



Symposium Article

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*)

Eric Garcia^o, W. Brian Simison, and Giacomo Bernardi^o

From the Ecology and Evolutionary Biology Department, Long Marine Laboratory, University of California Santa Cruz, 1156 High St, Santa Cruz, CA 95064 (Garcia and Bernardi); Department of Biological Sciences, Old Dominion University, 5115 Hampton Blvd. Norfolk, VA 23529 (Garcia); and Center for Comparative Genomics, California Academy of Sciences, 55 Music Concourse Dr. San Francisco, CA 95064 (Garcia and Simison)

Address correspondence to E. Garcia at the address above, or e-mail: e1garcia@odu.edu

Received December 14, 2018; First decision March 31, 2019; Accepted November 14, 2019.

Corresponding Editor: Rosemary Gillespie

Abstract

Studying how isolation can impact population divergence and adaptation in co-distributed species can bring us closer to understanding how landscapes affect biodiversity. The Sargo, *Anisotremus davidsonii* (Haemulidae), and the Longjaw mudsucker, *Gillichthys mirabilis* (Gobiidae), offer a notable framework to study such mechanisms as their Pacific populations cross phylogeographic breaks at Point Conception, California, United States, and Punta Eugenia, Mexico, and are separated to those in the Sea of Cortez by the Baja California peninsula. Here, thousands of loci are genotyped from 48 Sargos and 73 mudsuckers using RADseq to characterize overall genomic divergence, and search for common patterns of putatively neutral and non-neutral structure based on outlier loci among populations with hypothesized different levels of isolation. We further search for parallels between population divergence and the total proportion of outliers, outlier F_{ST} distribution, and the proportion of outliers matching coding regions in GenBank. Statistically significant differentiation is seen across Point Conception in mudsucker ($F_{ST} = 0.15$), Punta Eugenia in Sargo ($F_{ST} = 0.02$), and on either side of the Baja California peninsula in both species ($F_{ST} = 0.11$ and 0.23 , in Sargo and mudsucker, respectively). Each species shows structure using neutral and non-neutral loci. Finally, higher population divergence yields a more even distribution of outliers along their differentiation range but does not always translate into higher outlier proportions or higher rates in which outliers are matched to coding regions. If repeated in similar systems, observed genomic patterns might reveal speciation signatures in diverse networks of population isolation.

Subject areas: Population structure and phylogeography, Molecular adaptation and selection

Keywords: adaptive evolution, Baja California, differential selection, ecological divergence, incipient speciation, peripheral populations

Under the biological species concept, speciation is the process by which 2 or more populations accumulate sufficient genetic differences to reach reproductive isolation (Mayr 1942). 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 (Coyne and Orr 2004; Rocha and Bowen 2008; Bernardi 2013; Bowen et al. 2013).

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 of gene flow, which at high levels may prevent the accumulation of advantageous alleles in populations (Coyne and Orr 2004; Rundle and Nosil 2005; Bernardi 2013; Sexton et al. 2014). Yet, the effects of gene flow are more complex than a simple unequivocal homogenization of genetic variation. 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 historical events and the interplay of processes as drift, selection, and the level of gene flow between populations. Studies examining divergence mechanisms can shed light on the evolutionary histories of species and the similarities and differences between the way distinct modes of speciation operate in populations and allow them to persist and adapt to different environments and landscapes.

Sympatric and Allopatric Populations of the Baja California Disjunct Fishes

Compared to freshwater systems, marine fishes with allopatric populations are scarce due to the paucity of hard physical 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 (Rocha et al. 2002; Rocha and Bowen 2008; Helfman et al. 2009; Bernardi 2013; Bowen et al. 2013). The Baja California disjunct species are a rare example of temperate marine fishes (19 species) whose populations have been isolated by specific vicariant and dispersal events (Brusca 1973; Riddle et al. 2000; Bernardi et al. 2003) (see Supplementary Table S1 for the full list of species). These species include sympatric populations that have no geographic barriers to dispersal throughout their Pacific distribution from central California to approximately Bahía Magdalena, in southern Baja California (Bernardi et al. 2003; Allen et al. 2006; Miller and Lea 1972). Likewise, populations occur without obvious geographic barriers in the northern and central zones of the Sea of Cortez (also known as the Gulf of California, and from here on, also referred as the Gulf) (Thomson et al. 2000). However, Gulf populations are separated from the Pacific by the Baja California peninsula and putatively maintained in isolation by factors such as seawater temperature, lack of suitable habitat, and salinity gradients at the entrance to the Sea of Cortez, where these species are either rare or absent (see Supplementary Figure S1 for a representation of the typical Baja California disjunct species distribution) (Miller and Lea 1972; Thomson et al. 2000; Bernardi et al. 2003; Allen et al. 2006; Inda-Díaz et al. 2014; Campbell et al. 2018; García-De León et al. 2018). Other potential isolating mechanisms for northern Sea of Cortez populations include current eddies in the

northern Gulf and oceanic fronts that are present across the Midriff archipelago region, where several islands are scattered from east to west in the northern Gulf (Bernardi et al. 2003; Inda-Díaz et al. 2014; García-De León et al. 2018). This study considers the studied Gulf sites to be allopatric to those in the Pacific. Comparisons among these sites will be called trans-peninsular, disjunct, or allopatric. Sites within the Pacific and Gulf distributions are here called cis-peninsular but referred as sympatric when discussing both focal species.

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 also affected by environmental factors (Huang and Bernardi 2001; Bernardi et al. 2003; Bernardi and Lape 2005; Allen et al. 2006). 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; Briggs and Bowen 2012; Bowen et al. 2013). 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 in this area potentially due to lack of habitat and strong upwelling north of this point (Bernardi 2000; Bernardi and Talley 2000; Terry et 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 around 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 different levels of gene flow and a noteworthy system to study the mechanisms of divergence and signals of selection under different scenarios of isolation.

Despite significant ecological differences, the Sargo, *Anisotremus davidsonii* (grunts, Haemulidae), and the Longjaw 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 between disjunct populations (Table 1). Moreover, 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 and Lape 2005). However, since divergence is not homogeneous along the genome, these patterns might differ when assessing divergence with an increased number of markers from coding and non-coding regions (Nosil et al. 2009).

The present study scans thousands of loci throughout the genome using Restriction site-Associated DNA sequencing (RADseq) to characterize the genomic structure between sympatric and allopatric Northeastern Pacific and Sea of Cortez populations of Sargo and Longjaw mudsucker. We compare genomic structure to previously documented mitochondrial divergence and predict higher population differentiation across phylogeographic breaks and around the peninsula using RADseq data. Subsequently, by analyzing differentiation in neutral and outlier loci separately, we search for indications of having multiple drivers, such as drift and selection, creating population structure across the different levels of isolation. As Pacific and Sea of Cortez populations experience the highest environmental contrast, and are previously proposed to be reciprocal monophyletic, we hypothesize seeing the highest structure in both, neutral and non-neutral datasets, in trans-peninsular comparisons. We further expect to see weaker structure in both

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

Family	Species	Distribution Pacific/Gulf	Habitat and depth	PLD (days)	Gene flow across Point Conception	Gene flow across Punta Eugenia	Gene flow around peninsula	Proposed Gulf/Pacific divergence time	References
Gobiidae	<i>Gillichthys mirabilis</i> (Longjaw mudsucker)	Tomales bay to Bahia Magdalena/upper & central gulf	Intertidal shallow soft bottoms, estuaries and sloughs.	Unknown, but larvae settle at 8-12mm	Moderate to high $F_{ST}=0.13$	Extremely low $F_{ST}=0.73$	Extremely low $F_{ST}=0.67$	0.76 to 2.3 mya.	Huang and Bernardi (2001) Bernardi et al. (2003)
Haemulidae	<i>Anisotremus davidsonii</i> (Sargo)	Monterey to central Baja / upper & central gulf	Rocky reef. Occasionally sand bottoms. Up to 60m	40 to 50	n/a	Low $*F_{ST}=0.39$	Extremely low $F_{ST}=0.65$	0.16 to 0.64 mya.	Bernardi et al. (2003) Bernardi and Lape (2005)

* $F_{ST} = 0.39$ is an average value across Punta Eugenia of 4 species including *A. davidsonii* in Bernardi et al. (2003). Individual species values not given in reference.

datasets among Pacific populations in response to mentioned breaks, but no structure within the Sea of Cortez. Finally, we compare trans- and cis-peninsular sites searching for signals of selection and common patterns in the total proportion of outliers, the distribution of outliers across their F_{ST} range, and the proportion of outliers that can be matched to a coding region. We hypothesize a positive relationship between site divergence and proportions of outliers and matches to coding regions, but similar outlier distribution patterns regardless of the level of differentiation.

Methods

Collections and DNA Extractions

Samples collections were done by previous (Huang and Bernardi 2001; Bernardi et al. 2003; Bernardi and Lape 2005) and current study (recent collections performed in 2017). Sargo specimens were collected by pole spear during SCUBA or free diving, and mudsucker individuals were collected using minnow traps. See Figure 1 and Table 2 site and sampling information.

From all sites, only the northernmost sampling site 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 Niño, it is generally rare, and does not have established populations north of Point Conception. Between Point Conception and Punta Eugenia, samples were collected in 3 sites for the mudsucker, and one site for the Sargo. South of Punta Eugenia, both species were collected from one site only. We sampled a total of 3 distinct locations for each species within the northern and central zones of the Sea of Cortez (Figure 1 and Table 2). Fin clips, gill, or muscle tissue, were extracted from specimens and stored in 95% ethanol at room temperature in the field, and at -80°C in the lab. The DNeasy 96 tissue kits for purification of DNA from animal tissue (QIAGEN, Valencia, California, United States) were utilized following the manufacturer's protocol to extract genomic DNA from subsampled tissue.

Marker Genotyping, Discovery, and Validation

We followed the original RADseq protocol utilizing the Sbf1 restriction enzyme to digest DNA (Miller et al. 2007; Baird et al. 2008) and produced 2 RAD libraries. Each genomic DNA sample contained a concentration of 400 ng, which was then physically sheared using a Covaris S2 sonicator using an intensity of 5 for 30 s, 10% duty cycle,

and cycles/burst of 200. Final amplification PCR was performed with 50 μl 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 2 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, 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. *Populations* scripts in STACKS used these reads as the input for the population genomic analysis. The analysis of trans-peninsular or allopatric sites in each species started by running one *populations* script with all the Pacific individuals pooled as a single population, and all the Gulf individuals as another. Thus, trans-peninsular analyses compare pooled Pacific (Pac) versus pooled Gulf sites (Gulf) within species (Table 2). For cis-peninsular sites, one *populations* script was run per species considering each site as a separate population without pooling. Cis-peninsular analyses follow pairwise comparisons of adjacent sites from north to south in the Pacific and all possible comparisons within the Gulf. "Pacific-only" and "Gulf-only" comparisons refer to Pacific and Gulf sites from cis-peninsular analyses, respectively.

Parameter optimization for de novo and *populations* scripts followed a road map as seen in Paris et al. (2017). This study found a minimum stack depth of 3 (m) to be optimal for the formation of alleles in the de novo when measuring the impact that varying this parameter had on the number of loci and polymorphism observed from datasets of 3 distantly related organisms. We followed this recommendation in de novo computations, but given the small chance of encountering the same sequencing error in multiples individuals (Schunter et al. 2014), and to restrained false positives, we increased the minimum stack depth in the *populations* module. The likelihood of encountering false positives was further reduced by implementing stringent filters to exclude low-quality reads (Miller et al. 2012). Subsequently, highly polymorphic loci were excluded by setting the value of 3 as the maximum number of mismatches between stacks (M) in the de novo analyses. This allows for the conservative capture of polymorphic loci and simultaneously offers an initial filtering of paralogs. We further minimized the probability of linkage in our

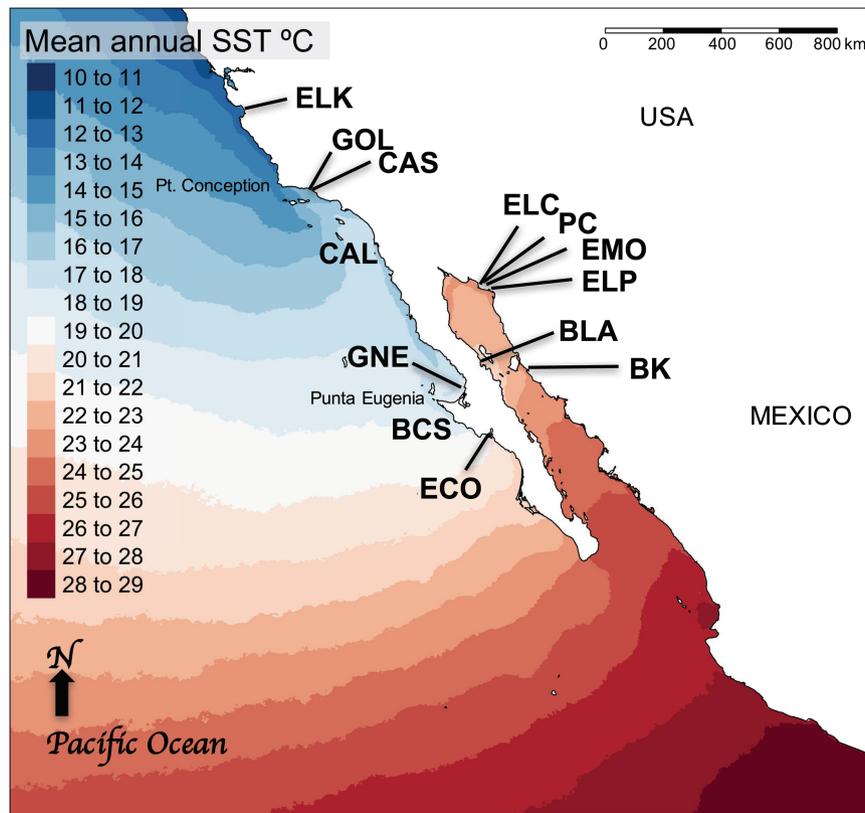


Figure 1. Map of Pacific and Sea of Cortez sampling sites and well-established phylogeographic breaks, Point Conception and Punta Eugenia, overlaid with mean annual Sea Surface Temperature (°C) from 2002 to 2010. The MARSPEC database (<http://marspec.weebly.com>) was used to obtain NASA Ocean Color Web temperature data. See Table 2 for the number of samples. ELK, Elkhorn Slough; GOL, Goleta Slough; CAS, Carpinteria Slough; CAL, southern California; GNE, Guerrero Negro; BCS, Baja California Sur; 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. Number of samples per site, region, and species

Region	Site	<i>Gillichthys mirabilis</i>	<i>Anisotremus davidsonii</i>
Pacific California	(Pac)	45	18
	Elkhorn Slough (ELK)	11	
	Goleta Slough (GOL)	10	
	Carpinteria Slough(CAS)	9	
	Southern California (CAL)		10
Baja California	Guerrero Negro (GNE)	8	
	Estero El Coyote (ECO)	7	
	Baja California Sur (BCS)		8
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

data by removing highly repetitive stacks in the de novo process, and by using the *write_single_SNP* flag in *populations* to select only the first SNP in reads that had more than one SNP. After substantial exploration of the different datasets, and to enable comparisons of

sites using the same parameters, only reads with a minimum depth coverage of $6\times$ ($m = 6$) that were present in at least 60% ($r = 0.60$) of the corresponding population were included in *populations* analyses. Reads passing the quality and population filters are presented as our total loci in Table 3.

Analysis of Genomic Divergence Between Sites

After running various *populations* analyses in STACKS to determine the optimal parameter set for our datasets, the final genepop output file from the *populations* run was converted into Arlequin format using PGDSpider (Lischer and Excoffier 2012). Arlequin version 3.5.1.2 (Excoffier and Lischer 2010) was then used to compute genomic indices of diversity for each site and determine genetic differentiation between them. Fixation index (F_{ST}) was calculated using the total number of loci, computing a distance matrix, and using the number of different alleles, under 10 000 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 the gene diversity or theta Θ (probability that 2 random homologous nucleotides are different), are also reported from the output of STACKS and 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 difference between the average polymorphism observed in cis- and trans-peninsular sites was statistically significant.

Table 3. Locus, polymorphism, and genetic diversity statistics of the Pacific and Sea of Cortez sites per species (after quality and population filters)

Species	Site	Samples	Total loci	Usable loci	Polym. loci	% Polym.	Theta Θ
<i>G. mirabilis</i>	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	15 058	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	15 058	168	121	72	0.05166
<i>A. davidsonii</i>	CAL	10	15 338	8182	2609	32	0.06383
	BCS	8	15 338	5780	2046	35	0.07524
	Pacific	18	4379	2107	1031	49	0.06214
	PC	10	15 338	3108	1111	36	0.05951
	BK	12	15 338	8464	3693	44	0.06592
	BLA	8	15 338	3191	1141	36	0.07564
	Gulf	30	4379	1391	1004	72	0.05963

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, Guerrero 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.

We further examined genetic structure in neutral and outlier loci, separately, to investigate the possibility that different factors have shaped population divergence in unique patterns. Presumed neutral loci were obtained by excluding the outlier loci from the total number of loci. The cataloging of the F_{ST} outlier loci, or presumed loci under selection, is explained below. After this classification, we ran *populations* scripts employing the blacklist and whitelist options to produce data files for neutral and outlier loci, respectively. We then used 2 different approaches to explore and visualize genomic structure in each dataset.

First, we performed a Discriminant Analysis of Principal Components (DAPC) (Jombart et al. 2010) using neutral and outlier loci individually. DAPC combines the benefits of discriminant and principal component analyses, and is particularly useful to study differences between clusters (i.e., sites or populations) as it utilizes a multivariate approach to explore the entire variation in the data while minimizing 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.

Second, structure files from the *populations* output 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 differs when analyzing neutral or outlier loci. For the neutral dataset, a range of K from 1 to 7 for *A. davidsonii*, and 1 to 10 for *G. mirabilis* (corresponding to the number of sites for each species plus 2 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 1 to 5 for both species. Only the plots with the K

obtaining the highest likelihood, according to the Evanno method (Evanno et al. 2005) implemented in Structure Harvester (Earl and 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 and Whitlock 2015), they are commonly used to show evidence of the diverging effects of selection between populations (Gaither et al. 2015; Longo and Bernardi 2015; Bernardi et al. 2016; Stockwell et al. 2016). Outlier loci were identified directly from the *phistats* output file in STACKS by selecting loci above a threshold value equal to 3 standard deviations above the mean AMOVA F_{ST} (or loci within the top 0.03% divergence values).

We then investigated the patterns of outlier divergence and abundance between sites. To begin the outlier analysis, a *populations* script was run using a whitelist containing the outlier identifications, and the resulting structure files were used to create DAPC and 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 intra- and inter-specific site comparisons. In both species, we tested if the proportion corresponding to the average number of outliers found in the cis-peninsular sites differed from that of the trans-peninsular sites, and whether differences in Pacific and Gulf outlier proportions were significant.

Outlier sequences were then uploaded into the GenBank database specifying an Expect threshold of 10^{-6} (an E-value which returns only matches with a one in a million chance to be paired with

a record by chance alone) to document matches to protein-coding regions as well as matches to regions without annotation. We 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 that were matched to coding genes was statically different between sites and species, and whether or not this was related to higher number of outliers.

Results

Loci and Polymorphism Statistics

Genomic DNA was sequenced from a total of 121 samples (73 *G. mirabilis* and 48 *A. davidsonii*) in 2 illumina lanes producing approximately 200 million 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 5852 SNPs per individual). The number of loci passing all filters in the *populations* scripts with Pacific and Gulf pooled populations of *G. mirabilis* and *A. davidsonii* were 15 058 and 4379, respectively (Table 3). Total loci resulting from the scripts with multiple populations were 4316 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 < 0.001, and z -value = 13.5, P -value < 0.001). Percent polymorphism ranged from 8% to 39% and from 32% to 49% in discrete mudsucker and Sargo sites, respectively. The average polymorphism in Gulf sites was also statistically higher than in Pacific sites for either species in the cis-peninsular analyses (z -value = 13.4 and P -value < 0.001 for mudsucker sites; z -value = 5.8 and P -value < 0.001 for Sargo sites). There was no discernable gene diversity (Theta Θ) pattern between cis-peninsular sites for any species but this was generally higher in the Gulf sites

of *G. mirabilis* compared to its Pacific sites. See Table 3 for full loci and polymorphism statistics.

Genomic Divergence and Structure Between Sympatric and Allopatric Sites

The first step in our site comparison was to characterize the genomic F_{ST} in Arlequin using the total number of loci. The disjunct Sargo and mudsucker populations (i.e., pooled Pacific sites versus pooled Gulf sites) respectively presented a genomic F_{ST} of 0.11 and 0.23, and both were statistically significant (using a significance level of 0.05). Pacific sites of the Sargo (CAS and BCS) diverged by a F_{ST} = 0.02, and this was significant as well. In contrast, differentiation between any pair of the Sargo Gulf sites only reached F_{ST} values lower than 0.01. Pairwise F_{ST} values between adjacent Pacific sites of the mudsucker from north to south ranged from 0 to 0.18, and all Gulf comparisons resulted in F_{ST} = 0 (See Table 4 for the complete distance matrix).

Subsequently, in order to determine if structure could have been produced by different factors, we dissected the observed genomic divergence by performing a DAPC analysis and STRUCTURE plots using neutral and outlier loci separately. Differences in the genetic composition of allopatric sites for both species resulted in separated sets of clusters in the DAPC graphs (Figure 2). Interestingly, Sargo Pacific sites and mudsucker Gulf sites are also well-separated within the plot for each species. DAPC patterns did not differ drastically between neutral and outlier loci. In contrast, STRUCTURE plots showed unique patterns of genomic structure when we analyzed either neutral or outlier loci (Figure 3). When using neutral loci in each species, Pacific individuals appeared to belong to a common cluster with Gulf individuals, but not without showing important differences. Plots utilizing outlier loci showed an extremely high probability that every Pacific individual belongs to a separate population compared to that of all Gulf individuals. Structure Harvester

Table 4. Genomic differentiation between cis-peninsular sites of *Anisotremus davidsonii* and *Gillichthys mirabilis*

<i>Gillichthys mirabilis</i>								
	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 F_{ST}	= 0.23 +							
<i>Anisotremus davidsonii</i>								
	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 F_{ST}	= 0.11 +							

Matrix reports F_{ST} values (estimated by computing a distance matrix in Arlequin; below diagonal) and whether these were found to be significant (+) or non-significant (-) (Significance level = 0.05; above diagonal). Results from pooled trans-peninsular populations (Pacific vs. Gulf) are given below matrices. ELK, Elkhorn Slough; GOL, Goleta Slough; CAS, Carpinteria Slough; CAL, 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.

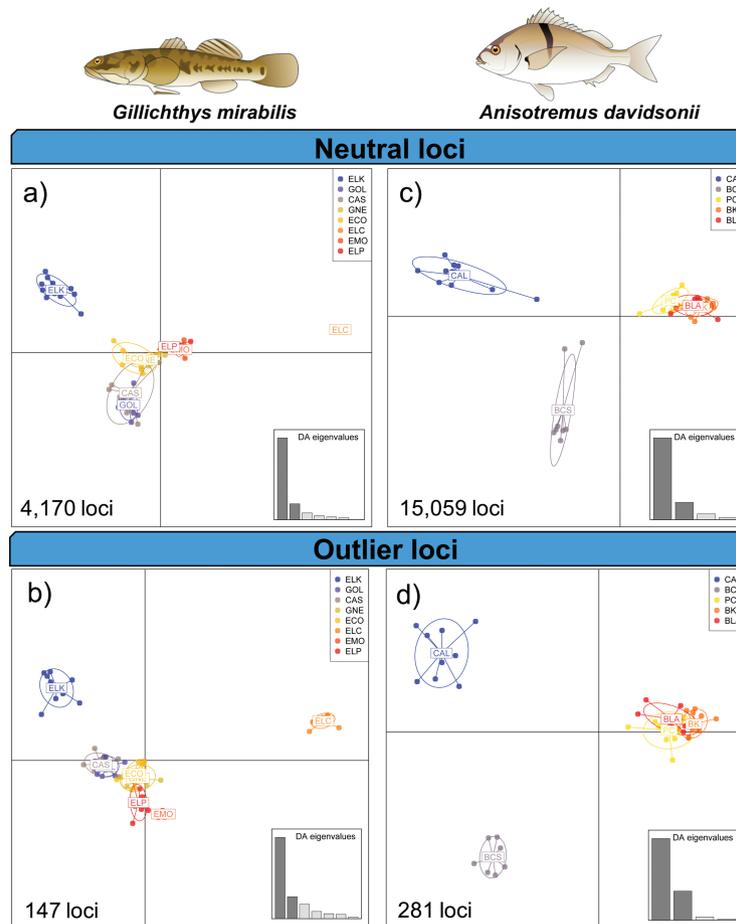


Figure 2. DAPC cluster plots of *Gillichthys mirabilis* and *Anisotremus davidsonii* cis-peninsular sites using putative neutral and outlier loci. For both species, Pacific sites (ELK, GOL, CAS, GNE, and ECO for *G. mirabilis*; CAL and BCS for *A. davidsonii*) are differentiated from Gulf sites (ELC, EMO, and ELP for *G. mirabilis*; PC, BK, and BLA for *A. davidsonii*), and the most northern Californian site (ELK for *G. mirabilis*; CAL for *A. davidsonii*) separates from Baja California clusters. Plots were created in R using the adegenet package using 6, 7, 4, and 4 discriminant functions, and retaining 30, 30, 30, and 15 principal components (selected by the validation tool xvalDapc) for panels a, b, c, and d, respectively. Each plot illustrates the first and second axes. See [Supplementary Figure S2](#) for percent cumulative variance per PC for each plot.

selected a K of 4 for the analysis of neutral loci in *A. davidsonii* and a K of 2 for every other treatment.

Outlier Loci Patterns and Evidence of Selection

Outliers were identified for trans-peninsular populations and for pairwise comparisons of cis-peninsular sites from north to south in the Pacific, and for all pairwise comparisons within the Gulf, by selecting loci with a differentiation higher than 3 standard deviations from the mean AMOVA F_{ST} . The number of outliers between allopatric populations 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 sites, 202 outlier loci were identified with F_{ST} values ranging from 0.24 to 0.77. In Sargo Gulf sites, pairwise comparisons resulted in a range from 192 to 224 outlier loci and F_{ST} values ranging from 0.19 to 0.81. Analyses of mudsucker cis-peninsular sites 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). No fixed loci were detected between any cis-peninsular comparisons for either species.

We then tested if the difference between the proportions of the total loci that were classified as outliers in each analysis (cis- and trans-peninsular) was significant within or between species. When analyzing both species, results indicated that the proportion of outliers found comparing trans-peninsular populations of the Sargo was statistically different from that between mudsucker trans-peninsular populations (z -value = 29.2, P -value < 0.001; Table 5) but the average outlier proportion among all cis-peninsular sites was similar in both species (z -value = 2.3, P -value = 0.02; Table 5). Yet, if we analyzed only Pacific sites, or only Gulf sites, the corresponding average outlier proportions were different between species. When analyzing each species individually, the proportion of outliers between Pacific sites was different from that between Gulf sites for the Longjaw mudsucker but not for the Sargo (Pacific-only and Gulf-only in Table 5). Finally, the average proportion of outliers among all cis-peninsular sites was different from the proportion of outliers between trans-peninsular population for each species.

Identified outliers were further categorized based on whether they matched known coding regions, other sequences without annotation, or produced no match at all in GenBank when specifying a e -value cutoff of 1×10^{-6} (which returns only matches with one in a million chance to be paired with a record by chance alone).

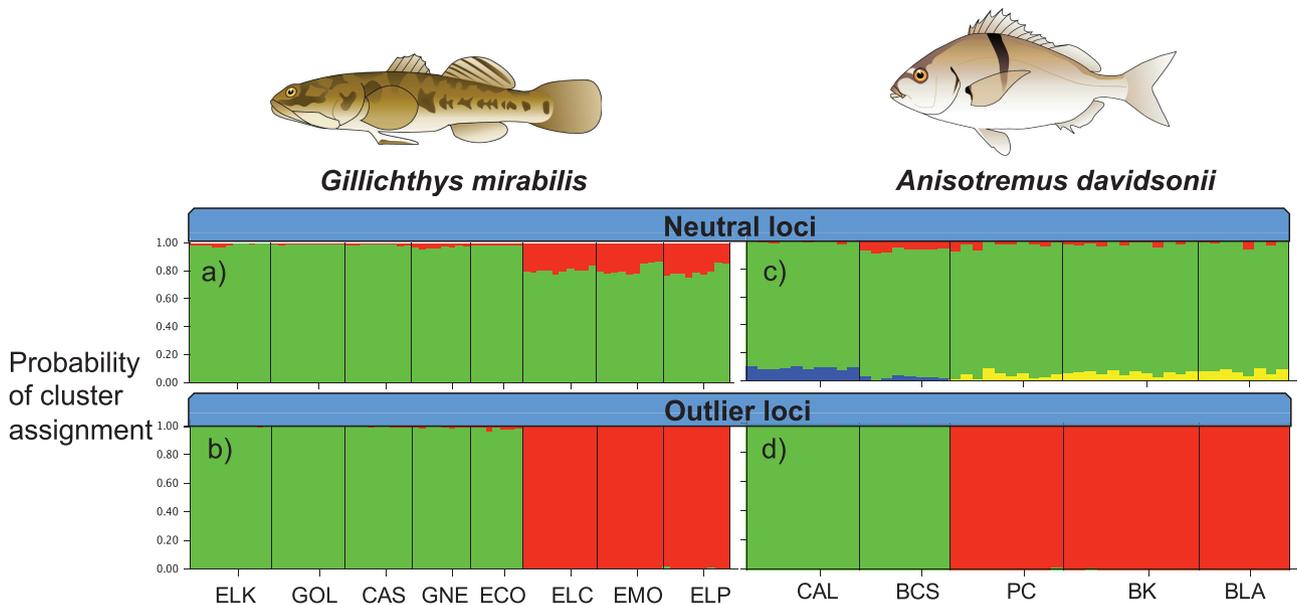


Figure 3. STRUCTURE plots with Bayesian assignment of individuals (vertical bars) from cis-peninsular sites into discernable genetic clusters or populations (green, red, blue, or yellow) based on presumed neutral loci (top panels) and outlier loci, or loci suspected to be under selection (bottom panels). 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 Table S2](#). For *G. mirabilis*, neutral loci = 4170 and outlier loci = 147. For *A. davidsonii*, neutral loci = 15 059 and outlier loci = 281. ELK, Elkhorn Slough; GOL, Goleta Slough; CAS, Carpinteria Slough; 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.

After this categorization, we tested if the proportion of outliers matching coding genes followed specific patterns within or between species. Out of the total outlier loci identified between allopatric Sargo and mudsucker sites, 25% and 17% were paired to coding regions by the BLASTn tool, respectively ([Figure 4](#)). The difference between these proportions was statistically significant (z -value = 3.6, P -value = 0.0004). On average, 38% of the outliers between Sargo Pacific sites and 27% between mudsucker Pacific sites produced matches to coding regions, but these 2 proportions were not significantly different (z -value = 0.9, P -value = 0.3424). The difference between the average proportion of matches occurring among Sargo and mudsucker Gulf sites, 35% and 16%, respectively, was significantly different (z -value = 2.8, P -value = 0.0046).

Sargo presented a statistically higher average proportion of outliers matching coding regions in cis-peninsular than in allopatric sites (z -value = 4.2, P -value = 0.0001). This difference was not significant in the Longjaw mudsucker. Percentages and numbers of loci falling into each category for every population analysis are given in [Figure 4](#).

Discussion

The Baja California disjunct species are a notable group that has a shared evolutionary history where genomic divergence, local adaptation, and modes of selections can be examined. This study analyzed 2 disjunct species that previously showed population structure and different levels of gene flow based on mtCYTB, and searched for common patterns of genomic divergence and outlier loci properties in trans- and cis-peninsular sites. Overall, the output from illumina sequencing yielded less reads in mudsucker compared to Sargo individuals, which could potentially due to older tissue samples, or lack of proper preservation ([Graham et al. 2015](#)). This is also reflected in the lower number of loci observed in mudsucker populations. The

lower number of loci in trans-peninsular compared to cis-peninsular analyses in the Sargo might be a product of pooling sites that already contained variation between them, impeding a large number of loci from passing established filters. Regardless of these limitations, substantial numbers of loci were obtained for meaningful comparative analysis of sites ([Table 3](#)).

Divergence and Structure Between Allopatric Populations

Similar to previous mitochondrial DNA studies, the current genomic results indicate that the Pacific and Gulf regions are highly differentiated from each other in either species. Genomic distance based on F_{ST} values is large and significant, DAPCs graphs separate Pacific from Gulf sites, and STRUCTURE plots show distinct genetic compositions in each region. The spatial placement of clusters does not differ drastically in the DAPCs graphs using the different type of loci. However, the contrasting probabilities of population assignment given to individuals in the different STRUCTURE panels ([Figure 3](#)) clearly show unique patterns of structure between disjunct populations when analyzing presumed neutral and loci under selection. When analyzing presumed neutral loci, Sargo Pacific and Gulf individuals show exclusive clusters (blue and yellow, respectively; [Figure 3c](#)), and mudsucker individuals from the Gulf possess a visibly higher probability of belonging to an alternate genetic pool (red) compared to Pacific individuals ([Figure 3a](#)). When analyzing the outlier loci in either species, the probability of any Pacific individual belonging to the same genetic pool than any Gulf individual, or vice versa, is extremely low or null ([Figure 3b,d](#)). These results indicate the presence of neutral and selective structure in disjunct populations of both species, but do not identify the source of such structure. Under previously suggested scenarios with very low gene flow between these regions and species, genetic drift and selection might be important forces shaping the observed structure. Yet, effects from historical

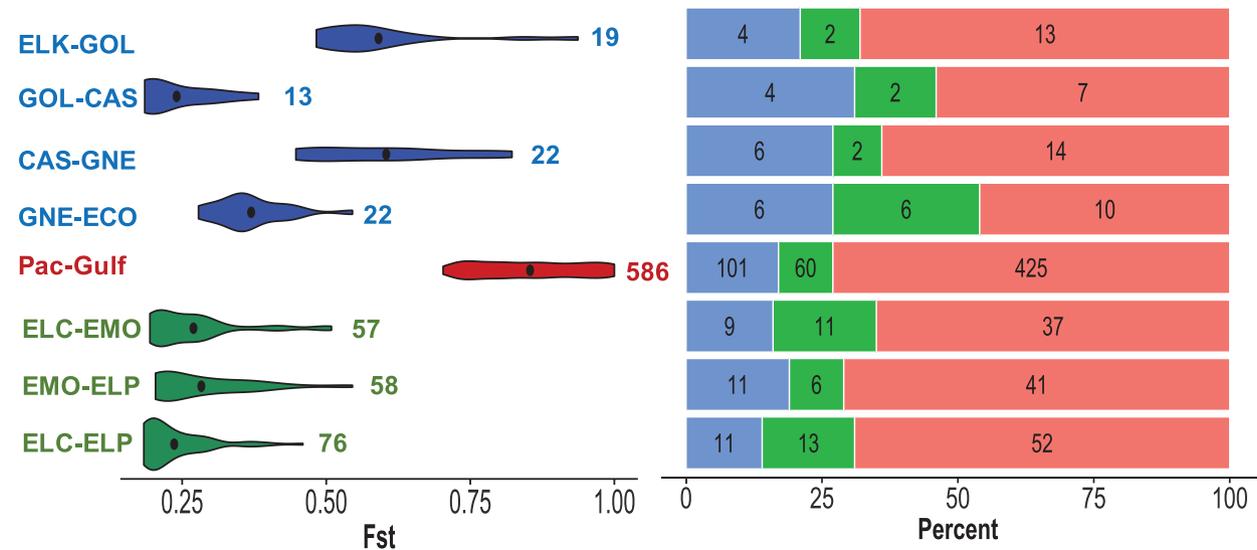
Anisotremus davidsonii*Gillichthys mirabilis*

Figure 4. Outlier loci statistics in trans-peninsular populations (Pac-Gulf) and pairwise comparisons of cis-peninsular sites of *Anisotremus davidsonii* and *Gillichthys mirabilis*. Violin plots (left panels) depict the relative density of outliers as a function of the overall outlier F_{ST} distribution. Shape size is not scaled to the number of outliers. Numbers at the end of shapes give the total number of outlier loci. Horizontal stacked bar plots (right panels) illustrate the percent of outlier loci matching to a protein-coding region (blue/left stacks), or a sequence with no annotation (green/middle stacks), or returning no matches using the BLASTn tool in GenBank (salmon-colored/right stacks). Numbers inside stacked bars give the actual number of outliers in each category. ELK, Elkhorn Slough; GOL, Goleta Slough; CAS, Carpinteria Slough; GNE, Guerrero 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.

Table 5. Analysis of the difference between the average proportion of outlier loci, and proportion of outliers matching a coding region in GenBank, between sites, compared by regions and species

Comparisons of sites	Outlier loci	Matches to coding regions
Interspecific		
ADA trans-peninsular vs. GMI trans-peninsular	<0.001*	0.0004*
ADA cis-peninsular vs. GMI cis-peninsular	0.02	<0.001*
ADA Pacific-only vs. GMI Pacific-only	<0.001*	0.34
ADA Gulf-only vs. GMI Gulf-only	0.48	0.0046*
Intraspecific		
ADA trans-peninsular vs. ADA cis-peninsular	<0.001*	<0.001*
ADA Pacific-only vs. ADA Gulf-only	0.94	0.48
GMI trans-peninsular vs. GMI cis-peninsular	<0.001*	0.46
GMI Pacific-only vs. GMI Gulf-only	<0.001*	0.28

Values report P -values from Z-tests of proportions (*significant difference at a 0.01 level). ADA, *Anisotremus davidsonii*; GMI, *Gillichthys mirabilis*. See Table 2 for site and region information.

colonization or range expansion events, as well as potential coupling of endogenous and exogenous factors, are other scenarios that could also explain such patterns (Klopstein et al. 2006; Bierne et al. 2011; Marko and Hart 2011). Likewise, pooling several Pacific and Gulf populations might have inflated the observed patterns. Genetic incompatibilities between southern trans-peninsular populations (that might be exclusive to the very different biology of each species but having similar effects), and contemporary mechanisms such as different seawater temperature, lack of proper habitat, current eddies in the north and oceanic fronts in the south of the Sea of Cortez, are among the factors that could be maintaining the divergence between observed genetic compositions in the neutral and outlier loci, respectively (Thomson et al. 2000; Bernardi et al. 2003; Bierne et al. 2011; Inda-Diaz et al. 2014; Garcia-De Leon et al. 2018). The observed higher levels of polymorphism in Gulf compared to Pacific populations, may be a reflection of the environmental heterogeneity in the Gulf (Thomson et al. 2000; Inda-Diaz et al. 2014; Garcia-De Leon et al. 2018), or for instance, the proposed older age and bigger effective population sizes of the Sargo populations in this region (Bernardi and Lape 2005).

Divergence and Structure Between Sympatric Populations

Point Conception and Punta Eugenia are suggested to be important discontinuities for gene flow between Pacific populations of both species (Huang and Bernardi 2001; Bernardi et al. 2003; Bernardi and Lape 2005). Analyses of genomic divergence in this study show statistically significant F_{ST} values across Point Conception for mudsucker ($F_{ST} = 0.12$ – 0.15 between ELK and the rest of Pacific sites), as well as across Punta Eugenia for Sargo ($F_{ST} = 0.02$ between CAL and BCS). Yet, GNE and ECO mudsucker sites do not show differentiation across Punta Eugenia, preventing the genomic divergence from mirroring previously observed mitochondrial F_{ST} patterns. Besides the comparisons of populations north and south of these phylogeographic points, results also report a low but significant F_{ST} value between the mudsucker sites from Carpinteria Slough (CAS) and Guerrero Negro (GNE). This result was not surprising since genetic divergence has also been reported between similar sites in other disjunct species (Stepien et al. 2001; Bernardi et al. 2003), and estuarine fish assemblages sampled close to Guerrero Negro in Laguna Ojo de Liebre (or Scammon's lagoon), are described as more representative of southern fish communities compared to those in the northern Pacific coast of Baja California (Bahia San Quintin), which themselves share more affinities with assemblages from southern California (Stepien et al. 2001; Allen et al. 2006). The large geographic distance between CAS and GNE, as well as current gyres, and the almost permanent upwelling regime between these sites (at the San Quintin region), might be responsible for the documented neutral and non-neutral divergence (Stepien et al. 2001; Allen et al. 2006; Iacchel et al. 2013). Yet, endogenous barriers between Pacific populations (either shared or species-specific), might have multiplicative effects with increasing distance and thus could also be responsible for some of the observed population divergence along this coast (Bierne et al. 2011).

Interestingly, DAPC reveals differentiation between the 3 Gulf sites within each species that was not detected by F_{ST} analyses alone. On one hand, the source generating the observed structure among goby populations in these 3 geographically close *Esteros* is not clear. However, more genetic divergence within the genus *Gillichthys* has been revealed from the northern Gulf than anywhere else as

the only other 2 recognized *Gillichthys* species (*Gillichthys seta* and *Gillichthys detrusus*) are endemic to this area (Barlow 1961; Thomson et al. 2000; Huang and Bernardi 2001; Allen et al. 2006; Swift et al. 2011). Potential hybridization between these species might be influencing diversity indexes and inflating the number of F_{ST} outliers observed in these populations (Bierne et al. 2011, 2013). Habitat choice along with the dynamic history of the northern Gulf has been hypothesized to have played an important role in the speciation of these lineages (Barlow 1961; Huang and Bernardi 2001; Swift et al. 2011). Currently, *G. mirabilis* may co-occur within the same estuaries with congeners but each species occupies different micro-habitats (Barlow 1961; Thomson et al. 2000; Huang and Bernardi 2001; Swift et al. 2011). In this study, Longjaw mudsuckers were only found in outcrops or small soft-bottom channels remarkably high in the intertidal zone, and this area was observed to be disconnected from any main flow in the estuaries due to the extremely large tidal fluxes prevalent in the northern Gulf. (Bernardi et al. 2003; Swift et al. 2011). Discerning the level of hybridization and just how much, if any, isolation is granted by the combination of the aforementioned factors requires additional studies. On the other hand, Sargo Gulf sites are relatively far from each other, but current eddies in the northern Gulf, as well as the series of islands between the peninsula and mainland Mexico, might be acting as mechanisms of larvae retention and stepping stone-dispersal across the northern Gulf (Inda-Diaz et al. 2014; Garcia-De Leon et al. 2018).

Outlier Loci Patterns and Signals of Selection

Results from this study strongly suggest the presence of differential selection in the Pacific and Gulf populations of each species. The 2 regions contain very different outlier pools as analyses identify large numbers of substantially diverged loci, including many fixed differences (40 for Sargo and 49 for mudsucker allopatric populations). Subsequent evidence of selection includes having higher proportions of outliers matching coding regions in GenBank than expected in a group of randomly selected loci (up to 25% and 38% of the outliers matched a gene in trans- and cis-peninsular analyses, respectively). The relative proportion of outliers, considering the initial total loci in each analysis, is higher in trans-peninsular than in cis-peninsular comparisons (Figure 4, Table 5). The selective divergence between the Pacific and Gulf is clearly illustrated in the outlier STRUCTURE plots, where the genetic constitution of individuals from one region is considerably different from that of individuals from the other. Moreover, our exploratory examination reveals that outliers from trans-peninsular sites displayed substantially higher ranges of differentiation compared to sites on either side of the peninsula, and these are distributed more evenly throughout such range (Figure 4). In contrast, the great majority of outliers diverging between cis-peninsular sites lie close to the lower end of the F_{ST} range instead. This trans-peninsular "outlier evenness" might be a byproduct distribution of groups of outliers with overall higher differentiation. However, this notion is not supported in current cis-peninsular comparisons as evenness does not increase with increasing outlier differentiation (except for CAS-GNE). The group of outlier loci with the highest differentiation between cis-peninsular sites (ELK-GOL) have a highly uneven distribution similar to that in other cis-peninsular comparisons with low differentiation ranges (Figure 4). Additionally, even when the Sargo trans-peninsular sites show an outlier loci distribution spanning from low to high F_{ST} , the scattering of loci remains fairly even. If these patterns remain constant in other disjunct species or other systems with allopatric and non-allopatric

populations, outlier evenness might represent a detectable signature of distinct mechanisms of differentiation responsive to different levels of isolation.

There was also an expectation for sites crossing Point Conception and Punta Eugenia (cis-peninsular Pacific-only) to harvest higher percentages of outlier loci than Gulf sites (cis-peninsular Gulf-only) given the higher average differentiation seen across the Pacific compared to the Gulf for each species. However, the percent of outlier loci is either higher in Gulf sites (for mudsucker), or not statistically different between the 2 regions (for Sargo). At the same, higher differentiation did not always translate into higher proportions of outliers matching coding regions. While trans-peninsular comparisons were more differentiated and produced overall higher numbers of outliers than cis-peninsular, there is no statistical difference in the average proportion of outliers matching genes between analyses. The same occurred when we compared Pacific to Gulf cis-peninsular sites, as the proportion of matches did not differ significantly between regions for neither species (Table 5). Outlier patterns are also likely the product of the interaction between the same endogenous, exogenous, and natural barriers, discussed as possible factors creating and maintaining population divergence in these species (see above).

Altogether, results suggest a complex nature of population structure that include neutral and selective divergence. Disjunct species experience unique and vastly diverse environments on either side of the peninsula. Upwelling regimes, habitat discontinuities, as well as the latitudinal temperature gradient, may be some of the present-day pressures maintaining observed divergence in Pacific populations. Similarly, environmental seasonality in the central and northern zones of the Sea of Cortez, and the extensive tidal range in combination with habitat choice, may do the same for Gulf 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 shed light on specific details identifying if, and how different types of speciation operate in these populations. It is important to restate that while we list some of the environmental stressors that could be affecting these populations, observed patterns are plausibly the product of the interplay of several factors not measured here including, but that may not be limited to drift, selection, historical vicariant and colonization events, past- and current gene flow, endogenous factors, and differences in effective population sizes and standing genetic variation, among the studied sites. A suite of other methods (including coalescence methods; see Marko and Hart 2011) are needed to assess these elements, and develop a better understanding the evolutionary histories of these populations. A higher number of populations within regions with larger and more consistent sampling than presented in some of the cis-peninsular sites might also be required for these analyses. The goal of the current study was, instead, to reveal spatial patterns of outlier distribution and genomic diverge while comparing the last to previously documented mtDNA divergence using similar sampling numbers. We have performed a power analysis to evaluate the variation in our estimates while decreasing the number of individuals in each cis-peninsular site by one at the time, for 3 times (final analysis decreased the number of individuals by 3 in each site). Calculated values for the proportion of polymorphic loci and theta Θ are considerably constant among the 4 tests (All current individuals, All -1, All -2, All -3, see Supplementary Material for full details). The actual number of total, usable, and polymorphic loci, follow a decreasing pattern with increasing number of individuals (from

All -3 to All), but the rate in which these numbers decrease slows down steeply with increasing individuals as well. This suggests that the number of loci can be expected to decrease by increasing the number of individuals to the current sampling scheme, but this difference is unlikely to be significant. Furthermore, measurements of differentiation (F_{ST}) between pairwise cis-peninsular sites were not significantly different across all tests in the power analysis (with the exception of comparisons including GNE in the mudsucker) suggesting that estimating divergence with higher number of individuals might not change the observed patterns, which are the main focus of this study.

Conclusion

The evolutionary history of the Sargo and Longjaw mudsucker around the Baja California Peninsula has created a deep genetic divergence between Pacific and Sea of Cortez populations. This differentiation is present in both, putative neutral and selective portions of the genome. While the sites within the Sea of Cortez appear to be well connected, Point Conception and Punta Eugenia display structure in the Pacific, but patterns of population divergence may differ depending on the utilized markers. Overall, patterns of genomic divergence are concordant, but do not mirror previous patterns based on mitochondrial sequences. Allopatric comparisons show higher numbers of outliers, and these are more evenly distributed across higher ranges of differentiation compared to analyses of sites throughout the sympatric distributions of these species. However, examining all site comparisons reveals that average differentiation in all loci does not always have a proportional relationship, and sometimes no relationship at all with the percentage of outliers and subsequent rate in which these are matched to coding regions. These patterns illustrate how speciation processes can operate in populations with different levels of isolation. Documenting the prevalence of these patterns in other natural populations with similar gradients of isolation can play an important role in understanding how biodiversity is shaped across landscapes.

Supplementary Material

Supplementary data are available at *Journal of Heredity* online.

Funding

This work was supported by Consejo Nacional de Ciencia y Tecnología (CONACYT), University of California Institute for Mexico and the United States (UC MEXUS), The Myers Trust, and Friends of the Long Marine Lab. All grants were provided to E.G.

Acknowledgments

We would like to thank Dr. Larry G. Allen for providing fish illustrations and Dr. Peter T Raimondi for his comments on the manuscript. We would like to thank the authorities for granting collection permits. Permiso de Pesca de Fomento “Estado de salud y Biodiversidad de los arrecifes del Golfo de California y Pacífico Mexicano” RNPA-0305004228, Dr. Carlos A. Sánchez Ortiz, UABCS (including GB). We thank the anonymous reviewers for helping improve this manuscript.

Conflict of Interest

The authors declare no conflict of interest.

Data Availability

Demultiplexed sequence data are archived online at Dryad (DOI <https://doi.org/10.5061/dryad.gmsbc2ht>)

References

- Allen LG, Pondella DJ, Horn MH. 2006. *The ecology of marine fishes: California and adjacent waters*. Berkeley (CA): University of California Press.
- Baird NA, Etter PD, Atwood TS, Currey MC, Shiver AL, Lewis ZA, Selker EU, Cresko WA, Johnson EA. 2008. Rapid SNP discovery and genetic mapping using sequenced RAD markers. *PLoS One*. 3:e3376.
- Barlow GW. 1961. Gobies of the genus *Gillichthys*, with comments on the sensory canals as a taxonomic tool. *American Society of Ichthyologists and Herpetologists*. 1961:423–437.
- Bernardi G. 2000. Barriers to gene flow in *Embiotoca jacksoni*, a marine fish lacking a pelagic larval stage. *Evolution*. 54:226–237.
- Bernardi G. 2013. Speciation in fishes. *Mol Ecol*. 22:5487–5502.
- Bernardi G, Azzurro E, Golani D, Miller MR. 2016. Genomic signatures of rapid adaptive evolution in the bluespotted cornetfish, a Mediterranean Lessepsian invader. *Mol Ecol*. 25:3384–3396.
- Bernardi G, Findley L, Rocha-Olivares A. 2003. Vicariance and dispersal across Baja California in disjunct marine fish populations. *Evolution*. 57:1599–1609.
- Bernardi G, Lape J. 2005. Tempo and mode of speciation in the Baja California disjunct fish species *Anisotremus davidsonii*. *Mol Ecol*. 14:4085–4096.
- Bernardi G, Talley D. 2000. Genetic evidence for limited dispersal in the coastal California killifish, *Fundulus parvipinnis*. *J Exp Mar Bio Ecol*. 255:187–199.
- Bierne N, Roze D, Welch JJ. 2013. Pervasive selection or is it...? Why are FST outliers sometimes so frequent? *Mol Ecol*. 22:2061–2064.
- Bierne N, Welch J, Loire E, Bonhomme F, David P. 2011. The coupling hypothesis: why genome scans may fail to map local adaptation genes. *Mol Ecol*. 20:2044–2072.
- Bowen BW, Rocha LA, Toonen RJ, Karl SA. 2013. The origins of tropical marine biodiversity. *Trends Ecol Evol*. 28:317–376.
- Briggs JC. 1974. *Marine zoogeography*. New York: McGraw-Hill.
- Briggs JC, Bowen BW. 2012. A realignment of marine biogeographic provinces with particular reference to fish distributions. *J Biogeogr*. 39:12–30.
- Brusca RC. 1973. *A handbook to the common intertidal invertebrates of the Gulf of California*. Tucson (AZ): University of Arizona Press.
- Campbell MA, Robertson DR, Vargas MI, Allen GR, McMillan WO. 2018. Multilocus molecular systematics of the circumtropical reef-fish genus *Abudefduf* (Pomacentridae): history, geography and ecology of speciation. *PeerJ*. 6:e5357.
- Catchen JM, Amores A, Hohenlohe P, Cresko W, Postlethwait JH. 2011. Stacks: building and genotyping Loci de novo from short-read sequences. *G3 (Bethesda)*. 1:171–182.
- Catchen J, Hohenlohe PA, Bassham S, Amores A, Cresko WA. 2013. Stacks: an analysis tool set for population genomics. *Mol Ecol*. 22:3124–3140.
- Cotoras DD, Bi K, Brewer MS, Lindberg DR, Prost S, Gillespie RG. 2018. Co-occurrence of ecologically similar species of Hawaiian spiders reveals critical early phase of adaptive radiation. *BMC Evol Biol*. 18:100.
- Coyne JA, Orr HA. 2004. *Speciation*. Sunderland (MA): Sinauer Associates.
- Dawson MN, Waples RS, Bernardi G. 2006. Phylogeography. In: Allen LG, Pondella DJ II, and Horn MH, editors. *The ecology of marine fishes: California and adjacent waters*. Berkeley (CA): University of California Press. p. 370–379.
- Earl D, VonHoldt B. 2012. Structure harvester: a website and program for visualizing structure output and implementing the Evanno method. *Conserv Genet Resour*. 4:359–361.
- Enderl JA. 1977. *Geographic variation, speciation and clines*. Princeton (NJ): Princeton University Press.
- Evanno G, Regnaut S, Goudet J. 2005. Detecting the number of clusters of individuals using the software STRUCTURE: a simulation study. *Mol Ecol*. 14:2611–2620.
- Excoffier L, Lischer HE. 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.
- Floeter SR, Bender MG, Siqueira AC, Cowman PF. 2018. Phylogenetic perspectives on reef fish functional traits. *Biol Rev Camb Philos Soc*. 93:131–151.
- Gaither MR, Bernal MA, Coleman RR, Bowen BW, Jones SA, Simison WB, Rocha LA. 2015. Genomic signatures of geographic isolation and natural selection in coral reef fishes. *Mol Ecol*. 24:1543–1557.
- García-De León FJ, Galvan-Tirado C, Sanchez-Velasco L, Silva-Segundo CA, Hernandez-Guzman R, Barriga-Sosa I de los A, Diaz-Jaimes P, Canino M, Cruz-Hernandez P. 2018. Role of oceanography in shaping the genetic structure in the North Pacific hake *Merluccius productus*. *PLoS One*. 13:1–26.
- Graham CF, Glenn TC, McArthur AG, Boreham DR, Kieran T, Lance S, Manzon RG, Martino JA, Pierson T, Rogers SM, et al. 2015. Impacts of degraded DNA on restriction enzyme associated DNA sequencing (RADSeq). *Mol Ecol Resour*. 15:1304–1315.
- Helfman GS, Collette BB, Facey DE, Bowen BW. 2009. *The Diversity of Fishes*, Second. Oxford (UK): Wiley-Blackwell.
- Huang D, Bernardi G. 2001. Disjunct sea of cortex-pacific ocean *Gillichthys mirabilis* populations and the evolutionary origin of their sea of cortex endemic relative, *Gillichthys seta*. *Mar Biol*. 138:421–428.
- Iacchi M, Ben-Horin T, Selkoe KA, Bird CE, García-Rodríguez FJ, Toonen RJ. 2013. Combined analyses of kinship and FST suggest potential drivers of chaotic genetic patchiness in high gene-flow populations. *Mol Ecol*. 22:3476–3494.
- Inda-Díaz EA, Sanchez-Velasco L, Lavin MF. 2014. The effects of a tidal-mixing front on the distribution of larval fish habitats in a semi-enclosed sea during winter. *J Mar Biol Assoc U K*. 94:1517–1530.
- Jombart T. 2008. adegenet: a R package for the multivariate analysis of genetic markers. *Bioinformatics*. 24:1403–1405.
- Jombart T, Devillard S, Balloux F. 2010. Discriminant analysis of principal components: a new method for the analysis of genetically structured populations. *BMC Genet*. 11:94.
- Klopfstein S, Currat M, Excoffier L. 2006. The fate of mutations surfing on the wave of a range expansion. *Mol Biol Evol*. 23:482–490.
- Lischer HE, 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. *Mol Phylogenet Evol*. 88:55–63.
- Losos JB. 2010. Adaptive radiation, ecological opportunity, and evolutionary determinism. American Society of Naturalists E. O. Wilson award address. *Am Nat*. 175:623–639.
- Lotterhos KE, Whitlock MC. 2015. The relative power of genome scans to detect local adaptation depends on sampling design and statistical method. *Mol Ecol*. 24:1031–1046.
- Love MS, Yoklavich M, Thorsteinson L. 2002. *The rockfishes of the Northeast Pacific*. Berkeley (CA): University of California Press.
- Marko PB, Hart MW. 2011. The complex analytical landscape of gene flow inference. *Trends Ecol Evol*. 26:448–456.
- Mayr E. 1942. *Systematics and the origin of species from the view-point of a zoologist*. New York: Columbia University Press.
- Miller MR, Brunelli JP, Wheeler PA, Liu S, Rexroad CE 3rd, Palti Y, Doe CQ, Thorgaard GH. 2012. A conserved haplotype controls parallel adaptation in geographically distant salmonid populations. *Mol Ecol*. 21:237–249.
- Miller MR, Dunham JP, Amores A, Cresko WA, Johnson EA. 2007. Rapid and cost-effective polymorphism identification and genotyping using restriction site associated DNA (RAD) markers. *Genome Res*. 17:240–248.
- Miller DJ, Lea RN. 1972. *Guide to the coastal marine fishes of California*. Berkeley (CA): Fish Bulletin 157, California Department of Fish and Game.

- NASA Goddard Space Flight Center & Processing Ocean Biology Group. 2014. *Sea-viewing Wide Field-of-view Sensor (SeaWiFS) Ocean Color Data*. Available from: http://doi.org/10.5067/ORBVIEWS-2/SEAWIFS_OC.2014.0 (accessed 5 June 2019).
- Nosil P, Funk DJ, Ortiz-Barrientos D. 2009. Divergent selection and heterogeneous genomic divergence. *Mol Ecol*. 18:375–402.
- Paris JR, Stevens JR, Catchen JM. 2017. Lost in parameter space: a road map for STACKS. *Methods Ecol Evol*. 8:1360–1373.
- Pritchard JK, 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.
- Rainey PB, Travisano M. 1998. Adaptive radiation in a heterogeneous environment. *Nature*. 394:69–72.
- Riddle BR, Hafner DJ, Alexander LF, Jaeger JR. 2000. Cryptic vicariance in the historical assembly of a Baja California peninsular desert biota. *Proc Natl Acad Sci U S A*. 97:14438–14443.
- Rocha LA, Bass AL, Robertson DR, Bowen BW. 2002. Adult habitat preferences, larval dispersal, and the comparative phylogeography of three Atlantic surgeonfishes (Teleostei: Acanthuridae). *Mol Ecol*. 11:243–252.
- Rocha LA, Bowen BW. 2008. Speciation in Coral-Reef Fishes. *J Fish Biol*. 72:1101–1121.
- Rocha LA, Robertson DR, Roman J, Bowen BW. 2005. Ecological speciation in tropical reef fishes. *Proc Biol Sci*. 272:573–579.
- Rundle HD, Nosil P. 2005. Ecological Speciation. *Ecology Lett*. 8:336–352.
- Schinske JN, Bernardi G, Jacobs DK, Routman EJ. 2010. Phylogeography of the diamond turbot (*Hypsopsetta guttulata*) across the Baja California Peninsula. *Mar Biol*. 157:123–134.
- Schluter D. 2000. *The ecology of adaptive radiation*. Oxford (UK): Oxford University Press.
- Schunter C, Garza JC, Macpherson E, Pascual M. 2014. SNP development from RNA-seq data in a nonmodel fish: how many individuals are needed for accurate allele frequency prediction? *Mol Ecol Resour*. 14:157–165.
- Sexton JP, Hangartner SB, Hoffmann AA. 2014. Genetic isolation by environment or distance: which pattern of gene flow is most common? *Evolution*. 68:1–15.
- Stepien CA, Rosenblatt RH, Bargmeyer BA. 2001. Phylogeography of the spotted sand bass, *Paralabrax maculatofasciatus*: divergence of Gulf of California and Pacific Coast populations. *Evolution*. 55:1852–1862.
- Stockwell BL, Larson WA, Waples RK, Abesamis RA, Seeb LW, Carpenter KE. 2016. The application of genomics to inform conservation of a functionally important reef fish (*Scarus niger*) in the Philippines. *Conservation Genetics*. 17:239–249.
- Swift CC, Findley LT, Ellingson RA, Flessa KW, Jacobs DK. 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.
- Tavera JJ, Acero PA, Balart EF, 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.
- 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 DA, Findley LT, Kersitch AN. 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. 2013. *MARSPEC: Ocean Climate Layers for Marine Spatial Ecology*. 94:27705.