UC Santa Cruz
UC Santa Cruz Electronic Theses and Dissertations

Title
Molecular Evolution in Embiotocid Surfperches

Permalink
https://escholarship.org/uc/item/243468bd

Author
Longo, Gary

Publication Date
2016

Supplemental Material
https://escholarship.org/uc/item/243468bd#supplemental

Peer reviewed|Thesis/dissertation

eScholarship.org Powered by the California Digital Library
University of California

UNIVERSITY OF CALIFORNIA
SANTA CRUZ

MOLECULAR EVOLUTION IN EMBIOTOCID SURFPERCHES

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

DOCTOR OF PHILOSOPHY

in

ECOLOGY AND EVOLUTIONARY BIOLOGY

by

Gary Charles Longo

December 2016

The Dissertation of Gary Charles Longo
is approved:

____
Professor Giacomo Bernardi, Chair

____
Associate Professor, Richard E. Green

____
Assistant Professor, Michael R. Miller

____
Professor, Peter Raimondi

____
Tyrus Miller
Vice Provost and Dean of Graduate Studies

Copyright © by

Gary Charles Longo

2016

Table of Contents

Lists of Figures and Tables **iv**

Abstract **vii**

Acknowledgments **ix**

Introduction **1**

Bibliography 3

Chapter 1: The evolutionary history of the embiotocid surfperch radiation based on genome-wide RAD sequence data **4**

Bibliography 11

Supporting Information 13

Chapter 2: The complete mitochondrial genome of the black surfperch, *Embiotoca jacksoni*: Selection and substitution rates among surfperches (Embiotocidae) **21**

Bibliography 25

Chapter 3: Genomic signatures of parallel selection in surfperches (Embiotocidae) **27**

Bibliography 48

Figures 54

Tables 59

Lists of Figures and Table

Figure 1.1 **5**
Embiotocid species richness map in the Eastern Pacific.

Figure 1.2 **8**
Phylogeny of Embiotocidae inferred using genome wide RAD markers with maximum likelihood and Bayesian methods.

Figure 1.3 **9**
Time calibrated phylogeny of Embiotocidae.

Figure 1.S1 **13**
RAD sequence read number per individual after passing quality filtering.

Figure 1.S2 **15**
Embiotocidae phylogenetic trees inferred with Bayesian methods from various sets of *populations* (Stacks v 1.13) filter parameters.

Figure 1.S3 **16**
Phylogeny of Embiotocidae inferred using genome wide RAD markers with maximum likelihood and Bayesian methods. Individual consensus sequences were used for phylogenetic inference here.

Table 1.1 **6**
Species, sample names, and locations for the 44 surfperch individuals (22 species) used in this study.

Table 1.S1 **18**
Unique stacks and polymorphic loci detected for each individual.

Table 1.S2 **19**
Number of loci, phylogenetically informative SNPs, and overall size of each concatenated supermatrix.

Table 1.S3 **20**
Sequence divergence between sister species of surfperches.

Figure 2.1 **22**
Organizational map of the complete mitochondrial genome of the black surfperch, *Embiotoca jacksoni*.

Figure 2.2 **24**
Pairwise divergence rates across mitochondrial protein-coding genes in *Cymatogaster aggregata*, *Ditrema temminckii*, and *Embiotoca jacksoni*.

Figure 2.3 **25**
Branch specific dN/dS ratios across mitochondrial protein-coding genes in *Cymatogaster aggregata*, *Ditrema temminckii*, and *Embiotoca jacksoni*.

Table 2.1 **23**
Organization of the mitochondrial genome of *Embiotoca jacksoni*.

Figure 3.1 **54**
Average SST for December 2013 along the California and northern Baja coastline and sampling sites.

Figure 3.2 **55**
STRUCTURE plots for pairwise population comparisons in *Embiotoca jacksoni*, *E. lateralis*, *Rhacochilus toxotes*, and *R. vacca*.

Figure 3.3 **56**
Balancing, divergent, and parallel selection observations in *Embiotoca jacksoni*, *E. lateralis*, *Rhacochilus* toxotes, and *R. vacca*.

Figure 3.S1 **57**
Average SST by month for 2013 along the California and northern Baja coastline with sampling sites in Monterey Bay, California, USA and Punta Banda, Baja California, Mexico. *See supplemental material*.

Figure 3. S2 **58**
Results from FDIST scans for selection using the complete RADseq data set for each species; *Embiotoca jacksoni*, *E. lateralis*, *Rhacochilus toxotes*, and *R. vacca*.

Table 3.1 **59**
RADseq loci counts for total loci, polymorphic loci, polymorphic loci identified as undergoing divergent selection and balancing selection for *Embiotoca jacksoni*, *E. lateralis*, *Rhacochilus toxotes*, and *R. vacca*.

Table 3.2 **60**
Pairwise F_{ST} comparisons between Monterey Bay and Punta Banda for *Embiotoca jacksoni, E. lateralis, Rhacochilus toxotes,* and *R. vacca*.

Table 3.S1 **61**
Sample name, species, location, number of quality filtered reads, number of unique stacks, and number of polymorphic loci for each sample used in the final analyses. *See supplemental material.*

Table 3.S2 **62**
Stacks sequence ID and 80 bp nucleotide sequence for all RADseq loci identified as under balancing or divergent selection, as well as the NCBI alignment score, NCBI protein reference number, and NCBI gene description for each sequence for *Embiotoca jacksoni*, *E. lateralis*, *Rhacochilus toxotes*, and *R. vacca*. *See supplemental material.*

Abstract

Molecular Evolution in Embiotocid Surfperches

Gary Charles Longo

Here, in embiotocid surfperches I have investigated diverse evolutionary processes on different genomic scales using advances in massively parallel DNA sequencing technologies. First I reconstructed phylogenetic relationships among all genera and 21 out of 23 embiotocid species using restriction-site associated DNA sequence (RADseq) markers. In Embiotocidae I found that RADseq supermatrices that retained 91% of orthologous markers across sampled species, which corresponded to 523 loci, yielded trees with the highest support values. The resulting phylogenetic hypotheses support a scenario where embiotocids first diverged into clades associated with sandy and reef habitats during the middle Miocene (13–18 million years ago) with subsequent invasions of novel habitats in the reef associated clade, and northern range expansion in the Northwest Pacific. In all cases, radiations occurred within specific habitats, a pattern consistent with niche partitioning. For my second chapter I characterized the complete mitochondrial genome of the black surfperch, *Embiotoca jacksoni*. The genome contains 13 protein-coding genes, two ribosomal RNA genes, 22 transfer RNA genes, and the non-coding control region (Dloop), the gene order of which is identical to that observed in most vertebrates. I then compared the protein-coding mitochondrial DNA gene sequences of *E. jacksoni* with two other embiotocid surfperches with available complete mitochondrial

genomes, *Cymatogaster aggregata* and *Ditrema temminckii*. Across all mitochondrial protein-coding genes in surfperches the weighted average substitution rate was 2.079% per million years and average dN/dS ratios for each protein-coding gene ranged from 0.016 in CytB to 0.608 in ND3. Substitution rates and dN/dS ratios were relatively high for ATP8 compared to other protein-coding genes. Although most protein-coding genes showed signals of purifying selection, I found evidence for positive selection in ND3 in *E. jacksoni*. Finally for chapter three I performed exploratory genome scans for selection in four species of Embiotocid surfperches between populations in Monterey Bay, California, USA and Punta Banda, Baja California, Mexico again using RADseq markers. These localities are on opposite sides of a well-known biogeographic break and experience significantly different sea surface temperatures (SST) as well heterogeneous species assemblages. Using a F_{ST} outlier approach, I detected strong signals of intraspecific balancing and divergent selection as well as clear evidence for interspecific parallel selection with some loci aligning to known proteins. These results suggest that California surfperches represent an ideal system for investigating different forms of selection in a natural marine environment.

Acknowledgements

There are many who contributed to this dissertation in one way or another and here I will only name a few. First and foremost I thank all my committee members who made this work possible. Most importantly I thank Giacomo Bernardi for giving me the opportunity to do science in his lab as an undergraduate and then as a graduate student. Giacomo has been an amazing mentor as well as a great friend. I look forward to working together as colleagues in the future. Pete Raimondi has been a shining example of what it means to be an amazing scientist and an amazing person. Spending time in the field with Giacomo and Pete teaching the marine ecology field quarter class has been one of the highlights of my graduate school career. You both are inspirations to many others besides me.

Mike Miller taught me almost everything I know about RADseq and next generation sequencing, thank you for your patience and generosity. Without that invitation to come participate in a library preparation I think my dissertation would have been much different. Sometimes playing a pickup game of basketball can lead to Ed Green being on your dissertation committee. That happened to me so keep that in mind next time you turn down an opportunity to shoot hoops. I thank Ed for his generosity and helping get my feet wet with genome sequencing. I would also like to specifically thank Stephen Hauskins for making my life easier with his unbelievable IT support over the years.

I would also like to acknowledge my funding sources and fellowships, which made this research possible. The American Museum of Natural History Lerner-Gray

Fund for Marine Research, the Earl and Ethel Myers Oceanographic and Marine Biology Trust, the Friends of the Long Marine Lab, and the UCSC Ecology and Evolutionary Biology Department have all contributed to funding this research. The Graduate Assistance in Areas of National Need and the UCSC Doctoral Student Sabbatical Fellowships both provided much appreciated quarter long breaks from serving as a teaching assistant.

I am extremely fortunate to come from a big, loving family that has provided me with loads of opportunities and support. My dad has always encouraged me to shoot for the stars and take pride in whatever I do while my mom has always been my biggest cheerleader from the sidelines of sporting events to the “homework” of graduate school. Thank you both for your unconditional love and support throughout my life. Vanessa, my sister, has always kept me in check and encouraged me to view the world in a different light. Your perspective has made me a better person and I thank you for it. My brother, Vince, is my best friend, my partner in crime, and the Luigi to my Mario. You are one of the hardest working people I know and your example has inspired me to push through many obstacles. I love you all very much.

I have many amazing friends who have supported me throughout my life but a special shout out to Emmett Ziegler who helped keep me motivated during the inevitable lows associated with portions of graduate school. I will always remember the amazing years we spent together in the Shanty while we didn't have jobs and our pet's heads were falling off.

I especially thank my amazing child bride, Brooke. You have been there with me through the highs and lows and on many amazing adventures. Your unwavering love, support, and optimism have undoubtedly helped me get through graduate school. I will return the favor as best I can as you continue your own adventure on the road to becoming a medical doctor. Home is whenever I'm with you. I love you.

Finally, I thank my grandparents. Although they are no longer here they live on in my family and me. Bob, Bee, Joe, and Marietta, I am but only a quarter of each of your genome's, thank you for everything.

Statement of Contributions

The text of this dissertation includes reprints of the following previously published material: Chapter One: Longo, G. & Bernardi, G. (2015) The evolutionary history of the embiotocid surfperch radiation based on genome-wide RAD sequence data. *Molecular Phylogenetics and Evolution* 88, 55–63.

Chapter Two: Longo, G.C., O'Connell, B., Green, R.E. & Bernardi, G. (2016) The complete mitochondrial genome of the black surfperch, Embiotoca jacksoni: Selection and substitution rates among surfperches (Embiotocidae). *Marine Genomics* 28, 107-112.

For Chapter One, I was responsible for some of the sample collections, all the lab work, all the data analyses, most of the manuscript preparation, and submission. Giacomo Bernardi contributed many of the samples and helped with the manuscript preparation.

For Chapter Two, I was responsible for the sample collection, most the data analyses, most the manuscript preparation, and submission. Ed Green and Brendan O'Connell performed the lab work and assembled the mitochondrial genome. Giacomo Bernardi assisted in manuscript preparation.

Introduction

The evolutionary history of species, populations, and individuals are recorded in the nucleotides of their genomes. Written in these genomes are evolutionary success stories that have navigated the gauntlet of survival and reproduction over the last 4 billion years. If diversity is an indicator of success then it is clear that certain lineages have faired better than others. Certainly stochastic processes have played a role in the diversification of some lineages, such as in the proliferation of mammals following the Cretaceous-Paleogene extinction event. However some especially successful lineages undoubtedly owe a portion of their diversity to revolutionary changes in their genomes. For instance the most speciose order of extant bony fishes, Cypriniformes, contains over 4,300 valid species while the order Amiformes persists with only one surviving member, *Amia calva* (Eschmeyer & Fong 2016). A portion of the Cypriniformes diversity is undoubtedly due to ancestral genomic innovations that led to hearing improvements via the Weberian apparatus (Bird & Hernandez 2007), although it is only recently that we have been able to investigate the genomic architecture responsible for these changes.

Advances in sequencing technology and chemistry have opened up doors and ushered us into the golden age of genetics where it is now economically feasible to generate and sequence thousands of genome wide markers in non-model organisms and even sequence full genomes. Computing power and analytical approaches have progressed as well allowing us to query genomes in novel ways. Indeed it is now possible to reconstruct an entire population's demographic history from the relative

diversity contained within an individual's genome (Li & Durbin 2011). Here in my dissertation we utilized genomic techniques to better understand evolutionary processes in embiotocid surfperches. Specifically for chapter one, we built the most robust and well supported phylogenetic hypothesis to date for Embiotocidae using genome wide restriction site associated DNA (RADseq) markers. For chapter two, we sequenced the full mitochondrial genome of the black surfperch, *Embiotoca jacksoni*, and compared substitution rates as well as the ratio of nonsynonymous to synonymous substitutions (dN/dS) among protein coding genes with the mitochondrial genomes of two other surfperches, *Cymatogaster aggregata* and *Ditrema temminckii*. Finally we performed genomic scans for signals of selection, again using RADseq markers, at the population level in four species of surfperches. From these genomic analyses in embiotocid surfperches we have learned a tremendous amount about the family's interspecific evolutionary relationships, uncovered and analyzed the full mitochondrial genome of the black surfperch, and gained insight into how selection and geneflow vary among different surfperch species.

Bibliography

Bird, N.C. & Hernandez, L.P. (2007) Morphological Variation in the Weberian Apparatus of Cypriniformes. 757, 739–757.

Eschmeyer, W.N. & Fong, J.D. (2016) SPECIES BY FAMILY/SUBFAMILY: CATALOG OF FISHES. Available from: http://researcharchive.calacademy.org/research/ichthyology/catalog/speciesbyfamily.asp (November 29, 2016).

Li, H. & Durbin, R. (2011) Inference of human population history from individual whole-genome sequences. *Nature* 475, 493–6.

Chapter 1

Contents lists available at ScienceDirect

Molecular Phylogenetics and Evolution

journal homepage: www.elsevier.com/locate/ympev

The evolutionary history of the embiotocid surfperch radiation based on genome-wide RAD sequence data

Gary Longo *, Giacomo Bernardi

Department of Ecology and Evolutionary Biology, University of California, Santa Cruz, CA, USA

ARTICLE INFO

Article history:
Received 19 December 2014
Revised 29 March 2015
Accepted 31 March 2015
Available online 7 April 2015

Keywords:
Evolutionary radiation
Embiotocidae
Surfperches
RAD-seq
Phylogenetics

ABSTRACT

The radiation of surfperches (Embiotocidae) in the temperate North Pacific has been suggested to be the product of ecological competition and niche partitioning. Surfperches are a family of viviparous marine fishes, which have been used to study multiple paternity, sperm competition, and population genetics. Phylogenetic inference is essential for interpreting the evolutionary context of embiotocid life history traits and testing alternative scenarios, yet previous studies have yielded phylogenies with low support and incongruent topologies. Here we constructed reduced representation genomic libraries using restriction-site associated DNA (RAD) sequence markers to infer phylogenetic relationships among all genera and 22 out of 24 embiotocid species. Orthologous markers retained across 91% of sampled species, corresponding to 523 loci, yielded trees with the highest support values. Our results support a scenario where embiotocids first diverged into clades associated with sandy and reef habitats during the middle Miocene (13–18 Mya) with subsequent invasions of novel habitats in the reef associated clade, and northern range expansion in the Northwest Pacific. The appearance of California kelp forests (Laminariales) in the late Miocene (8–15 Mya) correlates with further proliferation in the reef associated clade. In all cases, radiations occurred within specific habitats, a pattern consistent with niche partitioning. We infer fine scale species relationships among surfperches with confidence and corroborate the utility of RAD data for phylogenetic inference in young lineages.

© 2015 Elsevier Inc. All rights reserved.

1. Introduction

Adaptive radiations result from divergence of an ancestral population into an array of species that inhabit a variety of environments and that differ in traits used to exploit those environments (Schluter, 2000), such as the Galapagos finches (Grant, 1999; Lack, 1947), the Hawaiian silverswords (Baldwin and Sanderson, 1998), the Caribbean anoles (Losos, 2000), and the great African rift lake cichlids (Fryer and Iles, 1972). Adaptive radiations need to fulfill three requirements: multiplication of species and common descent, adaptation, and extraordinary diversification (Glor, 2010). There are relatively few described cases in marine fishes, such as the New Zealand triplefins (Hickey et al., 2009), California *Sebastes* rockfish (Johns and Avise, 1998), Antarctic notothenioids (Janko et al., 2011), Caribbean hamlets (Puebla et al., 2008) and South African clinids (von der Heyden et al., 2011). The paucity of adaptive radiations in marine fishes may be due to a number of factors including life history characteristics that are conducive to panmixia (Bernardi, 2013). Indeed most marine fishes have a bipartite life history where adults exhibit a mostly sedentary life while larvae remain pelagic for days to months (Leis, 1991). Protracted pelagic larval stages often result in high dispersal potential accompanied by high levels of gene flow (Reece et al., 2010; Selkoe and Toonen, 2011) potentially hindering adaptation to particular environments and therefore limiting local adaptation. Therefore, marine fishes that lack a pelagic larval stage may provide unique insights into such studies, such as the surfperches (Embiotocidae).

The family Embiotocidae comprises 24 species that are found in the temperate coastal waters of the North Pacific from Mexico to Japan but are absent from the higher latitudes along the Aleutian Islands. The center of distribution and the only known embiotocid fossils are located in central California, it has therefore been assumed that this is also their center of origin (David, 1943; Tarp, 1952) (Fig. 1). Reproductive courtship occurs in the winter and females retain sperm from multiple matings for up to several months before fertilization, and later give live birth from spring through late summer depending on the taxa (Bennett and

* Corresponding author at: Department of Ecology and Evolutionary Biology, University of California Santa Cruz, 100 Shaffer Road, Santa Cruz, 95060 CA, USA.
E-mail address: garycharleslongo@gmail.com (G. Longo).

http://dx.doi.org/10.1016/j.ympev.2015.03.027
1055-7903/© 2015 Elsevier Inc. All rights reserved.

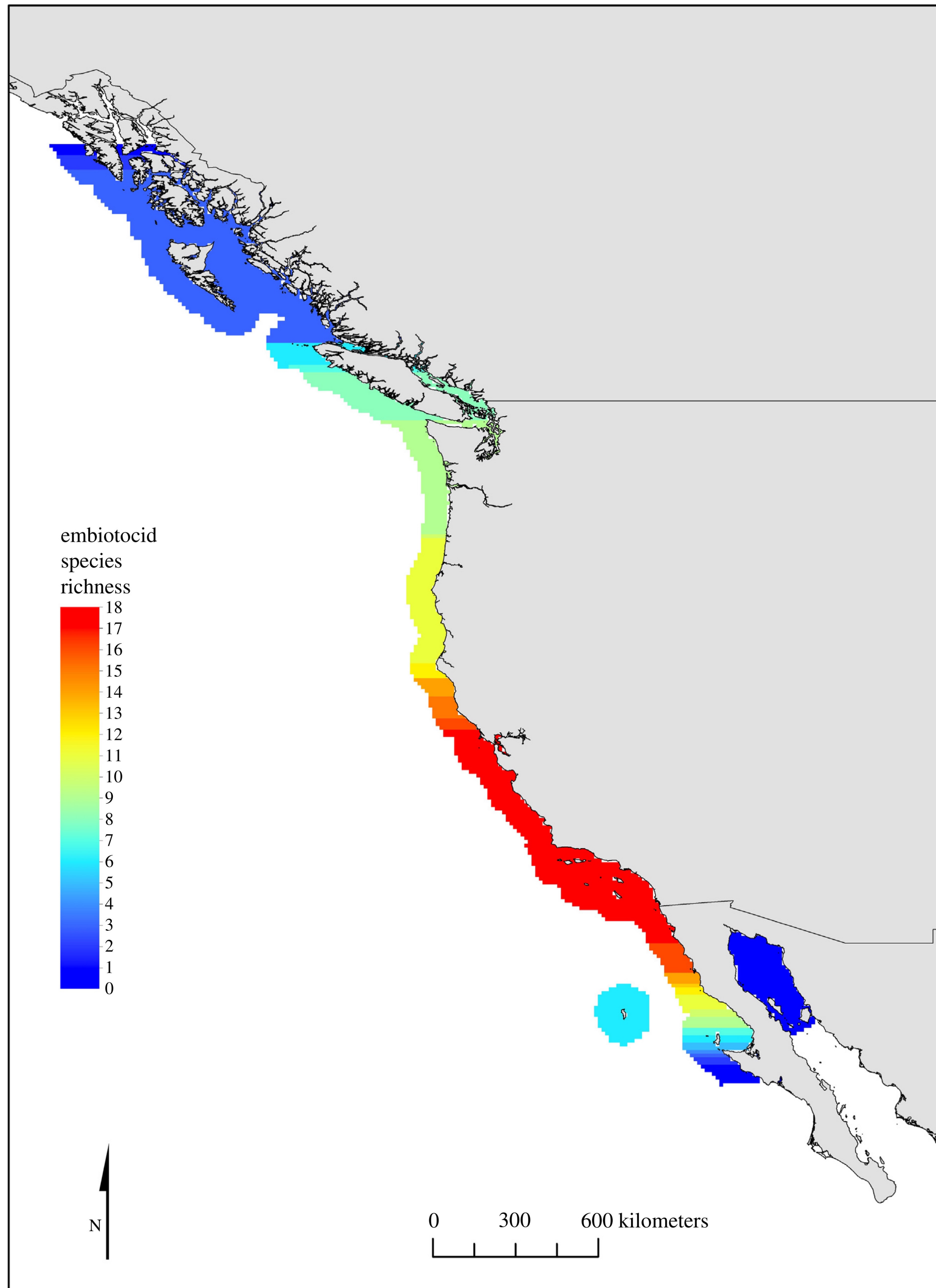


Fig. 1. Embiotocid species richness map in the eastern Pacific based off distributions from Love (2011).

Wydoski, 1977; LaBrecque et al., 2014; Liu and Avise, 2011; Reisser et al., 2009). Therefore, unlike most marine fishes, surfperches do not undergo a dispersive pelagic larval stage. As expected, this alternative life history strategy results in restricted gene flow and may increase the potential for local adaptation (Bernardi, 2005, 2000).

Surfperches along with damselfishes (Pomacentridae), cichlids (Cichlidae), and wrasses (Labridae) formed the traditionally recognized Labroids (Lauder and Liem, 1983). However, subsequent phylogenetic work rendered this group paraphyletic, separating wrasses from the closely related damselfishes, cichlids, and surfperches (Mabuchi et al., 2007) and grouping the latter taxa with a few other families in a larger cluster called Ovalentaria (Betancur-R et al., 2013; Wainwright et al., 2012). Compared to damselfishes (390+) and cichlids (1670+), embiotocids include very few species (24), yet surfperches have invaded very diverse habitats including deep zones, coastal pelagic zones, kelp forest rocky reefs, sandy bottoms, shallow seagrass beds, estuarine zones and even inland freshwater (Allen and Pondella, 2006; Tarp, 1952). These diverse environments may have served as adaptive zones where ancestral surfperches radiated to take advantage of under exploited niches. Generally, niche partitioning and habitat use can either arise from a single lineage invading an open niche and diversifying *in situ*, or from multiple lineage invasions. These processes are not mutually exclusive and a combination of scenarios can give rise to observed radiations.

In fishes, the first stage of adaptation often includes ecological niche partitioning, as seen in threespine sticklebacks along the benthic–limnetic axis and later in several other systems (Schluter, 2000). Niche partitioning has also been suggested as a potential mechanism for diversification in surfperches (Ebeling and Laur, 1986). Accordingly, surfperches display an array of morphologically divergent mouth shapes, dentitions, and pharyngeal jaws, which allow for diverse diets and feeding mechanisms among species and has been characterized as an adaptive radiation previously (De Martini, 1969; Drucker and Jensen, 1991). Ecological competition between closely related species (Hixon, 1980; Holbrook and Schmitt, 1992) and sexual selection on color patterns (Cummings and Partridge, 2001; Froeschke et al., 2007; Nakazono et al., 1981; Westphal et al., 2011) have also played an important role in diversification of surfperches.

To better understand the underpinnings of this radiation, a robust species level phylogeny is necessary. Currently there are three family wide surfperch phylogenies based on morphological data, mitochondrial sequence, or a combination of both. Tarp proposed a species level morphological phylogeny in his thorough revision of surfperches, where the family was divided into two sub-families, the Amphistichinae and the Embiotocinae (Tarp, 1952). This division was later corroborated using osteological characters (Morris, 1981). More recently, a molecular phylogeny based on two mitochondrial markers was established for one representative of each embiotocid genus (Bernardi and Bucciarelli, 1999). Additional phylogenetic hypotheses based on morphological characters and sequence data were proposed (Cassano, 1998) as were relationships restricted to a single genus, *Embiotoca* and *Ditrema,* or a single subfamily, Amphistichinae (Bernardi, 2009; Katafuchi et al., 2010; Westphal et al., 2011). Although some topological consensus exists among these studies, definitive relationships among embiotocid taxa remain contentious. Here, we use hundreds of genome wide RAD markers to infer a robust species tree that sheds light on surfperch evolution that is consistent with patterns of niche partitioning and adaptive radiation, and provides a time-frame for those events.

2. Materials and methods

2.1. Sampling

The proposed phylogeny is based on full representation of species for the Amphistichinae and 16 of 18 species from the Embiotocinae. The amphistichine surfperches consist of six species divided into two genera: *Amphistichus argenteus*, *A. koelzi*, and *A. rhodoterus* and *Hyperprosopon anale*, *H. argenteum*, and *H. ellipticum.* Embiotocines are divided into 11 genera, with 18 species: *Brachyistius aletes* and *B. frenatus*; *Cymatogaster aggregata*; *Ditrema jordani*, *D. temminckii*, and *D. viride*; *Embiotoca jacksoni* and *E. lateralis*; *Hypsurus caryi*; *Hysterocarpus traskii*; *Micrometrus aurora* and *M. minimus*; *Neoditrema ransonnetii*; *Phanerodon atripes* and *P. furcatus*; *Rhacochilus toxotes* and *R. vacca*; and *Zalembius rosaceus.* Initially the western Pacific genus *Ditrema* was considered monotypic (*Ditrema temminckii*) but it was later split into three species: *Ditrema temminckii*, *D. jordani*, and *D. viride*, which have been shown to be very closely related taxa (Katafuchi et al., 2010). We included *D. temminckii* as a representative of this clade resulting in 22 out of 24 species of surfperches for phylogenetic inference. Two individuals of each species were sequenced (Table 1) for a total of 44 individuals in the complete data set.

2.2. Molecular methods

Tissue samples were stored in 95% ethanol and DNA was extracted from fin clips or liver tissue using DNeasy Blood & Tissue kits (Qiagen) according to manufacturer's protocol. We constructed RAD libraries using a variation of the original protocol

Table 1
Species, sample names, and locations for the 44 surfperch individuals (22 species) used in this study.

Species	Sample names	Location
Amphistichus argenteus	AAR_SCP & AAR_SLV	Monterey Bay, CA
Amphistichus koelzi	AKO_MBA1 & AKO_MBA2	Monterey Bay, CA
Amphistichus rhodoterus	ARH_CAP & ARH_NSS	Monterey Bay, CA & Nehalem sand spit, OR
Brachyistius aletes	BAL_GUA1 & BAL_GUA2	Isla Guadalupe, Mexico
Brachyistius frenatus	BFR_CAT7 & BFR_CAT8	Catalina Island, CA
Cymatogaster aggregata	CAG_CB1 & CAG_SD1	Monterey Bay, CA & San Diego, CA
Ditrema temminckii	DTE_J193 & DTE_J194	Japan
Embiotoca jacksoni	EJA_MBA & EJA_PBA	Monterey Bay, CA & Punta Banda, Baja CA
Embiotoca lateralis	ELA_MBA & ELA_PBA	Monterey Bay, CA & Punta Banda, Baja CA
Hyperprosopon anale	HAN_MIB1 & HAN_MIB2	Monterey Bay, CA
Hyperprosopon argenteum	HAR_MBA & HAR_SCH1	Monterey Bay, CA
Hyperprosopon ellipticum	HEL_MBA & HEL_PAC	Monterey Bay, CA & Pacifica, CA
Hypsurus caryi	HCA_SCH1 & HCA_SCH2	Monterey Bay, CA
Hysterocarpus traskii	HTR_01 & HTR_02	Monterey Bay, CA
Micrometrus aurora	MAU_MBA1 & MAU_MBA2	Monterey Bay, CA
Micrometrus minimus	MMI_MBA & MMI_MIB	Monterey Bay, CA & San Diego, CA
Neoditrema ransonnetii	NRA_J1 & NRA_J2	Japan
Phanerodon atripes	PAT_MON & PAT_NR1	Monterey Bay, CA & Santa Barbara, CA
Phanerodon furcatus	PFU_SCH1 & PFU_SCH2	Monterey Bay, CA
Rhacochilus toxotes	RTO_MBA & RTO_PBA	Monterey Bay, CA & Punta Banda, Baja CA
Rhacochilus vacca	RVA_MBA & RVA_PBA	Monterey Bay, CA & Punta Banda, Baja CA
Zalembius rosaceus	ZRO_MB12 & ZRO_MB13	Monterey Bay, CA

(Baird et al., 2008; Miller et al., 2007) with restriction enzyme SbfI as described in Miller et al., 2012 with minor revisions reported below. Initial genomic DNA concentrations for each individual were 400 ng. We physically sheared libraries on a Covaris S2 sonicator with an intensity of 5, duty cycle of 10%, cycles/burst of 200, and a cycle time of 30 s. The final PCR amplification step was carried out in 50 μl reaction volumes with 18 amplification cycles. For all size selection and purification steps we used Ampure XP beads (Agencourt). Samples used in this study were sequenced in one of two libraries, each containing 96 individually barcoded samples. Each library was sequenced in a single lane on an Illumina HiSeq 2000 at the Vincent J. Coates Genomics Sequencing Laboratory at UC Berkeley.

2.3. Quality filtering and marker discovery

Raw reads were trimmed to 92 bp on the 3′ end, quality filtered, and then split according to the 6 bp unique barcode using Miller et al. (2012) custom Perl scripts. Sequences were dropped if the product of quality scores for their 92 bases was below 80%. The barcode (6 bp) and restriction site residue (6 bp) were then removed from the 5′ end, resulting in a final sequence length of 80 bp.

We used the software program Stacks version 1.13 (Catchen et al., 2013, 2011) to identify orthologous sequences among surfperch taxa. We first ran the program *denovo_map.pl*, which runs all three Stacks components in a pipeline (i.e., ustacks, cstacks, and sstacks). We set a minimum stack depth (−m) of three, a maximum of three mismatches per loci for each individual (−M), and allowed up to seven mismatches when building catalog loci (−n). We excluded highly repetitive stacks, set the max number of stacks per locus at two, and disabled calling haplotypes from secondary reads. We then ran multiple iterations of the Stacks program *populations* to generate output files for input into downstream phylogenetic programs. This internal program allows for fine control of which markers will be exported by specifying the number of species (−p) and the percentage of individuals in each species (−r) that must possess that marker. Due to generally high coverage across individuals, we increased the minimum stack depth (−m) to ten for all *populations* runs. In order to generate datasets that reflected various levels of orthologous sequence retention, we ran the program with −p set at 14, 16, 18, 20 & 21, which corresponds to about 64%, 73%, 82%, 91%, and 96% of surfperch species retaining the marker, respectively. Additionally we ran −r set to 50% and 100% (i.e., one or both of the individuals in each species possesses the marker) for each of the five −p settings, resulting in 10 total datasets. Full sequence RAD markers of each individual were exported for downstream analyses. The quality filtered sequences are deposited at the National Center for Biotechnology Information short-read archive (accession no. SRP056799).

2.4. Phylogenetic inference

For each Stacks *populations* parameter set, we built supermatrices with complete RAD sequences (80 bp) and identified phylogenetically informative sites using FASconCAT-G (Kück and Longo, 2014). Supermatrices were generated both with individual sequence data and species consensus sequence data with IUPAC ambiguity codes for polymorphic data.

We used both maximum likelihood methods as implemented in PhyML (Guindon et al., 2010) and Bayesian phylogenetic inference as implemented in MrBayes (Ronquist et al., 2012) to assess relatedness within Embiotocidae. For MrBayes analyses we partitioned the dataset by each 80 bp locus, in FASconCAT-G, which allowed each locus to be assigned its own GTR + Γ + I model and parameters. We selected a Markov Chain Monte Carlo (MCMC) search algorithm with a chain length of 1,000,000 using four chains with a sampling frequency of 1000. In PhyML we selected the GTR model of sequence evolution, six substitution rate categories, set the initial tree to random, and performed 100 bootstrap replicates. Phylogenetic trees and corresponding support values were visualized using FigTree v1.4.0 (Rambaut, 2014). Trees were midpoint-rooted due to the constraints of using an outgroup with RAD data, which relies on retaining orthologous restriction sites.

2.5. Estimating divergence times

Divergence times were estimated using standard models of evolution implemented in BEAST (Drummond et al., 2012) assuming mutual independence among sites. Exploratory runs showed a random local clock (RLC) model, in combination with a birth–death (BD) prior for rates of cladogenesis (Drummond and Rambaut, 2007) as appropriate for our data set. Three runs were conducted with 20 million generations each, with sampling every 1000 generations. The software Tracer v1.5 (Rambaut and Drummond, 2007) was used to quantify effective sample sizes (ESS) for model parameters, and the 'compare' command in AWTY (Nylander et al., 2008) was used to assess convergence, with 10% of each run discarded as burn-in. Runs were combined using LogCombiner v1.7.5 (Drummond et al., 2012), and a time tree was obtained using TreeAnnotator v1.7.5 (Drummond et al., 2012).

Internally calibrating divergence in surfperches is complicated because very few fossils are available. Three late Miocene (5.3 Mya) fossils were found by Eric Knight Jordan in Lompoc, California, and described by his father David Starr Jordan, as *Eriquius plectrodes* (one fossil) (Jordan, 1924), and *Erisceles pristinus* (two fossils) (Jordan, 1925). Upon reexamination in 1941, all three fossils were assigned to the single species *Eriquius plectrodes* (David, 1943) and represent the only reliable surfperch fossil remains (David, 1943; Tarp, 1952). David (David, 1943) proposed the fossils most closely resembled the genus *Embiotoca,* however, upon our own examination of all three specimens, we could not substantiate that claim. Although dorsal ray counts are more similar to embiotocin surfperches, other characters (e.g., body depth, caudal peduncle length, and overall shape) approximate to amphistichin surfperches. We therefore declined to use these embiotocid fossil remains for internal calibration due to their uncertain taxonomic assignment. Instead we used external calibration based on published molecular phylogenies that include diverse embiotocid taxa. Specifically, two molecular studies on bony fishes and one on pomacentrids, which all used several molecular markers, allowed us to estimate the crown age of Embiotocidae to approximately 13–18 Mya (Betancur-R et al., 2013; Frédérich et al., 2013; Near et al., 2013). These dates are consistent with a published molecular phylogeny of the family based on mitochondrial (Bernardi and Bucciarelli, 1999). Therefore this prior with a normal distribution was used as a calibration point, with the minimum and maximum bounds implemented with the 95th percentile of the distribution.

3. Results

3.1. RAD sequences

The final filtered library contained 57,124,651 reads among 44 individuals. Coverage ranged from 293,878 to 4,945,582 with an average of 1,298,287 filtered reads per individual (median = 1,042,690) (Fig. S1). A potential source of variance in coverage could be due to variable quality of starting genomic DNA of each individual as well as variability in the amount of data generated from different Illumina runs. The *denovo_map.pl* program detected

between 34,439 and 104,494 unique, genome wide RAD markers (i.e. loci) in each individual, which corresponded to between 568 and 5758 polymorphic loci in each individual (Table S1). Individuals with low coverage typically yielded lower numbers of loci. As we increased the stringency of *populations* filter parameters in Stacks (i.e., higher −p and −r values), the number of markers and overall size of the concatenated supermatrix decreased. Among the different datasets, the number of loci retained ranged from 116 to 30,629 while the number of parsimony informative nucleotide sites pulled from those loci ranged from 304 to 96,368 (Table S2).

3.2. Phylogenetic relationships

Phylogenetic reconstruction using more stringent datasets resulted in nearly identical topologies while low stringency datasets resulted in poorly resolved trees with low support (Fig. S2). The best supported tree (Fig. 2 & Fig. S3) was generated with the *populations* parameters −p 20 & −r 1 dataset (Table S2) (nexus file available at <http://dx.doi.org/10.6084/m9.figshare.1365521>). The inferred topology recovered monophyletic groups for the subfamilies Amphistichinae and Embiotocinae with high support. Within Amphistichinae, the genus *Hyperprosopon* was found to be paraphyletic with *H. anale* branching first from the rest of the amphistichines with high bootstrap support. However, the remaining Amphistichinae included two clades, which comprised on one hand the sister species *Hyperprosopon argenteum* and *H. ellipticum*, and on the other hand the *Amphistichus* clade: *A. rhodoterus, A. argenteus,* and *A. koelzi.* The sub-family *Embiotocinae* consists of two major clades. The first is comprised of four species with small body sizes including, *Micrometrus aurora* and *M. minimus* joined as sister taxa, and another sub-clade with *Cymatogaster aggregata* joined with the freshwater Tule perch, *Hysterocarpus traskii*. The second major clade indicates a divergence between the western Pacific species, *Ditrema temminckii* and *Neoditrema ransonnetii,* and a more diverse clade of eastern Pacific species. The eastern Pacific species separate into two sub-clades. One sub-clade includes the silvery open-water species, loosely associated with kelp (Laminariales) and rocky reefs, in the genera *Phanerodon, Rhacochilus*, and the deep water species *Zalembius rosaceus*, and the other sub-clade includes the species more strongly associated with kelp reefs in the genera *Embiotoca, Hypsurus, and Brachyistius.* Within this clade, the two species of *Brachyistius* are very closely related with a sequence divergence of only 0.025%. This is slightly larger than the divergence between individuals within a species sampled at large distances from each other (Monterey, California, and Punta Banda, Mexico) such as *Embiotoca jacksoni* (0.015%), *E. lateralis* (0.012%), *Rhacochilus toxotes* (0.007%), and *R. vacca* (0.010%). It is, however, a value that is about one order of magnitude smaller than the average sequence divergence observed in other surfperch sister species (0.317%, Table S3).

3.3. Habitat partitioning

We simplified surfperch habitat use into broad categories and color-coded accordingly on the phylogenetic tree described above (Fig. 2). Phylogenetic clusters strongly partitioned by habitat: sandy, seagrass, rocky, and kelp. Interestingly the shiner perch, which has the widest range of salinity tolerance and is known to inhabit estuaries, and sporadically freshwater (Love, 2011), partitioned with the only freshwater species, the Tule perch *Hysterocarpus traskii*.

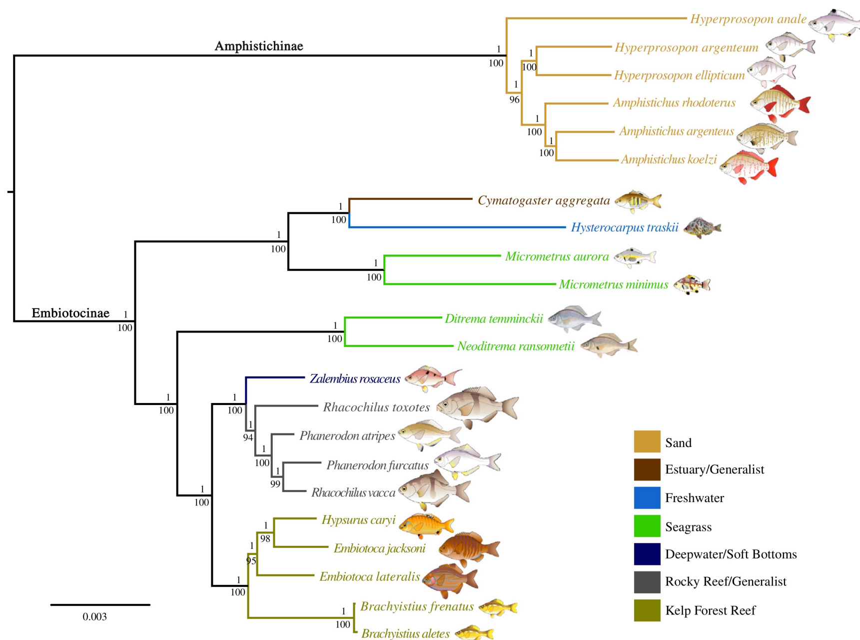


Fig. 2. Phylogeny of Embiotocidae inferred using genome wide RAD markers (p20_r1 supermatrix) with maximum likelihood and Bayesian methods. Species consensus sequences were used for phylogenetic inference here. Node values represent posterior probability and bootstrap support (top and bottom, respectively). Taxa are colored coded based on habitat preference.

3.4. Estimating divergence times

In order to estimate divergence times among embiotocid taxa we used a Bayesian multispecies coalescent approach, as implemented in BEAST (Drummond et al., 2012), based on a 13–18 Mya external calibration for absolute age of the family (Fig. 3). Our data suggest the crown Amphistichinae clade diverged approximately 5 Mya. The earliest Embiotocinae divergence occurred approximately 10 Mya and split the group into the smaller surfperches (i.e., *Micrometrus, Cymatogaster, Hysterocarpus*) on one hand and the relatively larger embiotocines on the other hand. The larger surfperches subsequently diverged approximately 7 Mya to give rise to the western Pacific and the eastern Pacific species. Roughly 5 Mya the eastern Pacific species diverged into two groups, the rocky reef species and the kelp reef species, each of which diversified approximately 2.5–3 Mya.

4. Discussion

4.1. Taxonomic notes

We generated a very robust phylogenetic hypothesis for Embiotocidae based on thousands of genome wide, phylogenetically informative DNA bases drawn from the most complete representation of the family to date. Our results are mostly in agreement with previously published phylogenies (Bernardi and Bucciarelli, 1999; Tarp, 1952; Westphal et al., 2011) but allow for reassessment of taxonomic issues due to better resolution and support.

In the subfamily Amphistichinae, the genus *Hyperprosopon* is likely paraphyletic, with *H. anale* branching first, and sister to the two genera *Amphistichus* and *Hyperprosopon.* One year after its description as *Hyperprosopon anale* (Agassiz, 1861), the species was re-described as *Hypocritichthys anale* (Gill, 1862), an available name that could be used in light of our new findings. In the subfamily Embiotocinae, the sister species *Brachyistius frenatus* and *B. aletes* may either be considered two different valid species or two diverging populations experiencing incipient speciation. The clade that includes *Phanerodon* and *Rhacochilus* reveals that both genera are paraphyletic. A simple way to resolve this issue would be to include all members of this clade in a single genus. Using a single genus for such a diverse group is consistent with other embiotocid examples such as the genus *Micrometrus*, which contains two species that are more divergent than any of the species contained in this clade. The genera *Phanerodon* and *Rhacochilus* were both described in 1854, however *Rhacochilus* was described earlier (May vs. October) and would therefore have precedence (Agassiz, 1854; Girard, 1854). Finally, the rainbow seaperch, *Hypsurus caryi,* originally described as *Embiotoca caryi,* was found to be the sister species of *Embiotoca jacksoni.* Therefore its original name, *Embiotoca caryi,* should be used, as previously suggested (Bernardi, 2009).

4.2. Ecological speciation and niche partitioning

One of the most salient features of the phylogenetic hypothesis presented here is the strong correlation between habitat use and phylogenetic relationships (Fig. 2). As mentioned earlier, diversification in a given habitat may happen via a single invasion followed by a radiation *in situ*, or by repeated invasions of different evolutionary lineages. Our data suggest that the former process repeatedly occurred during surfperch evolution, while there is no evidence of the latter process ever happening. Indeed, ancestral surfperches invaded different habitats and speciation followed

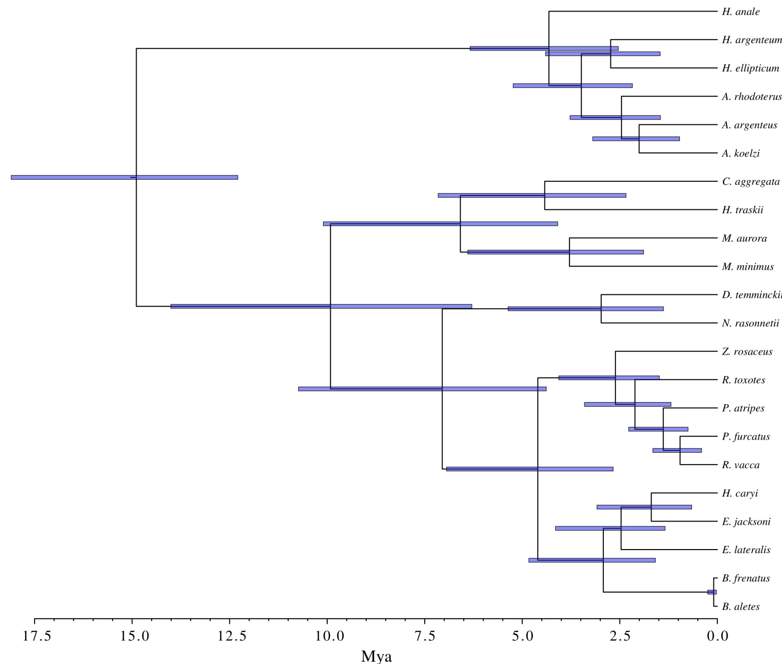


Fig. 3. Time calibrated phylogeny of Embiotocidae using an external time calibration based on two recently published, large-scale molecular phylogenies. Horizontal blue bars at the nodes represent the 95% confidence intervals for each date estimate.

within those habitats, resulting in the patterns observed here where habitat and phylogenetic clusters are perfectly correlated (Fig. 2).

The phylogeny proposed here suggests early divergence based on sandy versus shallow reef habitat followed by further specialization within each lineage. In the Amphistichinae, the spotfin surfperch, *Hyperprosopon anale,* is sister to the remaining lineages, which was previously proposed by (Westphal et al., 2011). Little work has been done on *H. anale,* but their diet consists largely of zooplankton (Love, 2011), while other amphistichines, such as the barred surfperch, *Amphistichus argenteus,* feed primarily on benthic invertebrates (Carlisle et al., 1960). It is therefore plausible that the extant benthic, sand dwelling amphistichin surfperches radiated from a limnetic ancestor that fed on zooplankton. The remaining taxa sort into *Amphistichus* and *Hyperprosopon* clades. Both of these *Hyperprosopon* species are widely distributed from Washington or Oregon, to Baja California, Mexico (one species, *H. argenteum*, is also found on Guadalupe Island, off Baja California). In contrast, the distribution of *Amphistichus* is variable. While *A. koelzi* is also widely distributed from Washington to Baja California, *A. argenteus* and *A. rhodoterus*, the barred and redfin surfperch respectively, are nearly allopatric. The redfin surfperch is distributed from British Columbia to Central California, while the barred surfperch prevails from Central California to northern Baja California, Mexico. Therefore in this group, allopatric speciation may have played a prominent role.

While the ancestors of Amphistichinae radiated over sandy habitats, the Embiotocinae lineage invaded and diversified in the near shore reef environment of the temperate North Pacific. Within this group, the clade containing the genera *Micrometrus*, *Hysterocarpus*, and *Cymatogaster* is sister to the rest of the lineage. *Cymatogaster* is a habitat generalist. However, these small surfperch tend to preferentially be found in shallow waters and often associate with seagrass (Byerly, 2001; De Martini, 1969; Love, 2011). Notably, *C. aggregata* has a high salinity tolerance permitting it to frequent estuaries and even enter freshwater for short periods of time (Love, 2011). *C. aggregata* is sister to the only obligate freshwater embiotocid, the Tule perch, *H. traskii*, indicating that ecological speciation likely drove their most recent common ancestor to split and invade freshwater as noted by Bernardi and Bucciarelli (1999). *Micrometrus spp.* are mostly restricted to shallow waters and are commonly found over seagrass or algae beds (Love, 2011). The next branching event led to the western Pacific surfperches, *Ditrema* and *Neoditrema*, which often associate with nearshore environments with low lying biological structure such as seagrass beds. Based on the common association with seagrass in extant embiotocines just discussed, it is plausible the most recent common ancestor of the Embiotocinae also associated with seagrass.

The next group comprises the most diverse surfperch group, the reef associated species, which is divided into two major ecological clusters. One includes the primarily rocky reef associated species (i.e., *Rhacochilus + Zalembius* clade), and the other includes the kelp associated species (i.e., *Embiotoca + Brachyistius* clade). *Zalembius rosaceus* differs from the rest of the clade as it is most commonly found over soft bottoms in depths greater than 50 m (Love, 2011). *Rhacochilus* and *Phanerodon spp.* are structure-oriented generalists that can be found in and around rocky reefs and kelp forests, as well as manmade structures such as pilings. The species of the last clade, *Embiotoca, Hypsurus,* and *Brachyistius,* are most typically associated with kelp. These species are very specialized and show high levels of ecological competition (Hixon, 1980). A number of feeding specializations evolved within this clade, such as winnowing (sorting food within the mouth). Winnowing is found in *R. toxotes, E. jacksoni*, *H. caryi*, and conspicuously absent in *E. lateralis,* which is consistent with the sister relationship of *E. jacksoni* and *E. caryi*.

4.3. Tempo and mode of evolution

Embiotocidae diverged relatively recently, approximately 13–18 Mya. The family likely originated off the coast of California as the only known fossils are from the Lompoc deposits (southern CA) (David, 1943; Jordan, 1924) and the center of distribution is central California (Tarp, 1952) (Fig. 1). Within the subfamily Embiotocinae, the group of smaller species (i.e., *Cymatogaster*, *Hysterocarpus*, and *Micrometrus*) diverged approximately 10 Mya. Within this group, divergence times between taxa are much greater compared to other surfperch groups. For example the two *Micrometrus* species, reef and dwarf surfperches, diverged between 3.5 to 4 Mya (Fig. 3). The next branching event, which occurred approximately 7 Mya, led to the western Pacific surfperches, *Ditrema* and *Neoditrema*, which may have migrated across the northern Pacific during a warmer climatic period (David, 1943; Tarp, 1952). During that time, the northern Pacific was dominated by seagrass, as evidenced by seagrass feeding marine mammals and invertebrates (Estes and Steinberg, 1988). The ocean cooling in the late Miocene, together with an influx of nutrients, led to a shift from a seagrass dominated system to a kelp dominated system (Bolton, 2010; Brasier, 1975; Estes and Steinberg, 1988). The expansion of kelp resulted in an increased breadth of niches and resources, which likely contributed to a radiation in the later surfperches. Indeed, the greatest diversity of extant embiotocin surfperches can be found in or around kelp forests. Within this new habitat, two major embiotocin groups evolved within the past 5 Mya. On one hand the rocky reef associated species (the *Rhacochilus + Zalembius* clade), and on the other hand the kelp associated species (the *Embiotoca + Brachyistius* clade).

At the same time, approximately 3.5–4 Mya, the Amphistichinae diverged into the two current genera *Hyperprosopon* and *Amphistichus.* Therefore most of the embiotocid diversification occurred between 5 and 3.5 Mya ago. It is interesting to note that the sole fossil representative of the surfperches, *Eriquius*, dated at 5.3 My, is also from that general era, which might suggest that this was a time of great diversification in surfperches.

4.4. Conclusion

Empirical and theoretical support for the applicability of RAD data for phylogenetic inference is growing (Emerson et al., 2010; Hipp et al., 2014; Rubin et al., 2012; Viricel et al., 2014; Wagner et al., 2013). Here we corroborate the applicability of RAD data and infer the most complete and fine scale Embiotocidae phylogeny with high support values. Embiotocids likely radiated ~13–18 Mya (Frédérich et al., 2013; Wainwright et al., 2012), making this one of the older lineages RAD data has been applied to empirically.

Surfperches diverged relatively recently and comprise comparatively few species (24). Although older, the other two major Ovalentaria families *sensu* Betancur-R et al. (2013), Pomacentridae and Cichlidae, comprise one to two orders of magnitude more species. Compared to oviparous groups, it is likely that the life history strategy of the livebearing surfperches does slow down speciation rates, yet, even within Embiotocidae, we can see some remarkable ecological diversity that in many respects encompasses what is observed in more diverse families.

The adaptive radiations of cichlids in the great African lakes show remarkable examples of convergent evolution, where mouth, teeth, and pharyngeal jaw shapes are strikingly similar among similar ecotypes (Kocher et al., 1993). Such convergence is not

completely surprising and recent theoretical models have shown that relatively high levels of convergence should be expected in adaptive radiations (Muschick et al., 2012; Scheffer and van Nes, 2006; terHorst et al., 2010). Morphological convergence of feeding structures between fish families has been shown before as well (Norton and Brainerd, 1993). As in cichlids, surfperches mouth, teeth, and pharyngeal jaw shape have been correlated to their diverse feeding mechanisms and ecological niches (De Martini, 1969). Recently, genetic and genomic approaches have been used to pinpoint the regions responsible for the evolution of structures involved in feeding specializations in cichlids (Albertson et al., 2003a,b; Brawand et al., 2014; Muschick et al., 2012) It will be interesting to determine if the same genomic regions are also involved in the surfperch radiation.

It appears that major divergences in surfperches, which strongly correspond to habitat, occurred relatively recently and were followed by subsequent specialization. Dispersal during favorable climatic conditions likely led to the divergence of eastern and western Pacific clades with the remaining clades likely arising through niche specialization. As for many other lineages, the diversity of surfperches has arisen through distinct ecological and evolutionary processes over space and time. Whether or not surfperch meet the three criteria of an adaptive radiation as outlined in Glor (2010) (i.e., multiplication of species and common descent, adaptation, and extraordinary diversification) remains uncertain. Our data and previous phylogenetic analyses support multiplication of species and common descent, while other work has shown potential ecomorphological adaptations (De Martini, 1969), however more definitive work is needed. The 24 species of surfperches may or may not qualify as extraordinary diversification, although it is comparable with the 15 species of Galapagos finches (Grant and Grant, 2008; Grant, 1999). The robust phylogeny presented here provides the necessary evolutionary framework to conduct field and laboratory studies that will be required to address the nature of the adaptations at the ecological and genomic levels.

Acknowledgments

We would like to thank Michael R. Miller for his generosity and insight on RAD-seq techniques. We would like to thank Karen Crow, Michael Westphal, and Steve Morey (team surfperch) for ongoing discussion on the ecology and evolution of Embiotocidae as well as for providing some valuable amphistichin samples. We would like to thank Larry Allen for permission to use his drawings for the illustration of Fig. 2, José Tavera for helping us with BEAST runs, and Max Tarjan for helping with ArcGIS. We would like to thank the California Academy of Sciences and in particular Jean Demouthe, for granting us access to the David Starr Jordan fossil fish collection. We would like to thank Jennifer Liberto for assisting in the exploration stages of analysis pipelines for phylogenetic RAD data. We would also like to thank Stephen Hauskins for his ongoing IT support. This project was funded by the Lerner-Gray fund for Marine research, Meyers Trust, and the UCSC Ecology and Evolutionary biology research and travel fund. This work used the Vincent J. Coates Genomics Sequencing Laboratory at UC Berkeley, supported by NIH S10 Instrumentation Grants S10RR029668 and S10RR027303.

Appendix A. Supplementary material

Supplementary data associated with this article can be found, in the online version, at http://dx.doi.org/10.1016/j.ympev.2015.03.027.

References

Agassiz, L., 1854. Additional notes on the Holconoti. Am. J. Sci. Arts 17, 365–369.
Agassiz, A., 1861. Notes on the described species of Holconoti, found on the western coast of North America. Proc. Bost. Soc. Nat. Hist. 8, 122–134.
Albertson, R.C., Streelman, J.T., Kocher, T.D., 2003a. Directional selection has shaped the oral jaws of Lake Malawi cichlid fishes. Proc. Natl. Acad. Sci. USA 100, 5252–5257. http://dx.doi.org/10.1073/pnas.0930235100.
Albertson, R.C., Streelman, J.T., Kocher, T.D., 2003b. Genetic basis of adaptive shape differences in the cichlid head. J. Hered. 94, 291–301. http://dx.doi.org/10.1093/jhered/esg071.
Allen, L.G., Pondella, D.J., 2006. Surf zone, coastal pelagic zone, and harbors. In: The Ecology of Marine Fishes: California and Adjacent Waters. Berkeley, pp. 149–166.
Baird, N.A., Etter, P.D., Atwood, T.S., Currey, M.C., Shiver, A.L., Lewis, Z.A., Selker, E.U., Cresko, W.A., Johnson, E.A., 2008. Rapid SNP Discovery and Genetic Mapping Using Sequenced RAD Markers. PLoS One 3, e3376. doi: 10.1371/journal.pone.0003376.
Baldwin, B.G., Sanderson, M.J., 1998. Age and rate of diversification of the Hawaiian silversword alliance (Compositae). Proc. Natl. Acad. Sci. USA 95, 9402–9406.
Bennett, D.E., Wydoski, R.S., 1977. Biology of the Redtail Surfperch (*Amphistichus Rhodoterus*) from the Central Oregon coast. U.S. Fish Wildl. Serv.
Bernardi, G., 2000. Barriers to Gene Flow in Embiotoca Jacksoni, a Marine Fish Lacking a Pelagic Larval Stage. Evolution (N. Y). 54, pp. 226–237.
Bernardi, G., 2005. Phylogeography and Demography of Sympatric Sister Surfperch Species, Embiotoca Jacksoni and E. Lateralis Along the California Coast: Historical versus Ecological Factors. Evolution (N. Y). 59, pp. 386–394.
Bernardi, G., 2009. The name of the father: conflict between Louis and Alexander Agassiz and the Embiotoca surfperch radiation. J. Fish Biol. 74, 1049–1055. http://dx.doi.org/10.1111/j.1095-8649.2008.02127.x.
Bernardi, G., 2013. Speciation in fishes. Mol. Ecol. 22, 5487–5502. http://dx.doi.org/10.1111/mec.12494.
Bernardi, G., Bucciarelli, G., 1999. Molecular phylogeny and speciation of the surfperches (Embiotocidae, Perciformes). Mol. Phylogenet. Evol. 13, 77–81.
Betancur-R, R., Broughton, R.E., Wiley, E.O., Carpenter, K., López, J.A., Li, C., Holcroft, N.I., Arcila, D., Sanciangco, M., Ii, J.C.C., Zhang, F., Campbell, M.A., Ballesteros, J.A., Roa-varon, A., Willis, S., Borden, W.C., Hough, D.J., Lu, G., 2013. The Tree of Life and a New Classification of Bony Fishes. PLOS Curr. Tree Life. doi: 10.1371/currents.tol.53ba26640df0ccaee75bb165c8c26288.Abstract.
Bolton, J.J., 2010. The biogeography of kelps (Laminariales, Phaeophyceae): a global analysis with new insights from recent advances in molecular phylogenetics. Helgol. Mar. Res. 64, 263–279. http://dx.doi.org/10.1007/s10152-010-0211-6.
Brasier, M., 1975. An outline history of seagrass communities. Palaeontology 18, 681–702.
Brawand, D., Wagner, C.E., Li, Y.I., Malinsky, M., Keller, I., Fan, S., Simakov, O., Ng, A.Y., Lim, Z.W., Bezault, E., Turner-Maier, J., Johnson, J., Alcazar, R., Noh, H.J., Russell, P., Aken, B., Alföldi, J., Amemiya, C., Azzouzi, N., Baroiller, J.-F., Barloy-Hubler, F., Berlin, A., Bloomquist, R., Carleton, K.L., Conte, M.A., D'Cotta, H., Eshel, O., Gaffney, L., Galibert, F., Gante, H.F., Gnerre, S., Greuter, L., Guyon, R., Haddad, N.S., Haerty, W., Harris, R.M., Hofmann, H.A., Hourlier, T., Hulata, G., Jaffe, D.B., Lara, M., Lee, A.P., MacCallum, I., Mwaiko, S., Nikaido, M., Nishihara, H., Ozouf-Costaz, C., Penman, D.J., Przybylski, D., Rakotomanga, M., Renn, S.C.P., Ribeiro, F.J., Ron, M., Salzburger, W., Sanchez-Pulido, L., Santos, M.E., Searle, S., Sharpe, T., Swofford, R., Tan, F.J., Williams, L., Young, S., Yin, S., Okada, N., Kocher, T.D., Miska, E.A., Lander, E.S., Venkatesh, B., Fernald, R.D., Meyer, A., Ponting, C.P., Streelman, J.T., Lindblad-Toh, K., Seehausen, O., Di Palma, F., 2014. The genomic substrate for adaptive radiation in African cichlid fish. Nature 513, 375–381. http://dx.doi.org/10.1038/nature13726.
Byerly, M.M., 2001. The Ecology of Age-1 Copper Rockfish (Sebastes caurinus) in Vegetated Habitats of Sitka Sound. University of Alaska Fairbanks at Juneau, Alaska.
Carlisle, J.G., Schott, J.W., Abramson, N.J., 1960. Fish Bulletin No. 109: The Barred Surfperch (*Amphistichus Argenteus* Agassiz) in Southern California.
Cassano, V.P.F. (UCLA), 1998. The Phylogeny of the Embiotocidae (Teleostei:Perciformes): Cladistic Analyses of Comparative Morphometry and 16s rDNA Mitochondrial Sequences. University of California Los Angeles.
Catchen, J.M., Amores, A., Hohenlohe, P.A., Cresko, W.A., Postlethwait, J.H., 2011. Stacks: building and genotyping Loci De Novo from short-read sequences. G3 genes. Genomes Genet. 1, 171–182. http://dx.doi.org/10.1534/g3.111.000240.
Catchen, J., Hohenlohe, P.A., Bassham, S., Amores, A., Cresko, W.A., 2013. Stacks: an analysis tool set for population genomics. Mol. Ecol. 22. http://dx.doi.org/10.1111/mec.12354.
Cummings, M.E., Partridge, J.C., 2001. Visual pigments and optical habitats of surfperch (Embiotocidae) in the California kelp forest. J. Comp. Physiol. A Sensory, Neural. Behav. Physiol. 187, 875–889. http://dx.doi.org/10.1007/s00359-001-0258-6.
David, L.R., 1943. Miocene fishes of Southern California. Geol. Soc. Am., 43
De Martini, E.E., 1969. A correlative study of the ecology and comparative feeding mechanism morphology of the Embiotocidae (surf-fishes) as evidence of the Family's adaptive radiation into available ecological niches. Wasmann J. Biol. 27, 177–247.
Drucker, E.G., Jensen, J.S., 1991. Functional analysis of a specialized prey processing behavior: winnowing by surfperches (Teleostei: Embiotocidae). J. Morphol. 210, 267–287. http://dx.doi.org/10.1002/jmor.1052100306.

Drummond, A.J., Rambaut, A., 2007. BEAST: Bayesian evolutionary analysis by sampling trees. BMC Evol. Biol. 7, 214. http://dx.doi.org/10.1186/1471-2148-7-214.

Drummond, A.J., Suchard, M.A., Xie, D., Rambaut, A., 2012. Bayesian phylogenetics with BEAUti and the BEAST 1.7. Mol. Biol. Evol. 29, 1969–1973. http://dx.doi.org/10.1093/molbev/mss075.

Ebeling, A.W., Laur, D.R., 1986. Foraging in surfperches: resource partitioning or individualistic responses? Environ. Biol. Fishes 16, 123–133. http://dx.doi.org/10.1007/BF00005165.

Emerson, K.J., Merz, C.R., Catchen, J.M., Hohenlohe, P.A., Cresko, W.A., Bradshaw, W.E., Holzapfel, C.M., 2010. Resolving postglacial phylogeography using high-throughput sequencing. PNAS 107, 16196–16200. http://dx.doi.org/10.1073/pnas.1006538107/-/DCSupplemental.www.pnas.org/cgi/doi/10.1073/pnas.1006538107.

Estes, J., Steinberg, P., 1988. Predation, herbivory, and kelp evolution. Paleobiology 14, 19–36.

Frédérich, B., Sorenson, L., Santini, F., Slater, G.J., Alfaro, M.E., 2013. Iterative ecological radiation and convergence during the evolutionary history of damselfishes (Pomacentridae). Am. Nat. 181, 94–113. http://dx.doi.org/10.1086/668599.

Froeschke, B., Allen, L.G., Ii, D.J.P., 2007. Life history and courtship behavior of black perch, Embiotoca jacksoni. Pacific Sci. 61, 521–531.

Fryer, G., Iles, T.D., 1972. The Cichlid Fishes of the Great Lakes of Africa. Edinburgh, UK.

Gill, T., 1862. Notice of a collection of the fishes of California presented to the smithsonian institution by Mr. Samuel Hubbard. Proc. Acad. Nat. Sci. Philadelphia 14, 274–282.

Girard, C.F., 1854. Observations upon a collection of fishes made on the Pacific coast of the United States, by Lieut. W. P. Trowbridge, USA, for the museum of the Smithsonian Institution. Proc. Acad. Nat. Sci. Philadelphia 7, 142–156.

Glor, R.E., 2010. Phylogenetic insights on adaptive radiation. Annu. Rev. Ecol. Evol. Syst. 41, 251–270. http://dx.doi.org/10.1146/annurev.ecolsys.39.110707.173447.

Grant, P.R., 1999. Ecology and Evolution of Darwin's Finches. Princeton University Press, Princeton, N.J..

Grant, P.R., Grant, B.R., 2008. How and Why Species Multiply: The Radiation of Darwin's Finches. Princeton University Press, Princeton, N.J..

Guindon, S., Dufayard, J.-F., Lefort, V., Anisimova, M., Hordijk, W., Gascuel, O., 2010. New algorithms and methods to estimate maximum-likelihood phylogenies: assessing the performance of PhyML 3.0. Syst. Biol. 59, 307–321. http://dx.doi.org/10.1093/sysbio/syq010.

Hickey, A.J.R., Lavery, S.D., Hannan, D.A., Baker, C.S., Clements, K.D., 2009. New Zealand triplefin fishes (family Tripterygiidae): contrasting population structure and mtDNA diversity within a marine species flock. Mol. Ecol. 18, 680–696. http://dx.doi.org/10.1111/j.1365-294X.2008.04052.x.

Hipp, A.L., Eaton, D.A.R., Cavender-Bares, J., Fitzek, E., Nipper, R., Manos, P.S., 2014. A framework phylogeny of the american oak clade based on sequenced RAD data. PLoS One 9, e93975. http://dx.doi.org/10.1371/journal.pone.0093975.

Hixon, M.A., 1980. Competitive interactions between california reef fishes of the genus Embiotoca. Ecology 61, 918–931.

Holbrook, S., Schmitt, R., 1992. Causes and consequences of dietary specialization in surfperches: Patch choice and intraspecific competition. Ecology 73, 402–412.

Janko, K., Marshall, C., Musilová, Z., Van Houdt, J., Couloux, A., Cruaud, C., Lecointre, G., 2011. Multilocus analyses of an Antarctic fish species flock (Teleostei, Notothenioidei, Trematominae): phylogenetic approach and test of the early-radiation event. Mol. Phylogenet. Evol. 60, 305–316. http://dx.doi.org/10.1016/j.ympev.2011.03.008.

Johns, G.C., Avise, J.C., 1998. Tests for ancient species flocks based on molecular phylogenetic appraisals of Sebastes rockfishes and other marine fishes. Evolution (N.Y). 52, 1135–1146.

Jordan, D.S., 1924. Miocene Fishes from Southern California. Los Angeles.

Jordan, D.S., 1925. The Fossil Fishes of the Miocene of Southern California. Standford University Press Library.

Katafuchi, H., Kai, Y., Nakabo, T., 2010. Genetic divergence in Ditrema jordani (Perciformes: Embiotocidae) from the Pacific coast of southern Japan, as inferred from mitochondrial DNA sequences. Ichthyol. Res. 58, 90–94. http://dx.doi.org/10.1007/s10228-010-0190-2.

Kocher, T.D., Conroy, J.A., McKaye, K.R., Stauffer, J.R., 1993. Similar morphologies of cichlid fish in Lakes Tanganyika and Malawi Are Due to Convergence. Mol. Phylogenet. Evol. 2, 158–165.

Kück, P., Longo, G.C., 2014. FASconCAT-G: extensive functions for multiple sequence alignment preparations concerning phylogenetic studies. Front. Zool. 11, 81. http://dx.doi.org/10.1186/s12983-014-0081-x.

LaBrecque, J.R., Alva-Campbell, Y.R., Archambeault, S., Crow, K.D., 2014. Multiple paternity is a shared reproductive strategy in the live-bearing surfperches (Embiotocidae) that may be associated with female fitness. Ecol. Evol. 4, 2316–2329. http://dx.doi.org/10.1002/ece3.1071.

Lack, D., 1947. Darwin's Finches. Cambridge University Press, Cambridge.

Lauder, G.V., Liem, K.F., 1983. The evolution and interrelationships of the actinopterygian fishes. Bull. Museum Comp. Zool. 150, 95–197.

Leis, J., 1991. The pelagic stage of reef fishes. In: Sales, P. (Ed.), The Ecology of Fishes on Coral Reefs. Academic Press Inc., San Diego, CA, pp. 182–229.

Liu, J.-X., Avise, J.C., 2011. High degree of multiple paternity in the viviparous Shiner Perch, Cymatogaster aggregata, a fish with long-term female sperm storage. Mar. Biol. 158, 893–901. http://dx.doi.org/10.1007/s00227-010-1616-0.

Losos, J.B., 2000. Ecological character displacement and the study of adaptation. Proc. Natl. Acad. Sci. USA 97, 5693–5695.

Love, M., 2011. Certainly More Than You Want to Know About Fishes of the Pacific Coast: A Post Modern Experience, first ed. Really Big Press, Santa Barbara.

Mabuchi, K., Miya, M., Azuma, Y., Nishida, M., 2007. Independent evolution of the specialized pharyngeal jaw apparatus in cichlid and labrid fishes. BMC Evol. Biol. 7, 10. http://dx.doi.org/10.1186/1471-2148-7-10.

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

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

Morris, S.L., 1981. The Osteology and Relationships of the Embiotocidae (Pisces). Oregon State University.

Muschick, M., Indermaur, A., Salzburger, W., 2012. Convergent evolution within an adaptive radiation of cichlid fishes. Curr. Biol. 22, 2362–2368. http://dx.doi.org/10.1016/j.cub.2012.10.048.

Nakazono, A., Tateda, Y., Tsukahara, H., 1981. Mating Habits of the Surfperch, *Ditrema temminckii*. Japanese J. Ichthyol. 28, 122–128.

Near, T.J., Dornburg, A., Eytan, R.I., Keck, B.P., Smith, W.L., Kuhn, K.L., Moore, J.A., Price, S.A., Burbrink, F.T., Friedman, M., Wainwright, P.C., 2013. Phylogeny and tempo of diversification in the superradiation of spiny-rayed fishes. Proc. Natl. Acad. Sci. USA 110, 12738–12743. http://dx.doi.org/10.5061/dryad.d3mb4.

Norton, S.F., Brainerd, E.L., 1993. Convergence in the feeding mechanics of ecomorphologically similar species in the Centrarchidae and Cichlidae. J. Exp. Biol. 176, 11–29.

Nylander, J.A.A., Wilgenbusch, J.C., Warren, D.L., Swofford, D.L., 2008. AWTY (are we there yet?): a system for graphical exploration of MCMC convergence in Bayesian phylogenetics. Bioinformatics 24, 581–583. http://dx.doi.org/10.1093/bioinformatics/btm388.

Puebla, O., Bermingham, E., Guichard, F., 2008. Population genetic analyses of Hypoplectrus coral reef fishes provide evidence that local processes are operating during the early stages of marine adaptive radiations. Mol. Ecol. 17, 1405–1415. http://dx.doi.org/10.1111/j.1365-294X.2007.03654.x.

Rambaut, A., 2014. FigTree v1.4.0 [WWW Document]. URL <http://tree.bio.ed.ac.uk/software/figtree/>.

Rambaut, A., Drummond, A.J., 2007. Tracer v1.5. [WWW Document]. URL <http://tree.bio.ed.ac.uk/software/tracer/>.

Reece, J.S., Bowen, B.W., Joshi, K., Goz, V., Larson, A., 2010. Phylogeography of two moray eels indicates high dispersal throughout the indo-pacific. J. Hered. 101, 391–402. http://dx.doi.org/10.1093/jhered/esq036.

Reisser, C.M.O., Beldade, R., Bernardi, G., 2009. Multiple paternity and competition in sympatric congeneric reef fishes, Embiotoca jacksoni and E-lateralis. Mol. Ecol. 18, 1504–1510. http://dx.doi.org/10.1111/j.1365-294X.2009.04123.x.

Ronquist, F., Teslenko, M., van der Mark, P., Ayres, D.L., Darling, A., Höhna, S., Larget, B., Liu, L., Suchard, M.a., Huelsenbeck, J.P., 2012. MrBayes 3.2: efficient Bayesian phylogenetic inference and model choice across a large model space. Syst. Biol. 61, 539–542. http://dx.doi.org/10.1093/sysbio/sys029.

Rubin, B.E.R., Ree, R.H., Moreau, C.S., 2012. Inferring phylogenies from RAD sequence data. PLoS One 7, e33394. http://dx.doi.org/10.1371/journal.pone.0033394.

Scheffer, M., van Nes, E.H., 2006. Self-organized similarity, the evolutionary emergence of groups of similar species. Proc. Natl. Acad. Sci. USA 103, 6230–6235. http://dx.doi.org/10.1073/pnas.0508024103.

Schluter, D., 2000. The Ecology of Adaptive Radiation. Oxford University Press, New York.

Selkoe, K., Toonen, R., 2011. Marine connectivity: a new look at pelagic larval duration and genetic metrics of dispersal. Mar. Ecol. Prog. Ser. 436, 291–305. http://dx.doi.org/10.3354/meps09238.

Tarp, F.H., 1952. Fish Bulletin No. 88. A Revision of the Family Embiotocidae (The Surfperches), State of California Department of Fish and Game.

terHorst, C.P., Miller, T.E., Powell, E., 2010. When can competition for resources lead to ecological equivalence? Evol. Ecol. Res. 12, 843–854.

Viricel, A., Pante, E., Dabin, W., Simon-Bouhet, B., 2014. Applicability of RAD-tag genotyping for interfamilial comparisons: empirical data from two cetaceans. Mol. Ecol. Resour. 14, 597–605. http://dx.doi.org/10.1111/1755-0998.12206.

Von der Heyden, S., Bowie, R.C.K., Prochazka, K., Bloomer, P., Crane, N.L., Bernardi, G., 2011. Phylogeographic patterns and cryptic speciation across oceanographic barriers in South African intertidal fishes. J. Evol. Biol. 24, 2505–2519. http://dx.doi.org/10.1111/j.1420-9101.2011.02382.x.

Wagner, C.E., Keller, I., Wittwer, S., Selz, O.M., Mwaiko, S., Greuter, L., Sivasundar, A., Seehausen, O., 2013. Genome-wide RAD sequence data provide unprecedented resolution of species boundaries and relationships in the Lake Victoria cichlid adaptive radiation. Mol. Ecol. 22, 787–798. http://dx.doi.org/10.1111/mec.12023.

Wainwright, P.C., Smith, W.L., Price, S.A., Tang, K.L., Sparks, J.S., Ferry, L.A., Kuhn, K.L., Eytan, R.I., Near, T.J., 2012. The evolution of pharyngognathy: a phylogenetic and functional appraisal of the pharyngeal jaw key innovation in labroid fishes and beyond. Syst. Biol. 61, 1001–1027. http://dx.doi.org/10.1093/sysbio/sys060.

Westphal, M.F., Morey, S.R., Uyeda, J.C., Morgan, T.J., 2011. Molecular phylogeny of the subfamily Amphistichinae (Teleostei: Embiotocidae) reveals parallel divergent evolution of red pigmentation in two rapidly evolving lineages of sand-dwelling surfperch. J. Fish Biol. 79, 313–330. http://dx.doi.org/10.1111/j.1095-8649.2011.03011.x.

Supporting information for (Figures and Tables)
The evolutionary history of the embiotocid surfperch radiation based on genome-wide RAD sequence data

Gary Longo & Giacomo Bernardi

Figure S1. RAD sequence read number per individual after passing quality filtering.

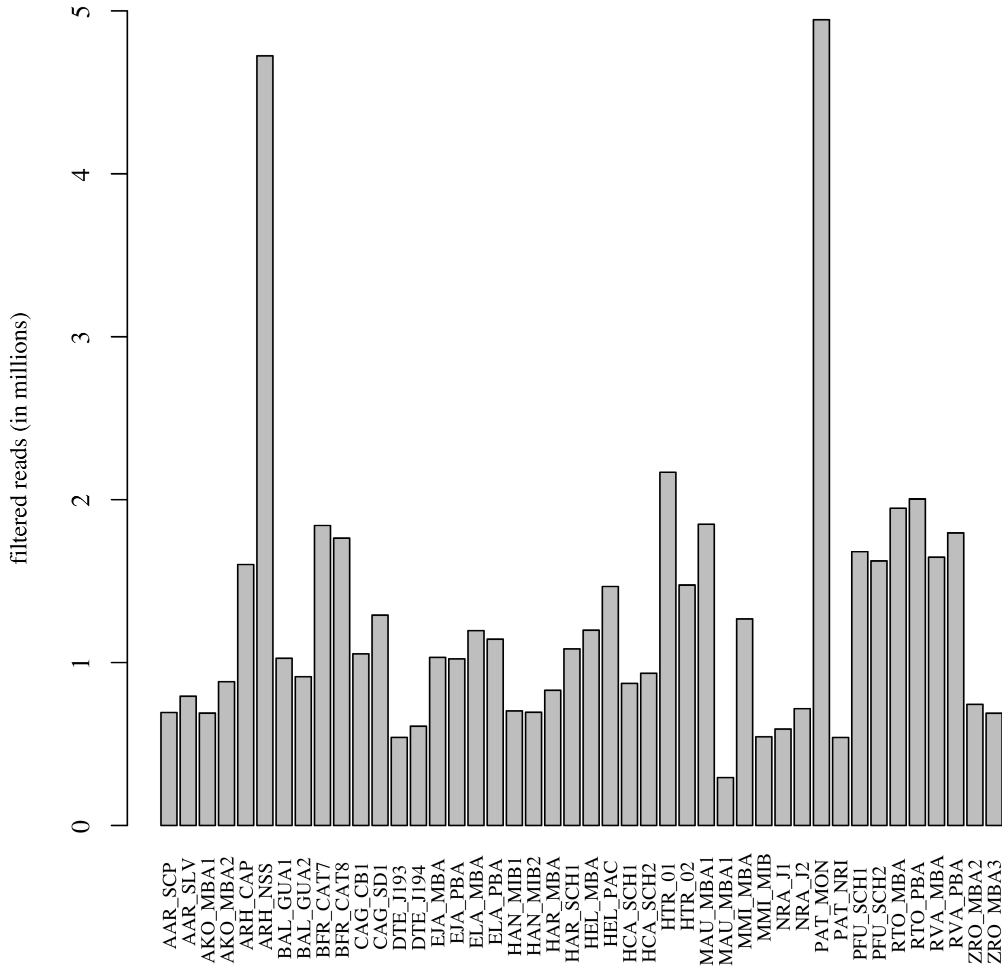
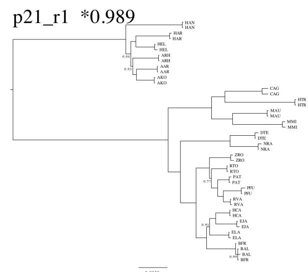
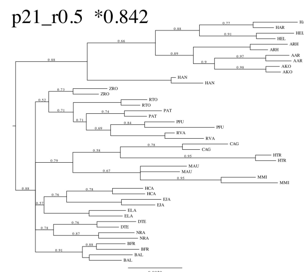


Figure S2. Embiotocidae phylogenetic trees inferred with Bayesian methods from various sets of *populations* (Stacks v 1.13) filter parameters "-p" (the minimum number of species that must possess that marker) and "-r" (the minimum percentage of individuals in each species that must possess that marker). Posterior probability was 1 at all nodes except where indicated. * denotes mean posterior probability values across all bifurcating branches for each tree.

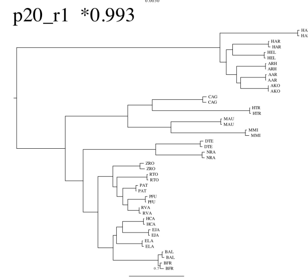
p21_r1 *0.989



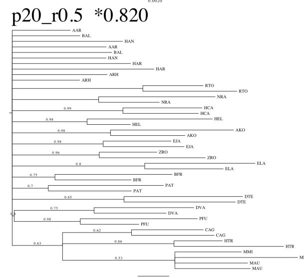
p21_r0.5 *0.842



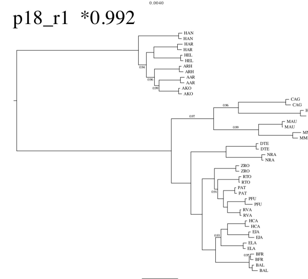
p20_r1 *0.993



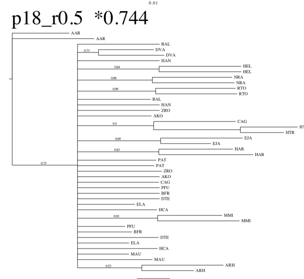
p20_r0.5 *0.820



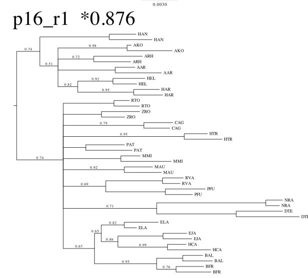
p18_r1 *0.992



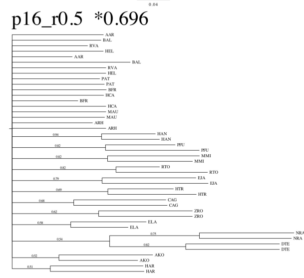
p18_r0.5 *0.744



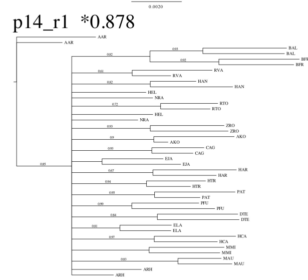
p16_r1 *0.876



p16_r0.5 *0.696



p14_r1 *0.878



p14_r0.5 *0.703

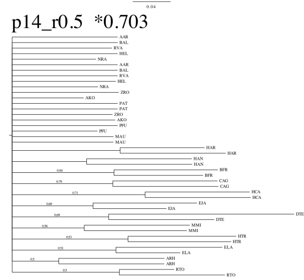


Figure S3. Phylogeny of Embiotocidae inferred using genome wide RAD markers (p20_r1 supermatrix) with maximum likelihood and Bayesian methods. Individual consensus sequences were used for phylogenetic inference here. Node values represent posterior probability (top) and bootstrap support (bottom) and are 1 and 100, respectively, except where indicated.

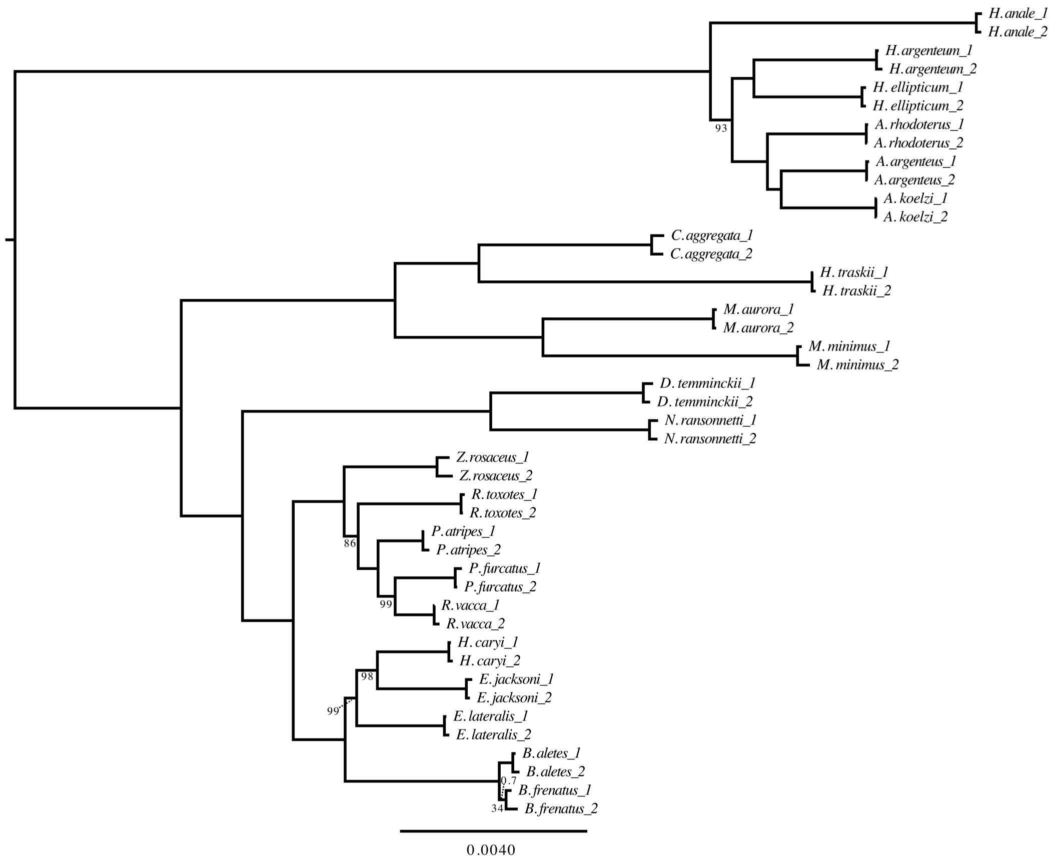


Table S1. Unique stacks (i.e., unique loci) and polymorphic loci detected by the *denovo_map.pl* program (Stacks version 1.13) for each individual.

sample name	unique stacks	polymorphic Loci
AAR_SCP	45168	1161
AAR_SLV	47755	1126
AKO_MBA1	45987	814
AKO_MBA2	47231	924
ARH_CAP	84872	4003
ARH_NSS	55232	2495
BAL_GUA1	63946	2973
BAL_GUA2	52854	2746
BFR_CAT7	53896	4778
BFR_CAT8	53010	4667
CAG_CB1	81237	4625
CAG_SD1	104494	5758
DTE_J193	42088	1818
DTE_J194	44276	1827
EJA_MBA	47319	2061
EJA_PBA	48155	3273
ELA_MBA	48214	2331
ELA_PBA	47974	2204
HAN_MIB1	43602	1397
HAN_MIB2	43377	1419
HAR_MBA	47789	2315
HAR_SCH1	50636	2742
HEL_MBA	46807	1562
HEL_PAC	50657	1971
HCA_SCH1	46896	1755
HCA_SCH2	47958	1918
HTR_01	50621	1511
HTR_02	48791	1369
MAU_MBA1	48139	697
MAU_MBA1	34439	568
MMI_MBA	48610	2402
MMI_MIB	41946	1949
NRA_J1	41686	1384
NRA_J2	43932	1684
PAT_MON	53020	3887
PAT_NRI	48762	2681
PFU_SCH1	50625	2364
PFU_SCH2	52540	2409
RTO_MBA	48610	1526
RTO_PBA	48940	1975
RVA_MBA	48943	2502
RVA_PBA	49379	2553
ZRO_MBA2	47572	3430
ZRO_MBA3	44484	3392

Table S2. Number of loci and phylogenetically informative SNPs as well as the overall size of the concatenated supermatrix obtained from a given set of *populations* (Stacks v 1.13) filter parameters "-p" (the minimum number of species that must possess that marker) and "-r" (the minimum percentage of individuals in each species that must possess that marker).

filter parameters	loci number	phylogenetically informative SNPs	total length (bp)
p21_r1	116	304	9,280
p21_r0.5	8,517	25,622	681,360
p20_r1	523	1,426	41,840
p20_r0.5	12,932	40,370	1,034,560
p18_r1	3,125	8,290	250,000
p18_r0.5	19,505	62,846	1,560,400
p16_r1	8,503	22,386	680,240
p16_0.5	24,941	79,908	1,995,280
p14_r1	15,025	39,421	1,202,000
p14_r0.5	30,629	96,368	2,450,320

Table S3. Sequence divergence between sister species of surfperches. Values correspond to uncorrected distances within a species and between sister species. Values within *E. lateralis* and *R. toxotes* are also given for information, but they are not sister species. Averages are provided for species except for the species in the genus *Brachyistius* that is shown at the bottom.

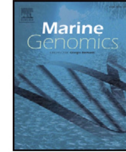
Species	intra-	inter-
A. argenteus	0.006	0.346
A. koelzi	0.000	
H. argenteum	0.013	0.443
H. ellipticum	0.016	
P. furcatus	0.012	0.195
R. vacca	0.010	
H.caryi	0.010	0.282
E. jacksoni	0.015	
E. lateralis	0.010	
R. toxotes	0.007	
Average	**0.010**	**0.317**
B. frenatus	0.015	0.025
B. aletes	0.012	

Chapter 2

Contents lists available at ScienceDirect

Marine Genomics

journal homepage: www.elsevier.com/locate/margen



The complete mitochondrial genome of the black surfperch, *Embiotoca jacksoni*: Selection and substitution rates among surfperches (Embiotocidae)

Gary C. Longo [a,*], Brendan O'Connell [b], Richard E. Green [b], Giacomo Bernardi [a]

[a] Department of Ecology and Evolutionary Biology, University of California, Santa Cruz, California 95060, USA
[b] Department of Biomolecular Engineering, University of California, Santa Cruz, California 95064, USA

ARTICLE INFO

Article history:
Received 13 January 2016
Received in revised form 11 March 2016
Accepted 12 March 2016
Available online 16 March 2016

Keywords:
Embiotoca jacksoni
Mitochondrial genome
Embiotocidae
Mitochondrial divergence
Black surfperch

ABSTRACT

The complete 16,515 bp nucleotide sequence of the mitochondrial genome was determined for the black surfperch, *Embiotoca jacksoni* (Perciformes: Embiotocidae). The black surfperch mitochondrial genome contains 13 protein-coding genes, two ribosomal RNA genes, 22 transfer RNA genes, and the non-coding control region (D-loop), the gene order of which is identical to that observed in most vertebrates. The protein-coding gene sequences of *E. jacksoni* mitochondrial DNA were compared with two other embiotocid surfperches with available complete mitochondrial genomes, *Cymatogaster aggregata* and *Ditrema temminckii*. Across all mitochondrial protein-coding genes in surfperches the weighted average substitution rate was 2.079% per My and average dN/dS ratios for each protein-coding gene ranged from 0.016 in CytB to 0.608 in ND3. Substitution rates and dN/dS ratios were relatively high for ATP8 compared to other protein-coding genes. Although most protein-coding genes showed signals of purifying selection, we found evidence for positive selection in ND3 in *E. jacksoni*.

© 2016 Elsevier B.V. All rights reserved.

1. Introduction

The black surfperch, *Embiotoca jacksoni*, is a reef dwelling marine fish in the family Embiotocidae (Perciformes) ranging from Northern California, USA to Central Baja California, Mexico. Embiotocid surfperches are endemic to the temperate North Pacific and over the last 13–18 My have radiated into diverse habitats including seagrass, kelp forests reefs, sandy bottoms, estuaries, and even freshwater (Bernardi and Bucciarelli, 1999; Longo and Bernardi, 2015; Tarp, 1952). Interestingly, surfperches have evolved a rare reproductive strategy in teleosts where they mate via internal fertilization and give birth to fully developed live young (viviparity). This natural history trait has made surfperches an attractive system for studying sperm competition, population genetics, and courtship displays (Bernardi, 2000; Cummings and Partridge, 2001; Reisser et al., 2009).

Mitochondrial DNA (mtDNA) has commonly been used for phylogenetic and population genetic analyses due to maternal inheritance, low effective population size, and high substitution rates compared to the nuclear genome (Avise, 2004). Although the vertebrate mitochondrial genome is inherited as a single unit and rarely undergoes recombination, various regions exhibit markedly different substitution rates (Naylor and Brown, 1998). Consequently, regions with relatively high substitution rates (i.e., hypervariable regions) have been used to infer relatedness in taxa with shallow divergence times, while markers that exhibit slower substitution rates (e.g., 16S rRNA locus) are better suited for resolving deeper taxonomic questions (Bernardi and Crane, 1999; Pardo et al., 2005; Sturmbauer and Meyer, 1993). The analysis of a species' complete mitochondrial genome best illustrates the variability observed among loci. Here we present the complete mitochondrial genome of the black surfperch, *E. jacksoni*, and utilize two other available surfperch mitogenomes, *Cymatogaster aggregata* and *Ditrema temminckii*, to estimate substitution rates and selective pressure by inferring nonsynonymous (dN) and synonymous (dS) substitution rates across protein-coding genes.

2. Materials and methods

2.1. Sample collection and DNA extraction

The mitogenomes of *C. aggregata* (NC_009059) and *D. temminckii* (NC_009060) were downloaded from NCBI (Mabuchi et al., 2007). We

* Corresponding author at: Department of Ecology and Evolutionary Biology, University of California Santa Cruz, 100 Shaffer Road, Santa Cruz, California 95060, USA.
E-mail address: garycharleslongo@gmail.com (G.C. Longo).

http://dx.doi.org/10.1016/j.margen.2016.03.006
1874-7787/© 2016 Elsevier B.V. All rights reserved.

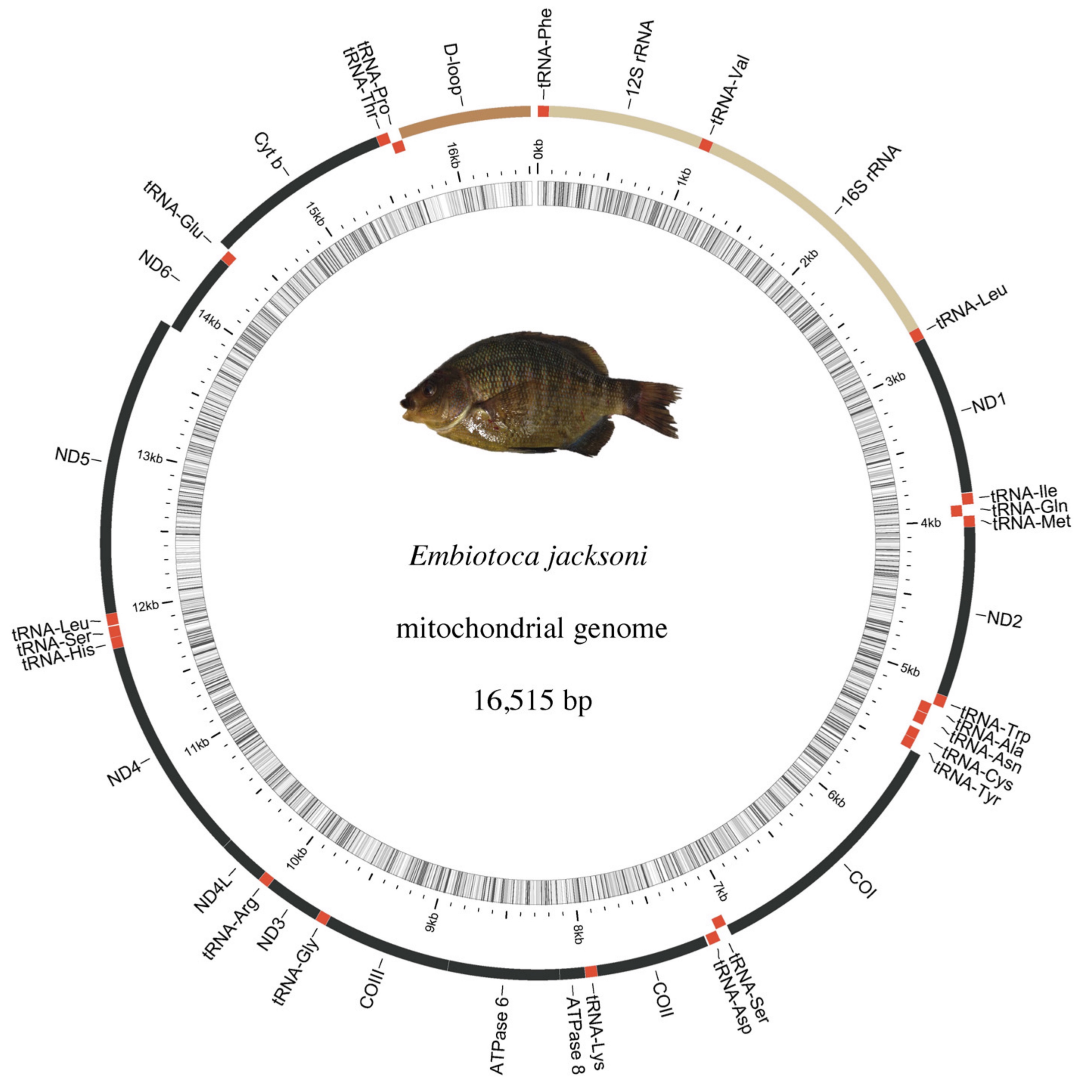


Fig. 1. Organizational map of the complete mitochondrial genome of the black surfperch, *Embiotoca jacksoni* (photograph of individual from which DNA was extracted). On the outer ring protein coding genes are shown in black, tRNAs in red, rRNAs in light brown, and the control region (D-loop) is in dark brown. Genes encoded on the light strand are inset relative to those on the heavy strand. The inner ring shows GC density with darker colors representing higher GC content.

generated a *de novo* mitochondrial genome of *E. jacksoni* using gill and liver tissue from a single individual freshly collected in Monterey, California (Fig. 1). Total DNA was extracted with Qiagen Blood and Cell Midi Kit following the manufacturer's protocol.

2.2. Sequencing and assembly

Genomic DNA was sheared and used to make an Illumina sequencing library (Meyer and Kircher, 2010). Briefly, the DNA was sonicated to approximately 300 bp and end repaired. Next, sequencing adapters were ligated to both ends of the DNA. The adapters were filled-in, and the DNA was amplified in an indexing PCR. After library preparation, the library size distribution was confirmed by agarose gel electrophoresis, and the library was size-selected with a Sage Science BluePippin with a 2% agarose gel cassette. The library was sequenced on an Illumina HiSeq 2500 with 2×150 PE rapid run chemistry. From these raw data we assembled the mitochondrial genome using a reference-guided, iterative assembly approach (https://github.com/mpieva/mapping-

iterative-assembler) designed to accurately assemble short, circular genomes from sequence data which may have many irrelevant reads in it (Briggs et al., 2009). The final assembly is supported by an average of 42.3 fold sequence coverage. No position was supported by fewer than 15 reads. The complete mitochondrial genome sequence of the black surfperch has been deposited in GenBank (accession number KU530212).

2.3. Sequence analyses

We compared substitution rates across protein-coding genes among the three available surfperch mitogenomes, *C. aggregata* (CAG), *D. temminckii* (DTE), and *E. jacksoni* (EJA). Individual genes were aligned and for consistent comparisons, one substitution model (Kimura 2-parameter) was used to estimate pairwise distances for all protein-coding loci using PAUP v4.0a (Swofford, 2003). Substitution rates (distance per My) were calculated for each species pair using the following estimated times to most recent common ancestor: 10 My for CAG/EJA and CAG/DTE, 7 My for DTE/EJA (Longo and Bernardi, 2015).

We estimated branch specific dN/dS ratios in a phylogenetic framework for the entire sequence of each protein-coding gene in the mitochondrial genome. A pruned surfperch phylogeny was constructed using the tree editing software Archaeopteryx (Han and Zmasek, 2009) and appended to each sequence alignment with branch lengths from a previously published embiotocid phylogeny based on 523 RADseq markers (Longo and Bernardi, 2015). The Branch-site random effects likelihood (Branch-site REL) method was used to compute average dN/dS for each branch in the pruned phylogeny as implemented in HyPhy (Kosakovsky Pond et al., 2011). Branch-site REL method detects lineages where a proportion of sites evolve with dN/dS > 1 but makes no assumptions about which lineages those are or about what happens to the rest of the lineages.

3. Results and discussion

3.1. Genome organization

The complete mitochondrial genome of *E. jacksoni* was determined to be 16,515 bp and contained 22 tRNA genes, 13 protein-coding genes, two rRNA genes, and a control region (D-Loop) in the same organization as for most fish genomes (Fig. 1, Table 1) including the two other published surfperch genomes (Iwasaki et al., 2013; Mabuchi et al., 2007). As commonly found in other vertebrates, most genes were encoded on the heavy strand except for eight tRNAs (Gln, Ala, Asn, Cys, Tyr, Ser, Glu, and Pro) and NADH dehydrogenase subunit 6 (ND6) (Fig. 1, Table 1). Black surfperch mitochondrial open reading frames overlapped in five locations for a total 23 bp (Table 1). Nucleotide frequencies in the *E. jacksoni* mitochondrial genome were A = 27.67%, C = 27.76%, G = 16.48%, and T = 28.09% (GC content = 44.24%).

3.2. Substitution rates and dN/dS

Pairwise substitution rates across embiotocid mitochondrial genes varied from 1.38% per My in CO1 (CAG × EJA) to 5.28% per My in ATP8 (DTE × EJA) (Fig. 2). The average weighted substitution rate across all mitochondrial protein-coding genes in surfperches was 2.079% per My (Fig. 2). In animal mitochondrial genomes, ATP8 generally exhibits a high substitution rate and dN/dS compared to other protein-coding genes (Castellana et al., 2011; Oliveira et al., 2008; Zardoya and Meyer, 2007). This elevated substitution rate can either be explained by positive selection or relaxed selection on ATP8 (Castellana et al., 2011). For example, ATP8 is missing in some metazoan mitochondrial genomes (Breton et al., 2010), which would support elevated substitution rates due to a relaxation in selective pressure (potentially followed by complete loss) in some lineages. In surfperches ATP8 yields the highest average substitution rate (3.49% per My) and second highest dN/dS (0.198) among mitochondrial protein-coding genes (Figs. 2, 3). These relatively high values for ATP8 in surfperches are likely due to a relaxation in selective pressure as opposed to positive selection as dN/dS < 1 in all lineages.

Branch-site REL analyses for positive selection in all mitochondrial protein-coding genes resulted in dN/dS ranging from 0 in NDL4 of *C. aggregata* to 1.76 in ND3 of *E. jacksoni* (Fig. 3). The relatively high dN/dS value observed in ND3 of the black surfperch suggests this protein may be experiencing positive selection with dN/dS > 1 (Fig. 3). ND3, as well as ND1, ND2, ND4, ND4L, ND5, and ND6, each code for a single protein component in NADH dehydrogenase complex 1, which consists of around 45 total subunits (~39 proteins are derived from nuclear genes) in Metazoa (Carroll et al., 2006; Castellana et al., 2011). Complex 1 is the first of several components in the oxidative phosphorylation pathway that generates ATP and is crucial for cellular energy production. Therefore variations in mitochondrial protein coding genes involved in this pathway potentially affect metabolic performance. ND3 has been shown to be under selection in several systems (Kennedy and Nachman, 1998). For example, a recent study analyzed mtDNA sequences from hundreds of Atlantic salmon individuals from many populations and found a pattern of positive selection in ND1, ND3, and ND4 that correlated with latitudinal gradients (Consuegra et al., 2015). Interestingly, similar results were

Table 1Please check Table 1 header if presented correctly and amend as necessary
Organization of the mitochondrial genome of *Embiotoca jacksoni*.

Gene	Position		Intergenic nucleotides (bp)	Strand
	Start	End		
tRNA-Phe	1	69	0	H
12S rRNA	70	1017	0	H
tRNA-Val	1018	1089	0	H
16S rRNA	1090	2780	0	H
tRNA-Leu	2781	2854	0	H
ND1	2855	3829	0	H
tRNA-Ile	3834	3903	4	H
tRNA-Gln	3903	3973	−1	L
tRNA-Met	3973	4041	−1	H
ND2	4042	5087	0	H
tRNA-Trp	5088	5159	0	H
tRNA-Ala	5161	5229	1	L
tRNA-Asn	5232	5304	2	L
tRNA-Cys	5339	5404	34	L
tRNA-Tyr	5405	5472	0	L
COI	5474	7069	1	H
tRNA-Ser	7070	7140	0	L
tRNA-Asp	7144	7215	3	H
COII	7225	7915	9	H
tRNA-Lys	7916	7989	0	H
ATP8	7991	8155	1	H
ATP6	8146	8828	−10	H
COIII	8829	9613	0	H
tRNA-Gly	9614	9683	0	H
ND3	9684	10,032	0	H
tRNA-Arg10033	10,101	0	H	
ND4L	10,102	10,398	0	H
ND4	10,392	11,772	−7	H
tRNA-His11773	11,841	0	H	
tRNA-Ser(2)	11,842	11,909	0	H
tRNA-Leu(2)	11,916	11,988	6	H
ND5	11,989	13,827	0	H
ND6	13,824	14,345	−4	L
tRNA-Glu	14,346	14,414	0	L
CYTB	14,420	15,560	5	H
tRNA-Thr	15,561	15,633	0	H
tRNA-Pro	15,635	15,704	1	L
D-Loop	15,705	16,515	0	H

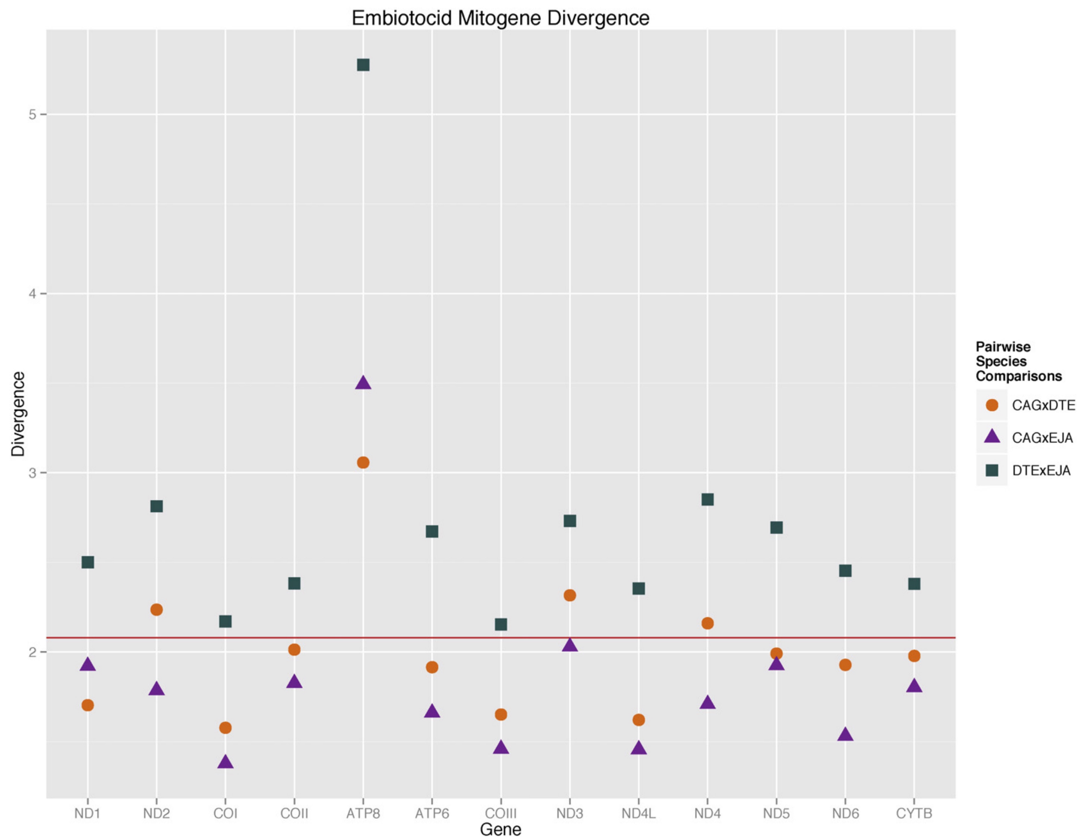


Fig. 2. Pairwise divergence rates across mitochondrial protein-coding genes in three embiotocid surfperches: *Cymatogaster aggregata* (CAG), *Ditrema temminckii* (DTE), and *Embiotoca jacksoni* (EJA). The red line represents the weighted average substitution rate across all mitochondrial protein-coding genes. Gene order reflects position in mitogenome.

also found in Pacific salmon (Garvin et al., 2011) and killer whales (Foote et al., 2011) suggesting these patterns of positive selection in mitochondrial genes could be driven by different metabolic demands in the lower temperatures of higher latitudes. Surfperches are certainly a good system to examine this pattern as they range from warmer waters off Southern California and Baja, Mexico to cooler waters off Alaska, but here we only have data from three individuals each representing separate species. Dissimilar metabolic demands across surfperch species may also be driving the observed dN/dS >1 in *E. jacksoni*, although additional data will be needed to confirm this pattern. On the other hand, purifying selection seems to be particularly strong in CytB and COIII, where dN/dS is low and narrow among comparisons (Fig. 3). Indeed, due to their crucial role in cellular respiration, mitochondrial protein-coding genes are expected to be mostly under purifying selection (dN/dS < 1), which we observe here.

4. Conclusion

While the mitochondrial genome is often seen as a single non-recombining unit, striking differences in substitution rates within the genome are observed, which in the case of surfperches translates into a more than twofold difference between the slowest and fastest genes (1.71%–3.94% per My). A number of explanations have been proposed to explain these differences, including differential selective pressure (Castellana et al., 2011; Consuegra et al., 2015; Popadin et al., 2012). Here, we examined dN/dS ratios among embiotocids and found most protein-coding genes exhibiting signals of purifying selection with some evidence for positive selection in the protein-coding gene ND3 in *E. jacksoni* when compared to *C. aggregata* and *D. temminckii*. Our results, however, should be interpreted with caution because the availability of only three surfperch mitochondrial genomes may have resulted in low power to detect purifying selection.

Acknowledgments

This project was funded by the Lerner-Gray Fund for Marine Research, Myers Trust, and the UCSC Ecology and Evolutionary biology research and travel fund. We would like to thank Andrew Kim for his dive support in Monterey.

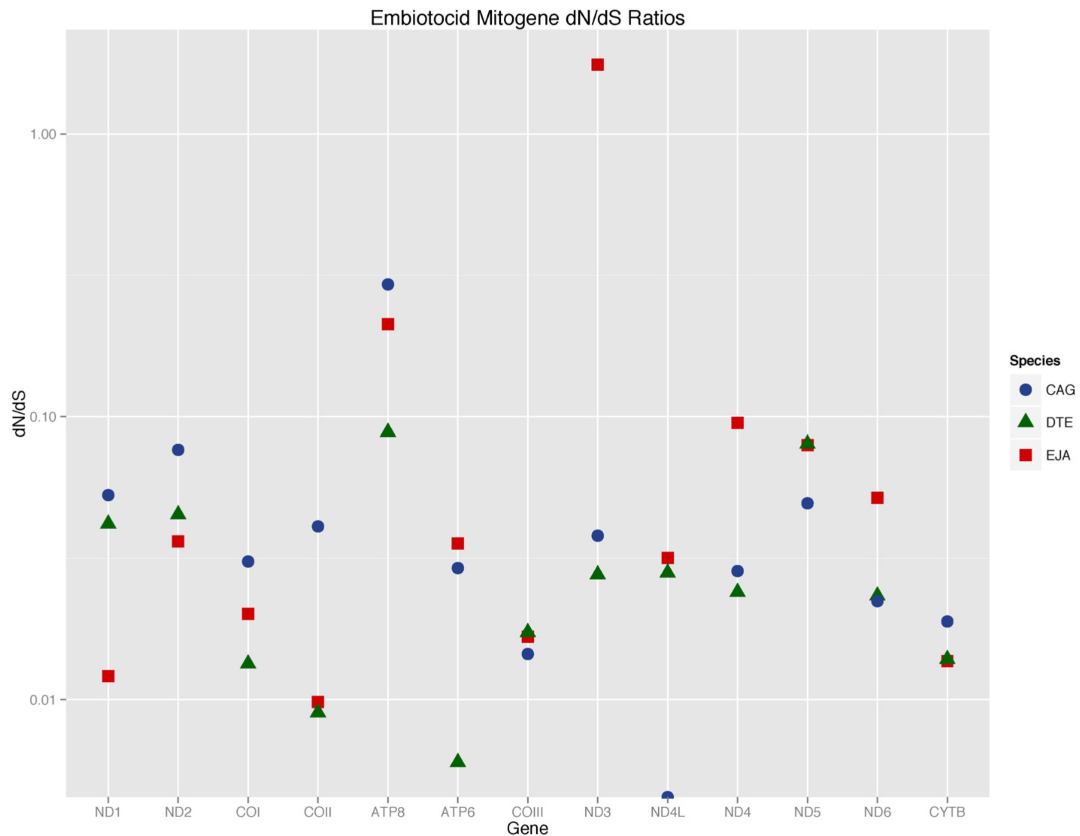


Fig. 3. Branch specific dN/dS ratios ($\log_{10}$) across mitochondrial protein-coding genes in three embiotocid surfperches: *Cymatogaster aggregata* (CAG), *Ditrema temminckii* (DTE), and *Embiotoca jacksoni* (EJA). Gene order reflects position in mitogenome.

References

Avise, J.C., 2004. Molecular Markers, Natural History, and Evolution. second ed. Chapman and Hall, New York.

Bernardi, G., 2000. Barriers to gene flow in *Embiotoca jacksoni*, a marine fish lacking a pelagic larval stage. Evolution (N. Y) 54, 226–237.

Bernardi, G., Bucciarelli, G., 1999. Molecular phylogeny and speciation of the surfperches (Embiotocidae, Perciformes). Mol. Phylogenet. Evol. 13, 77–81.

Bernardi, G., Crane, N.L., 1999. Molecular phylogeny of the humbug damselfishes inferred from mtDNA sequences. J. Fish Biol. 54, 1210–1217.

Breton, S., Stewart, D.T., Hoeh, W.R., 2010. Characterization of a mitochondrial ORF from the gender-associated mtDNAs of Mytilus spp. (Bivalvia: Mytilidae): identification of the "missing" ATPase 8 gene. Mar. Genomics 3, 11–18. http://dx.doi.org/10.1016/j.margen.2010.01.001.

Briggs, A.W., Good, J.M., Green, R.E., Krause, J., Maricic, T., Stenzel, U., Lalueza-Fox, C., Rudan, P., Brajkovic, D., Kucan, Z., Gusic, I., Schmitz, R., Doronichev, V.B., Golovanova, L.V., de la Rasilla, M., Fortea, J., Rosas, A., Pääbo, S., 2009. Targeted retrieval and analysis of five Neandertal mtDNA genomes. Science 325, 318–321. http://dx.doi.org/10.1126/science.1174462 (80-.).

Carroll, J., Fearnley, I.M., Skehel, J.M., Shannon, R.J., Hirst, J., Walker, J.E., 2006. Bovine complex I is a complex of 45 different subunits. J. Biol. Chem. 281, 32724–32727. http://dx.doi.org/10.1074/jbc.M607135200.

Castellana, S., Vicario, S., Saccone, C., 2011. Evolutionary patterns of the mitochondrial genome in metazoa: exploring the role of mutation and selection in mitochondrial protein-coding genes. Genome Biol. Evol. 3, 1067–1079. http://dx.doi.org/10.1093/gbe/evr040.

Consuegra, S., John, E., Verspoor, E., de Leaniz, C.G., 2015. Patterns of natural selection acting on the mitochondrial genome of a locally adapted fish species. Genet. Sel. Evol. 47, 58. http://dx.doi.org/10.1186/s12711-015-0138-0.

Cummings, M.E., Partridge, J.C., 2001. Visual pigments and optical habitats of surfperch (Embiotocidae) in the California kelp forest. J. Comp. Physiol. A Sensory, Neural, Behav. Physiol. 187, 875–889. http://dx.doi.org/10.1007/s00359-001-0258-6.

Foote, A.D., Morin, P.A., Durban, J.W., Pitman, R.L., Wade, P., Willerslev, E., Gilbert, M.T.P., da Fonseca, R.R., 2011. Positive selection on the killer whale mitogenome. Biol. Lett. 7, 116–118. http://dx.doi.org/10.1098/rsbl.2010.0638.

Garvin, M.R., Bielawski, J.P., Gharrett, A.J., 2011. Positive darwinian selection in the piston that powers proton pumps in Complex I of the mitochondria of Pacific salmon. PLoS One 6. http://dx.doi.org/10.1371/journal.pone.0024127.

Han, M.V., Zmasek, C.M., 2009. phyloXML: XML for evolutionary biology and comparative genomics. BMC Bioinf. 10, 356. http://dx.doi.org/10.1186/1471-2105-10-356.

Iwasaki, W., Fukunaga, T., Isagozawa, R., Yamada, K., Maeda, Y., Satoh, T.P., Sado, T., Mabuchi, K., Takeshima, H., Miya, M., Nishida, M., 2013. MitoFish and MitoAnnotator: A mitochondrial genome database of fish with an accurate and automatic annotation pipeline. Mol. Biol. Evol. 30, 2531–2540. http://dx.doi.org/10.1093/molbev/mst141.

Kennedy, P., Nachman, M.W., 1998. Deleterious mutations at the mitochondrial ND3 gene in South American marsh rats (Holochilus). Genetics 150, 359–368.

Kosakovsky Pond, S.L., Murrell, B., Fourment, M., Frost, S.D.W., Delport, W., Scheffler, K., 2011. A random effects branch-site model for detecting episodic diversifying selection. Mol. Biol. Evol. 28, 3033–3043. http://dx.doi.org/10.1093/molbev/msr125.

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. http://dx.doi.org/10.1016/j.ympev.2015.03.027.

Mabuchi, K., Miya, M., Azuma, Y., Nishida, M., 2007. Independent evolution of the specialized pharyngeal jaw apparatus in cichlid and labrid fishes. BMC Evol. Biol. 7, 10. http://dx.doi.org/10.1186/1471-2148-7-10.

Meyer, M., Kircher, M., 2010. Illumina sequencing library preparation for highly multiplexed target capture and sequencing. Cold Spring Harb. Protoc. 2010. http://dx.doi.org/10.1101/pdb.prot5448 (pdb.prot5448–pdb.prot5448).

Naylor, G.J.P., Brown, W.M., 1998. Amphioxus M itochond rial DNA, Chordate Phylogeny, and the limits of inference based on comparisons of sequences. Syst. Biol. 47, 61–76.
Oliveira, D.C.S.G., Raychoudhury, R., Lavrov, D.V., Werren, J.H., 2008. Rapidly evolving mitochondrial genome and directional selection in mitochondrial genes in the parasitic wasp nasonia (hymenoptera: pteromalidae). Mol. Biol. Evol. 25, 2167–2180. http://dx.doi.org/10.1093/molbev/msn159.
Pardo, B.G., Machordom, A., Foresti, F., Porto-Foresti, F., Azevedo, M.F.C., Bañón, R., Sánchez, L., Martínez, P., 2005. Phylogenetic analysis of flatfish (Order Pleuronectiformes) based on mitochondrial 16 s rDNA sequences. Sci. Mar. 69, 531–543. http://dx.doi.org/10.3989/scimar.2005.69n4531.
Popadin, K.Y., Nikolaev, S.I., Junier, T., Baranova, M., Antonarakis, S.E., 2012. Purifying selection in mammalian mitochondrial protein-coding genes is highly effective and congruent with evolution of nuclear genes. Mol. Biol. Evol. 30, 347–355. http://dx.doi.org/10.1093/molbev/mss219.
Reisser, C.M.O., Beldade, R., Bernardi, G., 2009. Multiple paternity and competition in sympatric congeneric reef fishes, *Embiotoca jacksoni* and *E. lateralis*. Mol. Ecol. 18, 1504–1510. http://dx.doi.org/10.1111/j.1365-294X.2009.04123.x.
Sturmbauer, C., Meyer, A., 1993. Mitochondrial phylogeny of the endemic mouthbrooding lineages of cichlid fishes from Lake Tanganyika in Eastern Africa. Mol. Biol. Evol. 10, 751–768.
Swofford, D.L., 2003. PAUP*. Phylogenetic Analysis Using Parsimony (*and Other Methods).
Tarp, F.H., 1952. Fish Bulletin No. 88. A Revision of the Family Embiotocidae (The Surfperches). State of California Department of Fish and Game.
Zardoya, R., Meyer, A., 2007. Phylogenetic Performance of Mitochondrial Protein-Coding Genes in Resolving Relationships among Vertebrates.

Chapter 3
Genomic signatures of parallel selection in surfperches (Embiotocidae)

Gary C. Longo* and Giacomo Bernardi

Department of Ecology and Evolutionary Biology, University of California, Santa Cruz, California 95060, USA.

*corresponding author
Department of Ecology and Evolutionary Biology
University of California Santa Cruz
100 Shaffer Road, Santa Cruz, California, 95060, USA
email: garycharleslongo@gmail.com

Abstract

Cases of parallel selection in closely related taxa provide excellent opportunities for better understanding the genetic component of selection and adaptation. Controlled experiments with *E. coli* and studies utilizing both terrestrial and freshwater systems have progressed our understanding of parallel selection, however there are fewer examples from the marine environment. Here, we perform exploratory genome scans for selection in four species of Embiotocid surfperches between populations in Monterey Bay, California, USA and Punta Banda, Baja California, Mexico. These localities are on opposite sides of a well-known biogeographic break and experience significantly different sea surface temperatures (SST) as well heterogeneous species assemblages. Using a F_{ST} outlier approach, we detect strong signals of intraspecific balancing and divergent selection as well as clear evidence for interspecific parallel selection with some loci aligning to NCBI proteins. These results suggest that California surfperches represent an ideal system for investigating different forms of selection in a natural marine environment.

Introduction

A tenet to Darwinian natural selection is that when selective pressure differs over space or time, it is reflected by expected changes in a population's collective genomes. Population genetics and the modern synthesis laid the mathematical groundwork for formally describing these expectations (Dobzhansky 1937; Mayr 1942; Simpson 1944; Stebbins 1950). Increased availability of molecular data resulted in the proposal of the neutral theory of molecular evolution, where genetic drift accounted for a significant share of the observed intraspecific and interspecific variation (Kimura 1968). Although the relative role of neutrality and selection in genome evolution was debated, cases of convergent and parallel evolution present opportunities to better understand the selective component of the equation. The genomic bases of parallel and convergent selection operating on populations of *E. coli* were observed experimentally (Woods *et al*. 2006; Barrick *et al*. 2009). In natural systems, the cichlid flocks of the Great African Lakes were identified early on as an ideal system to detect signatures of parallel and convergent evolution (Fryer & Iles 1972; Barlow 2000). DNA sequences of the cichlid flocks, and the recent sequencing of key cichlid genomes further confirmed the role of convergence in these systems (Meyer 1993; Brawand *et al*. 2014). Likewise, cases of multiple invasions of glacial freshwater lakes by marine forms of three-spine sticklebacks provided unique insight into both the mode and expectations of convergent and parallel evolution (Schluter & Nagel 1995; Colosimo *et al*. 2005; Hohenlohe *et al*. 2010). Indeed, recent genomic advances have resulted in increased power to detect selection both at the population

and gene expression levels with directional, balancing and disruptive selection all being uncovered in several systems (Andrews *et al*. 2016). While cases of convergent and parallel evolution in terrestrial and freshwater systems are well developed, there remains a need for more marine study systems where knowledge tends to lag behind (Webb 2012).

While different approaches have been used to identify selection, population-based methods often use variations of outlier approaches based on Wright's fixation index (F_{ST}) (Wright 1978). There, F_{ST} is calculated at a genomic scale and then outlier loci, or regions, are identified as candidates for selection (Bersaglieri *et al*. 2004; Hohenlohe *et al*. 2010). Importantly, this method is robust to past demographic fluctuations such as bottlenecks and population expansions (Beaumont 2005). In other words, these genome scans identify selection between or among populations by recognizing significant differences or similarities in allele frequencies relative to the rest of the genome. Here we performed genome scans on embiotocid surfperches, a family of predominantly marine fishes endemic to the temperate North Pacific.

The unique nature of embiotocid viviparity restricts dispersal resulting in relatively low gene flow and high F_{ST} values compared with most other reef fishes (Bernardi 2000, 2005). Surfperches belong to the speciose group Ovalentaria, which includes cichlids and damselfishes, among others (Wainwright *et al*. 2012; Betancur-R *et al*. 2013). Pharyngognathy, a modification to the musculature and structure of the pharyngeal jaws, is a synapomorphy for the group that has allowed for increased diet specialization and has likely contributed to the group's impressive diversity as

exemplified in cichlids (Wainwright *et al*. 2012) and to a lesser extent in surfperches (Tarp 1952; De Martini 1969; Bernardi & Bucciarelli 1999; Longo & Bernardi 2015). Although less speciose, surfperches are unique within Ovalentaria, as they exhibit internal fertilization and direct development. This differs from most marine fishes, which fertilize externally and have a bipartite life cycle with sedentary adults and pelagic larvae, potentially dispersing hundreds of kilometers (Leis 1991).

At the center of surfperch diversity, a prominent marine biogeographic break occurs in central California at Point Conception, where the northern Oregonian assemblage meets the southern San Diegan assemblage (Briggs 1974; Dawson 2001; Allen *et al*. 2006). Local current conditions play a key role in this break, resulting in significantly different sea surface temperatures (SST) north and south of Point Conception (Fig. 1, Fig. S1). These dissimilar biotic and abiotic conditions likely result in different selective pressures for populations along the central California coastline, therefore providing an opportunity to compare the effects of selection on a group of closely related marine fishes.

In this study we investigate genomic signals of selection between Monterey Bay, CA, USA (MBA) and Punta Banda, Baja California, Mexico (PBA) populations in four species of closely related surfperches, the black surfperch (*Embiotoca jacksoni*), striped surfperch (*Embiotoca lateralis*), rubberlip surfperch (*Rhacochilus toxotes*), and pile perch (*Rhacochilus vacca*). Here we find several instances where selection was identified as playing a role on the same genomic regions across species, thus

suggesting that similar selective pressures on closely related species translate to similar genomic signatures.

Materials and Methods

Sample collection and SST data

Twenty to 26 individuals were collected for each species with spears while on SCUBA or freediving from MBA and PBA (Fig. 1). Fin clippings or liver tissue were stored in 95% ethanol and DNA was extracted from tissue samples using DNeasy Blood & Tissue kits (Qiagen) according to the manufacturer's protocol. SST data and images were downloaded from the NOAA CoastWatch Program, specifically the data set used was SST, NOAA POES AVHRR, LAC, 0.0125 degrees, West US, Day and Night (Fig. 1, Fig. S1).

RADseq data

Restriction site associated DNA sequence (RADseq) library preparations were based on Baird and Miller's protocol (Baird *et al*. 2008; Miller *et al*. 2012) with some modifications (Longo & Bernardi 2015). Samples used in this study were sequenced in one of three libraries, each containing 96 individually barcoded samples. Each library was sequenced in a single lane on an Illumina HiSeq 2000 at the Vincent J. Coates Genomics Sequencing Laboratory at UC Berkeley.

Raw reads were trimmed to 92 bp on the 3' end, quality filtered, and then split according to the unique 6 bp barcode using Miller *et al*. (2012) custom Perl scripts.

Sequences were dropped if the product of quality scores for their respective 92 bases was below 80%. The barcode (6 bp) and restriction site residue (6 bp) were then removed from the 5' end, resulting in a final sequence length of 80 bp. Sequences from 172 individuals resulted in 217 million filtered reads ranging from 23,854 to 6,345,608 reads per individual. Ten individuals with <200,000 reads were discarded from subsequent analyses.

Loci were detected using the software program STACKS version 1.29 (Catchen *et al*. 2011, 2013) by running the *denovo_map* program for each species. In each *denovo_map* we set a minimum stack depth (-m) of five, a maximum of two mismatches per loci for each individual (-M), and allowed up to one mismatch when building catalog loci (-n). We excluded highly repetitive stacks, set the maximum number of stacks per locus at two, and disabled calling haplotypes from secondary reads. For each species' *denovo_map* batch, we ran the STACKS *populations* program to further filter the data and generate output files for downstream analyses. Due to generally high coverage across individuals, we increased the minimum stack depth (-m) to 10 and required that 80% of individuals in each population retain the marker (-p 2 & -r 0.8). We used the --write_single_snp flag so only a single SNP from each RADseq locus was used for detecting loci under selection and population structure analyses. STRUCTURE output files from the STACKS *population* scripts were converted into appropriate formats for downstream population genetic analyses using PGDSPIDER 2.0 (Lischer & Excoffier 2012). Although individuals with relatively low coverage were removed prior to running STACKS, some individuals

with high coverage had a high percentage of missing data. We removed eight individuals who were missing data for more than 50% of loci and ran again the STACKS *population* scripts, which resulted in 154 samples used for all downstream analyses (*E. jacksoni:* MBA = 17 & PBA = 20; *E. lateralis*: MBA = 21 & PBA = 20; *R. toxotes*: MBA = 19 & PBA = 21; *R. vacca*: MBA = 16 & PBA = 20, Table S1). The quality filtered sequences were deposited at the National Center for Biotechnology Information (NCBI) short-read archive (accession no. XXXXXX).

A Bayesian clustering approach as implemented in STRUCTURE 2.2 (Pritchard *et al.* 2000) was used to assess genetic partitioning for each species. We used an initial burnin of 10,000 with 100,000 MCMC iterations for K = 1-5 (assumed number of populations) with 10 replicates each. The most likely number of genetic partitions was then estimated using the Evanno method (Evanno *et al.* 2005) as implemented in STRUCTURE HAVESTER (Earl & vonHoldt 2012) and the final results were visualized using DISTRUCT (Rosenberg 2004).

To test for loci under selection between populations in each species independently we used the FDIST method (Beaumont & Nichols 1996) as implemented in ARLEQUIN v3.5 (Excoffier *et al.* 2005). This method detects signals of selection, either acting directly on the locus or indirectly on a closely linked locus (hitch-hiking), based on F_{ST} as a function of heterozygosity (Bierne *et al.* 2013; Lotterhos & Whitlock 2015). We ran 100 demes with 50,000 simulations and set the minimum and maximum expected heterozygosity to 0 and 1 respectively. Loci below the 1% quantile and above the 99% quantile were considered candidates for balancing and

divergent selection respectively. We investigated potential gene functions of loci undergoing selection by searching the NCBI database for sequence similarities using BLASTN 2.3.1 (Altschul *et al*. 1997) and report hits on predicted proteins with an E-value ≤1e-6 (Table S2).

Results

RADSeq data

Total number of loci output from STACKS *population* runs (*-m 10, -p 2, & -r 0.8*) for each species yielded between 19,761 to 34,252 loci in *R. vacca* and *E. lateralis* respectively, while polymorphic loci counts varied from 4,706 to 8,728 in *R. vacca* and *E. lateralis* respectively (Table 1).

Population structure

STRUCTURE HARVESTER results identified K=2 for all species as the most likely clustering model, however the Evanno method cannot detect a K=1 scenario, which seems to be the most likely case for *R. vacca*. STRUCTURE plots based on K=2 and a single SNP from each polymorphic locus showed variable levels of genetic structure between MBA and PBA in each species (Fig. 2). *E. lateralis* showed the strongest levels of genetic differentiation between individuals in MBA and PBA while almost no differentiation is observed between the populations in *R. vacca* corresponding to strong and weak population structure, respectively. However, population pairwise comparisons between MBA and PBA yielded a gradient of F_{ST}

values with 0.21699 in *E. jacksoni*, 0.1318 in *R. toxotes,* 0.11441 in *E. lateralis,* and 0.02634 in *R. vacca* (Table 2).

Loci under selection

FDIST approaches to identify loci under selection between MBA and PBA populations in each species yielded varying numbers of candidate loci. For *E. jacksoni,* out of 8,304 total polymorphic loci, 88 were classified as being under divergent selection (above 99% quantile) and 48 as balancing (below 1% quantile). For *E. lateralis,* out of 8,728 total polymorphic loci, 136 and 89 loci were classified as being under divergent and balancing selection respectively. For *R. toxotes,* out of the 4,706 total polymorphic loci, 58 and 44 loci were classified as being under divergent and balancing selection respectively. For *R. vacca,* out of the 5,702 total polymorphic loci, 59 and 420 loci were classified as being under divergent and balancing selection respectively (Figs. S2, Table 1).

Gene function and parallel selection

The proportion of loci under divergent selection that aligned with NCBI proteins ranged from 11.9% (7 of 59 loci) in *R. vacca* to 19.3% (17 of 88 loci) in *E. jacksoni* and from 16.7% (8 of 48 loci) in *E. jacksoni* to 32.6% (29 of 89 loci) in *E. lateralis* for loci under balancing selection (Fig. 3, Table S2). Remarkably, we observed several instances of parallel divergent and balancing selection across surfperch species in MBA and PBA populations (Fig. 3). In regards to parallel selection, we

refer to interspecific RADseq loci that share very high sequence similarity, likely due to orthologous SbfI restriction sites, as orthologous loci. On the other hand, RADseq loci that do not share high sequence similarity but align to the same NCBI protein in different physical locations are referred to as non-orthologous RADseq loci. Approximately half of the loci putatively undergoing parallel selection aligned to proteins on NCBI's database based on our BLASTN filter parameters.

One such case of parallel divergent selection involves non-orthologous RADseq markers from *E. jacksoni* and *R. toxotes* that align to the same predicted protein, transmembrane protease serine 9-like (Table S2), but in different locations. We also detected several instances of surfperch species sharing loci under balancing selection (Figure 3). Specifically, *R. vacca* and *R. toxotes* share an orthologous locus under balancing selection that aligns to a known centromeric protein. *R. vacca* and *E. lateralis* also share three loci under balancing selection that align to NCBI proteins, one of which is a trace amine-associated receptor protein (taar). Another involves non-orthologous markers that align to same known spectrin repeat containing, nuclear envelope 2 protein but in different locations. In the last case, non-orthologous *R. vacca* and *E. lateralis* loci align to the same predicted dopey family member protein (dopey 1) but in different fish taxa (Table S2). Finally, several loci under selection have predicted gene functions that may be related to surfperch life histories or distinct selective pressures between MBA and PBA, as briefly discussed below.

Discussion

Over the past few decades our understanding of parallel and convergent evolution in experimental (Woods *et al*. 2006; Barrick *et al*. 2009) and natural populations (Schluter & Nagel 1995; Colosimo *et al*. 2005; Hohenlohe *et al*. 2010) has grown. Here we present evidence from the marine environment that selection may act on the same genomic regions in closely related species subjected to the same selective environment. Additionally we discuss a select few cases of loci under selection with predicted gene functions that are particularly interesting given the system. Our results also show significant differences in gene flow and genetic structure among taxa between MBA and PBA populations, however further work is needed to better understand how these interspecific variations in gene flow affect genomic selection in such a complex natural system.

Population structure

Compared to most other marine fishes surfperches exhibit exceptionally reduced gene flow, which is largely attributed to their apelagic life history (Bernardi 2000, 2005). This unique reproductive strategy restricts dispersal to juvenile and/or adult movement, as seen in other systems (Planes *et al*. 2001; Bernardi & Vagelli 2004; Vagelli *et al*. 2009). Here, we show that closely related surfperch species exhibit significant variation in gene flow between the same locations, which is likely due to differences in life history traits such as habitat preference (Tarp 1952; Longo & Bernardi 2015). STRUCTURE results showed a very strong genetic break between MBA and PBA for both *E. jacksoni* and *E. lateralis*. There is a genetic break in *R.*

toxotes but it is less pronounced than for the *Embiotoca spp*. and exhibits some evidence of admixture. On the other hand, *R. vacca* shows signs of panmixia with almost complete admixture between populations (Fig. 2). Pairwise population comparisons produced F_{ST} values that differed by nearly an order of magnitude from 0.21699 in *E. jacksoni* to 0.02634 in *R. vacca* (Table 2). Previous population genetic work based on mitochondrial data has shown strong differentiation between northern and southern populations for *E. jacksoni* as well as for *E. lateralis* (Bernardi 2000, 2005). These signals of low admixture and strong genetic population structure from the *Embiotoca spp*. can be attributed to their affinity for physical structure, such as kelp forest reefs, and low probability of crossing sandy expanses (Bernardi 2000; Longo & Bernardi 2015). On the other hand, *R. vacca* populations exhibit the lowest observed F_{ST} in this study and show evidence for panmixia, which may be attributed to the species being habitat generalist that loosely associate with many types of structure and readily cross sandy expanses (Tarp 1952; De Martini 1969; Longo & Bernardi 2015). Although *R. toxotes* is also a habitat generalist, it is less likely to venture over sandy expanses than *R. vacca*, which would explain the stronger observed genetic structure.

Potential functions of loci under selection

Two conspicuous environmental disparities between MBA in central California and PBA in northern Baja California are the differences in SST (Fig. 1, Fig. S1) and the heterogeneous species assemblages. Although these dissimilarities would predict

strong divergent selection, balancing selection accounts for a large proportion of the observed genomic effects (Fig. 3, Fig. S2, Table 1). The proportion of loci under divergent and balancing selection that aligned with NCBI proteins varied (Fig. 3, Table S2). Encouragingly we find that all markers under divergent selection that pass our BLASTN filter parameters align to a predicted or known protein in a teleost fish (52/52). Although most markers under balancing selection also align to teleost proteins (119/122), three loci map to mammal genes (rat, horse, and polar bear), which all have highly conserved functions (Table S2). In general balancing selection is more likely to act on highly conserved genes such as those involved in cellular house keeping pathways. Here we discuss the predicted gene functions of outlier loci with BLASTN alignments that may be related to surfperch life histories or distinct selective pressure between MBA and PBA.

A particularly intriguing case of divergent selection seems to be occurring in *E. jacksoni* as a marker aligns to dynein axonemal heavy chain 2, a protein involved in sperm motility (Chapelin *et al*. 1997). Intense sperm competition likely occurs within female surfperches since they mate with multiple males and store sperm for up to several months until fertilization (Reisser *et al*. 2009). Another pertinent find in *E. jacksoni* populations involves a locus under divergent selection that aligns to mitogen-activated protein kinase-binding protein, a protein involved in a signaling pathway activated by cytokines and exposure to environmental stress (Ip & Davis 1998). With notable differences in SST and heterogeneous species assemblages MBA

and PBA populations likely encounter dissimilar environmental stressors, which could explain this signal of divergent selection.

Indeed MBA and PBA populations also likely face different immunological selective pressures as marine systems with higher temperatures are known to harbor higher pathogen loads than cooler systems (Harvell 1999). One would expect genes linked to immune responses to be strong candidates for divergent selection between populations and in fact a marker from *E. lateralis* aligns to a T-cell receptor alpha-chain protein (Table S2). The alpha-chain is a variable subunit of the T cell that functionally recognizes foreign antigens bound to MHC molecules (Brigl & Brenner 2004). Likewise *R. toxotes* populations also showed evidence for divergent selection in immune response when a marker aligned to NOD-like receptor C3, a protein involved in T cell activation pathways (Chen *et al*. 2009).

Interestingly in *R. vacca* we found evidence for balancing and divergent selection co-occurring in different subunits of the same holoenzyme complex, the vacuolar (H^+)-ATPase (V-ATPase). Broadly, V-ATPase is an evolutionarily conserved proton pump that functionally alters pH across lipid membranes and is involved in diverse processes such as lysosomal protein degradation, bone reabsorption by osteoclasts, sperm maturation and storage, and early embryonic left-right patterning (Blake-Palmer *et al*. 2007; Jansen & Martens 2012). The V-ATPase holoenzyme has two major domains, an integral membrane protein complex that contains a rotary transmembrane pore (V_0) and a cytoplasmic complex with an ATP hydrolysis site (V_1). Regulation and distribution of V-ATPase is at least partly controlled by V-

ATPase accessory proteins. A *R. vacca* locus under divergent selection aligned to the accessory subunit S1/ac45 of V-ATPase, which has been shown to act as a regulator in the neuroendocrine secretory pathway (Jansen & Martens 2012). S1/ac45 also plays a pivotal role in vertebrate development as knockout mice resulted in early embryonic lethality (Jansen & Martens 2012) and insertional mutagenesis in zebra fish resulted in pigmentation defects and failure to develop swim bladders in most embryos (Amsterdam *et al*. 1999). On the other hand, a locus under balancing selection in *R. vacca* aligned to an E1 subunit isoform of the V_0 subunit, which is essential for proper proton pump function in yeast and is ubiquitously expressed across tissues (Blake-Palmer *et al*. 2007). This finding illustrates the relatively small scale at which divergent and balancing selection can operate across the genome. With such a wide range of functions, further work is needed to assert what environmental pressures may be driving divergent selection in the regulatory accessory subunit of the V-ATPase complex between MBA and PBA populations of *R. vacca*.

Parallel selection

Observations of parallel selection across closely related taxa in a single system are rare. In such instances, selection favors the same phenotype in independently evolving lineages usually through parallel evolution or in some cases convergent evolution. Although links between genotype and phenotype from polymorphic RADseq data have been found in sticklebacks (Hohenlohe *et al*. 2010), much work remains to substantiate these associations in surfperches. Therefore coupling

genotype to phenotype, which is necessary to properly identify both parallel and convergent evolution, is beyond the scope of this study and we simply qualify observations of interspecific homologous loci or genes under balancing or divergent selection as parallel selection. Here we expand on cases where loci under parallel selection align to NCBI proteins and speculate on selective pressures that may be driving these observations.

In one instance, we observe parallel divergent selection in *E. jacksoni* and *R. toxotes* (Fig. 3) where non-orthologous loci align to transmembrane serine protease 9 (TMPSS9)/ polyserase-1 but in different physical locations along the nucleotide sequence (Table S2). Specifically the *E. jacksoni* RADseq query aligns to nucleotide positions 1837-1901 while the *R. toxotes* query aligns to nucleotide positions 1772-1828. Type II TMPSS are involved in diverse physiological functions such as digestion, blood pressure control, and hearing (Fontanil *et al*. 2014). Recently TMPSS9 has been linked to promotion of pro-tumor activities in mammals (Fontanil *et al*. 2014), although a more refined function has yet to be determined. Further work is needed to understand the potential phenotypes under selection in teleosts and vertebrates broadly.

We also detected several cases of parallel balancing selection between surfperch species pairs in MBA and PBA populations (Fig. 3). *R. toxotes* and *R. vacca* share an orthologous locus under balancing selection that aligns to trace amine-associated receptor (taar) 13-c like protein (Table S2). Class II and III taar gene products are known to function in olfactory sensory systems, although the latter are unique to

teleosts in which several gene families have been shown to be under positive Darwinian selection (Hussain *et al*. 2009). However the class II taar 13-c like gene, which is expressed in olfactory receptor epithelium in teleosts, shows signals of balancing selection in both global and site-by-site positive selection analyses (Hussain *et al*. 2009). Class II taar gene families arose early in jawed vertebrates and are more likely to be conserved, which may explain our finding of parallel balancing selection occurring across *R. toxotes* and *R. vacca* populations. However, with a targeted approach we would expect at least some scale of interspecific divergent selection in class III taar genes due to dissimilar natural histories requiring unique olfactory sensitivity (e.g., heterogeneous diet, habitat, hormones, etc.).

E. lateralis and *R. vacca* share three instances of parallel balancing selection. In one case, an orthologous locus aligns to centromeric protein A (CENP-A), which is a histone 3 variant and epigenetically responsible for centromere formation and kinetochore assembly (Allshire & Karpen 2008). Most regions of CENP-A are highly conserved due to its crucial role in cell division, although the N-terminal tail, which can be highly divergent both in length and AA composition among taxa, shows signs of positive selection in percid fishes (Abbey & Kral 2015). However, this orthologous locus aligns upstream (nucleotide positions 952-1031) of the first exon (nucleotide positions 2651-2774). One possibility is this upstream sequence could be an important promoter region and therefore conserved between these surfperch species.

E. lateralis and *R. vacca* also share non-orthologous loci that align to the dopey family member 1 protein (dopey1) but in different fish taxa, *Oreochromis niloticus*

and *Larimichthys crocea*, respectively (Table S2). Dopey1, a member of the leucine zipper-like family, is involved in processes such as endosome to golgi transport, organization of the endoplasmic reticulum, and cell morphogenesis (Pascon & Miller 2000). It has also been shown to be crucial for proper myelination in the central nervous system of mammals (Tanaka *et al*. 2014) and likely so in other vertebrates except hagfishes and lampreys, which lack myelin (Bullock *et al*. 1984). The critical nature of this highly conserved protein likely drives balancing selection in *E. lateralis* and *R. vacca* populations. The final case of parallel balancing selection in these two species involves non-orthologous RADseq markers that align to synaptic nuclear envelope 2 protein (Syne-2) but in different nucleotide positions (28,192 to 28,270 in *E. lateralis* and 13,029 to 13,107 in *R. vacca*). Syne proteins generally function to tether organelles to the cytoskeleton. Syne-2 has been shown to play a crucial role in properly anchoring neuromuscular nuclei, which is critical for motor neuron innervation and respiration in mammals (Zhang *et al*. 2007). Like dopey 1, Syne-2 is also crucially involved in basic cellular and organismal processes such as neuromuscular movement making it another excellent candidate for balancing selection. Overall these RADseq data have highlighted genes likely involved in various modes of selection in surfperches and suggest this promising system warrants further investigation.

Although powerful, genome scans based on F_{ST} outliers are limited to detecting SNPs directly under selection or those hitch-hiking. Without the inclusion of other data, such as QTL mapping experiments, phenotypes associated with the various

alleles cannot be deduced from genome scans alone. However the ability to detect selection with genome scans has been corroborated in systems with comprehensive QTL data, such as sticklebacks. Furthermore these studies identified previously unknown SNPs and candidate pathways involved in freshwater adaptation (Hohenlohe et al. 2010). Although surfperches lack any such phenotypically informative data, these exploratory genome scans detected strong signals of intraspecific divergent and balancing selection as well as interspecific parallel selection. However, our inference into the targets of selection is limited to the subset of loci that map to NCBI proteins. Undoubtedly some outlier loci did not map to known proteins but are in fact associated with novel genes or pathways in local adaptation to the dissimilar environments of MBA and PBA. In these scenarios, a high quality and well annotated genome would allow for mapping to better understand what genes and pathways may be driving these observations.

Our findings suggest this system warrants further investigation. For one we have shown that surfperches exhibit significant differences in gene flow among species, making this system attractive for studying the effects of dissimilar gene flow on selection and local adaptation. Additionally the differences in SST between MBA and PBA replicate selective pressures many marine populations will face as the climate continues to warm. Progressing our understanding of what standing genetic variability may be better suited for adaptation to increasing temperatures could be crucial for fisheries management and conservation efforts.

Conclusion

California surfperches present an ideal marine system for studying different modes of selection. Here, in four surfperch species, we identified signatures of balancing and divergent selection within species as well as cases of parallel selection between species. Although our study lacks the power to identify what phenotypes may be associated with a given outlier locus, some loci bearing signatures of selection aligned to NCBI proteins with high confidence allowing for insight into the types of gene pathways potentially under selection. The next step in better understanding these genomic signals is to map each marker, and its corresponding F_{ST} value, to a well-annotated genome in order to identify what additional genes are potentially driving these patterns of selection on the genomic scale.

Acknowledgements

We would like to thank Stephen Hauskins for his ongoing IT support. This work was supported through a grant from the Department of Ecology and Evolutionary Biology at University of California, Santa Cruz; Lerner-Gray Grants for Marine Research; and Myers Trust (to G.C.L.). This work used the Vincent J. Coates Genomics Sequencing Laboratory at UC Berkeley, supported by NIH S10 Instrumentation Grants S10RR029668 and S10RR027303.

Bibliography

Abbey HNA, Kral LG (2015) Adaptive Evolution of CENP-A in Percid Fishes. *Genes*, **6**, 662–71.

Allen L, Horn M, Pondella D (2006) *The Ecology of Marine Fishes: California and Adjacent Waters*. University of California Press, Berkeley.

Allshire RC, Karpen GH (2008) Epigenetic regulation of centromeric chromatin: old dogs, new tricks? *Nature reviews. Genetics*, **9**, 923–937.

Altschul SF, Madden TL, Schäffer a a *et al*. (1997) Gapped BLAST and PSI-BLAST: a new generation of protein database search programs. *Nucleic Acids Res*, **25**, 3389–402.

Amsterdam A, Burgess S, Golling G *et al*. (1999) A large-scale insertional mutagenesis screen in zebrafish. *Genes and Development*, **13**, 2713–2724.

Andrews KR, Good JM, Miller MR, Luikart G, Hohenlohe PA (2016) Harnessing the power of RADseq for ecological and evolutionary genomics. *Nat Rev Genet*, **advance on**, 81–92.

Baird NA, Etter PD, Atwood TS *et al*. (2008) Rapid SNP Discovery and Genetic Mapping Using Sequenced RAD Markers. *PLoS One*, **3**, e3376.

Barlow GW (2000) *The Cichlid Fishes: Nature's Grand Experiment In Evolution*. Perseus Publishers, Cambridge.

Barrick JE, Yu DS, Yoon SH *et al*. (2009) Genome evolution and adaptation in a long-term experiment with Escherichia coli. *Nature*, **461**, 1243–1247.

Beaumont M a. (2005) Adaptation and speciation: What can Fst tell us? *Trends in Ecology and Evolution*, **20**, 435–440.

Beaumont M a, Nichols R a (1996) Evaluating loci for use in the genetic analysis of population structure. *Proceedings of the Royal Society B: Biological Sciences*, **263**, 1619–1626.

Bernardi G (2000) Barriers to gene flow in Embiotoca jacksoni, a marine fish lacking a pelagic larval stage. *Evolution*, **54**, 226–237.

Bernardi G (2005) Phylogeography and demography of sympatric sister surfperch species, Embiotoca jacksoni and E. lateralis along the California coast: Historical versus ecological factors. *Evolution*, **59**, 386–394.

Bernardi G, Bucciarelli G (1999) Molecular phylogeny and speciation of the surfperches (Embiotocidae, Perciformes). *Molecular Phylogenetics and Evolution*, **13**, 77–81.

Bernardi G, Vagelli A (2004) Population structure in Banggai cardinalfish, Pterapogon kauderni, a coral reef species lacking a pelagic larval phase. *Marine Biology*, **145**, 803–810.

Bersaglieri T, Sabeti PC, Patterson N *et al*. (2004) Genetic signatures of strong recent positive selection at the lactase gene. *American Journal of Human Genetics*, **74**, 1111–1120.

Betancur-R R, Broughton RE, Wiley EO *et al*. (2013) The Tree of Life and a New Classification of Bony Fishes. *PLOS Currents Tree of Life*.

Bierne N, Roze D, Welch JJ (2013) Pervasive selection or is it …? why are F ST outliers sometimes so frequent ? *Molecular Ecology*, **33**, 2061–2064.

Blake-Palmer KG, Su Y, Smith AN, Karet FE (2007) Molecular cloning and characterization of a novel form of the human vacuolar H+-ATPase e-subunit: An essential proton pump component. *Gene*, **393**, 94–100.

Brawand D, Wagner CE, Li YI *et al*. (2014) The genomic substrate for adaptive radiation in African cichlid fish. *Nature*, **513**, 375–381.

Briggs JC (1974) *Marine zoogeography*. McGraw-Hill, New York.

Brigl M, Brenner MB (2004) CD1: Antigen Presentation and T Cell Function. *Annu. Rev. Immunol*, **22**, 817–90.

Bullock TH, Moore JK, Fields RD (1984) Evolution of Myelin Sheaths : Both Lamprey and Hagfish Lack Myelin. *Neuroscience Letters*, **48**, 145–148.

Catchen JM, Amores A, Hohenlohe PA, Cresko WA, Postlethwait JH (2011) Stacks: Building and Genotyping Loci De Novo From Short-Read Sequences. *G3: Genes, Genomes, Genetics*, **1**, 171–182.

Catchen J, Hohenlohe PA, Bassham S, Amores A, Cresko WA (2013) Stacks: an analysis tool set for population genomics. *Molecular ecology*, **22**.

Chapelin C, Duriez B, Magnino F *et al*. (1997) Isolation of several human axonemal dynein heavy chain genes: Genomic structure of the catalytic site, phylogenetic analysis and chromosomal assignment. *FEBS Letters*, **412**, 325–330.

Chen G, Shaw MH, Kim Y-G, Nuñez G (2009) NOD-like receptors: role in innate immunity and inflammatory disease. *Annual review of pathology*, **4**, 365–398.

Colosimo PF, Hosemann KE, Balabhadra S *et al*. (2005) Widespread Parallel Evolution in. *Science*, **307**, 1928–1933.

Dawson MN (2001) Phylogeography in Coastal Marine Animals : A Solution from California ? *Journal of Biogeography*, **28**, 723–736.

Dobzhansky T (1937) *Genetics and the Origin of Species*. Columbia University Press, New York.

Earl DA, vonHoldt BM (2012) STRUCTURE HARVESTER: A website and program for visualizing STRUCTURE output and implementing the Evanno method. *Conservation Genetics Resources*, **4**, 359–361.

Evanno G, Regnaut S, Goudet J (2005) Detecting the number of clusters of individuals using the software STRUCTURE: a simulation study. *Molecular ecology*, **14**, 2611–20.

Excoffier L, Laval G, Schneider S (2005) Arlequin (version 3.0): An integrated software package for population genetics data analysis. *Evolutionary Bioinformatics*, **1**, 47–50.

Fontanil T, Mohamedi Y, Esteban MM, Obaya AJ, Cal S (2014) Polyserase-1/TMPRSS9 induces pro-tumor effects in pancreatic cancer cells by activation of pro-uPA. *Oncology Reports*, **31**, 2792–2796.

Fryer G, Iles TD (1972) *The cichlid fishes of the Great Lakes of Africa* (Oliver & Boyd, Ed,). Edinburgh, UK.

Harvell CD (1999) Emerging Marine Diseases--Climate Links and Anthropogenic Factors. *Science*, **285**, 1505–1510.

Hohenlohe P a, Bassham S, Etter PD *et al*. (2010) Population Genomics of Parallel Adaptation in Threespine Stickleback using Sequenced RAD Tags. *Plos Genetics*, **6**, e1000862.

Hussain A, Saraiva LR, Korsching SI (2009) Positive Darwinian selection and the birth of an olfactory receptor clade in teleosts. *Proceedings of the National Academy of Sciences of the United States of America*, **106**, 4313–8.

Ip YT, Davis RJ (1998) Signal transduction by the c-Jun N-terminal kinase (JNK)--from inflammation to development. *Current opinion in cell biology*, **10**, 205–

219.

Jansen EJ, Martens GJ (2012) Novel insights into V-ATPase functioning: distinct roles for its accessory subunits ATP6AP1/Ac45 and ATP6AP2/(pro) renin receptor. *Curr Protein Pept Sci*, **13**, 124–133.

Kimura M (1968) Evolutionary rate at the molecular level. *Nature*, **217**, 624–626.

Leis J (1991) The Pelagic Stage of Reef Fishes. In: *The Ecology of Fishes on Coral Reefs* (ed Sales P), pp. 182–229. Academic Press Inc., San Diego, CA.

Lischer HEL, Excoffier L (2012) PGDSpider: An automated data conversion tool for connecting population genetics and genomics programs. *Bioinformatics*, **28**, 298–299.

Longo G, Bernardi G (2015) The evolutionary history of the embiotocid surfperch radiation based on genome-wide RAD sequence data. *Molecular Phylogenetics and Evolution*, **88**, 55–63.

Lotterhos KE, Whitlock MC (2015) The relative power of genome scans to detect local adaptation depends on sampling design and statistical method. *Molecular Ecology*, **24**, 1031–1046.

De Martini EE (1969) A Correlative Study of the Ecology and Comparative Feeding Mechanism Morphology of the Embiotocidae (Surf-fishes) as Evidence of the Family's Adaptive Radiation into Available Ecological Niches. *The Wasmann Journal of Biology*, **27**, 177–247.

Mayr E (1942) *Systematics and the Origin of Species, from the Viewpoint of a Zoologist*. Harvard University Press, Cambridge.

Meyer a (1993) Phylogenetic relationships and evolutionary processes in East African cichlid fishes. *Trends in ecology & evolution*, **8**, 279–84.

Miller MR, Brunelli JP, Wheeler P a *et al*. (2012) A conserved haplotype controls parallel adaptation in geographically distant salmonid populations. *Molecular ecology*, **21**, 237–49.

Pascon RC, Miller BL (2000) Morphogenesis in Aspergillus nidulans requires Dopey (DopA), a member of a novel family of leucine ripper-like proteins conserved from yeast to humans. *Molecular Microbiology*, **36**, 1250–1264.

Planes S, Doherty PJ, Bernardi G (2001) Strong genetic divergence among populations of a marine fish with limited dispersal, Acanthochromis

polyacanthus, within the Great Barrier Reef and the Coral Sea. *Evolution*, **55**, 2263–2273.

Pritchard JK, Stephens M, Donnelly P (2000) Inference of population structure using multilocus genotype data. *Genetics*, **155**, 945–959.

Reisser CMO, Beldade R, Bernardi G (2009) Multiple paternity and competition in sympatric congeneric reef fishes, Embiotoca jacksoni and E. lateralis. *Molecular Ecology*, **18**, 1504–10.

Rosenberg NA (2004) DISTRUCT: A program for the graphical display of population structure. *Molecular Ecology Notes*, **4**, 137–138.

Schluter D, Nagel LM (1995) Parallel Speciation by Natural Selection. *The American Naturalist*, **146**, 292–301.

Simpson GG (1944) *Tempo and Mode in Evolution*. Columbia University Press, New York.

Stebbins GL (1950) *Variation and Evolution in Plants*. Columbia University Press, New York.

Tanaka M, Izawa T, Yamate J *et al*. (2014) The VF rat with abnormal myelinogenesis has a mutation in Dopey1. *Glia*, **62**, 1530–1542.

Tarp FH (1952) *Fish Bulletin No. 88. A Revision of the Family Embiotocidae (The Surfperches)*.

Vagelli A, Burford M, Bernardi G (2009) Fine scale dispersal in Banggai Cardinalfish, Pterapogon kauderni, a coral reef species lacking a pelagic larval phase. *Marine Genomics*, **1**, 129–134.

Wainwright PC, Smith WL, Price S a *et al*. (2012) The evolution of pharyngognathy: a phylogenetic and functional appraisal of the pharyngeal jaw key innovation in labroid fishes and beyond. *Systematic biology*, **61**, 1001–27.

Webb TJ (2012) Marine and terrestrial ecology: Unifying concepts, revealing differences. *Trends in Ecology and Evolution*, **27**, 535–541.

Woods R, Schneider D, Winkworth CL, Riley MA, Lenski RE (2006) Tests of parallel molecular evolution in a long-term experiment with Escherichia coli. , **103**, 9107–9112.

Wright S (1978) *Evolution and the genetics of populations*. University of Chicago

Press, Chicago.

Zhang X, Xu R, Zhu B *et al*. (2007) Syne-1 and Syne-2 play crucial roles in myonuclear anchorage and motor neuron innervation. *Development (Cambridge, England)*, **134**, 901–908.

Figure 1. Average SST for December 2013 along the California and northern Baja coastline from the NOAA CoastWatch Program data with sampling sites in Monterey Bay, California, USA (MBA) and Punta Banda, Baja, Mexico (PBA).

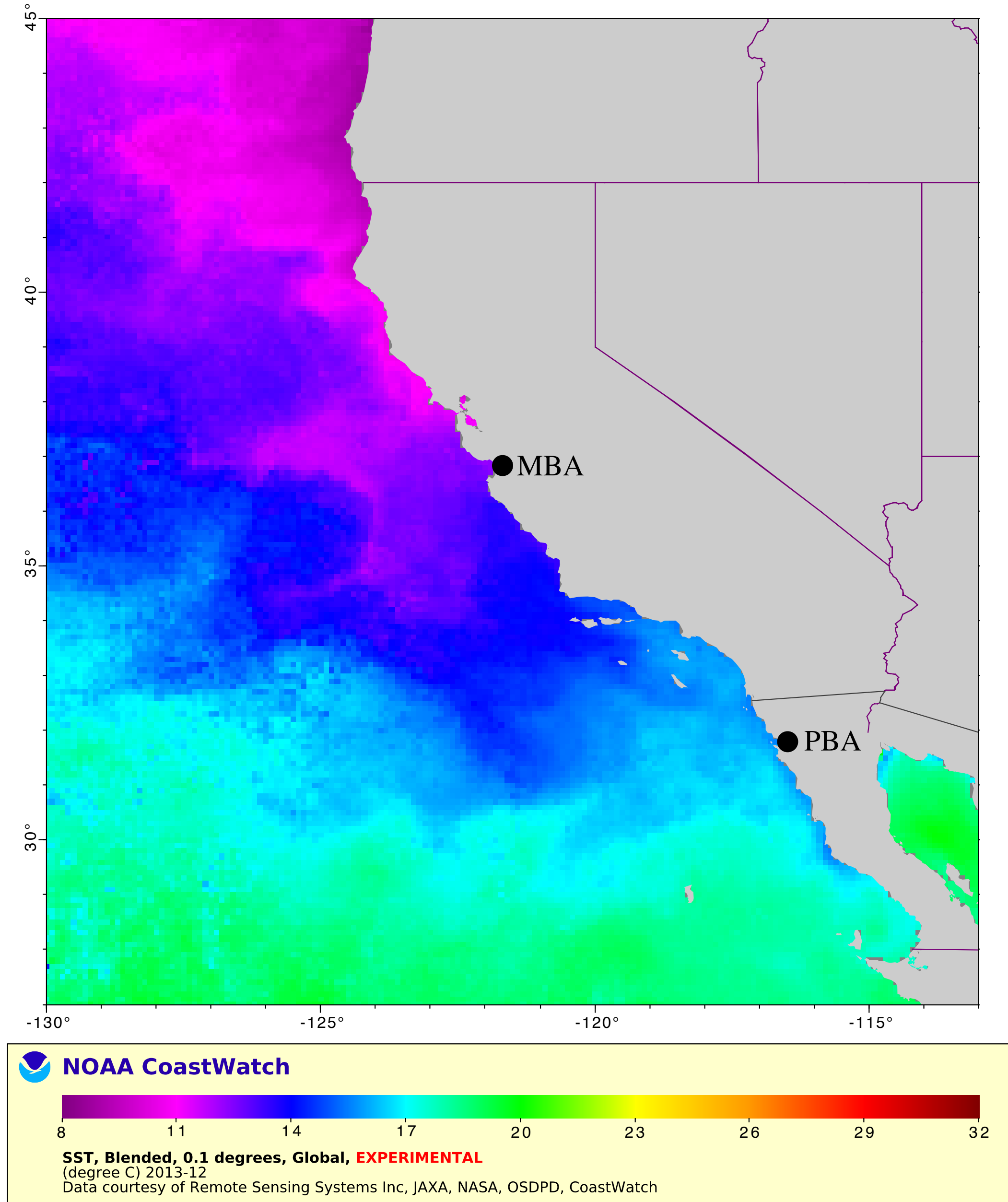


Figure 2. STRUCTURE plots for pairwise population comparisons between Monterey Bay, California, USA (MBA) and Punta Banda, Baja California, Mexico (PBA) in *Embiotoca jacksoni*, *E. lateralis*, *Rhacochilus toxotes*, and *R. vacca* (surfperch illustrations by Val Kells © 2016).

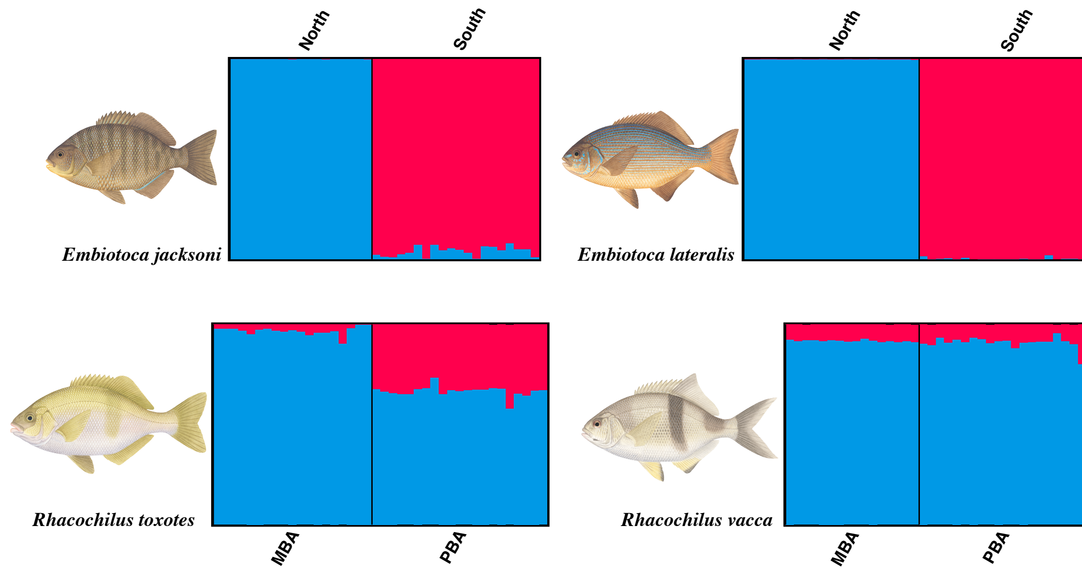


Figure 3. Balancing, divergent, and parallel selection observations in *Embiotoca jacksoni*, *E*. lateralis, *Rhacochilus* toxotes, and *R*. vacca (surfperch illustrations by Val Kells © 2016).

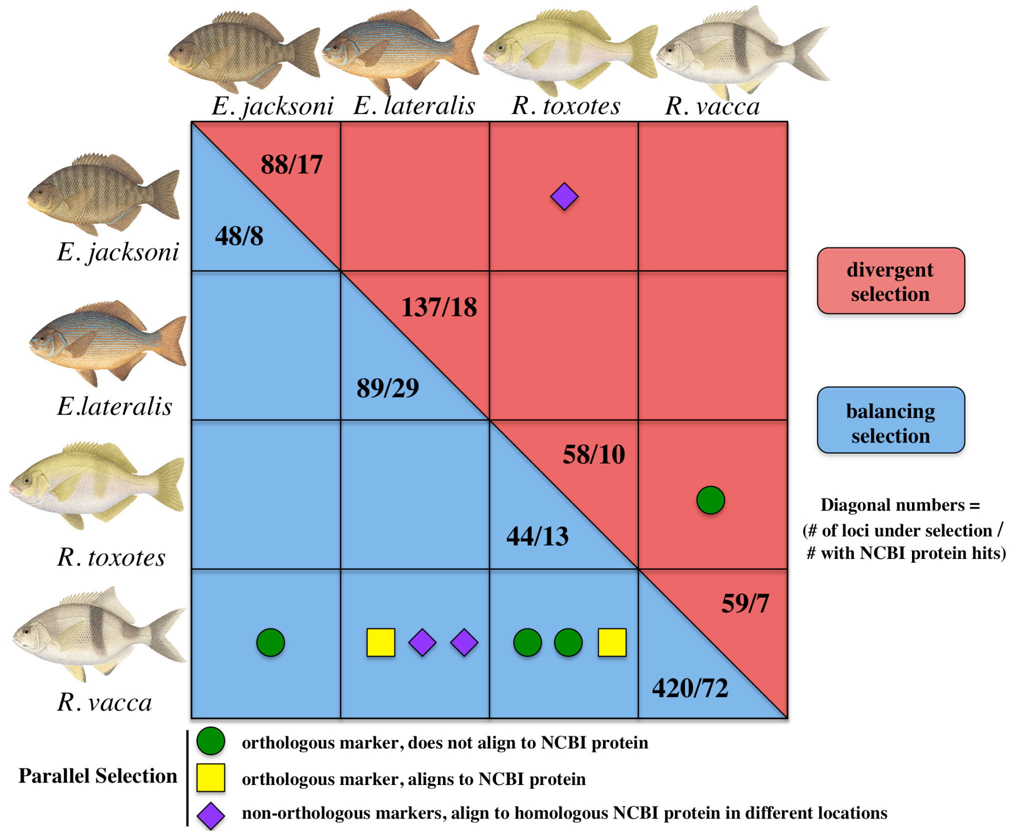


Figure S1. Average SST by month for 2013 along the California and northern Baja coastline from the NOAA CoastWatch Program data with sampling sites in Monterey Bay, California, USA (MBA) and Punta Banda, Baja California, Mexico (PBA).

See supplemental material

Figure S2. Results from FDIST scans for selection using the complete RADseq data set for each species; *Embiotoca jacksoni* (EJA), *E. lateralis* (ELA), *Rhacochilus toxotes* (RTO), and *R. vacca* (RVA). Loci below the 1% quantile (bottom red dotted line) and above the 99% quantile (top red dotted line) were considered candidates for balancing and divergent selection respectively.

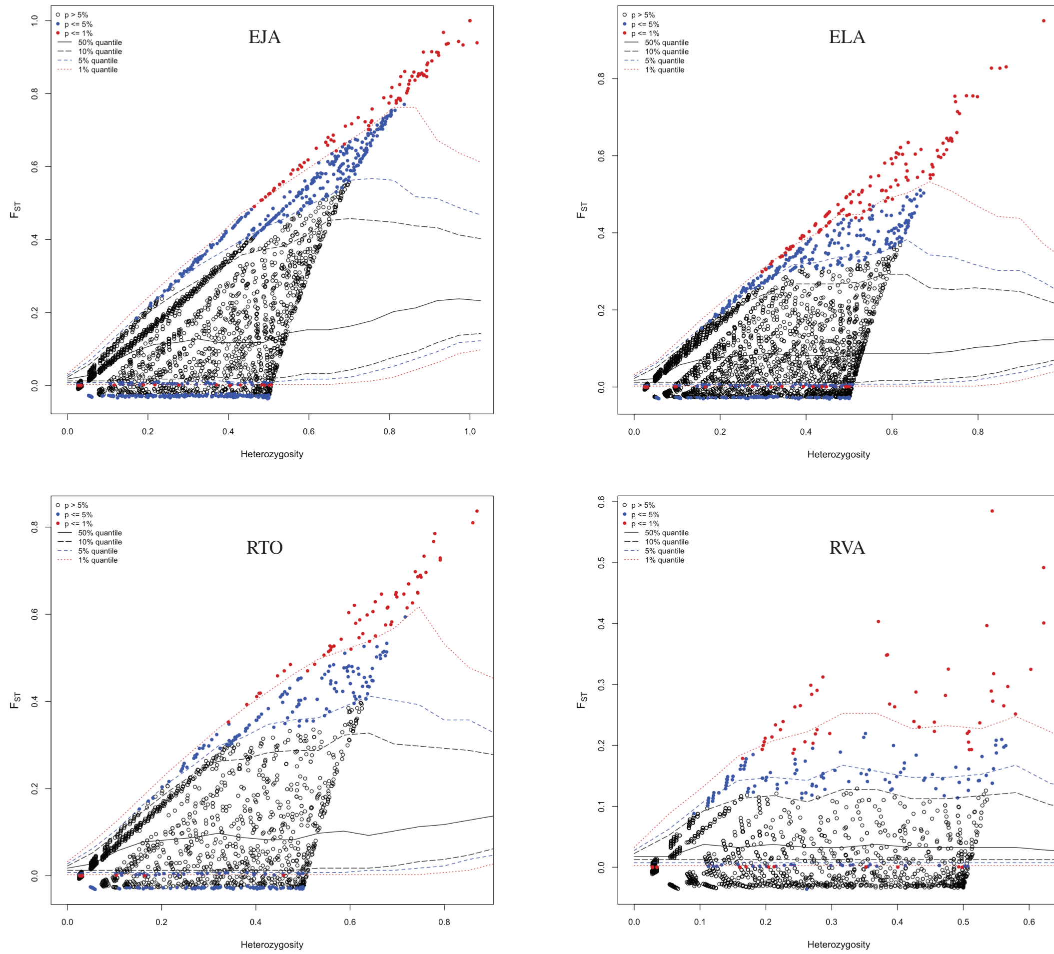


Table 1. RADseq loci counts for total loci, polymorphic loci, polymorphic loci identified through FDIST analyses as undergoing divergent selection and balancing selection for *Embiotoca jacksoni*, *E. lateralis*, *Rhacochilus toxotes*, and *R. vacca*.

Species	Total	Polymorphic	Divergent	Balancing
EJA	25318	8304	88	48
ELA	34252	8728	136	89
RTO	25711	4706	58	44
RVA	19761	5702	59	420

Table 2. Pairwise F_{ST} comparisons between Monterey Bay and Punta Banda for *Embiotoca jacksoni*, *E. lateralis*, *Rhacochilus toxotes,* and *R. vacca*. All p values were highly significant (i.e., every value was 0).

	E. jacksoni	*E. lateralis*	*R. toxotes*	*R. vacca*
F_{ST}	0.21699	0.11441	0.1318	0.02634

Table S1. Sample name, species, location, number of quality filtered reads (QF_reads), number of unique stacks (i.e., total loci count), and number of polymorphic loci for each sample used in the final analyses.

See supplemental material

Table S2. Stacks sequence ID and 80 bp nucleotide sequence for all RADseq loci identified as under balancing or divergent selection, as well as the NCBI alignment score (E value), NCBI protein reference number, and NCBI gene description for each sequence that passed our NCBI alignment filter parameters (E value ≤1e-6) for *Embiotoca jacksoni* (EJA), *E. lateralis* (ELA), *Rhacochilus toxotes* (RTO), and *R. vacca* (RVA).

See supplemental material