Ontology, UniProt, GTEx, and/or Entrez Gene.

Category	**Function - Pathway**	**Gene family or gene-protein type**	**SPU**	**KAZ**	**GMI**	**ADA**
Metabolism	Degradation pathways of several amino acids. Lipid processing	Dehydrogenase	0.28			0.38-0.46
	Olfactory transduction. Transport of salt, glucose, bile & acids. Functioning of cardiac, skeletal & smooth muscles	Anoctamin	0.25	0.26		
	Respiration, heat production	Ubiquinol-cytochrome c 1	0.2	0.4		
Development	Cell growth, signaling, cytoskeleton remodeling	ArfGAP		0.28	0.81	
	Cell adhesion, differentiation, migration & signaling	Laminin			0.79-0.98	0.54
	Cell shape, membrane proteins, organization of organelles. Gene expression and neural activity	Spectrin			0.84	0.63
	Extracellular matrix structural constituent	Collagen		0.41		0.4-0.64

Table 4 (*Continued.)*

Category	Function - Pathway	Gene family or gene-protein type	SPU	KAZ	GMI	ADA
Develop. or Reproduction	Pituitary development	LIM homeobox		0.27	0.95	
Immune response	Pathogen-associated molecular patterns	NLRC3	0.16	0.26		
	Cell cycle as well as dendrite growth and neuronal migration	F-box		0.33		0.38
	Apoptosis, integrin-mediated signal transduction, and cell transformation	Guanine	0.16			0.39-0.94
	Transmembrane signaling receptor activity and carbohydrate binding	Sialoadhesin	0.15		0.97	
	Apoptosis. Obsolete signal transducer activity and palmitoyltransferase activity Nucleic acid binding and chromatin binding.	Zinc finger	0.21	0.34		0.52-0.97
	Mediate proteolysis, autocatalytic degradation suppresses cellular proliferation	E3 ubiquitin	0.2		0.9-0.95	0.76
Multiple	Synaptic transmission: Transmission across chemical synapses. Calcium ion and calcium-dependent phospholipid binding	Synaptotagmin		0.27		0.57

Furthermore, 15 outlier loci for sargo, 8 for mudsucker, 9 for zebraperch, and 8 for sheephead, were matched the same gene or gene family than another loci from a different species (Table 4). All together, these 40 loci matched 15 different genes or gene families in GenBank. GeneCards' function annotations for these genes provided numerous cellular processes and gene pathways generally classified within metabolism, development, immune response, and possibly reproduction (see discussion). KEGG's orthology and annotation results classified convergent genes into 9 different functional categories (Figure 6). More than half of all loci were classified into two categories: genes used to process genetic information, and genes processing environmental information.

Discussion

The Baja California peninsula offers a rare opportunity to study the effects of isolation in populations of multiple marine fishes with diverse ecologies that have been subjected to similar selective pressures. Here, we obtained RADseq data from two disjunct fishes with hypnotized low gene flow, mudsucker and sargo, and two with high gene flow across the peninsula, zebraperch and California sheephead, to explore the genomic signatures of different micro-evolutionary processes (drift and selection) and search for patterns of convergent evolution in these species. The established pipeline was successful in producing substantial number of loci in which to scanned these populations and investigate our questions. While previous analyses studied disjunct species with a handful of markers (Present 1987; Tranah & Allen 1999; Terry et al. 2000; Huang & Bernardi 2001; Stepien et al. 2001; Bernardi et al.

2003), this study compared populations with approximately 5,000 to 10,000 (depending on the species) unlinked loci that passed rigorous filters. (Table 3). An important number of these loci, approximately 300 to 3000, were also polymorphic in the different populations.

Differentiation, Gene Flow, and Polymorphism

Genomic analyses agreed with previously observed genetic divergence across the peninsula in the sargo and mudsucker but uncovered undetected patterns of differentiation in the zebraperch and California sheephead. The high average of genomic differentiation from thousands of loci in the sargo and mudsucker (F_{ST}= 0.1 and 0.3, respectively; Table 3) was consistent with the hypothesis that these species are in the initial stages of allopatric speciation (Bernardi et al. 2003). Moreover, the higher level of differentiation in mudsucker than in sargo was also concordant with a proposed older divergence between Pacific and Gulf goby populations (Bernardi & Lape 2005; Huang & Bernardi 2001; Bernardi et al. 2003).

While no divergence between zebraperch populations across the peninsula was previously found and high levels of gene flow were predicted (Bernardi et al. 2003), our results revealed a low but significant differentiation (i.e. higher than expected, $F_{ST=}$0.03, p-value <0.05) and refuted the potential panmixia between disjunct zebraperch populations. Similarly, when analyzing microsatellite and mitochondrial variation throughout the distribution of the sheephead, Poortvliet et al. (2013) found that even in the lack of divergence between populations in their data,

the partition between Pacific and Gulf clusters explained 9.1% of the total variance. Authors concluded that genetic divergence could be present but it has not been detected yet. Although high structure was here seen in outlier loci, our results rather confirm high levels of gene flow as indicated by the lack of structure in neutral loci and the extremely low and non-significant average divergence between sheephead populations.

Interestingly, in every species except for the sheephead, the polymorphism was higher in the Gulf than in Pacific populations (especially for zebraperch and mudsucker). This might be a reflection of older origins for the Gulf populations and the environmental variability within the Sea of Cortez. Bernardi and Lape (2005) estimated based on coalescence analysis applied to mitochondrial cyt b that Gulf populations of the sargo might be 330-641 thousand years old and Pacific populations 163-317 thousand years old. While congenerics of the sheephead only occur in temperate latitudes, ancestral lineages for the zebraperch, sargo, and mudsucker, were present in tropical waters. Gulf populations in the last three species have likely maintained high proportions of standing variation from their lineages. These species might have been migrating north and the uplifting of the peninsula could have produced peripheral populations in the Pacific with less genetic variation.

Population Structure and Signatures of Drift

Results showed very distinct genetic pools for Pacific and Gulf sargo and mudsucker populations providing more evidence that these populations are in the

process of speciation. Their genetic variation clustered with no overlap in our DAPC analyses and both, neutral and outlier loci, showed high structure between the two regions (Figure 2). Sargo populations illustrated deeper structure in neutral loci than the mudsucker, which might be a sign of stronger drift due to smaller effective population size in this species (Figure 3). Yet, the structure seen in the mudsucker was also significant. In contrast, the genetic variation of Pacific and Gulf populations of both zebraperch and sheephead, showed some differentiation but their clusters do overlap in the DAPC plots. Interestingly, zebraperch also illustrated high structure when using suspected neutral loci. Thus, the separation of Pacific and Gulf zebraperch populations might have happened at a relatively recent time, enough for drift to create the observed structure but too short to produce divergence that can detected by first generation genetic tools. Concordant with the proposed high levels of gene flow, sheephead does not show discernable patterns of structure between its populations in neutral loci. Pacific and Gulf populations for all species appeared to belong to distinct genetic pools when analyzing outlier loci.

Outlier Loci and Signatures of Convergent Selection

Our analyses support the presence of unique selective pressures in the Pacific and Gulf. On one side of the Baja California peninsula, the Pacific distributions of these species cross multiple upwelling regions and geographic points such as Point Conception and Punta Eugenia where oceanographic conditions change drastically (Allen et al. 2006). On the other side, the Sea of Cortez represents a very dynamic

and diverse environment for fishes as well. It holds multiple upwelling regimes, average salinity is higher than in the Pacific, and temperature changes drastically with the seasons and from north to south (Thomson et al. 2000). Oceanographic parameters are also affected by more than one hundred islands throughout the Gulf, deep trenches in the middle, and a shallow shelf as well as exaggerated tides in the northern Gulf (Thomson et al. 2000). Thus, it is not surprising that very high levels of structure were observed in the outlier loci of every species (Figure 3). Moreover, we detected a large number of outlier loci and these contained high F_{ST} ranges. The separation of disjunct populations of sargo and mudsucker might be older than in zebraperch (sheephead populations might have never been separated at all) producing many fixed differences (see results section) and noticeably higher average differentiation not only in outlier but all loci.

Strong evidence of selection is presented by the fact that almost half of all the outlier loci in sargo (95 loci or 46%), 74 (22%) in mudsucker, 42 (35%) in zebraperch, and 49 (19%) in sheephead, were directly matched to coding genes by the Blastn tool in GenBank. Many other outlier loci (13 to 18% per species) were similarly matched to a chromosomal location but further investigation is needed to determine whether or not these regions in the genome are adaptive. More importantly, we identified 40 loci diverging convergently in two or three of the four studied species (Supplementary Table 2). The pairing of sargo with zebraperch, which are the two species with the closest ecologies, produced a slightly larger number of diverging genes than any other species pairing. Nonetheless, assertively associating groups of

genes to particular ecologies requires analyze a larger number of species as diverging genes were well-distributed among almost every possible species combination (Table 4).

KEGG annotation classified 25% these loci into genes used to process environmental information (Figure 6). This was KEGG's second largest classification, only after the expected differences in "housekeeping genes" or genes used to process genetic information. General categories of suggested gene pathways and functions for these genes included metabolism, development, immune response and possibly reproduction. The last one refers to two loci, one in zebraperch and one in mudsucker, that were matched to LIM Homeobox genes which have been found to be overexpress in the pituitary and testes (National Institutes of Health's data base "Genotype-Tissue Expression, GTEx"). These genes might be involved in the development of the pituitary (GeneCards) which regulates many fish reproductive processes and mediates the interaction between the environment and reproductive organs (Schreibman et al. 1973; Yamazaki 1965). Other potential pathways of these genes worth noticing include lipid processing, olfactory transduction, salt transport, and heat production (Table 4). However, the annotation search engines used in this study and much of the knowledge of the gene functions focus in providing information pertinent to humans. While the function of orthologous genes is often maintained a large variety of organisms, the particular manner in which genes matched to our identified outlier loci partake in cellular pathways and provide adaptive advantages to these fishes is outside the scope of our analyses. Overall, we

interpret our findings as evidence of convergent evolution in these four species and suspect that other disjunct species might be sharing similar evolutionary trajectories as well.

Isolation in the Disjunct Populations

The warmer seawater in the southern portions of the peninsula is probably limiting the gene flow between Pacific and Gulf disjunct populations as commonly assumed. Yet, temperature alone cannot be responsible for the isolation (Bernardi et al. 2003). During the summer, the surface temperature in the north of the Gulf can be the same and sometimes even higher than in the south of the peninsula. Observed divergence has probably resulted from the combination of several biotic and abiotic factors such as competition, diet and habitat specializations, as well as differences in salinity and oceanographic regimes. For instance, sargo and zebraperch both showed comparable structure trends in neutral and outlier loci suggesting that they might be following related evolutionary routes. The lack of tide pools (which they both use as juveniles) and the presence of congeneric species that are more numerous at the southern portion of the Sea of Cortez might be affecting dispersal in both of these species. As an exclusive herbivore, zebraperch might be more dependent on food availability or also displaced by competition from other species of herbivores such as the opaleye, *Girella simplicidens*, which is very abundant in the south.

Furthermore, tidal influences and the choice of habitat might be strong sources of isolation in the mudsucker to the extent that we see substantial genetic separation

between our three Gulf *Esteros,* which are geographically really close to each other. In fact, the other two recognized species in the genus (the shortjaw mudsucker, *G. seta,* and the delta mudsucker, *G. detrusus*) likely speciated from the longjaw mudsucker by specializing to distinct habitats (Swift et al. 2011; Barlow 1961). The longjaw mudsucker is often found very high in the intertidal in small soft bottom channels of very large estuaries (as in many of our sites). Thus, it is not uncommon for this habitat to get completely isolated from the rest of the estuaries at low tides. The northern Sea of Cortez is a particularly extreme case where the ocean is very shallow, estuaries very large, and tides can be very long (Thomson et al. 2000). This along with the paucity of estuaries systems in the southern coast of the peninsula can help explain the isolation between the disjunct populations as well as the substantial higher differentiation seen within the Gulf than the Pacific. In the case of the California sheephead, it has been suggested that this species might migrate around the peninsula following deep reefs as records of its presence in the south are rare but do exist from deep collections (Bernardi et al. 2003; Poortvliet et al. 2013). In this manner, this species is able to maintain gene flow across the peninsula without conspicuous shallow populations. Therefore, the classification of this species as a Baja California disjunct might have to be reconsidered.

Conclusion

The purpose of this study was to search for common patterns of evolution among the studied Baja California disjunct fishes which illustrate very different

ecologies and life histories but at the same time share a very similar distribution. All of our analyses are concordant with the formerly proposed hypothesis where sargo and mudsucker disjunct populations are in the process of allopatric speciation. Signatures of prolonged isolation in these species were evident as deep genomic divergence, strong structure in neutral loci, and high differentiation in outlier loci. Results further uncovered previously unseen structure in disjunct zebraperch populations which might have only recently become isolated. All species, except for sheephead, show signs of genetic drift in the form of structure in neutral loci. California sheephead Pacific and Gulf populations did not show evidence implying that they ever separated. Instead, this species presented the highest levels of gene flow, most like achieved by migrating around the peninsula by following deep reefs. Disjunct species experience very different environments along their distributions within and between the northern Eastern Pacific and the Sea of Cortez. Our results strongly support the presence of unique selective pressures in the Pacific and Gulf environments. We identified specific adaptive genomic regions that are simultaneously contributing to the divergence across the peninsula and that might be assisting local adaptation of species with diverse ecologies. While exemplifying the utility of RADseq data, this study provides a useful framework to characterize genomic patterns of neutral and non-neutral population structure, and highlights the potential of outlier loci analysis to pinpoint markers that can identify specific selective pressures in isolated populations that might be in the process of allopatric speciation.

References

Allen, L. G., Pondella, D. J. & Horn, M. h. (2006) *The Ecology of Marine Fishes: California and Adjacent Waters*. UC Press.

Baird, N. A., Etter, P. D., Atwood, T. S., Currey, M. C., Shiver, A. L., Lewis, Z. A., Selker, E. U., Cresko, W. A. & Johnson, E. A. (2008) Rapid SNP Discovery and Genetic Mapping Using Sequenced RAD Markers. *PLoS ONE* **3**, 1–7.

Barlow, G. W. (1961) Gobies of the Genus Gillichthys , with Comments on the Sensory Canals as a Taxonomic Tool. *American Society of Ichthyologists and Herpetologists* 423–437.

Bay, L. K., Crozier, R. H. & Caley, M. J. (2006) The Relationship between Population Genetic Structure and Pelagic Larval Duration in Coral Reef Fishes on the Great Barrier Reef. *Marine Biology* **149**, 1247–1256.

Bernardi, G. & Lape, J. (2005) Tempo and Mode of Speciation in the Baja California Disjunct Fish Species Anisotremus Davidsonii. *Molecular Ecology* **14**, 4085–4096.

Bernardi, G., Findley, L. & Rocha-Olivares, A. (2003) Vicariance and Dispersal across Baja California in Disjunct Marine Fish Populations. *Evolution; international journal of organic evolution* **57**, 1599–1609.

Bernardi, G., Beldade, R., Holbrook, S. J. & Schmitt, R. J. (2012) Full-Sibs in Cohorts of Newly Settled Coral Reef Fishes. *PLoS ONE* **7**.

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**, 3384–3396.

Bierne, N., Welch, J., Loire, E., Bonhomme, F. & David, P. (2011) The Coupling Hypothesis: Why Genome Scans May Fail to Map Local Adaptation Genes. *Molecular Ecology* **20**, 2044–2072.

Bierne, N., Roze, D. & Welch, J. J. (2013) Pervasive Selection or Is It.? Why Are FSToutliers Sometimes so Frequent? *Molecular Ecology* **22**, 2061–2064.

Bowen, B. W., Bass, A. L., Muss, A., Carlin, J. & Robertson, D. R. (2006) Phylogeography of Two Atlantic Squirrelfishes (Family Holocentridae): Exploring Links between Pelagic Larval Duration and Population Connectivity. *Marine Biology* **149**, 899–913.

Catchen, J., Hohenlohe, P. A., Bassham, S., Amores, A. & Cresko, W. A. (2013) Stacks: An Analysis Tool Set for Population Genomics. *Molecular Ecology* **22**, 3124–3140.

Catchen, J. M., Amores, A., Hohenlohe, P., Cresko, W. & Postlethwait, J. H. (2011) *Stacks* : Building and Genotyping Loci *De Novo* From Short-Read Sequences. *G3: Genes|Genomes|Genetics* **1**, 171–182.

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

Earl, D. & VonHoldt, B. (2012) STRUCTURE HARVESTER: A Website and Program for Visualizing STRUCTURE Output and Implementing the Evanno Method. *Conservation Genetics Resources* **4**, 359–361.

Endler, J. A. (1977) *Geographic Variation, Speciation and Clines*. Princeton University Press.

Evanno, G., Regnaut, S. & Goudet, J. (2005) Detecting the Number of Clusters of Individuals Using the Software STRUCTURE: A Simulation Study. *Molecular Ecology* **14**, 2611–2620.

Excoffier, L. & Lischer, H. E. L. (2010) Arlequin Suite Ver 3.5: A New Series of Programs to Perform Population Genetics Analyses under Linux and Windows. *Molecular Ecology Resources* **10**, 564–567.

Gaither, M. R., Bernal, M. A., Coleman, R. R., Bowen, B. W., Jones, S. A., Simison, W. B. & Rocha, L. A. (2015) Genomic Signatures of Geographic Isolation and Natural Selection in Coral Reef Fishes. *Molecular Ecology* **24**, 1543–1557.

Helfman, G. S., Collette, B. B., Facey, D. E. & Bowen, B. W. (2009) *The Diversity of Fishes*, Second. Wiley-Blackwell.

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**, 421–428.

Jombart, T. (2008) Adegenet: A R Package for the Multivariate Analysis of Genetic Markers. *Bioinformatics* **24**, 1403–1405.

Jombart, T., Devillard, S., Balloux, F., Falush, D., Stephens, M., Pritchard, J., Pritchard, J., Stephens, M., Donnelly, P., Corander, J., et al. (2010) Discriminant Analysis of Principal Components: A New Method for the Analysis of Genetically Structured Populations. *BMC Genetics* **11**, 94.

Jones, G. P., Jones, G. P., Milicich, M. J., Milicich, M. J., Emslie, M. J., Emslie, M. J., Lunow, C. & Lunow, C. (1999) Self-Recruitment in a Coral Reef Fish Population. *Nature* **402**, 802–804.

Kanehisa, M., Araki, M., Goto, S., Hattori, M., Hirakawa, M., Itoh, M., Katayama, T., Kawashima, S., Okuda, S., Tokimatsu, T., et al. (2008) KEGG for Linking Genomes to Life and the Environment. *Nucleic Acids Research* **36**, 480–484.

Leis, J. M. & Carson-Ewart, B. M. (2000) Behaviour of Pelagic Larvae of Four Coral-Reef Fish Species in the Ocean and an Atoll Lagoon. *Coral Reefs* **19**, 247–257.

Leis, J. M. & Carson-Ewart, B. M. (2002) In Situ Settlement Behaviour of Damselfish (Pomacentridae) Larvae. *Journal of Fish Biology* **61**, 325–346.

Leis, J. M. & Lockett, M. M. (2005) Localization of Reef Sounds by Settlement Stage Larvae of Coral Reef Fishes (Pomacentridae). *Bullentin of Marine Science* **76**, 715–724.

Leis, J. M., Carson-Ewart, B. M., Hay, A. C. & Cato, D. H. (2003) Coral-Reef Sounds Enable Nocturnal Navigation by Some Reef-Fish Larvae in Some Places and at Some Times. *Journal of Fish Biology* **63**, 724–737.

Lischer, H. E. L. & Excoffier, L. (2012) PGDSpider: An Automated Data Conversion Tool for Connecting Population Genetics and Genomics Programs. *Bioinformatics* **28**, 298–299.

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.

Lotterhos, K. E. & Whitlock, M. C. (2015) The Relative Power of Genome Scans to Detect Local Adaptation Depends on Sampling Design and Statistical Method. *Molecular Ecology* **24**, 1031–1046.

Medina, M. & Walsh, P. J. (2000) Comparison of Four Mendelian Loci of the California Sea Hare (Aplysia Californica) from Populations of the Coast of California and the Sea of Cortez. *Marine Biotechnology* **2**, 449–455.

Miller, D. J. & Lea, R. N. (1972) *Guide to the Coastal Marine Fishes of California.* Berkeley, CA.

Miller, M., Dunham, J., Amores, a, Cresko, W. & Johnson, E. (2007) Genotyping Using Restriction Site Associated DNA (RAD) Markers. *Genome Research* **17**, 240–248.

Miller, M. R., Brunelli, J. P., Wheeler, P. A., Liu, S., Rexroad, C. E., Palti, Y., Doe, C. Q. & Thorgaard, G. H. (2012) A Conserved Haplotype Controls Parallel Adaptation in Geographically Distant Salmonid Populations. *Molecular Ecology* **21**, 237–249.

Ogata, H., Goto, S., Sato, K., Fujibuchi, W., Bono, H. & Kanehisa, M. (1999) KEGG: Kyoto Encyclopedia of Genes and Genomes. *Nucleic Acids Research* **27**, 29–34.

Palumbi, S. R. (1992) Marine Speciation on a Small Planet. *Trends in Ecology & Evolution* **7**, 114–118.

Planes, S., Lecaillon, G., Lenfant, P. & Meekan, M. (2002) Genetic and Demographic Variation in New Recruits of Naso Unicornis. *Journal of Fish Biology* **61**, 1033–1049.

Poortvliet, M., Longo, G. C., Selkoe, K., Barber, P. H., White, C., Caselle, J. E., Perez-Matus, A., Gaines, S. D. & Bernardi, G. (2013) Phylogeography of the California Sheephead, Semicossyphus Pulcher: The Role of Deep Reefs as Stepping Stones and Pathways to Antitropicality. *Ecology and Evolution* **3**, 4558–4571.

Present, T. M. C. (1987) Genetic Differentiation of Disjunct Gulf of California and Pacific Outer Coast Populations of Hypsoblennius Jenkinsi Genetic Differentiation of Disjunct Gulf of California and Pacific Outer Coast Populations of Hypsoblennius Jenkinsi. *Copeia* **1987**, 1010–1024.

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

Pujolar, J. M., Maes, G. E. & Volckaert, F. A. M. (2006) Genetic Patchiness among Recruits in the European Eel Anguilla Anguilla. *Marine Ecology Progress Series* **307**, 209–217.

R Core Team. (2013) R: A Language and Environment for Statistical Computing. Vienna, Austria: R Foundation for Statistical Computing 2013, doi:ISBN 3-900051-07-0.

Rocha, L. A., Bass, A. L., Robertson, D. R. & Bowen, B. W. (2002) Adult Habitat Preferences, Larval Dispersal, and the Comparative Phylogeography of Three Atlantic Surgeonfishes (Teleostei: Acanthuridae). *Molecular Ecology* **11**, 243–252.

Schreibman, M. P., Leatherland, J. F. & McKeown, B. A. (1973) Functional Morphology of the Teleost Pituitary Gland. *American Zoologist* **13**, 719–742.

Selkoe, K. A., Gaines, S. D., Caselle, J. E. & Warner, R. R. (2006) Current Shifts and Kin Aggregation Explain Genetic Patchiness in Fish Recruits. *Ecology* **87**, 3082–3094.

Shulman, M. & Bermingham, E. (1995) Early Life Histories , Ocean Currents , and the Population Genetics of Caribbean Reef Fishes. *Evolution* **49**, 897–910.

Stepien, C. A., Rosenblatt, R. H. & Bargmeyer, B. A. (2001) Phylogeography of the Spotted Sand Bass , Paralabrax Maculatofasciatus : Divergence of Gulf of California and Pacific Coast Populations. *Evolution* **55**, 1852–1862.

Stockwell, B. L., Larson, W. A., Waples, R. K., Abesamis, R. A., Seeb, L. W. & Carpenter, K. E. (2016) The Application of Genomics to Inform Conservation of a Functionally Important Reef Fish (Scarus Niger) in the Philippines. *Conservation Genetics* **17**, 239–249.

Swearer, S. E., Shima, J. S., Hellberg, M. E., Thorrold, S. R., Jones, G. P., Robertson, D. R., Morgan, S. G., Selkoe, K. a, Ruiz, G. M. & Warner, R. R. (2002) Evidence of Self Recruitment in Demersal Marine Populations. *Bulletin of Marine Science* **70**, 251–271.

Swift, C. C., Findley, L. T., Ellingson, R. a, Flessa, K. W. & Jacobs, D. K. (2011) The Delta Mudsucker, Gillichthys Detrusus, a Valid Species (Teleostei: Gobiidae) Endemic to the Colorado River Delta, Northernmost Gulf of California, Mexico. *Copeia* **2011**, 93–102.

Terry, A., Bucciarelli, G. & Bernardi, G. (2000) Restricted Gene Flow and Incipient Speciation in Disjunct Pacific Ocean and Sea of Cortez Populations of a Reef Fish Species , Girella Nigricans. *Evolution* **54**, 652–659.

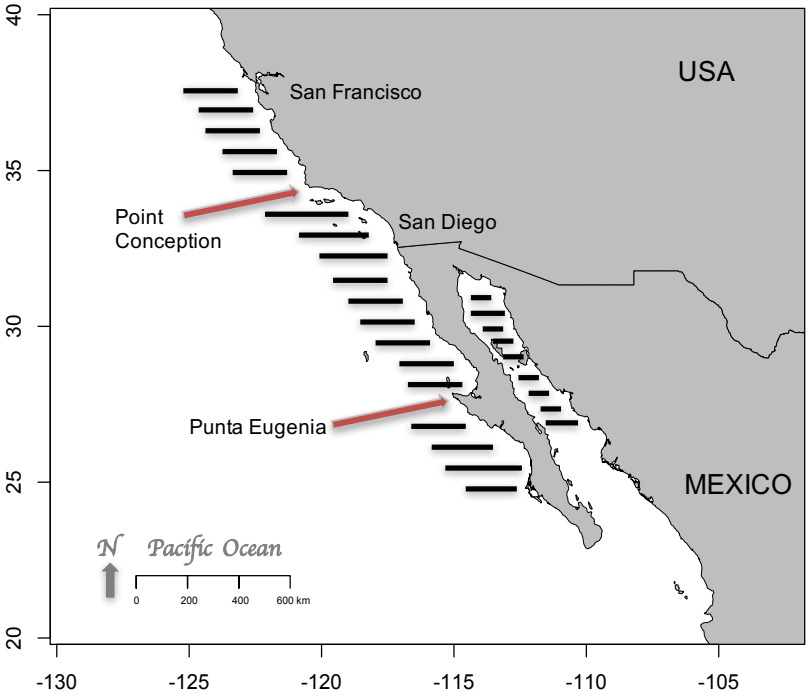
Thomson, D. A., Findley, L. T. & Kersitch, A. N. (2000) *Reef Fishes of the Sea of Cortez: The Rocky Shore Fishes of the Gulf of California.* Austin, TX: The University of Texas Press.

Tranah, G. J. & Allen, L. G. (1999) *Morphologic and Genetic Variation among Six Populations of the Spotted Sand Bass, Paralabrax Maculatofasciatus, from Southern California to the Upper Sea of Cortez*. Los Angeles, CA.

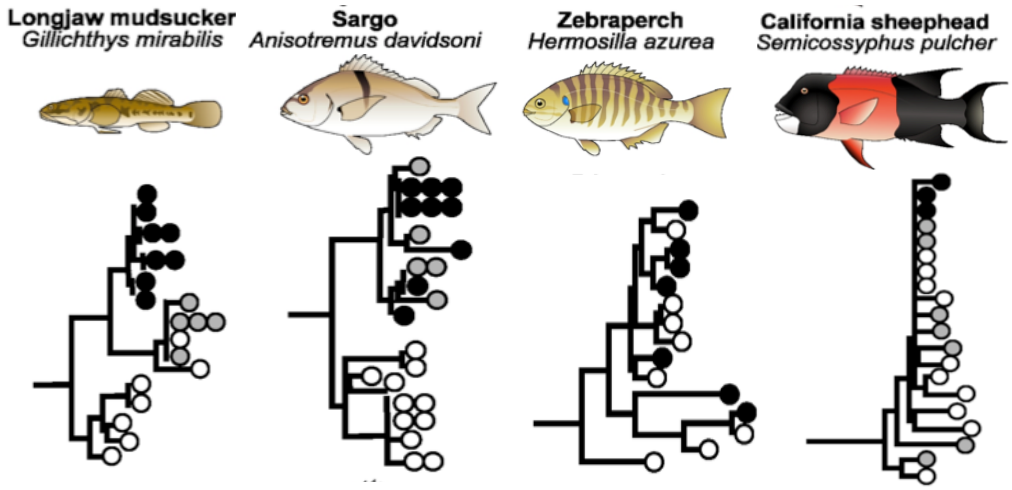
Wickham, H. (2016) *Ggplot2: Elegant Graphics for Data Analysis*. New York.

Yamazaki, F. (1965) *Endocrinological Studies on the Reproduction of the Femail Goldfish, Carassius Auratus L., with Special Reference to the Function of the Pituitary Gland*. Vol. 13.

Supplemental Information:



Supplementary Figure 1. Typical Baja California disjunct distribution and relevant phylogeograpic breaks in the Northeastern Pacific Ocean (Point Conception and Punta Eugenia).



Supplementary Figure 2. Phylogenetic relationships (mtDNA) between populations of the studied Baja California disjuncts. Pacific populations north of Punta Eugenia (black circles), south of Punta Eugenia (gray circles), and in the Sea of Cortez (white circles). Note reciprocal monophyly by Pacific and Sea of Cortez populations in sargo and mudsucker. The two white circles within the mudsucker gray clade are suspected anthropogenic releases by sport fishermen (Bernardi et al. 2003).

Supplementary Table 1**.** The 19 Baja California disjunct species (Bernardi et al. 2003).

Family	**Species**	**Common name**
Atherinidae	*Leuresthes tenuis/ L. sardina*	grunion
Girellidae	*Girella nigricans / G. simplicidens*	opaleye
Haemulidae	*Anisotremus davidsonii*	sargo
Blenniidae	*Hypsoblennius jenkinsi*	mussel blenny
Chaenopsidae	*Chaenopsis alepidota*	oragethroat pikeblenny
Serranidae	*Paralabrax maculatofasciatus*	spotted sand bass
Gobiidae	*Gillichthys mirabilis*	longjaw mudsucker
Gobiidae	*Lythrytpnus dalli*	blue banded boby
Blenniidae	*Hypsoblennius gentilis*	bay blenny
Kyphosidae	*Kyphosus azureus*	zebraperch
Labridae	*Halichoeres semicinctus*	rock wrasse
Labridae	*Semicossyphus pulcher*	California sheephead
Embiotocidae	*Zalembius rosaceus*	pink surfperch
Scorpaenidae	*Sebastes macdonaldi*	Mexican rockfish
Scorpaenidae	*Scorpaena guttata*	scorpionfish
Polyprionidae	*Stereolepis gigas*	giant seabass
Agonidae	*Xenetremus ritteri*	flagfin poacher
Pleuronectidae	*Hypsopsetta guttulata*	diamond turbot
Pleuronectidae	*Pleuronichthys verticalis*	hornyhead turbot

Supplementary Figure 3. Structure Harvester Output:

Semicossyphus pulcher

10,273 neutral loci (10532 total loci -259 outlier loci). K=2 was selected.

K	Reps	Mean LnP(K)	Stdev LnP(K)	Ln'(K)	|Ln''(K)|	Delta K
1	10	-354276.510000	55.834148	—	—	—
2	10	-305608.620000	80.766451	48667.890000	361094.08000	4470.84
3	10	-618034.810000	165002.207277	-312426.190000	2376.310000	0.01440
4	10	-932837.310000	706850.782204	-314802.500000	1720110.51222	2.43348
5	9	-2967750.3222	2130977.65850	-2034913.01222	—	—

259 outlier loci (potentially under selection). K=2 was selected.

K	Reps	Mean LnP(K)	Stdev LnP(K)	Ln'(K)	|Ln''(K)|	Delta K
1	10	-11003.310000	1.568757	—	—	—
2	10	-9736.060000	1.581279	1267.250000	715.790000	452.665110
3	10	-9184.600000	73.169605	551.460000	218.750000	2.989629
4	10	-8851.890000	8.444518	332.710000	608.250000	72.028976
5	10	-9127.430000	1883.600076	-275.540000	—	—

Kyphosus azureus

4738 neutral loci (4858 total loci -120 outlier loci). K=2 was selected.

K	Reps	Mean LnP(K)	Stdev LnP(K)	Ln'(K)	|Ln''(K)|	Delta K
1	10	-115762.090000	29.875871	—	—	—
2	10	-131925.930000	17087.715398	-16163.840000	2401051.460	140.513311
3	10	-2549141.2300	1680878.7350	-2417215.3000	1992290.290	1.185267
4	10	-2974066.2400	1793501.1344	-424925.01000	5426858.530	3.025846
5	10	-8825849.7800	5254278.8003	-5851783.5400	—	—

120 outlier loci (potentially under selection). K=2 was selected.

K	Reps	Mean LnP(K)	Stdev LnP(K)	Ln'(K)	|Ln''(K)|	Delta K
1	11	-4032.263636	1.145664	—	—	—
2	11	-3226.272727	1.842331	805.990909	660.236364	358.370200
3	11	-3080.518182	60.214497	145.754545	109.163636	1.812913
4	11	-2825.600000	2.618778	254.918182	167.009091	63.773664
5	11	-2737.690909	33.018130	87.909091	—	—

Gillichthys mirabilis

8,370 neutral loci (8700 total loci -330 outlier loci). K=2 was selected.

K	Reps	Mean LnP(K)	Stdev LnP(K)	Ln'(K)	|Ln''(K)|	Delta K
1	10	-394436.1400	830.231446	—	—	—
2	10	-545660.7100	476498.123	-151224.5700	397687.250	0.8346
3	10	-299198.0300	479.341210	246462.6800	3172066.25	6617.5
4	10	-3224801.600	1706842.710	-2925603.5700	5429106.71	3.1807
5	8	-11579511.88	6499308.876	-8354710.287	—	—

330 outlier loci (potentially under selection). K=2 was selected.

K	Reps	Mean LnP(K)	Stdev LnP(K)	Ln'(K)	|Ln''(K)|	Delta K
1	10	-31118.480000	2.626277	—	—	—
2	10	-17679.410000	1.779794	13439.070000	10190.760000	5725.808731
3	10	-14431.100000	4.053805	3248.310000	3203.300000	790.195917
4	10	-14386.090000	396.973090	45.010000	2480.440000	6.248383
5	10	-16821.520000	3373.243943	-2435.430000	—	—

Anisotremus davidsonii

8,031 neutral loci (8,238total loci -207 outlier loci). K=3 was selected.

K	Reps	Mean LnP(K)	Stdev LnP(K)	Ln'(K)	|Ln''(K)|	Delta K
1	10	-338481.170000	32.570063	—	—	—
2	10	-291762.360000	7195.96301	46718.8100	38095.27000	5.293978
3	10	-283138.820000	116.141167	8623.540000	9843788.000	84757.09
4	10	-10118303.370	5180965.32	-9835164.55	18794462.72	3.627599
5	10	-38747930.640	13001531.73	-28629627.2	—	—

207 outlier loci (potentially under selection). K=2 was selected.

K	Reps	Mean LnP(K)	Stdev LnP(K)	Ln'(K)	|Ln''(K)|	Delta K
1	10	-13941.350000	2.055480	—	—	—
2	10	-10560.720000	0.814862	3380.630000	1883.990000	2312.035795
3	10	-9064.080000	1.209040	1496.640000	1801.310000	1489.867966
4	10	-9368.750000	1473.669507	-304.670000	391.240000	0.265487
5	10	-10064.660000	1613.562619	-695.910000	—	—

Supplementary Table 2. List of genes which were matched to outlier loci diverging simultaneously in disjunct populations of more than one species. Gene names are actual GenBank descriptions. F_{ST} values show the divergence between Pacific and Gulf populations of *Gillichthys mirabilis* (GMI), *Anisotremus davidsonii* (ADA), *Kyphosus azureus* (KAZ), or *Semicossyphus pulcher* (SPU).

Gene	**Species**	**Fst: Pac-Gulf**
Anoctamin 5 (ano5), mRNA	KAZ	0.2558
Anoctamin-1-like (LOC109997874), mRNA	SPU	0.2473
ArfGAP with SH3 domain, ankyrin repeat and PH domain 1 (asap1), mRNA	GMI	0.8058
ArfGAPP with dual PH domain-containing protein 1-like (LOC111671840), mRNA	KAZ	0.2838
Collagen alpha-4(IV) chain-like (LOC111654436), mRNA	ADA	0.6367
Collagen alpha-5(IV) chain-like (LOC111663337), mRNA	ADA	0.5631
Collagen type VII alpha 1 chain (col7a1), transcript variant X4, mRNA	KAZ	0.4079
Collagen type XIX alpha 1 chain (col19a1), transcript variant X6, mRNA	ADA	0.4478
Collagen type XXIV alpha 1 chain (col24a1), mRNA	ADA	0.4044
Cytochrome b-c1 complex subunit 1, mitochondrial-like (LOC109980887), mRNA	SPU	0.1978
Cytochrome b-c1 complex subunit 2, mitochondrial-like (LOC109985804), mRNA	KAZ	0.397
Dehydrogenase E1 and transketolase domain containing 1 (dhtkd1), mRNA	ADA	0.3765
Dehydrogenase/reductase SDR family member on chromosome X-like (LOC109139865), partial mRNA	ADA	0.4571
Dehydrogenase/reductase X-linked (dhrsx), mRNA	SPU	0.2844
E3 ubiquitin-protein ligase HECW2-like (LOC110001920), mRNA	SPU	0.1978
E3 ubiquitin-protein ligase Midline-1 (LOC104924095), mRNA	GMI	0.9507
E3 ubiquitin-protein ligase RBBP6-like (LOC106531352), transcript variant X2, mRNA	ADA	0.7601
E3 ubiquitin-protein ligase SMURF2-like (LOC110175789), partial mRNA	GMI	0.9023
F-box protein 16 (fbxo16), mRNA	KAZ	0.3278
F-box protein 31 (fbxo31), transcript variant X2, mRNA	ADA	0.3819

Supplementary Table 2 *(Continued)*

Gene	Species	Fst: Pac-Gulf
Guanine nucleotide exchange factor 1 (sos1), SOS Ras/Rac , mRNA	ADA	0.9382
Guanine nucleotide exchange factor 5-like, rap (LOC111668812), mRNA	SPU	0.1552
Guanine nucleotide exchange factor DBS-like (LOC109951381), partial mRNA	ADA	0.3907
Laminin subunit alpha 1 (lama1), mRNA	ADA	0.5385
Laminin subunit alpha 5 (lama5), mRNA	GMI	0.976
Laminin subunit gamma-3-like (LOC110157350), mRNA	GMI	0.7902
LIM homeobox 2 (lhx2), transcript variant X2, mRNA	KAZ	0.2684
LIM homeobox transcription factor 1-alpha-like (LOC110169526), mRNA	GMI	0.9543
NLRC3-like protein (LOC104930296), mRNA	SPU	0.1607
NLRC3, NLR family CARD domain-containing protein 3-like (LOC111670284), mRNA	KAZ	0.2558
Sialoadhesin (LOC105008515), transcript variant X4, mRNA	SPU	0.1464
Sialoadhesin-like (LOC110153737), mRNA	GMI	0.9749
Spectrin alpha, non-erythrocytic 1 (sptan1), transcript variant X3, mRNA	ADA	0.6268
Spectrin beta, non-erythrocytic 4 (sptbn4), mRNA	GMI	0.8411
Synaptotagmin 12 (syt12), mRNA	ADA	0.5689
Synaptotagmin 14 (syt14), mRNA	KAZ	0.269
Zinc finger DHHC-type containing 13 (zdhhc13), mRNA	ADA	0.5214
Zinc finger E-box-binding homeobox 2-like (LOC108874500), transcript variant X2, mRNA	ADA	0.9707
Zinc finger protein 385A-like (LOC111668975), transcript variant X4, mRNA	KAZ	0.3392
Zinc finger protein 385B (znf385b), mRNA	SPU	0.2127

GENERAL CONCLUSION

The field of evolutionary biology has undergone major shifts throughout its short history, from Darwin's theory of natural selection and Mendel's fundamental work, to the reconciliation of both in the modern synthesis. Likewise, the methodologies used to study evolution have also drastically changed, from the characterization of DNA as the unit of change in organisms, to the successional development of techniques to detect changes in allele frequencies in natural populations, and to the current parallel sequencing and high throughoutput technologies. The technological advances in last few decades have forever changed the field once again, as they allow scientists to examine the mechanisms of evolutionary change with tools more powerful than ever before. In this genomic era, finding appropriate natural systems where to apply these technologies is also more crucial than ever.

The Baja California disjunct fishes represent an extraordinary natural experiment where the formation of the Baja California peninsula divided the distribution of 19 temperate marine fishes into allopatric Pacific and Sea of Cortez populations, as well as sympatric populations within each region. This natural repetition of species with populations experiences different levels of gene flow presents an excellent framework to test hypothesis of biogeography and to search for genomic signatures of drift and selection.

This dissertation has employed Restriction Site-Associated DNA sequencing (RADseq) to four of these species and revealed previously unseen genetic structure and narrower estimates for the time of disjunction helping to pinpoint the specific events that isolated the populations of these species. This dissertation also analyzed these populations with large numbers of neutral and non-neutral loci and found evidence of convergent drift and selection. Results suggested that the Baja California peninsula acts as an effective barrier to gene flow between populations of the sargo, mudsucker and to a lesser extent zebraperch, but not for the sheephead which might be maintaining connectivity using deep reefs as stepping stones from Pacific to Sea of Cortez. Point Conception and Punta Eugenia appeared to also significantly impact gene flow between Pacific populations of mudsucker and sargo, respectively. Sympatric populations within the Pacific and Sea of Cortez, showed large number of outlier loci matching protein-coding genes and emulating a complex selective landscape where individual populations might be pursuing different adaptive peaks in the face of ongoing gene flow. Observed differentiation patterns in these loci, where most fell near the lower end of the range of their differentiation values, supported a possible prevalence of soft selective sweeps brining these populations closer to their respective selective summits.

The strongest evidence of selection, however, was seen between allopatric Pacific and Sea of Cortez populations of all four species. Results strongly supported the idea of differential selection in these regions as indicated by the observed highest non-neutral differentiation (including dozens of fixed differences in populations of

sargo and mudsucker), and the largest number of outlier loci matching genes (up to 49% of all outliers in sargo, for instance). These findings, along with the higher "outlier loci evenness" or the more even distribution of loci along differentiation values observed in disjunct populations, suggested that the Pacific and Sea of Cortez might represent two very distinct selective peaks and that a more even contribution of soft and hard selective sweeps might be involved in allopatric adaptation. Therefore, low levels of gene flow might facilitate or be required for the presence of hard sweeps in the adaptation of these populations. Furthermore, the examination of outlier loci matching genes among allopatric populations of these species, exposed 15 genomic regions, potentially involved in processing environmental information, metabolism, immune response, and possibly reproduction, diverging in more than one of these species at the same time. While the differentiation of these loci should be validated larger population samples, these results implied the presence of convergent evolution in these species in spite of their very different ecologies and life history strategies.

This dissertation highlights the potential of RADseq to reveal genomic signatures of evolutionary mechanisms, and study evolutionary histories of organisms as well as the interaction between populations and their environments. The unprecedented capabilities to generate information of modern genetic tools, offer a real opportunity to improve our ability to protect natural populations at a time when many organisms are endangered by anthropogenic activity, climatic changes, and other natural pressures.