ORIGINAL PAPER

Phylogeography of the diamond turbot (*Hypsopsetta guttulata*) across the Baja California Peninsula

Jeffrey N. Schinske · Giacomo Bernardi · David K. Jacobs · Eric J. Routman

Abstract We compared morphology and sequenced nuclear and mitochondrial genes from 11 populations of a previously genetically unstudied "Baja California disjunct" species, the diamond turbot (*Hypsopsetta guttulata*). This species exhibits very limited adult movement and restriction to soft-bottom habitats but has a moderately long pelagic larval duration. Therefore, if pelagic larval duration is correlated with gene flow between Gulf of California and Pacific populations, we expect a reduced level of genetic and morphological differentiation. However, if adult habitat and ecology have more effect on gene flow, we expect the populations in the two bodies of water to be more highly differentiated. We used logistic regression to compare morphological features and phylogenetic and population genetic analyses to compare nucleotide sequence data. Gulf of California *H. guttulata* are different from Pacific populations in morphology and both mitochondrial and nuclear gene sequences. MtDNA shows reciprocal monophyly, and nuclear sequences from the Gulf of California formed a monophyletic group. Population genetic analyses also suggest further population subdivision within the Pacific and within the Gulf of California. We argue that adult ecology has a significant effect on migration rates among populations in the Pacific Ocean and the Gulf of California.

J. N. Schinske · E. J. Routman (✉)
Department of Biology, San Francisco State University,
1600 Holloway Avenue, San Francisco, CA 94132, USA
e-mail: routman@sfsu.edu

G. Bernardi
Department of Ecology and Evolutionary Biology,
University of California Santa Cruz, 100 Shaffer Road,
Santa Cruz, CA 95060, USA

D. K. Jacobs
Department of Ecology and Evolutionary Biology, UCLA,
621 Young Drive South, Los Angeles, CA 90095-1606, USA

Introduction

At least 19 species of fish have populations in Mexico's Gulf of California (Sea of Cortez) that appear to be partially or fully disjunct from their respective Pacific coast populations (Present 1987; Bernardi et al. 2003). These "Baja California disjunct" species fall into 14 families and display a wide range of life history characteristics. The Baja California region therefore serves as an excellent natural laboratory for exploring connections between a marine species' life history characteristics and its evolutionary response to vicariance events (see discussion in Jacobs et al. 2004 and Dawson et al. 2006 for discussion of taxon distributions and the geologic history of the Gulf of California). The degree of genetic subdivision exhibited across the Baja California peninsula in these species ranges from undetectable in some species to very strong in others (Crabtree 1983; Orton and Buth 1984; Present 1987; Waples 1987; Tranah and Allen 1999; Terry et al. 2000; Stepien et al. 2001; Huang and Bernardi 2001; Bernardi et al. 2003; Bernardi and Lape 2005).

Traditionally, the degree of genetic subdivision in marine species is believed to relate to a species' ability to disperse via larval exchange (Palumbi 1992, 1994). Bernardi et al. (2003) searched for correlations between the pelagic larval duration (PLD) and genetic structure exhibited by Baja California disjunct species and found no evidence for a relationship between PLD and genetic subdivision. The authors instead suggested that adult

migration and habitat availability in the Gulf of California may be more important factors in determining the level of subdivision in Baja California disjunct species. This hypothesis is supported by studies finding evidence for larval retention in the Gulf of California (Brogan 1994). In addition, the northwestern Gulf of California includes primarily soft-bottom, estuarine adult habitats that are mostly absent from the rocky, southern tip of Baja California (Walker 1960). However, few studies exist to provide detailed information regarding the migratory abilities and other life history characteristics of most Baja California disjunct fishes. The diamond turbot (*Hypsopsetta guttulata*) is a Baja California disjunct species for which detailed life history information is available but has not been characterized genetically.

Hypsopsetta guttulata (see Cooper and Chapleau 1998 and Evseenko 2003 for a discussion of nomenclatural history) is an estuarine flatfish ranging from Cape Mendocino to Bahía Magdalena on the Pacific coast of North America, and from Bahía Concepción to Guaymas in the Gulf of California (Fig. 1) (Miller and Lea 1972; Lane 1975; Present 1987). Tagging studies and otolith trace element tagging fingerprints suggest that individuals of the species exhibit low overall adult migration between estuaries (Lane 1975; Swearer et al. 2003). *Hypsopsetta guttulata* settles at the smallest size of any near-shore California flatfish (Ahlstrom et al. 1984) and exhibits a moderate PLD of 5–6 weeks (Gadomski and Peterson 1988). If larval duration was the main factor determining gene flow, diamond turbot would have the potential for substantial migration between the Gulf of California and the Pacific populations. However, if adult habitat specificity is an effective limitation to migration gene flow between the Gulf of California and the Pacific Ocean may be greatly reduced.

The goal of this study is to estimate the genetic differentiation between Pacific and Gulf of California populations of *H. guttulata*, using mitochondrial and nuclear sequence data and external morphology. We hypothesized that although the diamond turbot has the potential for gene flow via pelagic larvae, the limited adult migratory ability of *H. guttulata* due to its restriction to soft-bottom habitats should limit gene flow and result in genetic divergence between Gulf of California and Pacific populations.

Materials and methods

Sample collection, morphometrics, and DNA extraction

We collected *H. guttulata* in the field using a beach seine. In total, we obtained 77 specimens from 9 sites along *H. guttulata*'s Pacific range and 20 specimens from 2 sites

in the Gulf of California (Fig. 1; Table 1a). Note that the Yavaros sample in the Gulf of California represents a southward extension of the published range. Individual specimens of *Pleuronichthys verticalis*, *P. ocellatus*, and *Paralichthys californicus* were sampled as outgroups. Samples were stored in 95% ethanol and/or frozen at −20°C after collection (one *H. guttulata* specimen, G1, was freeze-dried for preservation). We recorded external morphological characteristics by counting dorsal and anal fin rays and measuring distances between external landmarks (Fig. 2). We augmented our sample sizes for morphological analysis by measuring additional specimens from the Scripps Institution of Oceanography Marine Vertebrate Museum (SIO_Samples). This brought our total number of specimens for morphological analysis to 50 Gulf of California samples and 90 Pacific samples. We extracted genomic DNA from muscle tissue of each non-museum specimen using standard silicon capture extraction protocols.

PCR amplification and sequencing

We designed primers, JSCRF: 5′-CCCTAACTCCCAAA GCTAGGATTCTAG-3′ and JSCRR: 5′-GGCCCATCTT AACATCTTCAGTGT-3′, to amplify the mitochondrial control region in *H. guttulata*. PCR products were sequenced using the forward primer, JSCRF, or an internal reverse primer, JSSeqCRR: 5′-CCTTACCCGCTGGAGTG AAYG-3′. Every unique haplotype was sequenced in both directions.

We used the universal primers, S7RPEX1F and S7RPEX2R (Chow and Hazama 1998), to amplify the first intron of the low-copy, ribosomal-protein-coding nuclear gene, S7 (S7 1). In order to confirm preliminary genetic patterns observed in S7 1, we also sequenced the second intron of S7 (S7 2) for 22 Pacific and 20 Gulf of California individuals (Table 1). We initially used the universal primers, S7RPEX2F and S7RPEX3R (Chow and Hazama 1998), to amplify S7 2. Due to inconsistent PCR results, we designed new primers based on preliminary sequences: JSS72F: 5′-CATCTCCAGCTCGAGCAGAG-3′ and JSS72R: 5′-AAAGCCAGACGAGTTTGAGTCT-3′. We sequenced S7 1 products using S7RPEX1F and an internal primer, JSS7SeqR: 5′- CTGGACGCCTGAATGTC-3′. We sequenced S7 2 products using JSS72F and JSS72R. Unique haplotypes were sequenced in both directions. When both S7 1 and S7 2 sequences were available for a specimen, we concatenated the sequences prior to analysis. PCR and sequencing protocols are available from the authors.

We used CHB v. 1.0 (Niu et al. 2001; Zhang et al. 2006) and PHASE v. 2.1 (Stephens et al. 2001; Stephens and Scheet 2005) to infer the gametic phases of nuclear

Fig. 1 Collecting sites for *H. guttulata*, including site abbreviations for each locality. Additional relevant geographic points are labeled in smaller text. Previously described range of *H. guttulata* (Present 1987) is shown as *gray line* parallel to the Pacific and Gulf of California coasts

haplotypes. We performed initial runs in each program using default parameter values then performed additional runs with different random number seeds and a greater number of preliminary burn-in steps and post-burn-in iterations (respectively, 350/175 for PHASE and 1000/700 for CHB). We compared the results of these runs for consistency between programs, and between runs within a program. Haplotypes whose gametic phases could not be consistently determined with greater than an 85% posterior probability were cloned using the TOPO TA cloning kit for sequencing (Invitrogen Corporation). We screened 4–9 colonies from each product in PCRs using M13 vector primers. We cleaned and sequenced those products as indicated earlier.

Morphological analyses

We analyzed morphological data in SPSS (v. 11.5.0) using direct logistic regression analysis with Gulf of California/Pacific origin as outcome and morphological observations as predictors. Logistic analysis permits the inclusion of data that are not normally distributed. Prior to analysis of

measurements between landmarks, we performed a least squares regression between measurements taken in different geometric planes (i.e., between BL and BD as opposed to BL and HL, see Fig. 2). BL, TL, HL, PF, and ML were regressed against BD. PF and ML were also regressed against BL (Fig. 2). In order to control for allometry, we used the residuals from these regressions for further analysis (Santos et al. 2006). We initially performed the logistic regression with all variables entering the model together in the first step. We also searched for the most efficient model by using stepwise addition of predictor variables based on the likelihood ratio test. We raised the probability for stepwise entry to 0.2 to help ensure that all variables with coefficients different than 0 entered the model (Tabachnick and Fidell 2007). A 2×2 classification table was used to examine the success of the resulting model in assigning individuals to region of origin.

Estimation of phylogenetic trees

Phylogenies were estimated separately for control region sequences and S7 sequences, using the following

Table 1 Genetic diversity for diamond turbot genes

Site name (code)	N	CR	S7 1	N (for S7 2)	S7 (1 + 2)
(a) Genetic diversity for each collecting site					
Tomales Bay (T)	2	2 (0.032)	2 (0.001)	0	
Oakland (O)	18	9 (0.024)	9 (0.002)	8	9 (0.005)
Los Alamitos Bay (L)	15	15 (0.038)	13 (0.004)	1	2 (0.008)
Mission Bay (M)	2	2 (0.058)	3 (0.003)	0	
Estero Punta Banda-Ensenada (E)	5	5 (0.042)	7 (0.004)	2	3 (0.003)
Bahía San Quintin (S)	1	1	2 (0.004)	0	
Guerrero Negro (G)	1	1	2 (0.006)	0	
El Cuarenta (U)	15	15 (0.043)	15 (0.005)	0	
Cancun/Gallinitas (C)	18	18 (0.038)	17 (0.004)	11	17 (0.005)
Bahía de Los Ángeles (*B*)	8	7 (0.036)	3 (0.001)	8	5 (0.002)
Yavaros (*Y*)	12	8 (0.013)	3 (0.001)	12	5 (0.001)

Locus	Gulf of California-Pacific	Within Pacific	Within Gulf of California		
(b) Average within region and between region genetic distances for each locus					
Control region	0.1211	0.0360	0.0230		
S7 (Introns 1 and 2)	0.0029	0.0051	0.0015		

N number of individuals sequenced. For each locus the number of different haplotypes are shown followed in parentheses by average within-sample Tamura/Nei genetic distances. *CR* mitochondrial control region. *S7 1* S7 ribosomal protein exon 1, *S7 2* S7 ribosomal protein exon 2. Gulf of California collecting sites are in italics
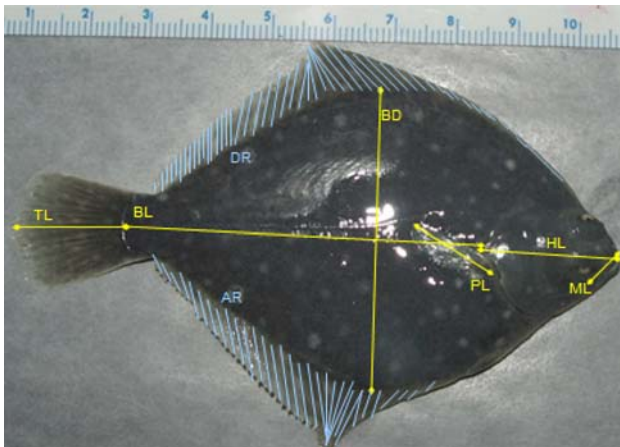

Fig. 2 Morphological measurements included body length (*BL*), tail length (*TL*), body depth (*BD*), head length (*HL*), maxilla length (*ML*), pectoral fin length (*PL*), number of anal rays (*AR*), and number of dorsal rays (*DR*)

procedures: sequences were aligned automatically in Sequencher (v. 3.1.1), and manually adjusted to ensure parsimonious alignment around gaps. We used PAUP 4.0b10 (Swofford 2003) to estimate maximum likelihood trees. The BIC in Modeltest 3.7 (Posada and Crandall 1998) was used to determine the least parameterized substitution model. In PAUP, we initially created a neighbor-joining tree and estimated model parameters on this topology using maximum likelihood and the substitution model selected using Modeltest. We then fixed these parameters values

and performed a maximum likelihood heuristic topology search. The new topology was then used to re-estimate the model parameters, and this process was repeated until model parameters and log likelihood scores stabilized. Maximum likelihood bootstrapping (100 replicates) was performed using the final estimated model parameters. We used these same parameters for Bayesian tree analysis in MrBayes (Huelsenbeck and Ronquist 2001; Ronquist and Huelsenbeck 2003). We ran 4 chains for 2,000,000 generations, sampling every 500 generations, and monitored the average deviation of split frequencies between independent runs for convergence. In calculating posterior probabilities, we used only the trees from generations after the average deviation of split frequencies between independent runs dropped below 0.05. All previous trees were deleted as burn-in. We calculated posterior probabilities by creating a 50% majority rule consensus tree in PAUP from all post-burn-in trees.

Population genetic analysis

To estimate *H. guttulata* population structure across the Baja peninsula and along their Pacific range, we estimated F_{st}, Φ_{st} (between populations), Φ_{ct} (between regions) and AMOVA statistics with Arlequin v.3.01 (Schneider et al. 2000) using the TrN+G model with gamma distribution shape parameter set to the values estimated for each dataset in PAUP.

When high levels of haplotype diversity were encountered, we created haplotype networks using TCS v.1.21

(Clement et al. 2000). (Links to certain control region haplotypes could not be established in haplotype networks due to their unusually high sequence divergence from the other haplotypes. Those haplotypes were treated as separate clusters.) Closed loops in the networks were broken by preserving evolutionary relationships between haplotypes suggested by our maximum likelihood trees. Haplotypes were clustered as in Nested Clade Analysis (Templeton et al. 1995), and Arlequin analyses were repeated at various nesting levels to examine the population subdivision for related clusters of alleles.

We used LAMARC v.2.1 (Kuhner 2006) to estimate the population parameter Θ and the number of migrants per generation between populations. We employed Bayesian searches on combined control region and S7 1 data, since recent studies suggested these strategies produce more precise, and potentially more accurate, estimates (Beerli 2006; Kuhner and Smith 2007). Only populations with greater than five individuals (O, L, U, C, B, Y, Fig. 1) were used. We set the effective population size scalar to 4 for nuclear sequences and 1 for control region sequences. We set the relative mutation rate to 1 for nuclear sequences and 7.3 for control region sequences, determined by examining relative model-corrected distances among nuclear haplotypes and among control region haplotypes. The transition/transversion ratio for each locus was set at the values estimated by PAUP. We performed iterative analyses in LAMARC, similar to maximum likelihood tree estimation. All search parameter values were left at default levels for the initial run, except that we increased the number of long chains to 3. All additional runs used only 1 long chain, and no short chains, with 3,000 burn-in steps, 20,000 samples, and a sampling interval of 40. Parameters were averaged over five independent sub-runs. In addition, adaptive heating was used with initial temperatures set to 1.0, 1.2, 1.5, and 3.0, and a swap interval of 1. After the initial run, starting values of Θ and migration for run i were set to the values estimated in run $i\text{-}1$. We performed 3 runs, tracking the most probable estimates and confidence intervals of parameters for consistency and convergence over runs. After 3 runs, 95% confidence intervals for estimated parameters remained relatively constant between runs, so analysis was concluded. We additionally plotted the contents of the curve files produced for each parameter to determine whether the program estimated a single optimum for each parameter or whether the data suggested multiple potentially optimal values. Since we set the control region effective population size scalar to 1 and the S7 scalar to 4, LAMARC estimated Θ on the mitochondrial scale ($\Theta = N\mu$). We multiplied this value by 4 in order to report Θ estimates on the scale of an autosomal gene. We also multiplied the migration estimates [M(LAMARC) = m/μ] by the Θ estimate (4 Nμ) for the recipient population to report migration as M = 4 Nm. We calculated 95% confidence intervals for M by multiplying the upper and lower confidence values for Θ by the respective upper or lower values for M(LAMARC). We used these same LAMARC protocols to estimate migration between the Gulf of California and Pacific regions. For these analyses, the input file contained combined S7 1 and S7 2 data grouped by region, along with corresponding control region sequences for those individuals.

Results

Morphological analyses

Average body depth was not significantly different between the Gulf of California sample [29.39 mm ± 16.75 (SD)] and the Pacific Ocean sample (30.95 mm ± 26.63), suggesting that our samples did not differ in overall body size. The initial logistic regression model with all variables entered was statistically better than a constant-only model (Chi-square 55.249, df 10, $P < 0.001$). This shows that the morphological variables, as a set, reliably distinguish between Gulf of California and Pacific samples. The stepwise logistic regression indicates that 5 variables (number of anal fin rays, head length, pectoral fin length, maxilla length, and tail length) contribute most to the ability of the model to predict region of origin (Table 2). Including additional morphological variables in the model does not significantly improve the model's ability to predict the outcome. This 5-variable model correctly assigns 77.27% of Gulf of California samples and 88.16% of Pacific samples to their region of origin. Head length exhibits the largest odds ratio (4.1097), indicating that assignment to Gulf of California or Pacific is most highly influenced by changes in size-corrected head length. The values of the 5 morphological variables included in the model overlap significantly in Pacific and Gulf of California samples, so we cannot assign a strict cutoff for separating Pacific and Gulf of California specimens based on these features.

MtDNA sequence analysis

We obtained between 426 and 429 bps of sequence from the 5′ end of the control region in *H. guttulata*. This locus includes 143 variable sites and 6 single-base insertions/deletions (indels) between sequences. Sequences exhibit 10 fixed differences between Gulf of California and Pacific populations, including 1 fixed indel in Gulf of California sequences at base 349. The average, corrected pairwise genetic distance between Pacific and Gulf of California regions derived from Arlequin using the GTR+I+G model described later is 0.1211. This contrasts with an average

Table 2 Results of logistic regression

Variable	B	Wald.	Sig.	Exp(B)
(a) Logistic regression variables				
Anal fin rays	−0.453	12.353	0.0004	0.636
Head length-body depth	1.413	13.484	0.0002	4.110
Pectoral fin length-body depth	0.749	4.238	0.0395	2.115
Maxilla length-body depth	−4.056	11.819	0.0006	0.017
Tail length-body depth	−0.771	6.238	0.0125	0.463
Constant	24.050	13.134	0.0003	2.80E+10

Observed	Predicted by model		
	Gulf of California	Pacific	Percentage correct (%)
(b) Classification table: percentage of samples correctly assigned to their region of origin using the model from the logistic regression			
Cortez	34	10	77.27
Pacific	9	67	88.16
Overall	–	–	84.17

B is the logistic regression coefficient for each predictor variable. Wald. and Sig. describe the results and significance of Wald Tests, respectively. Exp(B) is the odds ratio for each predictor. The degrees of freedom for each Wald value is 1

pairwise distance of 0.0360 among Pacific sequences and 0.0230 among Gulf of California sequences (Table 1b). All control region sequences are available from GenBank with accession numbers FJ155668–FJ155748.

MtDNA phylogenetic tree

The large between-region genetic distances of control region sequences are reflected in the strong statistical support for reciprocal monophyly between Gulf of California and Pacific haplotypes in the maximum likelihood phylogenetic tree (Fig. 3). Modeltest (BIC) chose the GTR+I+G substitution model for control region sequences. The final maximum likelihood model included a ti/tv of 10.1246, a proportion of invariant sites of 0.5145, and a shape parameter of 1.1429. We did not obtain clean control region sequence for *P. verticalis* or *P. ocellatus*, and *P. californicus* control region exhibits an uncorrected genetic distance of 0.57 from the nearest *H. guttulata* haplotype. In contrast, the maximum uncorrected distance between *H. guttulata* haplotypes is 0.12. This extreme level of sequence divergence between *H. guttulata* and the outgroup prevented outgroup estimation of the proper tree root. The tree is therefore shown with midpoint rooting.

MtDNA population genetic analyses

Clustering populations into 2 regions for AMOVA results in a Φ_{ct} of 0.774 between the Pacific and the Gulf of California regions. All pairwise Φ_{st} values between
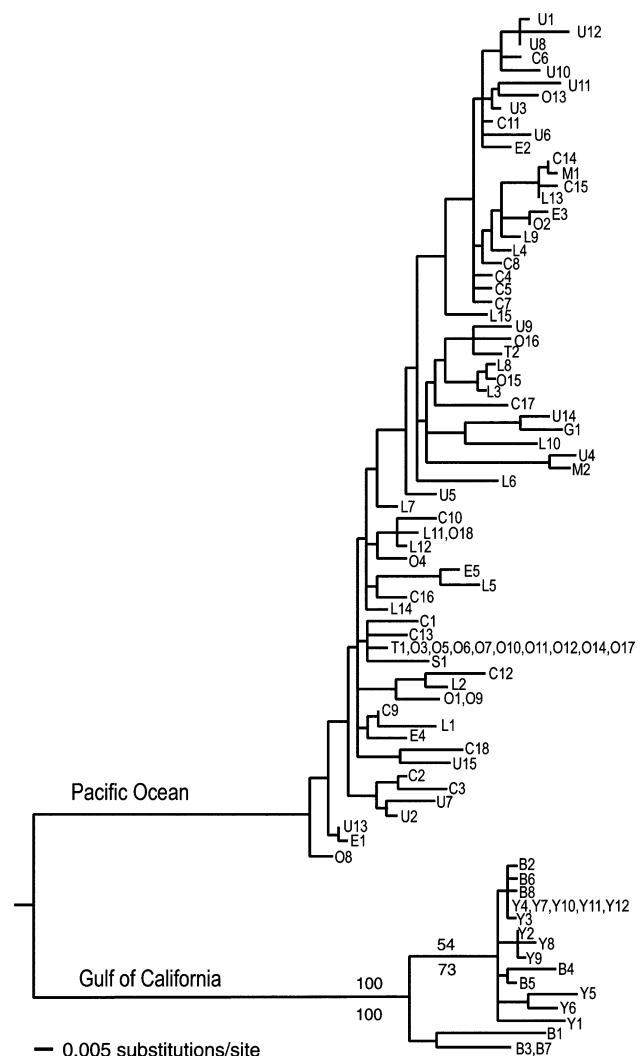

Fig. 3 Maximum likelihood control region haplotype phylogeny. Bayesian posterior probabilities are shown above branches and maximum likelihood bootstraps are below branches. Haplotypes are labeled with sample numbers (site code + sample no. from site) of the individuals possessing the haplotype

populations from the Pacific and Gulf of California populations (with $N > 5$) were found to be significantly different than 0, ranging from 0.740 to 0.869 (Table 3). Between-population variation was much lower within regions. Pairwise Φ_{st} ranges from 0.000 to 0.130 in the Pacific and equals 0.148 between the two Gulf of California sites. However, some significant variation was found. Significant pairwise Φ_{st} values were estimated between Oakland and 2 southern Pacific sites, and between the Gulf of California sites of Bahía de Los Ángeles and Yavaros (Table 3).

Because of the large number of control region haplotypes, the high within-population variation (Table 1) could mask between-population variation in Φ_{st} analyses for the Pacific populations. Therefore, we clustered haplotypes by

Table 3 Population structure comparisons for Diamond Turbot

	Oakland	Los Alamitos	Ensenada	El Cuarenta	Cancun	*BahíaLA*	*Yavaros*
Oakland	–	0.155**	0.113**	0.177**	0.041**	0.431**	0.468**
Los Alamitos	0.050	–	0.007	0.007	0.040*	0.430**	0.474**
Ensenada	0.048	0.000	–	0.000	0.002	0.433**	0.513**
El Cuarenta	0.130**	0.022	0.000	–	0.060**	0.404**	0.448**
Cancun	0.074*	0.003	0.000	0.005	–	0.352**	0.389**
Bahía LA	0.807**	0.758**	0.745**	0.740**	0.751**	–	0.068
Yavaros	0.869**	0.830**	0.862**	0.815**	0.818**	0.148*	–

Values of Φ_{st} analysis of control region (below diagonal) and S7 1 (above diagonal)

Note that the sample numbers (in parentheses) for the nested analysis are lower than those in Table 1. This is because, in forming haplotype networks, TCS could not determine the links to some highly diverged control region haplotypes, and those haplotypes were excluded from this analysis

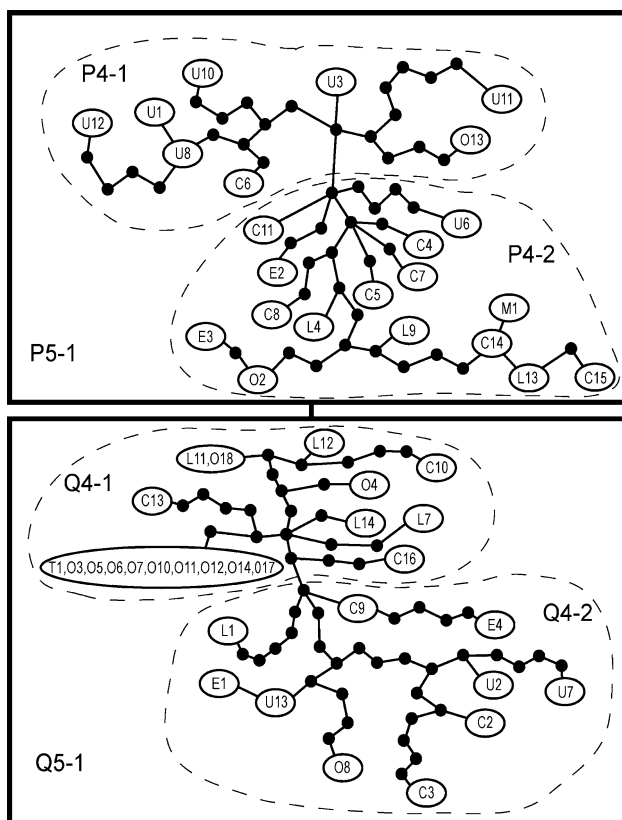* $P < 0.05$ and ** $P < 0.01$. Gulf of California collecting sites are in italics

Fig. 4 Haplotype networks for Pacific control region sequences ("P" network and "Q" network). The exact connection between these networks is unknown. Nestings are shown only for higher level clades. Each branch represents a single nucleotide difference between adjacent haplotypes. Haplotype names are shown as *ovals* containing sample names. Unsampled but inferred intermediate haplotypes are shown as *small filled dots* in branches

relationship by recoding the haplotypes for Pacific individuals as belonging to one of 4 clades: P4-1, P4-2, Q4-1, and Q4-2 (Fig. 4). The recoded data set was then used to estimate traditional (frequency only) F_{st} values in Arlequin.

High and statistically non-zero F_{st} estimates between northern and southern Pacific sites support the conclusion of restricted gene flow within this region (Table 4).

Autosomal intron sequence analysis

As expected, rpS7 intron sequences are less variable than control region sequences. We identified 27 polymorphic sites out of 508 total bases in S7 1. One individual from Oakland additionally appears heterozygous for a 2-base deletion in S7 1. Preliminary comparisons between Pacific and Gulf of California populations suggested that only one very common allele was shared between those regions. We sequenced S7 2 in order to determine whether additional sequence data would provide evidence that this shared allele is in fact different between the Pacific and the Gulf of California regions. We obtained 280 bps of sequence from S7 2, which included 11 polymorphic sites and no indels. Concatenated S7 sequences revealed that some individuals with the shared S7 1 haplotype are indeed different when additional S7 gene sequence is added. There is still a single, shared haplotype between Pacific and Gulf of California populations, but that 2-intron haplotype (h1) is not as prevalent (occurring only twice in Pacific samples and 5 times in Gulf of California samples) as the shared S7 1 haplotype. The average, corrected pairwise genetic distance between Pacific and Gulf of California S7 sequences derived from Arlequin (S7 1 and 2 combined) is 0.0029. The within Pacific distance is 0.0051, and the within Gulf of California distance is 0.0015 (Table 1b). All rpS7 sequences are available from GenBank with accession numbers FJ155749–FJ155799.

Autosomal intron phylogenetic tree

Here, we present only the tree estimated from combined S7 1 and S7 2 data (Fig. 5). Since individual S7 1 or S7 2 trees

Table 4 Results of F_{st} analysis of nested Pacific haplotype data

	Oakland (14)	Los Alamitos (8)	Ensenada (4)	El Cuarenta (10)
Los Alamitos	0.079			
Ensenada	0.487**	0.102		
El Cuarenta	0.475**	0.311*	0.218	
Cancun (14)	0.292**	0.000	0.000	0.223*

Note that the sample numbers (in parentheses) for the nested analysis are lower than those in Table 1. This is because, in forming haplotype networks, TCS could not determine the links to some highly diverged control region haplotypes, and those haplotypes were excluded from this analysis

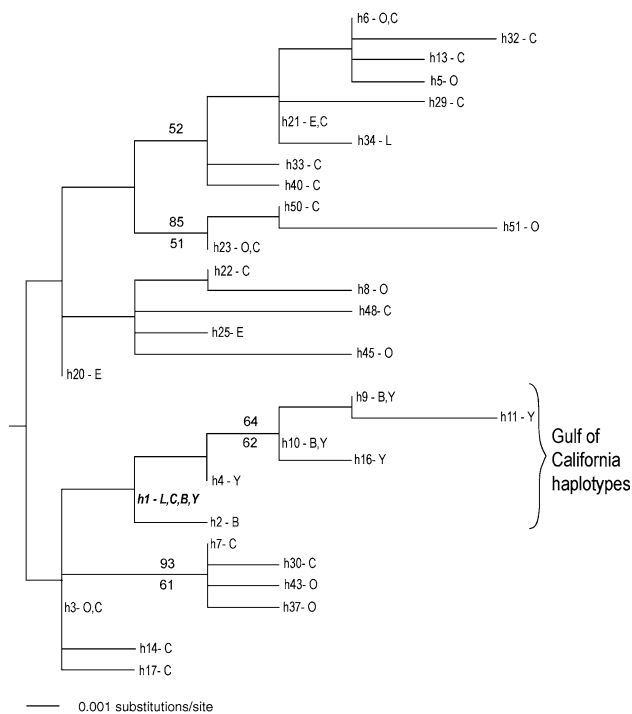* $P < 0.05$ and ** $P < 0.01$

Fig. 5 Maximum likelihood S7 haplotype phylogeny. Bayesian posterior probabilities are shown above branches and maximum likelihood bootstraps are below branches. *Letters* next to haplotype numbers are site codes indicating sites where that haplotype was found

incorporate less sequence data, they include fewer haplotypes and the resolution of clades is poor. Again, divergence between *H. guttulata* and outgroup S7 sequences is large. The least diverged outgroup from *H. guttulata* exhibits an uncorrected genetic distance of 0.19 from *H. guttulata* specimens. That contrasts with a maximum uncorrected distance of 0.017 between *H. guttulata* specimens. As explained above for control region, the phylogenetic tree for S7 is therefore midpoint rooted. Modeltest chose the HKY+I substitution model for S7 sequences. After heuristic searches, the final model included a ti/tv of 1.5402, a proportion of invariant sites of 0.9429, and a shape parameter of 3.4358. The maximum likelihood tree places the only haplotype shared between the Pacific and

the Gulf of California (h1) at the base of a monophyletic clade that includes all the other Gulf of California haplotypes (h2, h4, h9, h10, h11, h16). Although the monophyly of Gulf of California haplotypes is not strongly supported by bootstrap values or Bayesian posterior probabilities, it is interesting to note that maximum likelihood analyses of both control region and S7 grouped haplotypes from the Gulf of California as monophyletic groups.

Autosomal intron population genetic analyses

All S7 1 pairwise Φ_{st} values between Pacific and Gulf of California populations are statistically distinguishable from 0. As mentioned above, Φ_{st} values are reported only for populations with 5 or more individuals sampled. Pairwise Φ_{st} values between the Gulf of California sites and Pacific sites range from 0.352 to 0.513 (Table 3). By comparison, pairwise Φ_{st} values among populations within the Pacific range from 0.000 to 0.177 and equals 0.068 between the Gulf of California sites. Clustering populations into regions in the overall AMOVA results in a Φ_{ct} of 0.345 between the Pacific and the Gulf of California. Combining S7 1 data with S7 2 data in AMOVA produces a Gulf of California-Pacific Φ_{ct} of 0.456.

Migration between populations

Using combined control region and S7 data, LAMARC estimated very restricted migration between the Gulf of California and the Pacific, and relatively high levels of migration between most Pacific populations (Table 5; Fig. 6). Migration into Oakland and out of Oakland to El Cuarenta and Cancun appear relatively reduced compared to migration between other sites. Additionally, migration from Cancun to Los Alamitos and El Cuarenta (northward) was estimated at several times the rates of migration into Cancun from those sites (southward). However, the 95% confidence intervals for these migration rates are quite high, and the differences are not statistically significant. Analysis of curve files produced by LAMARC indicates that the Θ and migration estimates for each locus are the

Table 5 Estimates of *H. guttulata* migration produced by LAMARC analysis

From	To			
	Oakland	Los Alamitos	El Cuarenta	Cancun
(a) Within regions				
Pacific Ocean				
Oakland	–	1.52 (**12.10**) 36.45	0.02 (**4.99**) 27.72	0.32 (**3.43**) 17.74
Los Alamitos	0.91 (**4.49**) 8.40	–	4.04 (**26.46**) 75.24	3.85 (**11.09**) 42.24
El Cuarenta	0.01 (**1.41**) 6.14	3.44 (**23.38**) 55.52	–	4.75 (**10.93**) 41.99
Cancun	0.15 (**1.92**) 7.22	4.39 (**29.36**) 56.83	6.08 (**34.80**) 82.51	–

From	To	
	Bahía LA	*Yavaros*
Gulf of California		
Bahía LA	–	0.22 (**3.39**) 12.48
Yavaros	0.26 (**4.72**) 23.97	–

From	To	
	Pacific	Gulf of California
(b) Between regions		
Pacific	–	0.004 (**1.339**) 12.286
Gulf of California	0.021 (**0.763**) 4.721	–

Most probable estimates (MPE) of Nm (=number of migrant individuals per generation) are shown in bold flanked by upper and lower 95% confidence values in regular text. Gulf of California collecting sites are in italics

results of single, optimal peaks in the data, and the overall estimate for combined data appears to reconcile differences in estimates between the two loci. It should be noted that interpretation of Nm estimates as migration assumes that retained ancestral polymorphism is not responsible for shared alleles or related sets of alleles. Given the reciprocal monophyly of the mtDNA, retention of ancestral polymorphism may be a better interpretation of the LAMARC results than high rates of gene flow.

Discussion

Gulf of California range for *H. guttulata*

Present (1987) included *H. guttulata* in the group of Baja California disjunct species whose Gulf of California range does not extend south of Guaymas Bay on the coast of mainland Mexico (Fig. 1). Since the Yavaros samples in this study originate from a site over 150 km south of Guaymas Bay, the Gulf of California range for *H. guttulata* should be extended to at least that point.

Gulf of California: Pacific disjunction and speciation

Strong evidence of reciprocal monophyly between Gulf of California and Pacific mitochondrial control region

sequences (Fig. 3) and the implication of ongoing sorting toward monophyly in nuclear sequences (Fig. 5) both suggest that *H. guttulata* has significantly reduced gene flow between Pacific and Gulf of California populations. Like some other Baja California disjunct species (Bernardi et al. 2003; Bernardi and Lape 2005), Gulf of California and Pacific *H. guttulata* may be in the process of speciation. Given a 1:1 sex ratio, the effective population size for mitochondrial loci is ¼ that of autosomal loci because of maternal inheritance and effective haploidy. This means that allele sorting between allopatric populations is expected to proceed more rapidly for mitochondrial sequences (N generations on average, where N is the inbreeding effective population size) than for nuclear sequences (4 N generations) (Moore 1995). Our data suggest that Pacific and Gulf of California populations of *H. guttulata* have not been separated long enough for full reciprocal monophyly for nuclear loci. Another Baja California disjunct species, *A. davidsonii*, has been studied using both mitochondrial and nuclear sequence data (Bernardi and Lape 2005). Gulf of California and Pacific populations of *A. davidsonii* showed reciprocal monophyly for mtDNA, but a significant amount of sharing of rpS7 sequences.

Additional morphological and molecular analyses support the possibility of allopatric, incipient speciation. External morphological features show statistically
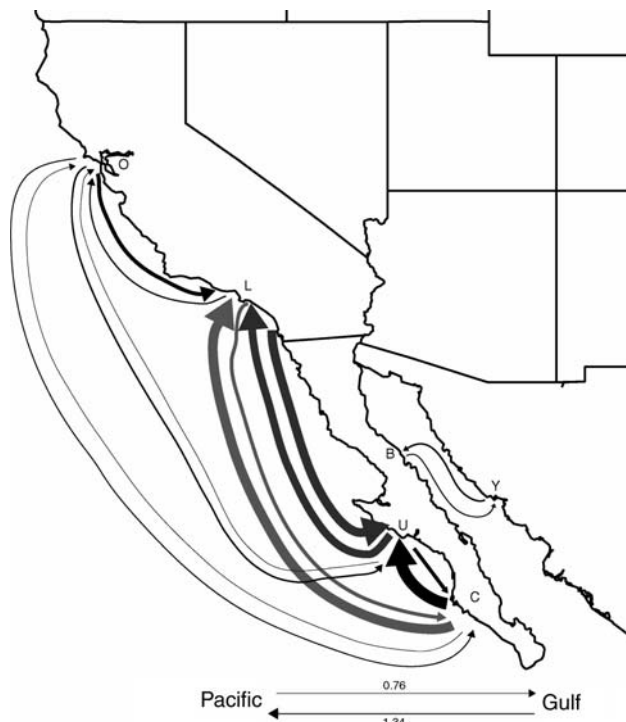
Fig. 6 Estimates of migration rate between 4 Pacific sites (O, L, U, C), the 2 Gulf of California sites (B, Y), and between the Gulf of California and Pacific regions. Relative thickness of *arrows* corresponds to relative levels of migration between sites. Color differences between *arrows* are only to improve readability. *Numbers on arrows* indicate number of migrants per generation

detectable differences between Gulf of California and Pacific *H. guttulata* (Table 2). Φ_{st} and migration analyses of control region and S7 sequences both additionally detect significant restrictions in gene flow between the Pacific and the Gulf of California regions (Tables 3, 4, 5). Further study of *H. guttulata* using a wider range of more quickly evolving nuclear genes could potentially detect reciprocal monophyly for some loci.

Phylogeography within regions

Φ_{st} analyses and migration estimates suggest varying levels of reduced gene flow between sites within the Gulf of California and Pacific regions. In the Gulf of California, two analyses (control region Φ_{st}, migration estimation) detect structure between Bahía de Los Ángeles and Yavaros (Tables 3 and 5; Fig. 6). With only two populations sampled from the Gulf of California, we cannot assess details of within Gulf differentiation that are known for some species (Riginos 2005). Φ_{st} analyses, migration estimates, and control region F_{st} analysis of clustered haplotypes also detect genetic structure among Pacific sites. All Pacific analyses suggest restricted gene flow between northern Pacific and southern Pacific sites. Point

Conception and Punta Eugenia (Fig. 1) have traditionally been recognized as important phylogeographic barriers for marine fishes (Dawson et al. 2006). Our data are consistent with the possibility of these general regions as phylogeographic barriers for *H. guttulata*. The S7 1 Φ_{st} analysis (Table 3) suggests genetic structure between Oakland and Los Alamitos. Additionally, when nested control region data (data not shown) are grouped as either north or south of Point Conception, the resulting frequency-based F_{st} is significant ($F_{st} = 0.262$). Analyses also imply genetic structure in the region of central Baja California. The S7 1 Φ_{st}, control region Φ_{st}, and migration estimates (Tables 3, 4, and 5; Fig. 6) all point to restricted gene flow between Oakland and the 2 sites south of Punta Eugenia (El Cuarenta and Cancun). The S7 1 Φ_{st} analysis shows restricted gene flow between Los Alamitos and Cancun (Table 3), and the control region F_{st} analysis on nested haplotypes (Table 4) detects restricted gene flow between sites north and south of Punta Eugenia. Taken together, these results seem to indicate that central Baja California represents an important area in the phylogeography of *H. guttulata*. Two population genetic analyses, control region nested F_{st} and S7 1 Φ_{st}, also indicate structure between the southern Pacific sites of El Cuarenta and Cancun.

Ecology/life history and trans-Baja California differentiation

Results from our study indicate that *H. guttulata* conforms to expectations of population genetic structure based on its adult life history. As an example of a Baja California disjunct species with a low adult migratory ability and a dependence on soft-bottom habitats, *H. guttulata* exhibits relatively strong genetic differentiation between Gulf of California and Pacific populations. Examination of habitat associations, life history characteristics, and museum collection records for *H. guttulata* and other Baja California disjunct species supports the hypothesis that adult migration and ecological factors play important roles in determining whether Baja California disjunct species are genetically distinct or homogeneous across Baja California. For example, all of the presently studied Baja California disjunct species exhibiting little genetic differentiation across Baja California (zebraperch, rock wrasse, California sheephead, Mexican rockfish) are associated with conspicuous rocky habitats (Allen and Pondella 2006). In contrast, those Baja California disjunct species associated with either soft-substratum habitats (grunion, spotted bass, diamond turbot, orangethroat pikeblenny, longjaw mudsucker) or cryptic rocky reef habitats (mussel blenny, blue banded goby) (Allen and Pondella 2006) exhibit high Gulf of California/Pacific genetic divergence (for genetic references see Intro). This matches well with the hypothesis

that the non-uniform distribution of soft-bottom habitat in the Gulf of California (Walker 1960) contributes to allopatry and incipient speciation in these species. A full review of the literature is in progress (Schinske in preparation).

Ecological studies on some Baja California disjunct fishes additionally suggest relatively limited adult migration in species with genetic differentiation across Baja California. Adult movement appears restricted for the diamond turbot (Lane 1975; Swearer et al. 2003), mussel blenny (Losey 1968; Stephens et al. 1970), spotted bass (Allen et al. 1995), blue banded goby (Steele 1997, 1998), orangethroat pikeblenny (Thresher 1997), longjaw mudsucker (Barlow 1961; Bernardi et al. 2003), and opaleye (Valle 1989). While few studies have examined migration in the panmictic Baja California disjunct species, collection records indicate the possible occasional appearance of those species near the tip of Baja California (CICIMAR Sample 1693; SIO Sample 65-346; SIO Sample 84-73; SIO Sample 59-219; 65-230; Present 1987; Bernardi et al. 2003). These studies and observations further support the concept that life history considerations restrict the populations of certain Baja California disjunct species to particular areas in the Baja California region, resulting in allopatric speciation between those geographically isolated populations. However, drawing firm conclusions will require additional long-term ecological studies of these species' habitat associations and adult movement (especially in panmictic species). The present study demonstrates the necessity of both ecological and genetic data when considering the forces that affect marine evolutionary divergence and speciation.

References

Ahlstrom EH, Amaoka K, Hensley DA, Moser HG, Sumida BJ (1984) Pleuronectiformes: development. In: Moser HG, Richards WJ, Cohen DM, Fahay MP, Kendall AW, Richardson SL (eds) Ontogeny and systematics of fishes. Allen Press Inc., Lawrence, pp 640–670 (Special Publications of the American Society of Ichthyologists and Herpetologists)

Allen LG, Pondella DJ (2006) Ecological classification. In: Allen LG, Pondella DJ, Horn MH (eds) The ecology of marine fishes: California and adjacent waters. University of California Press, Berkeley, pp 81–113

Allen LG, Hovey TE, Love MS, Smith JTW (1995) The life history of the spotted sand bass (*Paralabrax maculatofasciatus*) within the southern California bight. Calif Co-op Fish Investig Rep 36:193–203

Barlow GW (1961) Gobies of the genus *Gillichthys*, with comments on the sensory canals as a taxonomic tool. Copeia 1961:423–437

Beerli P (2006) Comparison of bayesian and maximum-likelihood inference of population genetic parameters. Bioinformatics 22:341–345

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, Findley L, Rocha-Olivares A (2003) Vicariance and dispersal across Baja California in disjunct marine fish populations. Evolution 57:1599–1609

Brogan MW (1994) Distribution and retention of larval fishes near reefs in the Gulf of California. Mar Ecol Prog Ser 115:1–13

Chow S, Hazama K (1998) Universal pcr primers for s7 ribosomal protein gene introns in fish. Mol Ecol 7:1255–1256

CICIMAR Colección Ictiológica del Centro Interdisciplinario de Ciencias Marinas. Collection database queried via FishBase (http://www.fishbase.org/museum/SearchFishCollections.php)

Clement M, Posada D, Crandall KA (2000) Tcs: a computer program to estimate gene genealogies. Mol Ecol 9:1657–1660

Cooper JA, Chapleau F (1998) Monophyly and intrarelationships of the family pleuronectidae (pleuronectiformes), with a revised classification. Fish Bull 96:686–726

Crabtree CB (1983) Assessment of genetic differentiation within the atherinid genus *leuresthes*. Isozyme Bull 16:77

Dawson MN, Waples RS, Bernardi G (2006) Phylogeography. In: Allen LG, Pondella DJ, Horn MH (eds) The ecology of marine fishes: California and adjacent waters. University of California Press, Berkeley

Evseenko SA (2003) New finding of accessory lateral line canals in the caudal fin of flatfishes: *Hypsopsetta guttulata* (girard, 1856) (pleuronectidae sensu chapleau and keast, 1988). J Ichthyol 43:603–606

Gadomski DM, Peterson JH (1988) Effects of food deprivation on the larvae of two flatfishes. Mar Ecol Prog Ser 44:103–111

Huang D, Bernardi G (2001) Disjunct sea of cortez-pacific ocean *gillichthys mirabilis* populations and the evolutionary origin of their paedomorphic relative, *gillichthys seta*. Mar Biol 138:421–428

Huelsenbeck JP, Ronquist F (2001) Mrbayes: Bayesian inference of phylogenetic trees. Bioinformatics 17:754–755

Jacobs DK, Haney TA, Louie KD (2004) Genes, diversity, and geologic process on the pacific coast. Annu Rev Earth Planet Sci 32:601–652

Kuhner MK (2006) Lamarc 2.0: maximum likelihood and bayesian estimation of population parameters. Bioinformatics 22(76):768–770

Kuhner MK, Smith LP (2007) Comparing likelihood and bayesian coalescent estimation of population parameters. Genetics 175:155–165

Lane ED (1975) Quantitative aspects of the life history of the diamond turbot, *hypsopsetta guttulata* (girard), in Anaheim Bay. In the marine resources of Anaheim Bay. Calif Depar Fish Game Fish Bull 165:153–174

Losey GS (1968) The comparative behavior of some Pacific fishes of the genus *Hypsoblennius* Gill (Blenniidae). Ph.D. Thesis, Department of Zoology, San Diego

Miller D, Lea R (1972) Guide to the coastal marine fishes of California. Calif Depar Fish Game Fish Bull 157:249

Moore WS (1995) Inferring phylogenies from mtdna variation: mitochondrial-gene trees versus nuclear-gene trees. Evolution 49:718–726

Niu T, Qin ZS, Xu X, Liu JS (2001) In silico haplotype determination of a vast set of single nucleotide polymorphisms. Technical Report, Department of Statistics, Harvard University, Cambridge

Orton RD, Buth DG (1984) Minimal genetic differentiation between *Girella nigricans* and *Girella simplicidens* (perciformes: Kyphosidae). Isozyme Bull 17:66

Palumbi SR (1992) Marine speciation on a small planet. Trends Ecol Evol 7:114–118

Palumbi SR (1994) Genetic divergence, reproductive isolation, and marine speciation. Annu Rev Ecol Syst 25:547–572

Posada D, Crandall KA (1998) Modeltest: testing the model of DNA substitution. Bioinformatics 14:817–818

Present T (1987) Genetic differentiation of disjunct Gulf of California and pacific outer coast populations of *hypsoblennius jenkinsi*. Copeia 1987:1010–1024

Riginos C (2005) Cryptic vicariance in Gulf of California fishes Parallels Vicariant patterns found in baja California mammals and reptiles. Evolution 59:2678–2690

Ronquist F, Huelsenbeck JP (2003) Mrbayes 3: Bayesian phylogenetic inference under mixed models. Bioinformatics 19:1572–1574

Santos S, Hrbek T, Farias IP, Schneider H, Sampaio I (2006) Population genetic structuring of the king weakfish, macrodon ancylodon (sciaenidae), in Atlantic coastal waters of South America: deep genetic divergence without morphological change. Mol Ecol 15:4361–4373

Schneider S, Roessli D, Excoffier L (2000) Arlequin: a software for population genetics data analysis. Ver 2.000. Genetics and Biometry Laboratory, Department of Anthropology, University of Geneva, Geneva

Sio_Samples Scripps Institution of Oceanography Marine Vertebrate Museum Sample Nos. 61-183, 61-186, 62-223, 62-737, 63-484, 64-69, 64-71, 65-31, 68-383

Steele MA (1997) Population regulation by post-settlement mortality in two temperate reef fishes. Oecologia 112:64–74

Steele MA (1998) The relative importance of predation and competition in two reef fishes. Oecologia 115:222–232

Stephens M, Scheet P (2005) Accounting for decay of linkage disequilibrium in haplotype inference and missing data imputation. Am J Hum Genet 76:449–462

Stephens JS, Johnson RK, Key GS, McCosker JE (1970) The comparative ecology of three sympatric species of California blennies of the genus *Hypsoblennius* gill (Teleostomi, Blenniidae). Ecol Monogr 40:213–233

Stephens M, Smith NJ, Donnelly P (2001) A new statistical method for haplotype reconstruction from population data. Am J Hum Genet 68:978–989

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

Swearer SE, Forrester GE, Steele MA, Brooks AJ, Lea DW (2003) Spatio-temporal and interspecific variation in otolith trace-elemental fingerprints in a temperate estuarine fish assemblage. Estuar Coast Shelf Sci 56:1111–1123

Swofford DL (2003) Paup*. Phylogenetic analysis using parsimony (*and other methods), version 4.0b10. Sinauer Associates, Sunderland

Tabachnick BG, Fidell LS (2007) Using multivariate statistics, 5 edn. Allyn & Bacon, Boston

Templeton AR, Routman E, Phillips CA (1995) Separating population structure from population history: a cladistic analysis of the geographical distribution of mitochondrial DNA haplotypes in the tiger salamander, Ambystoma tigrinum. Genetics 140:767–782

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

Thresher RE (1997) Field observations on the behavior and ecology of the orangethroat pikeblenny, *Chaenopsis alepidota*, in the Sea of Cortez. Bull Mar Sci 60:1195–1204

Tranah GJ, Allen LG (1999) Morphologic and genetic variation among six populations of the spotted sand bass, *Paralabrax maculatofasciatus*, from southern California to the upper sea of cortez. Bull South Calif Acad Sci 98:103–118

Valle CF (1989) The homing behavior and intertidal movements of the opaleye, *Girella nigricans* (Pisces: Kyphosidae). Master's Thesis, Department of Biology, Long Beach

Walker BW (1960) The distribution and affinities of the marine fish fauna of the Gulf of California. Syst Zool 9:123–133

Waples RS (1987) A multispecies approach to the analysis of gene flow in marine shore fishes. Evolution 41:385–400

Zhang Y, Niu T, Liu JS (2006) A coalescence-guided hierarchical bayesian method for haplotype inference. Am J Hum Genet 79:313–322