UC Merced

UC Merced Electronic Theses and Dissertations

Title

Genomic approaches to identify chromosomal regions responding to divergent selection in multiple Sebastes species (subgenus: Sebastosomus)

Permalink

https://escholarship.org/uc/item/4965j7x6

Author

Martinez, Andres Nathaniel

Publication Date

2014

Peer reviewed|Thesis/dissertation

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

UNIVERSITY OF CALIFORNIA, MERCED

Genomic approaches to identify chromosomal regions responding to divergent selection in multiple Sebastes species (subgenus: *Sebastosomus*)

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

in

Quantitative and Systems Biology

by

Andres Nathaniel Martinez

Chair: Dr. Miriam Barlow
Advisor: Andres Aguilar
Professor: Michael Dawson
Professor: Monica Medina
Professor: Grant Pogson

Copyright
Andres Nathaniel Martinez, 2014
All rights reserved

This dissertation of Andres Nathaniel Martinez is approved and is acceptable in quality and form for publication electronically

_____

Advisor Andres Aguilar

_____

Professor Monica Medina

_____

Professor Michael Dawson

_____

Professor Grant Pogson

_____

Chair: Professor Miriam Barlow

_____

Date

University of California, Merced

2014

DEDICATION

To my little family, Nathaniel and Susana. I love you both very much.

TABLE OF CONTENTS

LIST OF ABBREVIATIONS

ABI	Applied Biosystems Incorporated
AK	Alaska
ATP	Adenosine Triphosphate
BF	Bayes Factor
BSA	Bovine Serum Albumin
CA	California
CAP3	Contig Assembly Program 3
cDNA	Complementary DNA
CDR	Cluster Derived Reference
Chl-a	Chlorophyl a
CI	Confidence Interval
DNA	Deoxyribonucleic Acid
EST	Expressed Sequence Tag
gDNA	Genomic Deoxyribonucleic Acid
HWE	Hardy Weinberg equilibrium
Max-hom	Maximum minor allele frequency in Homozygotes
Min-het	Minimum minor allele frequency in Heterozygotes
NCBI	National Center for Biotechnology Information
NFW	Nucelase Free Water
OR	Oregon
PCR	Polymerase Chain Reaction
PLD	Pelagic Larval Duration
RAD	Restriction site Associated Deoxyribonucleic Acid
SAM	S-Adenosyl-Methionine
SE	Standard Error
SNP	Single Nucleotide Polymorphism
SSR	Simple Sequence Repeat
SST	Sea Surface Temperature
UTR	Untranslated Region
WA	Washington

LIST OF TABLES

LIST OF FIGURES

ACKNOWLEDGEMENTS

I would first like to thank my parents Dolores and Hector for their encouragement and foresight when my thoughts were unclear. I am very lucky to have you as my parents and I could not have gotten this far without understanding that anything that is worth doing is worth doing with love. Thank you for leaving an empty space on your living room wall for my Ph.D. certificate to complete the trilogy. I would like to thank my sister Sabrina who came to my rescue and helped me with Nathaniel while I was writing up the final portions of my dissertation, you helped find me time. To my sister Anna, thank you for breaking up my time as a doctoral candidate with trips to Oakland A's games and making me a very happy uncle. To the Santiago family, thank you for expecting so much from me and reminding me how much of a role model I am to the kids. To Andres Santiago who reminded me that education is the only thing that no man could take away from me. To Susana my beautiful wife and mother of my son, you make me feel that I am much more than what I think I am. You are a strong and inspirational woman, thank you for all of your sacrifices I am glad I went through this with you. To Nathaniel, you have taught me a whole new aspect of love and I hope that one day you ask me to tell you stories about my time as a Ph.D. student.

My experience here at UC Merced couldn't have been possible without the support from Dr. Andres Aguilar. Thank you for taking me on as your graduate student, you have been a role model for me since before I knew what a null allele was. Thank you for providing the guidance and curiosity to continue studying evolution. Most importantly, thank you for letting me know that you were dedicated to seeing me through this dissertation that meant a lot. You always kept your cool even though I was constantly losing mine. To my committee Monica, Mike, and Grant, thank you for agreeing to take the time to assess my work and providing valuable feedback to make it as best as possible. I am very fortunate to have a committee with the amount of experience you all bring. My interactions with you have all been positive and I felt inspired by your words and scientific research. To other UC Merced faculty, I would like to thank Dr. David Ardell who was always there to provide me with support and motivation during my time here. Thank you for inviting me to the arena of professionalism and using any opportunity to teach me how to use my skill set to solve any problems I faced. Thank you to Dr. Miriam Barlow for taking me on as your student in my last year. You made a difficult time much more bearable with your concern and support. Thank you for Dean Juan Meza who provided me with financial support in my last semester, which is a major reason why I completed my dissertation.

To Joseph and Jason, my fellow lab mates. I am glad to have come into the Ph.D. program with you, there was never a dull moment in any aspect. I respect you both as scientists and appreciate you as my friends. Thank you for helping me learn more about evolution, a subject we all enjoy. You made my time as a PhD. student fun, from the time we entered and into the future we will always be known as Aguilar lab. Thank you to the Dawson lab who were always around and willing for a casual or scientific conversation. Thank you for inviting us to your journal club.

I would like to thank the people who contributed tissue samples to my studies: John Hess, David Wagmann, Nate Hall, Scott Blankenship, Martha Burford, and John Carlos Garza. I would like to thank Josh Uolla and John Flicker for their laboratory

assistance. I would also like to thank Chris's Fishing trips and Stagnaro's Sport Fishing for allowing me to work as a deckhand in exchange for access to rockfish carcasses.

VITA

Education

Ph.D.-Quantitative and Systems Biology University of California, Merced. Expected graduation August 2014

Advisor: Dr. Andres Aguilar

- Coursework in systems biology, evolution, molecular ecology, population genetics, and functional genomics

M.S.-Ocean sciences, University of California, Santa Cruz 2008

Advisor: Dr. John Carlos Garza

- Coursework in marine geology, chemical oceanography, physical oceanography, biological oceanography, and evolutionary ecology
- Project: Genomic screen of the *O. mykiss* genome to detect regions under selection using anonymous and EST-linked microsatellite markers

B.S. -Marine biology University of California, Santa Cruz 2005

Advisor: Dr. John Carlos Garza

- Coursework in marine biology, ecology, evolution, biochemistry, marine zoology, and scientific diving
- Project: Measuring microsatellite mutation rate in steelhead trout (*O. mykiss*) from the San Lorenzo River (CA)

Publications

- A.N. Martinez & A. Aguilar. In prep. A microsatellite genome screen to identify the environmental features influencing local adaptation in two species of rockfish from the Northeast Pacific Ocean. Molecular Ecology
- A.N. Martinez & A. Aguilar. 2011. Development of fifty-one novel EST-SSR loci for use in rockfish (genus *Sebastes*). Conservation Genetics Resources. 3: 335-340
- A.N. Martinez, D. Pearse, J.C. Garza. 2011 A microsatellite genome screen identifies chromosomal regions under differential selection in steelhead and rainbow trout (*Oncorhynchus mykiss*). Transactions of the American Fisheries Society 140(3) 829-842.

Awards

- Society for the Advancement of Chicanos and Native Americans in Science (SACNAS) Genome scholars fellowship award
- Smithsonian institution minority internship program- Funding for fellowship at the National Zoo In Washington D.C.
- SACNAS travel fellowship to attend 2006 national conference in Tampa, FL, 2007 travel fellowship in Kansas City MO

Presentations

- Participation award for the 2004 undergraduate research symposium at the University of California, Santa Cruz.
- SSB/SSE grant for minorities attending Evolution 2004 meeting in Ft. Collins, Colorado
- Poster presentation at the 2006 SACNAS national conference in Tampa, Florida
- Poster presentation at the 2006 Coastwide Salmonid Genetics meeting, Santa Cruz, Ca
- Oral presentation at the 2007 SACNAS national conference in Kansas City, MO
- Poster presentation at the 2013 Evolution conference in Snowbird, UT

Research Experience

University of California, Merced CA (August 2008-present)
Advisor: Dr. Andres Aguilar
Studied population genomics and genetics in *Sebastes*, identifying genomic regions involved in adaptive evolution, with EST-linked and anonymous microsatellites. Using identified sequences in comparative genomics to identify molecular pathways affected by selection

Smithsonian Institution/National Zoo, Washington D.C. (May 2005-August2005)
Advisor: Dr. Jesus Maldonado
Participated in minority internship program and contributed to projects which included population genetics of the endangered shrew *Sorex ornatus*. Also established techniques of non-invasive sampling for molecular parentage analysis for captive wild-dog (*Lycaon pictus*) litters. Participated in outreach programs for local elementary school children from local areas educating them on Zoological sciences.

National Marine Fisheries service, Santa Cruz CA (2002-2008)
Advisor: Dr. John Carlos Garza
Laboratory assistant and graduate student, running PCR reactions, electrophoresis gels, sequencing gels, general lab maintenance and inventory, genotyping, identifying rockfish, running various population software programs. Also participated in outreach programs with an outdoor school as a mentor, educating elementary school children from Salinas, Ca. on the importance of maintaining clean waterways and the effects of pollution on salmonid populations.

Teaching experience

Genetics, University of California, Merced. Prof. William Turner. Teaching assistant (2014).
Genome biology University of California, Merced. Prof. Carolin Frank. Teaching assistant (2013).
Cell and Molecular biology, University of California, Merced, Prof. David Ardell. Teaching assistant (2011-2012).
Cell and Molecular biology, University of California, Merced, Prof. Kamal Dulai. Teaching assistant (2011).

Evolution, University of California, Merced, Prof. Andres Aguilar. Teaching/Lab assistant (2010-2011).
Introductory biology, University of California, Merced, Prof. Kamal Dulai. Teaching/Lab assistant (2008-2009).
Ecology and Evolution, University of California, Santa Cruz, Prof. Donald Potts & Giaccomo Bernardi Teaching assistant (2007-2008).

Dissertation Abstract

The role of selection during speciation is a key component towards fully diverged species, yet the processes and time required for genomic divergence to evolve new species remains an active area of research in evolution. A population genomics approach provides a promising means to understanding evolutionary processes influencing genomic divergence. To determine the effects selection has on the genome, I used *Sebastes* as a model system due to its rapid adaptive radiation in the North Pacific Ocean. Species in *Sebastes* inhabit a wide range of environments providing a unique system to study a range of divergence from local adaptation between populations to incipient speciation. I identified genomic regions diverging due to local adaptation and candidate genomic regions involved in the incipient speciation of a lineage within the subgenus *Sebastosomus*. I developed 67 microsatellites from expressed sequence tagged (EST) libraries from three different *Sebastes* species to identify gene function responding to selection. I then implemented a low-density genome scan on a species pair (*S. flavidus* & *S. melanops*) and identified outlier loci with two outlier detection methods. The two species are distributed sympatrically but differ in pelagic larval duration (PLD). *S. flavidus* has a three to four month PLD had a higher number of identified candidate outlier loci while *S. melanops* which has a four to six month PLD had significantly fewer suggesting that gene flow may reduce local adaptation. I identified two outliers in *S. flavidus* associated with oceanic variables influencing genetic structure while none were found in *S. melanops*. Finally, I assembled a restriction site associated (RAD) library and sequenced using the Illumina Hiseq technology from two divergent *S. mystinus* morphotypes. I sampled 21,318 single nucleotide polymorphic (SNP) loci across six populations from Oregon to San Francisco. Population structure analyses identified two morphotypes previously described as genetically distinct groups and highlighted a population maintaining signatures of hybridization. I identified 209 SNP loci diverging between morphotypes providing an initial set of genomic regions diverging due to speciation. My dissertation highlights the importance of integrating novel genomic techniques to address fundamental evolutionary questions that seek to identify genomic responses to selection.

Chapter 1: Introduction

Identifying the location and number of genomic regions responsible for adaptive divergence between populations and recently diverged species provide valuable evolutionary information as to how genomes respond to natural selection (Rogers & Bernatchez 2007; Stinchcombe & Hoekstra 2007; Via & West 2008; Bernatchez et al. 2010). Rapidly changing environments, environmental gradients, or introduction to new environments will relax selection on previously selected regions in the genome while imposing selection on novel regions not previously responsible for adaptation in the local habitat (Schluter 2001; Egan et al. 2008; Schwarz et al. 2009). Understanding the specific genetic responses to selection in local habitats has important evolutionary implications that will reveal population dynamics and the effect of divergence and speciation at the genomic level (Turner et al. 2005; Via & West 2008). Differentiation between recently split populations will have a notable effect on genomic regions involved in local adaptation, which may lead to phenotypic divergence in key behavioral and life history traits (Stinchcombe & Hoekstra 2007; Via & West 2008; Bernatchez et al. 2010). On the other hand, the effects of local adaptation that lead to reproductive isolation and differentiation between populations will have little to no effect at regions not responsible for locally adapted traits leaving similar genomic regions relatively homogeneous between populations (Emelianov et al. 2004; Turner et al. 2005). This mechanism of selection at specific genomic regions is defined as the "genetic mosaic of speciation" (Via & West 2008). The "genetic mosaic of speciation" concept refers to incipient speciation between two reproductively isolated populations, but the early effects of speciation is detectable in two or more ecologically isolated populations. The difference is that the genomic regions involved in early speciation mechanisms that determine population differentiation will be linked to resource use or mate choice (Via & West 2008). These genes are sufficient in causing phenotypic variation and reproductive isolation, both key situations that can lead to speciation.

Increased differentiation within species at a population level due to local adaptation will lead to reproductive isolation, which can segregate populations and initiate early speciation processes. Genomic regions involved in early speciation mechanisms will promote elevated levels of population differentiation and are linked to traits involved with resource use, responses to the local environment, or mate choice (Via & West 2008) these genomic regions can be differentiated from genomic regions not involved in early speciation mechanisms. The process of speciation is not always direct and the effects it has on a genome will target different genes at differential times during the speciation process. It is important to identify genomic regions involved in early differentiation within a species that cause population divergence since genes linked to those regions are responsible for initial reproductive isolating mechanisms, which can establish the circumstances for speciation. Similarly, identifying the genomic regions linked to genes involved in early speciation between species will reveal mechanisms reinforcing the speciation process. A promising approach to identify genomic regions involved in population differentiation due to local adaptation and genes involved in speciation processes is to perform an inter-and intra-species genomic scan from a closely related group of species inhabiting similar but varying habitats. This approach has identified genomic regions responding to selection due to local adaptation in habitats

within species and those genomic regions controlling differentiation and regions involved in speciation.

Identifying genetic signatures causing population differentiation at key genomic regions is facilitated by analyzing the genomes of organisms with notable phenotypic variation (Storz 2005), inhabit contrasting environments (Scotti-Saintagne et al. 2004; Oetjen et al. 2007; Martinez et al. 2011), or inhabit a notable physical gradient (Bonin et al. 2006; Jump et al. 2006). A genomic scan across multiple populations inhabiting differing habitats will reveal genomic regions linked to differentiation resulting from adaptation to local habitats (Schemske 2000; Via 2009). The evolution of reproductive barriers between populations located in contrasting environments may lead to speciation when the genes leading to differentiation are linked to local adaptation, mate choice, or genetic incompatibilities (Butlin 2008). Similarly identifying genes that are involved in early speciation processes between recently diverged sister species will reveal genomic regions that are continuously maintaining differentiation between them. It is of importance to study those species pairs that have recently diverged or are approaching speciation.

Since it is difficult to directly identify which genes are responsible for divergence, the idea of hitchhiking provides identification of genomic areas rather than specific genes under selection. The underlying mechanism by which genetic hitchhiking works is that beneficial mutations or variations in coding regions of the genome are subject to a selective sweep when introduced to a new environment or the environment changes rapidly. The selective sweep will then remove much of the variability of that selected region leaving a signature of reduced variability including the neutral areas surrounding the genomic region that encountered the selective sweep (Schlötterer 2003). The effect of the selective sweep in the neutral linked regions depends highly on several factors. These factors are strength of selection, probability of recombination between the selected gene and the neutral regions, distance from the selected genomic region, initial frequency of the selected genomic region, mutation rate, and time of fixation (Maynard et al. 1974; Kaplan et al. 1989; Stephan 1992; Wiehe et al. 1993; Wiehe 1998). These factors are important to this study because the markers used to sample the *Sebastes* genome are linked to areas of expressed sequences but lie outside of the coding region of the gene. As a result linkage disequilibrium will decrease as physical distance to the selected gene increases.

Certain natural systems are more conducive to directly observing the effects of selection on the genome while other systems introduce a unique challenge in addressing similar questions. Terrestrial systems display physical boundaries that establish reproductive isolation which are more easily observed than in marine systems (Palumbi 1994). Furthermore marine systems are considered “open” systems due to the lack of apparent physical boundaries between populations and high assumed gene flow leading to lack of genetic structure (Palumbi 1994; Berntson & Moran 2009). High amounts of gene flow in marine populations could prevent differentiation at all adaptive and neutral genomic regions reducing population heterogeneity along the species distribution.

Large-scale population divergence across the Northwest pacific coastal oceanographic basin is distinguished by larger gyres and eddies that influence population connectivity and denote dramatic faunal shifts (Kudela et al. 2008, Checkley et al. 2009a,

Checkley et al. 2009b). The Alaska current, California Current and the Southern California eddy yield the distinct species assemblages in each region. Within each region there are smaller mesoscale factors that influence differentiation, such as areas of strong upwelling around prominent bathymetric features and coastal freshwater inputs (Marchesiello et al. 2003). The smaller mesoscale features are significant in spatial and temporal genetic differentiation in fish populations over small scales across kilometers (Doherty et al. 1995; Cowen et al. 2000; Knutsen et al. 2003). Small-scale population divergence influenced by local oceanographic features control the extent of larval release, larval duration and dispersal distance of larvae. Physical boundaries or isolation by distance may influence levels of differentiation over large geographical scales while more recent speciation events include strong ecological factors and maintenance of mesoscale oceanographic boundaries.

Larval fish recruitment has proved to be a dynamic aspect of marine population differentiation in temperate reef fishes and invertebrates. In the case of marine invertebrates, Addison et al. (2008) and Palumbi & Wilson (1990) found that there is no population structure along the California coast using mitochondrial and allozyme markers. In the fish genus *Sebastes* there has been a gradient in population differentiation and connectivity that suggest in some cases strong isolation between populations and other instances populations did not display population structure over the majority of the species range. In one study, where population structure was observed in the Pacific Ocean perch (*S. alutus*), Withler et al. (2001) found that genetic structure was due to local oceanographic currents and eddies that self recruited larvae to the parental habitat. More recent studies on blue rockfish (*Sebastes mystinus*) observed strong breaks among populations that coincided with Cape Mendocino in Northern California (Cope 2004; Burford & Larson 2006; Burford & Bernardi 2008). Cape Mendocino was also seen as a significant genetic break for Vermillion (*Sebastes miniatus*) rockfish but, an additional very strong genetic beak was observed at Point Conception in Southern California (Hyde 2007). More recently the genetic break observed around Cape Menodcino CA was also observed in yellowtail rockfish (*Sebastes flavidus*) (Hess etl al. 2011). There have also been studies that have shown the opposite pattern, where the distribution does not appear to have any detectable structure. Investigations in the population structure for the Mexican rockfish (*Sebastes macdonaldi*), have not shown to be influenced by local oceanographic currents in Southern and Baja California (Bernardi et al. 2003; Rocha-Olivares et al. 2003).

The genus *Sebastes*, provides a unique system in which adaptive divergence due to environmental factors, and its effects on the genome, can be studied from recently diverged species and populations. There are instances of significant spatial and temporal genetic differentiation in fish populations over small scales (Doherty et al. 1995; Cowen et al. 2000; Knutsen et al. 2003). Gilbert-Horvath et al. (2006) found no genetic structure in young-of-year recruits and slight structure in adult populations of *Sebastes atrovirens* along the California coast. *S. atrovirens* belongs to a *Sebastes* subgenus *Pteropodus* characterized by a shortened (3-4 month) pelagic larval phase. Since population connectedness in the marine system is highly dependent on timing of recruitment, pelagic larval duration, and dispersal distance of larvae, the local population dynamics related to boundaries are more sensitive to other factors besides physical divisions between

Sebastes populations. The genus *Sebastes*, specifically *Sebastosomus,* provides a unique challenge and opportunity to study the effects of selection on the genome due to its rapid speciation in the marine environment and protracted generation times (Parker et al. 2000; Love et al. 2002). To assess the presence of adaptive divergence in the marine environment, I will focus on the subgenus *Sebastosomus* which includes five species: widow rockfish (*S. entomelas*), blue rockfish (*S. mystinus*), yellowtail rockfish (*S. flavidus*), olive rockfish (*S. serranoides*), and black rockfish (*S. melanops*) (Hyde & Vetter 2007).

The genus *Sebastes* includes a large number of species of rockfish that inhabit a diverse span of ecological marine niches. In the family *Scorpaenidae*, the genus *Sebastes* represents over 110 species of rockfish where 74 genera reside in the Northeastern boundary of the Pacific Ocean from Alaska to Costa Rica (Love et al. 2002; Berntson & Moran 2008). Surprisingly over 65 species can be found within the temporally and spatially dynamic California current (Bakun 1996). Species in the genus can be pelagic, inhabit the continental shelf past 1000m, or remain exclusive to intertidal or kelp forest environments near the coast (Love et al. 2002). These environments are further separated as some species will reside in soft versus rocky bottoms or in the case of coastal species, bottom dwelling demersal versus mid to top-level swimming.

The *Sebastosomus* subgenus is identified by inhabiting near shore habitats aggregating in large mid-water groups over rocky reefs and macro-algal environments, this feature distinguishes this subgenus from demersal rockfish species within the same coastal environment (Lenarz et al. 1995). *S. flavidus* and *S. melanops* can be found sympatrically throughout their distributions from the Aleutian Islands, AK to Southern California with highest densities from British Colombia Canada to Central California (Love et al. 2002). The species pair can commonly be found together amongst each other forming large groups shallower than 50 meters, but in general *S. melanops* are considered to inhabit shallower water (Laroche & Richardson 1980; Love et al. 2002). Yellowtail rockfish (*S. flavidus*) are categorized as a deeper water species ranging in depth from surface waters to 549 meters but are regularly observed in mid-water above rocky reefs from 90-180 meters from the Aleutian Islands, AK to southern California, having the highest abundance from British Colombia to Northern California (Love et al. 2002). Population analyses in *S. flavidus* have shown moderate structuring across their range with a significant barrier to gene flow around Cape Mendocino CA (Sivasundar &Palumbi 2010; Hess et al. 2011). *S. flavidus* larvae peak parturition time is during the winter months from November to January (Moser 1996) and remain as pelagic larvae for 3-4 months (Love et al. 2002). Larval settlement pulses occur during strong spring upwelling events from April to August (Love et al. 2002; Sivasundar & Palumbi 2010) Juveniles settle in the kelp canopy (Love et al. 2002; Ammann 2004) and gradually move to deeper substrates (Love et al. 2002). Near-shore species in *Sebastomus* include the Black rockfish (*S. melanops*) which occurs between the Aleutian Islands to Southern California, with highest densities from British Colombia in Canada to Central California (Love et al. 2002). Black rockfish can be found in surface waters down to 366 meters but are most commonly found at depths shallower than 55 meters (Laroche & Richardson 1980; Love et al. 2002). They can be found resting along rocky substrates but are frequently seen at mid-water depths among large groups of Yellowtail, and Widow

rockfish (Love et al. 2002). Studies that have attempted to analyze genetic differentiation of *S. melanops* have seen significant genetic breaks at strong physical barriers around the Colombia River plume (Miller et al. 2004) and slight differentiation in coarse scale sampling between populations in Monterey Bay California and Garibaldi Oregon (Sivasundar & Palumbi 2010) but not to the degree observed in *S. flavidus*. *S. melanops* has a parturition time during the months of January to March and remain pelagic for 4-6 months (Bobko and Berkeley 2004). Larval recruitment occurs between the months of May-July (Love et al. 2002; Sivasundar & Palumbi 2010). The Blue rockfish (*S. mystinus*) can be found in the near shore environment inhabiting surface waters along macroalgal assemblages to depths of 549 meters, with highest densities seen at depths between the surface and 90 meters (Love et al. 2002). Blue rockfish have been reported to have the highest levels of genetic differentiation in *Sebastosomus* around Cape Mendocino which has been suggested as incipient speciation as a result of secondary contact of two distinct expanding populations maintained prior to the last glacial maximum approximately 19,000 years ago (Burford & Bernardi 2008). Differences in depth habitats also have a strong effect on diet and mouth morphology, most notably in the Blue rockfish (*Sebastes mystinus*) which are omnivorous as a result of adaptation to picking off small crustaceans off of kelp fronds. The differences in parturition and pelagic larval duration in these closely related species provides a natural system in which I will determine the interaction between local adaptation and gene flow in a marine environment.

Investigations to the possible maintenance of population structure in *S. flavidus* and lower structure in *S. melanops* may be approached by analyzing the pelagic life history stage that can contribute most significantly to population connectivity. Larval parturition, length of the pelagic life phase, and timing of recruitment can all influence the degree of population structure. Selection at these stages favor larval cues and behavior that maximize successful larval recruitment to suitable habitat. By strongly selecting for timing of currents and environmental features that increase probability of survival and recruitment, local adaptation for these cues will significantly influence genes involved in successful recruitment. Identifying the environmental factors driving the divergence of populations at the pelagic larval stage should also associate with pelagic larval duration resulting in reduction of genetic structure for species with longer pelagic larval phases curtailing any effect of localized selective pressures.

Reproductive isolation through larval dispersal is another factor that can induce and maintain differentiation in marine organisms. In the subgenus *Sebastosomus*, larval duration can last anywhere between 3-6 months (Love et al. 2002; Shanks 2009; Sivasundar & Palumbi 2009). Yellowtail rockfish larvae peak parturition time is during the winter months from November to January (Moser 1996) and remain as pelagic larvae for 3-4 months (Love et al. 2002). Larval settlement pulses occur during strong spring upwelling events from April to August (Love et al. 2002; Sivasundar & Palumbi 2010). Juveniles settle in the kelp canopy (Love et al. 2002; Ammann 2004) and gradually move to deeper substrates (Love et al. 2002). Black rockfish (*S. melanops*) have highest parturition from the months of January to March (Bobko & Berkeley 2004) settling in the kelp canopy. Blue rockfish (*S. mystinus*) have parturition times between October and March, but the majority of larval release is between December and January (Love et al.

2002). Larval pelagic retention ranges between 3-5 months and juveniles settle in the kelp canopy where they later descend to deeper depths in the kelp environment.

Investigations in the connectivity and differentiation of marine organisms with a pelagic life phase have produced mixed results. In the case of marine invertebrates Addison et al. (2008) and Palumbi & Wilson (1990) found that there is no population structure along the California coast using mitochondrial and allozyme markers. In *Sebastes* there has been a mixture of population differentiation and connectivity that suggest in some cases strong isolation between populations and other instances populations did not display population structure over the majority of the species range. In one study, where population structure was observed in the Pacific ocean perch (*Sebastes alutus*), Withler et al. (2001) found that genetic structure was due to local oceanographic currents and eddies that self recruited larvae to the parental habitat. More recent studies on blue rockfish (*S. mystinus*) observed strong breaks among populations that coincided with Cape Mendocino in Northern California (Cope 2004; Burford & Larson 2006; Burford & Bernardi 2008). Cape Mendocino was also seen as a significant genetic break for Vermillion (*S. miniatus*) rockfish but, an additional very strong genetic beak was observed at Point Conception in Southern California (Hyde et al.2008). Investigations on the population structure for the Mexican rockfish (*S. macdonaldi*), have not shown to be influenced by local oceanographic currents in Southern and Baja California (Bernardi et al. 2003; Rocha-Olivares et al. 2003). The ecological factors that determine larval duration have a strong effect on larval self-recruitment which has a significant effect on population divergence.

Based on the rapid adaptive radiation and highly variable life history traits that the genus *Sebastes* has undergone, researchers are provided with a unique biological system to study divergence and speciation within the marine environment and its effects on the genome. The dynamic marine system does not always reveal apparent niche differences as seen on terrestrial systems where speciation can be observed based on several reproductive barriers such as habitat preference, mate preference, and reproductive timing. Factors influencing speciation in the terrestrial system can also be observed in the marine system with the use of sea surface current data taken at different locations and times maintaining physical separation between populations (Palumbi 1994; Gilg & Hilbish 2003). A major goal in my dissertation is to identify ecological factors influencing population divergence due to local adaptation and identify divergent genomic regions between an incipient species pair in the subgenus *Sebastosomus*. This will require the combination of identification of genomic regions linked to functionally important genes in conjunction with oceanographic current data identifying possible barriers to gene flow.

Biologists studying non-model organisms now have a wealth of information from next-generation high-throughput sequencing of genomes that have undergone recent divergence or speciation (Rogers et al. 2001; Rexroad et al. 2008; Via & West 2008). The benefits of these resources are crucial when a complete genome is unavailable and provide direction in pursuing future research of potentially interesting genomic regions. The development of markers to study the genome of non-model organisms is rapidly increasing with next generation sequencing technology. These resources provide researchers of non-model organisms with a tremendous number of genomic sequences

that can be mined for variable single nucleotide polymorphisms (SNPs) and microsatellites. In addition, the development of expressed sequence tagged-markers (EST) from EST libraries provide genomic sampling of genes presumably more likely to be identified as being influenced by selection. The genomic coverage of anonymous and EST-linked marker sequences provided by current sequencing technology provides greater coverage of regions linked to adaptive divergence as well as selectively neutral regions not influenced by divergence.

Given the large amounts of sequence data produced by next generation sequencing techniques and the dramatic reduction in costs, researchers studying non-model organisms have the ability of genomic sampling across hundreds-of-thousands of sites all of which may produce SNPs that will assist in identifying genomic regions contributing to population divergence or early speciation. One such method that combines restriction-site-associated DNA (RAD) preparation followed by next-generation sequencing has been recently implemented across non-model organisms (Luikart et al. 2003; Storz 2005; Slate et al. 2009; Hohenlohe et al. 2010; Amores et al. 2011). The development of these RAD-sequencing techniques, have evolved the field of population genetics to population genomics and I may begin to address evolutionary processes on a genomic level for non-model organisms. This dense sampling of the genome will assist in identifying entire genomic regions responding to selection through the identification of large genomic blocks in linkage disequilibrium when compared to the variation measured across the entire genomic dataset. Larger samples of genomic regions will assist in the annotation of the gene by providing increased sequence information to identify homologous sequences in other teleosts with a higher degree of annotation.

My dissertation will attempt to contribute to the understanding of the evolutionary mechanisms shaping the genomic architecture within species across populations and between recently diverged species. The shaping of genomic variance due to adaptive divergence via local adaptation between populations across the species range should reveal the interaction of gene flow and selection in the dynamic marine environment. Descriptions of genomic regions noted as “islands” or “continents” illustrate the mechanism by which specific regions responding to selection diverge in the face of gene flow (Turner et al. 2005; Via & West 2008). According to the ecological theory of adaptive radiation, organisms exposed to variable habitats will produce adaptive shifts to those habitats and populations will diverge for genotypes linked to optimal survival (Mayr 1963; Schluter 2000). In order to identify this detectable genomic divergence within a species the increased sampling of the genome is required and has been a growing trend in the study of population genomics, shifting the regime into the genomics realm. Like population genetics, population genomics consists of the analysis of hundreds-of-thousands of loci representing a more complete picture of an organism’s genome to obtain a better understanding of evolutionary mechanisms such as mutation, genetic drift, gene flow, and selection (Luikart et al. 2003). Fortunately, the application of population genomics to identify genomic regions responding to local adaptation may also be applied to the identification of genomic regions diverging during the early phases of speciation and surpass the initial identification of outliers. Several studies have utilized this technique in two intertidal snail morphotypes (*Littorina saxatilis*) (Wood et al. 2008;

Galindo et al. 2010), pea aphids (*Acyrthosiphon pisum*) on selective host plant association (Via & West 2008), dwarf and normal lake whitefish (*Coregonus* spp.) (Bernatchez et al 2010), phenotypic variants of butterflies (*Heliconius melpomene*) (Ferguson et al. 2010), and marine/freshwater forms of the three-spined stickleback (*Gasterosteus aculeatus*) (Hohenlohe et al. 2010). My research represents an inflection point in the direction of characterizing divergence through population genetics to the increased power and sensitivity of genome-wide sampling to identify the molecular basis of evolution in natural populations.

Chapter 2: Development of fifty-one novel EST-SSR loci for use in rockfish (genus Sebastes)

Abstract

I mined expressed sequence tag libraries from three species of rockfish (*Sebastes caurinus*, *S. goodei*, and *S. rastrelliger*) for microsatellites. A total of 111 novel sequences with repetitive elements were isolated, and I was able to successfully amplify from 67 of these in five different rockfish species. Population level surveys for two species (*S. flavidus* and *S. melanops*) revealed that 51 of these loci could be successfully scored and were polymorphic. 12 loci showed deviations from Hardy-Weinberg equilibrium, indicating the presence of null alleles or duplicated loci. 11 of the loci annotated to the SWISSPROT database. This suite of loci will prove useful in current and ongoing conservation and genomic studies of numerous rockfish species

Introduction

Identifying the location and number of genomic regions responsible for adaptation between diverging populations and recently diverged species provide valuable evolutionary information as to how genomes respond to natural selection in variable environments (Rogers & Bernatchez 2007; Stinchcombe & Hoekstra 2007; Via & West 2008; Bernatchez et al. 2010). Understanding the specific genetic responses to selection in local habitats has important evolutionary implications that will reveal population dynamics and the effect of speciation at the genomic level (Turner et al. 2005; Via & West 2008). Differentiation between recently split populations will have a notable effect on genomic regions involved in local adaptation, which leads to phenotypic divergence in key behavioral and life history traits (Stinchcombe & Hoekstra 2007; Via & West 2008; Bernatchez et al. 2010). On the other hand, the effects of local adaptation that lead to reproductive isolation and differentiation between populations will have little to no effect at regions not responsible for locally adapted traits leaving similar genomic regions relatively homogeneous between populations (Emelianov et al. 2004; Turner et al. 2005). The difference is that the genomic regions involved in early speciation mechanisms that determine population differentiation will be linked to resource use or mate choice (Via & West 2008). These genes are sufficient in causing phenotypic variation and reproductive isolation, both key situations that can lead to population divergence.

One of the first steps in identifying genomic responses to selection to local environmental variables is obtaining population sample sizes from across a broad geographic distribution effectively sampling populations and increasing the variation in environmental variables. Sampling large population sizes while attempting to detect selection on the genome is important because loci responding to selection will only be detected if the sampling size is large enough to reduce the variation due by drift (Luikart et al. 2003). Loci that are detected as being affected by selection are those with larger than the average genome-wide effects contributing to the measured variation (Hedrick 2000). In order to accurately sample genomic variation, genotyping tens to hundreds of loci linked to known genes and anonymously sampled across the genome. Increased degree of sampling presumably across a large proportion of the genome will provide the "neutral" genomic degree of variation which will be used to detect outlier loci. Lastly,

the identification of outlier behavior as detected statistically based on the baseline degree of genomic variation and confirmation using multiple rounds of testing or confirmation of multiple independent statistical methods.

Advances in molecular techniques and genomic data collection have redirected the attention towards determining the effects of selection on key genomic regions coding for adaptive traits in specific environments (Vasemägi et al. 2005; Bonin et al. 2006; Rogers & Bernatchez 2007; Bernatchez et al. 2010). Biologists studying non-model organisms now have a wealth of information in linkage maps, high-throughput sequencing and controlled crosses to identify important speciation genes (Rogers et al. 2007; Rexroad et al. 2008; Via & West 2008). The benefits of these resources are crucial when a complete genome is unavailable and provide direction in pursuing future research on potentially interesting genomic regions. These sequences provide researchers of non-model organisms with a tremendous number of genomic sequences that can be mined for variable single nucleotide polymorphisms (SNPs) and microsatellites (Ellegren 2008; Hudson 2008). Developing a set of genetic markers that densely samples the genome of a non-model organism with limited genomic information or lack of successful cross-species amplification can be expensive in both time and cost, but with the rapid advancement of high-throughput sequencing and lowering of cost, a wealth of genomic information can be rapidly obtained.

When accumulating the genetic resources necessary for determining genomic responses to adaptation, it is necessary to choose markers that can be successfully amplified and will produce genotypes in multiple species. By obtaining large numbers of markers we can observe how selection may influence the same genomic regions in closely related species with varying life histories. Among the many DNA markers used in population genetics, microsatellites have been highly used due to their ease of amplification and rapid genotyping providing population-level variation (Galindo et al. 2010). More specifically the development and use of expressed sequence tag (EST) linked microsatellites are of great utility in implementing questions about gene flow and selection in natural populations of non-model organisms. The discovery of EST-linked microsatellite sets may be created using large EST sequence dataset like those produced with next-generation sequencing projects targeting a specific species or developed from freely public sequence datasets like those submitted to NCBI/Genbank. Since the microsatellites are linked to transcribed regions, functional descriptions of the sequence may provide insight to genes responding to selection and a comparative genomic approach to closely related species with more comprehensive genomic information may reveal other linked genes responding similarly (Toth et al 2007; Dassanayake et al. 2009; Schwarz et al. 2009;). Cross-species amplification within a genus can be easily attained since the development of EST-linked microsatellites are developed from conserved transcribed sequences since they generally reside in the untranslated 5' and 3' UTRs (Primmer 2009).

The genus *Sebastes* is a large group of saltwater fishes that inhabit a diverse span of ecological niches in the marine environment. The genus *Sebastes* represents over 110 species of rockfish where 74 species reside in the Northeastern boundary of the Pacific Ocean from Alaska to Costa Rica (Love et al. 2002; Bernston & Moran 2008). Surprisingly over 65 species can be found within the temporally and spatially dynamic

California current (Bakun 1996). The large number of species and wide range of environments (from demersal continental shelf to coastal kelp forest environments) that *Sebastes* inhabit provide researchers a unique opportunity to study rapid speciation within the marine environment.

In addition to studying evolutionary questions, rockfish along the northwest Pacific are a commercially and recreationally important fishery that should be managed on a per-location per-species basis due to large mixed stocks of species. Stock identification can be facilitated using hypervariable genetic markers such as microsatellites linked to expressed sequences that would identify habitat and species-specific stocks in a fishery. Here I describe a set of expressed sequence tag microsatellites from two independent datasets from three different species of *Sebastes*. This set of microsatellites provides a significant number of expressed sequence tagged molecular markers that can be used in genetic mapping, identification of genomic responses to selection, and the use in marker-assisted selection for future mariculture approaches of fishery stock supplementation.

Methods

I used three independent data sets to develop EST-linked microsatellite primer sequences for the *Sebastes* genus. The first two sets of EST sequences were downloaded from GenBank submissions for two rockfish species: *S. rastrelliger* and *S. caurinus* (Genbank accession numbers given in Table 2.1). Given the close phylogenetic affinity of these two species (Hyde and Vetter 2007) I assembled all ESTs from *S. caurinus* and *S. rastrelliger* with CAP3 (Huang & Madan 1999) using default parameters. The third set of ESTs used for microsatellite discovery was developed from cDNA libraries constructed from mature gonadal tissue (ovary and testes) from *S. goodei* (Aguilar et al. in prep). All *S. goodei* ESTs were mined for microsatellites were previously checked for quality, trimmed, and assembled with CAP3 (Aguilar et al. in prep). Microsatellite searches were preformed with TANDEM REPEATS FINDER v 4.04 (Benson 1999) and primers were designed with PRIMER3 v0.4.0 (Rozen & Skaletsky 2000). All cDNA sequences were annotated using the Blast2go software (Conesa et al. 2005) to identify similarity to any functional expressed sequences in other species.

Microsatellites were amplified using a three-primer amplification which includes a fluorescently-labeled m13 primer as described by Schuelke (2000). The PCR reagent protocol was conducted in a 15µl reaction as follows: 1X ABI buffer, 2.5 mM of $MgCl_2$, 0.4 mM of dNTPs, 2.7 x 10^{-4} mg/ml of BSA, 0.3 µM of reverse primer, 0.3 µM of fluorescently-labeled M13 sequence (5'-CACGACGTTGTAAAACGAC-3') using Applied Biosystems labels (FAM, VIC, NED, PET), 0.07 µM of M13 5' end labeled forward primer, 0.2 units of ABI Taq polymerase, and 5µl of DNA(5-20 ng/µl). A touchdown thermal profile was used to amplify as many microsatellite loci as possible to account for the large range in annealing temperatures. The thermal profile was conducted as follows: An initial 95°C denaturing step for 3 minutes, annealing at 58°C for 2 minutes, and an extension step for 45 seconds. The following touchdown profile was repeated over 8 cycles with a one-degree decrease after each cycle that encompassed the annealing range of all the microsatellite primers (57°-49°). Initial denaturing at 94°C, 57°C annealing step, and a 72°C extension. To fully incorporate the fluorescently-

labeled m13 sequence, a cycle of 94°C for 30 seconds, a 48°C annealing step, and a 72°C extension step for 45 seconds was repeated 35 times followed by a final extension for 5 minutes.

An initial species panel was developed with the three rockfish species: *Sebastes goodei* (N=3; Localities: Monterey Bay [2], Cordell Bank [1]), *Sebastes flavidus* (N=3; Locality: Cape Mendocino [3]), and *Sebastes melanops* (N=2; Locality: Cape Mendocino [2]). The species panel of individuals was used to test the amplification across the genus with 111 novel microsatellite primer pairs developed from three distinct resources. The products were run on an agarose gel for confirmation of amplification and finally run on an ABI3100 automated capillary sequencer and genotyped using the Genemapper software version 4.0 (Applied Biosystems). Each microsatellite was tested for amplification in each species and given a score: 1-best, 2-better, 3-good. The score was based on peak quality, stutter ratio, ease of scoring the genotype, and intensity of amplification. Comments were also added for each successful and unsuccessful amplification denoting any features of the peak morphology, degree of polymorphism, and species-specific amplifications. The marker specific information was later used to optimize amplification in two population samples of *Sebastes*.

An additional set of 48 *S. flavidus* and 48 *S. melanops* samples from North and South of Cape Mendocino, CA were used to assess population level genetic variation and confirm successful amplification of microsatellite primers from the initial species panel. Departure from Hardy-Weinberg equilibrium, allele counts (k), and measured levels of heterozygosity (H_{exp}, H_{obs}) were calculated using the program GENEPOP v4.0 (Rousset 2008). A second thermal profile was used for the amplifications due to the narrow range of annealing temperatures of the previously amplified loci. The thermal profile followed a standard PCR protocol with a 52° annealing step for 30 cycles followed by a final 15 cycle profile with a 48° annealing step and a final extension time for 7 minutes at 72°C

Results

Initial amplification of the 111 primer sets (44 designed from *S. rastrelliger* and *S. caurinus* cDNA GenBank sequences; 67 designed from *S. goodei* EST library) produced 65 scorable genotypes (Table 2.1). Twenty-five of 44 primer pairs from the *S. caurinus S. rasterlliger* EST sequences gave successful amplification while 40 primer pairs out of 67 produced successful amplification products from the *S. goodei* EST library. I only present information from loci that I was able to successfully amplify and genotype and had at least a score of 3 (good). Eleven loci were annotated to the SWISSPROT database. The remaining 40 loci did not show any significant matches. Two of the loci (Sgoo10_911_1 and Sgoo2027_1_ annotated to the same protein (Protein LBH) but were not in linkage disequilibrium.

After omitting fixed loci allele counts for *S. flavidus* and *S. melanops* ranges from 2 to 19 and 2 to 13, respectively. Observed heterozygosity ranged from 0.042 to 1.0 *S. flavidus* and 0.042 to 0.913 in *S. melanops*. Seven microsatellite loci did not conform to Hardy–Weinberg expectations ($P \leq 0.05$) in *S. flavidus*, while five loci did not conform to Hardy–Weinberg expectations ($P \leq 0.05$) in *S. melanops* (Table 2.2). Two primer sets appear to be amplifying duplicated loci (Sgoo690_1 and Sgoo9_832_1) due to the fact all individuals that successfully amplified were heterozygous (Table 2.2).

Discussion

The development of gene associated markers such as these EST-linked microsatellite loci will assist in the identification of genomic regions responding to local adaptation within variable environments across a species range influencing population divergence. Microsatellite loci linked to polymorphisms associated with genes responding to selection across a variable environment provide an initial genomic location to associate environmental features possibly influencing the degree of genomic variation at the respective genes. With the benefit of sampling a larger proportion of the genome, researchers can now incorporate features of the marine environment and implement a seascape genomics approach towards understanding the evolutionary patterns of adaptation in the marine environment (Manel et al. 2003; Galindo et al. 2006; Hansen & Hemmer-Hansen 2007; Selkoe et al. 2008). Ideally, the environmental feature affecting the genomic divergence would be linked to habitat preference, mate preference, or reproductive timing making the association between the genome and environment more apparent.

This set of EST microsatellite loci display a successful degree of cross-species amplification providing valuable genomic samples across the *Sebastes* genome. This study displays the utility of the development of EST-linked microsatellite markers to produce a large dataset measuring genomic divergence within and among populations of *Sebastes*. Since these makers were developed from expressed sequences, the gene product may be identified via comparative genome mapping to closely related species or annotation to previously described databases such as SWISSPROT. The eleven loci that were successfully annotated to the SWISSPROT database display the limits to the approach of using annotation procedures alone to identify genes associated with EST markers. Therefore it is important to utilize other approaches toward identifying the genes of interest in such a study. Once these genomic regions are identified with EST-linked microsatellites in one species, the same makers can be used to genotype populations in other species of *Sebastes* and determine whether the genomic region is responding similarly to the same environmental features.

Another evolutionary valuable use of this set of markers would be the application of divergence among recently diverged species within *Sebastes*. Recent and rapid speciation has been identified previously in *Sebastes* in multiple instances, all at different stages in the speciation process. The blue rockfish (*S. mystinus*) has been described to have two distinct genetic groups caused by a historical vicariant event followed by secondary contact and currently reinforced by a oceanic barrier (Burford & Bernardi 2008). In Vermillion rockfish (*S. miniatus*), cryptic speciation is observed in two distinct depth specific groups (Hyde et al. 2008). In the case of the Rougheye (*S. aleutianus*) rockfish coloration in the light and dark sampled uniformly form an area along the coast of Alaska revealed two distinct groups (Gharrett et al. 2006, 2005). The study on black-and-yellow rockfish (*S. chrysomelas*) and gopher rockfish (*S. carnatus*), revealed that the two are recently diverged and reproductively isolated incipient species (Narum et al. 2004). Populations of an older *Sebastes* lineage in *S. mentella* in the Northern Atlantic has revealed a situation of incipient speciation along a depth gradient (Stefánsson et al. 2009). To date, the species pair *S. chrysomelas* and *S. carnatus* have had further investigation of genomic regions diverging between them by sequencing a 15kb region

around an identified microsatellite displaying elevated divergence between the species. All of these species pairs would benefit from such an analysis and would provide important evolutionary insight to genomic responses involved in early speciation and the genes that show little introgression between diverging groups of *Sebastes*. This provides researchers with a unique opportunity to apply a systemic approach towards answering how genomes are shaped by historical and current evolutionary mechanisms by integrating oceanographic, behavioral, molecular, and genomic datasets.

This is the first set of EST-linked microsatellites in *Sebastes* and shows that there is significant cross-species amplification among other species in the genus. As more genomic data is obtained investigation of this information will provide insight for both evoutionists and those interested in the practical application of the data in management strategies. It is vital that assessments of fishery stocks are conducted rapidly to determine stock identification groups and implement informed fishery management plans. With the rapid production of sequence data across non-model species and whole genome sequencing costs rapidly decreasing, the application of multiple genomic markers in natural populations will provide a rapid understanding of human-induced microevolutionary processes on economically important fishery stocks.

Table 2.1 PCR primers for 51 polymorphic EST-derived microsatellites. Genbank accession numbers, repeat motifs, observed allele size ranges and annotations to the SWISSPROT database are also reported.

Locus / Genbank Acc. no.	Primer set (5'-3')	Motif	Range	Annotation
Sras112 / EW976549.1	F-CGGCCATCTGGTTCTAATC R-CGTGGTCTTAAACACATTGC	$(TG)_2..(TG)_3..(TG)_3..(TG)_6$	152-196	-
Scau306 / GE806111.1	F-TCTCTTTCCCTTCCCTCAAT R-CCATGAAACCGGGTCTATTT	$(AC)_3..(AC)_{11}..(AC)_{10}$	164-176	s-phase kinase-associated protein 1
Sras320 / GE816499.1	F-GCCACTTATTTCAGGCATGT R-CATGGACACAATGAGGGTCT	$(AG)_{12}$	111-121	-
Sras345 / EW984164.1	F-GCCACTTATTTCAGGCATGT R-CATGGACACAATGAGGGTCT	$(TC)_{12}..(TC)_2..(TC)_2……(TC)_3$	111-121	-
Sras413 / EW984981.1	F-GAGTGGTGGTGGGGTTTAAT R-TACCTGCGCAGTAAATGCTT	$(AC)_7..(AC)_3..(AC)_5..(AC)_2..(AC)_7$	170-182	-
Scau506 / GE804468.1	F-AGCCAGATCTGATGATGTCC R-AGCTTGAGAGTGCGACAGAT	$(TC)_{11}..(TC)_4$	151-155	-
Sras1856 / EW982965.1	F-CAAGCTGCACCTTTAACCAC R-GCCCTTAAAGAGCATTAGGC	$(AAG)_{15}$	106-124	-
sras309791 / EW983522.1	F-CCGGCAAGATGATTAGTGAG R-TTGGACTTCAATTGGTGGAT	$(AC)_{14}$	138-138	-
sras309817 / EW983581.1	F-CCTGGACGTCAACCTTTTTA R-CACCGACTACAAAATGCACA	$(AC)_{18}$	103-115	-
sras310745 / EW985747.1	F-GCAAGTTTTGAGCTTTGGTT R-CTACATGGCAATCCCTACCA	$(TGTT)_7$	127-151	-
sras311101 / EW986599.1	F-GAGAATGTTTGGGACAATGC R-CCAGCGTTCAGTCAGAAGAC	$(CA)_{13}$	141-145	
sras311217 / EW986866.1	F-ACTCAGCTGTCTGCATCTCA R-GCAGCATGCATCCTTGTAT	$(CA)_2..(CA)_{11}..(CA)_4$	161-165	-
sras311560 / EW979523.1	F-TGTGGGTATGCGTCAGTATG R-CATGCGTCATTCAAAACTCA	$(TG)_{14}$	142-148	-
sras311860 / GE807252.1	F-ACGGAGGGGAACATAAAGG R-CAGGGAGGATGTAGGATGTG	$(CA)_{12}$	169-177	growth hormone-releasing factor

Sras311899 / EW980302.1	F-CGATCCATGAGTCAATCTCAC R-GGTTGGTGTGTGAATGTGTG	$(AC)_{18}$	160-170	-
sras312105 / EW980788.1	F-GAAGGATTTACGACATGACCA R-CATTTCGGTAGAGCGTCTGT	$(TTC)_{10}$	133-148	zinc finger five domain-containing protein 16
sras312558 / EW980263.1	F-ATTTTCATCCTGTTTGTCCA R-TTAAAGCCAGTCCCTGAAAA	$(TG)_{13}$	117-119	-
scau356962 / GE799327.1	F-CGTGTGTTTTATATGTGGCATT R-AAACGCCAACTTCATGACTG	$(TG)_{21}$	174-192	-
scau357863 / GE800773.1	F-ACTAATGCATGAAGGCCACT R-ATCAGTAGAGGGCAGCAATG	$(TTC)_{18}$	90-135	-
scau359573 / GE803879.1	F-TGCCTATACAACCTGCTGCT R-TCATATCGCCAAATTCTTGTT	$(TCTA)_{10}$	122-154	-
scau361657 / GE802559.1	F-AAGGTGTGACGACGACTACAA R-CGGATGTGCCTCTCATAAAC	$(TG)_{12}$	127-127	-
Sgoo3582a/HS573393	F-GAAGATGGTTTGTGCCACATT R-CCCTGTGATCTGTCTGTCTGTC	$(ACAG)_{28}$	127-139	-
Sgoo1801a/HS573394	F-CAGGGACTTGGTGTCACAGTT R-CTTTTCCTGTGCATTCTTGGA	$(AGAT)_{24}$	177-241	
Sgoo1193_3/HS573418	F-CAGTTCAAACATGCCACAACA R-CCTCTGTAAGCAACCATCAGC	$(AC)_{20}$	169-173	-
Sgoo690_1/HS573395	F-TGAAACAACAACACACTCCTGA R-CTCTAAAGGACTCCCCTGTGC	$(AC)_{23}$	68-108	-
Sgoo930_1/HS573419	F-GGCTGTGTGATCGTATCTGCT R-CTCCTTTATGGTGGATGCTGA	$(ACTG)_{10}$	106-178	Synaptobrevin homolog YKT6
Sgoo5909_1/HS573396	F-CCGCACATGTTACAACAACAA R-CTCATCCTCACCATGGCTTG	$(AGC)_{11}$	109-125	-
Sgoo3574_1/HS573397	F-GATTCAGTTGAGAACAAAGAGC R-CTTGATTAAGCCCCAATGTT	$(AC)_{17}$	115-131	-
Sgoo5968_1/HS573398	F-TCAACACCCTATTGAAGCACA R-ACTCCTTTGCATTAGGGGAGA	$(AGC)_{10}$	114-117	Protein Tob1
Sgoo10_911_1/HS573399	F-GTTTACGTGCTGTGCTGCTTT R-AATCTGCTGATTTAGGGGAATC	$(AAAG)_{9}$	107-107	Protein LBH

Sgoo1190_1/HS573420	F-CGTCCAGAGTGAAGAGTTGCT R-TGAAGCAGACATGAACCAGAA	$(ATC)_9$	127-235	Src substrate cortactin
Sgoo3931_1/HS573400	F-GACTGTTTCTGGATGTGAGACTG R-GCTGAGTTGGACCCTTCTCTC	$(AC)_{13}$	133-147	-
Sgoo6768_1/HS573401	F-GGGTTCCTTGTTCTCTTGGTT R-TTCTCGTGAAGGTTCTGGTTC	$(AAT)_9$	114-135	-
Sgoo642_1/HS573402	F-ATCTTTGCCAAGGCCGTAAT R-GGAGGTTATGTTTTCGGTTCG	$(AAT)_9$	172-187	-
Sgoo2027_1/HS573403	F-CAGACTTTCAGCCACTTTTGG R-ACACCTGCTGTCAAATCTGCT	$(AAAG)_7$	133-145	Protein LBH
Sgoo5877_1/HS573404	F-CGGCCAAACCCATATTCTTAT R-GGCGGGGACAGTATTCAGTTA	$(AAT)_9$	103-105	-
Sgoo583_1/HS573421	F-CAGGAACAGGTTAAACCAGCA R-AGTAAAGAATGGCGTCCAGGT	$(AG)_{13}$	93-109	-
Sgoo2408_1/HS573405	F-GCAGCAAAGGGAAGAAAAGAG R-GCTGTAGAGTCAAACCGCAGA	$(AG)_{13}$	167-209	Acyl-CoA desaturase
Sgoo7119_1/HS573406	F-GCTGACTGTGTGCACCAACTA R-CAAGCTGGCACCTTACTTGAG	$(AC)_{13}$	167-171	Synaptotagmin-like protein 1
Sgoo4874_1/HS573407	F-GCAACCAACAGAAGAACAAGC R-GCTGTGGGTTCCGTTCTTATT	$(ATC)_9$	149-164	-
Sgoo4_308_1/HS573408	F-CATGGCGTACAACCAGGAATA R-GGGGTTTCTCTCCATCAGAAG	$(AC)_{16}$	175-185	-
Sgoo9_832_1/HS573409	F-AAGACCTGAAGGAGCTCAAGG R-GTGGTTGAGGGTGTTTCAGAC	$(AGG)_{11}$	125-138	Semaphorin-3F
Sgoo1765_1/HS573410	F-TGCTTAATACCATGGGACAGG R-TCACAACACCAATGCAGACTT	$(AAT)_{11}$	137-137	-
Sgoo8_703_1a/HS573411	F-GTGAAGCCCTCAAACTTTCAA R-CTGTACAAGTAGCATCGGATTGTT	$(AAT)_{13}$	79-85	-
Sgoo2916_1a/HS573412	F-CTCTTGGCTGTTTTGTTGCTC R-CTCTCTCGTGCTTCACCACTC	$(AC)_{12}$	124-127	-
Sgoo1842_1a/HS573413	F-AGGAGAGAGCCAGGGATCTAA R-GAACACTGCCCATGATGTTG	$(AC)_{12}$	139-161	-
Sgoo742_1/HS573422	F-GCTGACACAAGTTTGTATGTGTTG	$(AC)_{15}$	157-171	-

	R-CTTTTCTGTGTCAAGGCCACT			
Sgoo3644_1/HS573414	F-TTCGCTGTGCAGTCAACTAGA R-GACCTCAGGAGTGGTGATTGA	$(AC)_{15}$	135-139	-
Sgoo4_310_1/HS573415	F-GACACTCCAGGACAGAGGACA R-TGACTGCTAGAGGAAGCAGGA	$(AC)_{11}$	148-160	-
Sgoo3699_1/HS573416	F-CTACATAGAGCAGCTTCATCGTC R-CGTTGTTTTGTCCACCGTTAT	$(AT)_{10}$	105-107	-
Sgoo813_1/HS573417	F-TTACAGCTCCAGTGGCTTCTC R-GCTACATGCACCCATGGTATT	$(AC)_{18}$	118-120	-

Table 2.2 Sample size (N), observed number of alleles (k), observed (H_o) and expected (H_e) heterozygosity for two populations of *S. flavidus* and *S. melanops*.

	S. flavidus-North				*S. flavidus*-South				*S. melanops*-North				*S. melanops*-South			
Locus	**N**	**k**	**H_o**	**H_e**	**N**	**k**	**H_o**	**H_e**	**N**	**k**	**H_o**	**H_e**	**N**	**k**	**H_o**	**H_e**
Scau306	23	6	0.565	0.637	24	4	0.542	0.568	23	5	0.609	0.58	24	7	0.708	0.683
Scau356962	24	3	0.208	0.294	22	3	0.409	0.413	23	2	0.522	0.502	23	2	0.348	0.502
Scau357863	20	4	0.150*	0.317*	23	6	0.334*	0.552*	23	5	0.478	0.454	21	5	0.667	0.548
Scau359573	24	9	0.75	0.848	23	7	0.826	0.781	23	5	0.87	0.819	21	5	0.81	0.803
Scau361657	24	1	0	0	23	1	0	0	24	2	0.042	0.042	24	2	0.167	0.156
Sras112	24	11	0.875	0.748	24	12	0.917	0.854	23	2	0.087	0.085	24	3	0.167	0.159
Sras320	23	5	0.783	0.689	23	5	0.609	0.682	24	9	0.667	0.707	22	6	0.591	0.617
Sras345	24	5	0.792	0.697	24	6	0.625	0.688	24	9	0.667	0.707	24	6	0.583	0.638
Sras413	23	4	0.652	0.596	24	6	0.667	0.606	24	4	0.417	0.48	24	5	0.583	0.567
Sras506	21	2	0.524	0.512	19	3	0.526	0.522	24	2	0.167	0.284	24	2	0.25	0.383
Sras1856	24	6	0.708	0.745	24	5	0.958	0.772	22	8	0.682	0.79	22	5	0.682	0.714
Sras309791	20	1	0	0	23	1	0	0	23	4	0.522	0.637	24	3	0.25	0.26
Sras309817	23	4	0.304	0.278	23	3	0.174	0.166	24	7	0.708	0.661	24	5	0.667	0.626
Sras310745	23	7	0.739	0.608	24	9	0.708	0.657	23	5	0.565	0.542	22	5	0.727	0.623
Sras311101	21	3	0.143	0.138	22	3	0.227	0.21	23	3	0.652	0.498	23	4	0.652	0.559
Sras311217	22	2	0.091	0.089	23	3	0.565	0.492	23	2	0.13	0.125	17	1	0	0
Sras311560	23	3	0.435	0.426	23	2	0.522	0.487	19	4	0.368	0.368	23	2	0.087	0.085
Sras311860	23	3	0.478	0.474	24	4	0.417	0.444	24	3	0.25	0.231	23	3	0.174	0.166
Sras311899	23	3	0.522	0.463	22	2	0.318	0.333	22	4	0.5	0.576	15	5	0.533	0.676
Sras312105	24	3	0.125	0.121	24	4	0.208	0.197	23	4	0.304*	0.530*	24	4	0.542*	0.621*
Sras312558	24	1	0	0	24	2	0.125	0.12	23	2	0.217	0.198	24	3	0.167	0.159
Sgoo3582a	23	4	0.783	0.666	23	3	0.739	0.68	24	3	0.0833	0.082	22	3	0.091	0.09
Sgoo1801a	24	9	1	0.887	22	11	0.864	0.904	24	9	0.708	0.814	23	9	0.913	0.853
Sgoo1193_3	22	3	0.273	0.246	22	2	0.091	0.169	18	2	0.056	0.056	24	3	0.25	0.29
Sgoo690_1[a]	24	2	1*	0.511*	24	2	1*	0.511*	–	–	–	–	–	–	–	–
Sgoo930_1	24	12	0.875	0.905	24	12	0.833	0.871	–	–	–	–	–	–	–	–
Sgoo5909_1	23	5	0.609	0.515	23	2	0.609	0.407	24	2	0.458	0.403	22	3	0.318	0.369
Sgoo3574_1	23	4	0.217	0.206	23	6	0.435	0.415	22	7	0.636*	0.781*	21	8	0.714	0.711
Sgoo5968_1	21	2	0.048	0.048	21	1	0	0	24	1	0	0	24	1	0	0
Sgoo10_911_1	22	1	0	0	23	1	0	0	24	1	0	0	24	2	0.042	0.042

Sgoo1190_1	23	19	0.957	0.935	22	11	0.773	0.617	–	–	–	–	–	–	–	–
Sgoo3931_1	22	4	0.455	0.563	24	7	0.833	0.713	24	7	0.667	0.681	23	7	0.739	0.729
Sgoo6768_1	22	4	0.227*	0.325*	20	6	0.4*	0.7*	–	–	–	–	–	–	–	–
Sgoo642_1	23	3	0.478	0.498	22	4	0.5	0.536	19	2	0.263	0.235	24	2	0.125	0.12
Sgoo2027_1	20	1	0	0	23	2	0.043	0.043	24	1	0	0	24	3	0.042*	0.121*
Sgoo5877_1	23	2	0.087	0.085	23	2	0.043	0.043	24	2	0.083*	*0.422	24	3	0.042*	0.26*
Sgoo583_1	17	6	0.588*	0.77*	20	6	0.6	0.715	24	6	0.792	0.736	23	5	0.652	0.636
Sgoo2408_1	20	14	0.9	0.912	23	13	0.826	0.877	22	15	0.864	0.928	18	12	0.778	0.778
Sgoo7119_1	22	3	0.455	0.935	21	3	0.667	0.617	–	–	–	–	–	–	–	–
Sgoo4874_1	23	4	0.696	0.596	22	5	0.545	0.5	19	5	0.158*	0.509*	24	4	0.208	0.197
Sgoo4_308_1	24	4	0.25	0.233	24	4	0.375	0.438	23	6	0.696	0.703	23	6	0.783	0.696
Sgoo9_832_1[a]	23	2	1*	0.511*	21	2	1*	0.512*	–	–	–	–	–	–	–	–
Sgoo1765_1	24	1	0	0	23	1	0	0	24	3	0.125	0.121	24	1	0	0
Sgoo8_703_1a	21	2	0.143*	0.438*	24	3	0.083*	0.494*	–	–	–	–	–	–	–	–
Sgoo2916_1a	24	2	0.042	0.042	24	2	0.083	0.082	24	2	0.083	0.082	22	2	0.045	0.045
Sgoo1842_1a	16	6	0.625	0.673	24	7	0.625	0.627	–	–	–	–	–	–	–	–
Sgoo742_1	24	3	0.167	0.159	23	3	0.174	0.165	24	2	0.208*	0.403*	24	3	0.458	0.434
Sgoo3644_1	24	2	0.417	0.422	21	3	0.619	0.452	24	2	0.25	0.383	24	2	0.25	0.284
Sgoo4_310_1	21	6	0.571*	0.741*	21	5	0.333*	0.621*	24	3	0.292	0.37	24	3	0.458	0.393
Sgoo3699_1	24	2	0.042	0.042	23	2	0.261	0.232	22	2	0.045	0.045	22	3	0.273	0.246
Sgoo813_1	24	2	0.458	0.488	21	2	0.524	0.494	24	2	0.25	0.223	24	2	0.083	0.082

*: significant deviation from HWE (P<0.05)

-: indicates no data was obtainable for this locus (either no amplification or locus was unscorable)

a: indicates a possible duplicated locus

Chapter 3: A microsatellite genome screen to identify the environmental features influencing local adaptation in two species of rockfish from the Northeast Pacific Ocean

Abstract

Marine organisms with pelagic life phases, on the scale of months, have a greater potential for increased levels of gene flow and reduced ability for local adaptation. Likewise, organisms with shorter pelagic life phases will have relatively lower levels of gene flow and display a higher proportion of genes adapted to local conditions. Selection at genomic regions associated with local adaptation in a marine species should display an elevated level of divergence when compared to the neutral background of genomic regions not responding to selection. I implemented a genome scan using two outlier detection methods, BayeScan and BayesFst, to determine how differences in pelagic larval duration affect local adaptation and gene flow in two sister-species in *Sebastes* (*S. flavidus* and *S. melanops*). A set expressed sequence tagged (EST) linked and anonymous microsatellite were amplified in *S. flavidus* and *S. melanops*. I identified candidate outlier loci in each data set and associated environmental effects on genetic structure. I found that outliers in *S. flavidus*, which has a shorter pelagic larval duration, produced a higher number of candidate outlier loci in contrast to *S. melanops* which has a longer pelagic larval duration across the same distribution. Measures of divergence (F_{ST} and G'st) were higher in *S. flavidus* than in *S. melanops,* but the differences were not statistically significant. Associations to environmental variables were only detected for three outlier loci in *S. flavidus* . These results empirically demonstrate the impact that gene flow can have on local adaptation in the marine realm.

Introduction

Marine organisms with protracted pelagic life phases are generally believed to have little genetic structure across their distributed ranges, while those with relatively shorter pelagic life phases can have higher genetic structure (Shanks 2009, Palumbi 1994). Understanding the factors influencing population divergence across marine environments is essential to revealing the processes that lead to contemporary patterns of genetic structure. Dispersal in the marine environment via pelagic larval stages is a prominent factor that can shape genetic structure on the scale of kilometers to hundreds of kilometers in the California current system in the north east Pacific Ocean (Burton 1998, Dawson 2001 & 2002, Sotka et al. 2004, Kelly & Palumbi 2010). Given the large variance in pelagic larval duration (PLD) and dispersal distances in marine species, alternative assessments of population connectivity are necessary to identify how divergence between populations arise in the face of gene flow within the marine environment (Treml et al. 2012).

Divergence of populations distributed across heterogeneous environments may lead to local adaptation in fitness related traits (Via 2001; Schluter 2001; Kawecki & Ebert 2004; Via & West 2008). Local adaptation in respective environments will confer higher fitness in the face of gene flow between different environments if selection is significant across the species distribution (Lenormand 2002; Sotka et al. 2004; Pespeni et al 2011; De Wit & Palumbi 2012). In some marine systems, divergence has been observed when gene flow is high and identification of the environmental factors influencing the divergence have been informative (Sotka et al. 2004; Sotka 2005; Sanford & Kelly 2010). Analyses of fish populations in the Atlantic Ocean have suggested evidence for local adaptation across a diverse range of habitats with some associations to environmental factors such as depth, salinity, and variation in temperature (Hemmer-Hansen et al. 2007; Gaggiotti et al. 2009; Neilsen et al. 2009; White et al. 2010). Considering that these environmental features may be highly correlated in the marine environment, approaches toward identifying the specific factor(s) influencing local adaptation leading to population divergence would require sampling populations across variable habitats which would increase divergence for locally adapted traits (Kawecki & Ebert 2004; Pespeni & Palumbi 2013).

Identifying the imprints of selection on the genetic level in response to differences in environmental features across the distribution of a species is an important aspect to understanding the genome-environment interaction. Advances in molecular techniques and genomic data collection have redirected the attention towards determining the effects of selection on key genomic regions coding for adaptive traits in specific environments (Vasemägi et al. 2005; Bonin et al. 2006; Rogers & Bernatchez 2007; Bernatchez et al. 2010; De Wit & Palumbi 2012). These resources provide researchers of non-model organisms with a tremendous number of genomic sequences that can be mined for variable single nucleotide polymorphisms (SNPs) and microsatellite loci potentially linked to genes under selection. Genomic regions involved in local adaptation will display elevated levels of divergence in response to selection and reduced gene flow among populations while the rest of the sampled genomic regions will show weak to no divergence (Via and West 2008). One approach that can increase the utility of genome

scans in non-model organisms is the use of expressed sequence tag (EST) linked microsatellite markers in conjunction with anonymous microsatellites as a method to directly test for divergence in known EST-linked sequences versus a neutral background provided by the anonymous markers (Galindo et al. 2010; Martinez et al. 2011; Vasemägi et al. 2012, 2005). The benefit of including genomic markers linked to known expressed sequences is that if a microsatellite marker is identified as an outlier, the sequence used to develop the microsatellite can then be used to identify potential gene candidates containing the microsatellite.

The genus *Sebastes*, specifically the *Sebastosomus* subgenus provides an opportunity to study the effects of selection on the genome due to the rapid speciation of the genus in the marine environment and protracted generation times (Parker et al. 2000; Love et al. 2002). Within *Sebastes* ranges in PLD have supported the correlation between pelagic larval duration and elevated genetic structure across multiple species, (Buonaccorsi et al. 2002; 2004 Burford & Bernardi 2008; Burford 2009; Hess et al 2011) but other studies report little to no structure across similar ranges (Miller et al. 2004; Rocha-Olivares et al. 1999, 2003). Near-shore *Sebastes* species that display a shorter pelagic larval duration or remain in slow moving water masses have displayed higher genetic structure across their ranges, but this is not always the case (Narum et al. 2004; Johansson et al. 2008; Sivasundar & Palumbi 2010). Gilbert-Horvath et al. (2006) found no genetic structure in young-of-year recruits and slight structure in adult populations of *S. atrovirens* along the California coast suggesting that the balance between gene flow during the pelagic larval phase may provide important clues towards understanding how the genetic structure of marine organisms is shaped across their distribution. Since genetic structure in the marine systems for organisms that have a pelagic life phase is highly dependent on PLD, population dynamics due to physical oceanographic boundaries are more sensitive to additional factors such as local adaptation besides physical divisions between *Sebastes* populations. To assess the interaction between gene flow and environmental features influencing local adaptation shaping population divergence in the marine environment, I will focus on two species with contrasting PLD within the subgenus *Sebastosomus*: *S. flavidus* (yellowtail rockfish) and *S. melanops* (black rockfish). These differences in pelagic larval duration provide a natural system in which I will assess the interaction between gene flow via larval dispersal and local adaptation in marine environment.

The first goal in this study is to identify the relationship between PLD and gene flow across a similar range distribution for a pair of sister-species in *Sebastosomus*. *Sebastes flavidus* (3-4 month PLD) should exhibit elevated genetic structure relative to *S. melanops,* which has a longer pelagic life phase (4-6 month PLD). My second goal is to determine whether differences in gene flow influences genetic signatures of local adaptation. If elevated levels of gene flow exist due to longer PLD, then the genetic signatures of local adaptation will be stifled by high migration rates. Conversely, shorter PLD, or barriers to gene flow, should result in the increased potential for local adaptation resulting increased F_{ST} outliers. In this study I utilize a set of anonymous and EST-linked microsatellites to determine the influence of PLD and gene flow has on population genetic structure. I will also conduct a genome scan using two outlier detection methods (Beaumont and Balding 2004; Foll & Gaggiotti 2008) and identify candidate outlier

makers potentially diverging due to local adaptation in *S. flavidus* and *S. melanops* across their range. Finally I attempt to associate candidate outlier loci with oceanic factors believed to influence population genetic structure and divergence from the results of the outlier detection methods between species.

Materials and Methods

Sampling

I collected and obtained fin clips from 144 *S. flavidus* individuals from six locations and 168 *S. melanops* individuals from seven locations (Figure 3.1). Samples were collected using hook and line sampling and salvaged carcasses from recreational vessels. Tissue samples were either air dried on filter paper or placed directly in 95% ethanol. The fine-scale sampled locations across Cape Mendocino CA represent previously identified genetic break in *S. flavidus* (Hess et al. 2011) and were collected to investigate the effects pelagic larval duration has on population connectivity between *S. flavidus* and *S. melanops*.

DNA extraction and amplification

Total DNA was extracted using the DNeasy Blood and Tissue DNA extraction kit (Qiagen Inc.) and quantified on a Nanodrop 1000. I amplified two sets of microsatellite markers. The first set of amplified markers consisted of previously published anonymous microsatellites (Miller et al. 1999; Roques et al. 1999; Sekino 1999; Wimberger et al. 1999; Gomez-Urchida 2003; Westerman et al. 2005; Yoshida et al, 2005; Suck An et al. 2009) and the second set of markers were amplified from a set of microsatellites isolated from ESTs (Martinez & Aguilar 2011). The six sampled *S.flavidus* populations were genotyped with twenty eight anonymous and fifty seven EST-linked. The seven *S.melanops* populations were genotyped with forty-six EST linked and twenty-eight anonymous microsatellites. Due to the identification and annotation of these markers, they are believed to have a higher probability of experiencing selective pressures since they are linked to expressed genes.

I utilized the approach of Schuelke (2000) to amplify and label microsatellite loci. Reactions were run in 15μL volumes with the following reagent concentrations: 1X buffer (ABI), 2.5mM $MgCl_2$, 0.4 mM dNTPs, 2.7 x 10^{-4} mg/ml BSA, 0.3μM reverse primer, 0.3μM fluorescently labeled M13 primer, 0.07μM M13 5' end labeled forward primer, 0.2 units Taq polymerase (ABI), and 5μL of extracted DNA (5-20ng/μL). To incorporate the shorter M13 labeled forward primer, I used the following modified thermal profile: 95˚C for 3 min, 10 cycles of 94^{O}C for 30 s, 58^{O}C for 45 s (dropping 1^{O}C each cycle, 72^{O}C for 45 s. The final cycle incorporated the forward M13 primer: 35 cycles at 94^{O}C for 30 s, 48^{O}C for 30 s, and a 72^{O}C for 45 s and a final extension of 72^{O}C for 7 min. Amplified PCR products were run on an ABI 3100 automated sequencer with the LIZ500 size standard. Fragment size analysis and genotypes were scored with the GENEMAPPER v4.0 (ABI) software.

Within Population

Tests for Hardy-Weinberg equilibrium for each locus in each population were calculated using the program GENEPOP v4.0 (Rousset 2008). GENEPOP v4.0 was also used to calculate observed heterozygosity (H_{obs}), expected heterozygosity (H_{exp}), and numbers of alleles (A). Allelic richness (A_R) was calculated using the rarefaction method to account for variance in sample sizes using the program HP-rare (Kalinowski 2005). The values used to calculate allelic richness were six for *S. flavidus* populations and 18

for *S. melanops* populations. Linkage disequilibrium was measured for each population according to the Cockerham and Weir (1978) method as implemented by GENEPOP v4.0 to determine the allelic association between all pairs of loci. Due to the high number of pair-wise comparisons between loci, I applied Holm-Bonferroni evaluation to determine the significance of p-values (Holm 1979).

Among Population

To test for population divergence between all populations, within species F_{ST} was calculated with GENEPOP v4.0. Comparing population divergence between species across the same sites requires the standardization of differentiation to accurately demonstrate differences between populations from different species. Standardization of population differentiation measurements were conducted according to Hedrick - G'_{ST} (2005). This represents the maximum level of differentiation between populations when there is a large variance in expected heterozygosity when using hypervariable loci. The values for maximal F_{ST} and by G'_{ST} were calculated with the program GENEPOP v4.0 before comparison between populations between species.

Outlier analysis

To identify highly divergent microsatellite loci used in this study I used the Bayesian approach implemented the program BayeScan 2.1 (Foll & Gaggiotti 2008). This method incorporates a Bayesian approach to calculate the posterior probability of 'significant' outlier loci. Additionally, this method can distinguish between loci diverging via directional or balancing selection, which was not a feature in similar methods (e.g. Fdist2 [Beaumont & Nichols 1996]; Detsel [Vitalis et al. 2003]). The program was run with the default parameters but with an increased number of burn-in (100,000), thinning interval (20), number of iterations (300,000), number of pilot runs (20), and length of each pilot run (10,000). Bayes factors were used to determine the 'significance' of specific outlier loci. A second program, BayesFst, that utilizes a Bayesian likelihood method modified from the method Fdist2, was also implemented to detect markers influenced by directional and balancing selection (Beaumont and Balding 2004). The program was compiled with an increased number of simulated values taken from the F_{ST} distribution from 2,001 to 50,001. Outliers were considered significant based on adjusted p-values as described in Beaumont & Balding (2004) at 95% and 99% confidence intervals. Candidate outlier loci that were independently identified by both methods were regarded as significantly influenced by directional or balancing selection. Within *S. melanops* I conducted analyses with two different groups of populations. I either analyzed all populations or a subset that did not include samples from Kodiak AK.

F_{ST} outliers and environmental influences

To determine whether environmental effects have significant influence on genetic structure between populations of *S.flavidus* and *S. melanops*, I implemented a hierarchical Bayesian approach with the GESTE 2.0 software package (Foll & Gaggiotti 2006). The program estimates F_{ST} values from genetic data in each population and associates these estimates with environmental factors using a generalized linear model. Each alternate model includes a unique set of environmental variables and their respective posterior

probabilities. The model with the highest posterior probability can then be used to associate environmental factors with genetic structure across populations. Environmental data were obtained from the National Oceanic and Atmospheric Administration coast watch browser for the west coast of North America and Alaska (http://coastwatch.pfel.noaa.gov). Due to the low migration rate of adult rockfish (Haldorson et al. 1994; Mitamura et al. 2002) the potential for gene flow between distant populations will be highly dependent on the potential of larvae to disperse over long distances. Associations between seasonality and pelagic larval phases in rockfish influence the timing of larval release and recruitment, optimizing the chances of settling juveniles to suitable substrates. The environmental variables I considered are mean sea surface temperature in the months of November (Nov-SST) and June (Jun-SST) and chlorophyll-a concentrations in the same months (Nov-Chl-a & Jun-Chl-a). The months of November and June are closely associated with pelagic larval release and recruitment respectively in the studied species and should influence population divergence in *S. melanops* and *S. flavidus*. These four factors produce sixteen alternative models, including a constant model which does not include any environmental factors. Each model is associated with a posterior probability and the model with the highest probability best associates the genetic divergence with the environmental factors.

Due to the high number of microsatellite loci in each dataset, and the four environmental factors, I focused on candidate outlier loci identified by both BayeScan 2.1 and BayesFst. I ran GESTE 2.0 with 10 pilot runs, 5,000 iterations, 5,000,000 iterations, thinning interval of 50, and a sample size of 24,000 simulated data points. The model with the highest posterior probability was considered as the model that best explained genetic structure between populations. To identify the environmental factor that best explains the genetic structure in each data set, I added the posterior probabilities from each model that included one of the four factors. The contribution and relationship that each environmental factor has on the model that best explains the genetic structure can be determined by examining the absolute value and sign of the regression coefficients from the respective model. If the sign of the regression coefficient is negative then I suggest that genetic divergence, given the environmental effect, decreases as the difference between populations of the environmental effect increases. If the sign of the regression coefficient is positive then I suggest that the genetic divergence increases as the difference of the environmental variable between locations also increases.

Results

Within Populations

A total of seventeen loci were out of Hardy-Weinberg expectations in at least two populations of *S. flavidus* at p-values ≤ 0.05. Six loci were out of HWE expectation in two populations (KSs16, KSs18a, Sras311217, Sras311560, Sgoo2408_1, Sgoo3574_1), one in three populations (Sras308784), three in four populations (Sgoo642_1, Sgoo1190_1, Sgoo8_703_1), three in five populations (KSs17, Scau357863, Sgoo6768_1, and four in all six populations (SEB46, Sgoo690_1, Sgoo4_310_1, Sg_Sgoo9_832_1). In *S. melanops*, eleven microsatellite loci were out of Hardy-Weinberg expectations at p-values ≤ 0.05. Three loci were out of expectations in at least two populations (Sg_CCAT6620_1, Sgoo642_1,Sgoo742_1), five loci in three populations (KSs17, Sgoo3574_1, Sgoo2027_1, Sgoo4_310_1, Sra16-5), one in four populations (KSs17), and two in five populations (Sgoo5877_1, Spi7). Linkage disequilibrium analysis in *S. melanops* produced two significant sets of loci (KSs11B & KSs16 and Sras320 & Sras345) adjusting for multiple pair wise comparisons with Holm-Bonferroni corrections. In *S. flavidus*, three sets of loci (Sras345 & Sras320, KSs7 & Sgoo2408_1, and Sras311217 & Sras311560) were identified as having significant linkage disequilibrium with Holm-Bonferroni corrections.

In *S. flavidus*, the mean percentage of successful amplifications across loci ranged from 89% in Santa Cruz to 95% in Vancouver Island WA. Mean expected heterozygosity across populations ranged from 0.415-0.471 in N. Mendocino CA and Newport OR respectively, while mean observed heterozygosity ranged from 0.394 in Santa Cruz CA and 0.449 in San Luis Obispo CA. Allele counts ranged from one allele in all populations, to nineteen alleles in N. Mendocino CA. Allelic richness ranged from 4.96 in Santa Cruz CA to 5.34 in Newport OR. *S. melanops* had a range of 78% in Santa Cruz CA to 95% mean percentage of successful amplification across all loci between populations. Mean expected heterozygosity ranged from 0.404 in Kodiak AK to 0.460 in Newport OR, mean observed heterozygosity ranged from 0.372 in Santa Cruz CA to 0.450 in Newport OR. Allele counts ranged from one allele in all populations to eighteen in Juan de Fuca WA and S Mendocino CA, allelic richness ranged from 3.147 in Kodiak AK to 3.698 in Newport OR.

F_{ST} and G'_{ST} interspecies comparisons

The mean F_{ST} value in *S. flavidus* was calculated as 0.029 (95% CI 0.019-0.041) while the G'_{ST} value was calculated as 0.062 (95% CI 0.040-0.086). The comparison of F_{ST} values between EST and anonymous microsatellites were not significantly different (two-tailed t-test; p = 0.98. The complete *S. melanops* dataset produced an F_{ST} value of 0.034 (95% CI 0.022-0.047) and the G'_{ST} value was calculated as 0.066 (95% CI 0.042-0.089). When the Kodiak Island population was removed, the F_{ST} value lowered substantially to 0.016 (95% CI 0.010-0.022) and the G'_{ST} value was calculated as 0.032 (95% CI 0.019-0.044). Estimates of F_{ST} between EST-linked and anonymous microsatellites were not significantly different when Kodiak AK was included or removed (two-tailed t-test; All populations p = =0.63, Without Kodiak AK population p = 0.87).

Sixty-two microsatellite loci were shared between all genotyped *S.flavidus* and *S. melanops* individuals. Mean F_{ST} value calculated across populations for these sixty-two loci in *S.flavidus* was 0.032 (95% CI 0.089-0.049) with one locus (Sras311560) removed due to non-overlapping allele distribution between populations. Using the Hedrick (2005) correction for F_{ST} (G'_{ST}) the level of divergence increased to 0.049 (95% CI 0.031-0.069) (Figure 3.2). The mean F_{ST} value for all *S. melanops* populations (including Kodiak Island, AK) among shared loci was calculated to be 0.04 (95% CI 0.026-0.054), and the G'_{ST} value was 0.069 (95% CI 0.043-0.096) When the Kodiak Island AK population was removed from the analysis the F_{ST} value was calculated as 0.022 (95% CI 0.013-0.030) (Figure 3.2), while the G'_{ST} value was 0.034 (95% CI 0.020-0.047). A t-test between the distribution of G'_{ST} values between *S. flavidus* and *S. melanops* without the Kodiak population was not significantly different (p=0.15).

F_{ST} outlier detection-S. flavidus

In total, all eighty-five microsatellite loci produced genotypes in at least two populations to allow for F_{ST} estimation, but markers with limited amplification across populations were noted in subsequent analyses and were removed. One locus was removed (Sras311560) from subsequent analyses due to non-overlapping allele ranges, producing an elevated F_{ST} and could be considered a false outlier. Three loci were identified as "substantial" outliers ($0.5 \leq BF \leq 1$) Sras413, Sgoo930_1, Sgoo2408_1, and were also reported by BayeScan as significantly influenced by balancing selection with α values -1.1838, -0.93424, and, -0.94844 respectively (Table 3.1 and Figure 3.3a). Four loci were identified as "strong" outliers ($1 \leq BF \leq 1.5$): Sgoo8_703_1a, Scau357863, Sras112, and Spi7 (Table 3.1 and Figure 3.3a). Of these four identified outliers Spi7 was identified as being influenced by balancing selection (α value=-1.1836), while the other three were identified to be influenced by diversifying selection. Six loci were identified as "decisive" outliers ($BF \geq 2$): KSs6, KSs16, Sgoo4874_1, Sgoo6768_1, Sras311217, and Sras308784 (Table 3.1 and Figure 3.3a). All six loci were identified as being influenced by divergent selection. The outlier Sras308784 was fixed for the same allele in N. and S Mendocino CA and was not considered an outlier for subsequent analyses. Additionally, the outlier Sgoo6768 was fixed in the Santa Cruz CA population elevating F_{ST} estimates and produced viable genotypes in only five individuals. These two inconsistent loci can be attributed to limited amplification and genotyping error due to poor amplification in heterozygous individuals. These loci were removed from subsequent analyses and were not considered candidate outliers due to lack of successful amplifications.

The additional outlier detection method BayesFst identified four loci at the transformed 95% confidence interval. Three loci were identified as being influenced by balancing selection (Sgoo2408_1, Sgoo930_1, and Sras413), while one (Sgoo8_703_1) was identified as being influenced by directional selection (Table 3.1 and Figure 3.3b). At the transformed 99% confidence interval six loci were identified as being influenced by directional selection (Sras311217, KSs16, Sgoo4874_1, KSs6, Sras112, Scau357863) and one locus (Spi7) influenced by balancing selection (Table 3.1 and Figure 3.3b).

F_{ST} outlier detection-S. melanops

There were seventy-one microsatellite loci that produced genotypes in at least two populations of *S.melanops*. Three loci were removed form the analysis due to non-overlapping allele frequencies (Sgoo3582a and KSs5) or failed to amplify in more than two populations while being fixed in the remaining populations. When Kodiak Island was removed from the analysis, two microsatellite markers were identified as outliers with BayeScan. The microsatellite Ssc23, was identified as a "substantial" ($0.5 \leq BF \leq 1$) outlier but was not considered due to lack of amplification in Santa Cruz CA and San Luis Obispo CA populations. The locus SEB46 was also identified as "decisive" ($BF \geq 2$) and influenced by balancing selection (α=-0.9455) (Table 3.1 and Figure 3.3c). When the *S. melanops* populations were analyzed without the Kodiak AK population with BayesFst at the 95% transformed confidence interval one locus (Sras309791) was identified as being influenced by directional selection while no loci were identified at the transformed 99% confidence interval (Table 3.1 and Figure 3.3d).
The BayeScan analysis that included Kodiak Island AK identified two "substantial" outliers ($0.5 \leq BF \leq 1$): Sgoo3574_1 and Ssc23. The microsatellite locus Ssc23 was identified as being influenced by balancing selection (α=-0.99539) but did not produce genotypes in the Santa Cruz and San Luis Obispo populations and was therefore not included as an outlier. The other locus Sgoo3574_1 was identified as diverging due to directional selection (Table 3.1 and Figure 3.3e). A third outlier, Sgoo1801a, was identified as a "very strong" outlier ($1.5 \leq BF \leq 2$) (Table 3.1 and Figure 3.3e) influenced by diversifying selection (α value=-1.2373). Two outliers were detected as "decisive" ($BF \geq 2$) where Sras309791 was diverging due to directional selection and SEB46 was identified due to balancing selection (α=-1.1525) (Table 3.1 and Figure 3.3e).
BayesFst identified eight loci when the Kodiak AK population was included. Four loci were identified as being influenced by directional selection at the 95% transformed confidence interval (KSs 11B, KSs16, Sras311560, Sgoo4874_1), and one locus (Sgoo1801a) was influenced by directional selection (Table 3.1 and Figure 3.3f). At the 99% transformed confidence interval, the outlier detection method identified two loci (Sras309791 and Sgoo3574_1) influenced by directional selection and one (SEB46) influenced by balancing selection (Table 3.1 and Figure 3.3f).

The interspecies dataset comparison revealed no significant difference in the number of candidate outliers detected by both methods between *S. flavidus* and *S. melanops* when the Kodiak Island AK population was included (Fisher's two-tailed exact test; p =0.1841). Removing the Kodiak Island AK population and providing an unbiased comparison across both species distributions produced a significant statistical difference (Fishers two-tailed exact test; p = 0.0382) between the number of candidate outliers detected between the two species. Interspecies comparisons of candidate outliers detected across both species revealed one locus (KSs16) shared between the datasets but was not identified as an outlier when the Kodiak Island AK population was removed.

The eleven candidate outlier loci identified by both methods in *S. flavidus* consisted of three anonymous and eight EST-linked microsatellite markers. The identification of candidate outlier loci between anonymous and EST-linked microsatellite markers was not significant (Fisher's two-tailed exact test; p=1.000). In either *S. melanops* data sets, identification of outlier loci between anonymous and EST-linked

microsatellite loci was not statistically significant (Fisher's two tailed exact test; p=1.000)

Environmental effects and F_{ST} outliers

Bayesian analysis implemented in GESTE was used to identify the environmental model that best described the genetic divergence out of eight possible combinations. Eight of eleven candidate outlier loci in *S. flavidus* had the constant model with the highest posterior probability and three candidate outlier loci had models with the highest posterior probabilities that included an environmental variable (Table 3.2). Two loci were anonymous microsatellite markers (KSs6, Kss16) while one (Sgoo_703_1) is an EST associated microsatellite. The model with the highest posterior probability that best associated the genetic divergence in KSs6 included: Nov-Chl-a, Jun-Chl-a, and Nov-SST (Table 3.2). The effect that each environmental factor contributed to the best model in KSs6 was determined by the regression coefficient and ranked by their absolute value as: Nov-SST (α_3=-1.55), Nov-Chl-a (α_1=-1.48), and Jun-Chl-a (α_1=-0.956) (Table 3.2). The genetic structure for candidate outlier Kss16 was best explained by Nov-Chl-a with the highest posterior probability of 0.258 and structure increased with Nov-Chl-a (Table 3.2). The third candidate outlier marker Sgoo_703_1 was best associated with Nov-SST with a posterior probability of 0.105, and genetic structure decreased with Nov-SST (Table 3.2). No environmental associations were identified in *S. melanops* populations that describe divergence among outlier loci.

Discussion

Population Structure

Differences in PLD between *S. flavidus* and *S. melanops* across their ranges displayed contrasting levels of population structure, highlighting the importance of gene flow via the pelagic larval stage in the marine environment. *Sebastes melanops*, which has a longer PLD, had lower genetic structure compared to *S. flavidus*, which has a shorter PLD and a relatively higher degree of genetic structure. These findings are in accordance with other marine spices with pelagic larval stages where the amount time spent in the pelagic life phase closely correlates to potential for dispersal (Shanks 2009). Due to the longer PLD of *S. melanops* and differences in the timing of parturition, *S. melanops* larvae encounter oceanographic features conducive to greater dispersal potential which leads to increases in gene flow across the species range. This difference in genetic structure may have led to the increase in outlier loci observed in *S. flavidus*.

Timing of larval parturition and duration are two factors in maintaining or restricting gene flow between populations in the marine environment. A possible restriction to gene flow that may contribute higher genetic structure in *S. flavidus* than *S. melanops*, may be attributed to the differences in timing of larval release, pelagic larval duration, and time of recruitment. In general, both *S. flavidus* and *S. melanops* release pelagic larvae in the winter months while upwelling is reduced and net current flow is generally towards the north in the California current system (Shanks & Eckert 2005; Petersen et al. 2009). Parturition for *S. flavidus* occurs between November-June across their range with the peak occurring during February-March (Westrheim 1975; Tagart 1991; Love 1996) and remain as pelagic larvae for 3-4 months (Love et al. 2002). On the

other hand parturition in *S. melanops* occurs from November-March across their range (Wyllie-Echeverria1987; Bobko & Berkeley 2004) and remain as pelagic larvae for 5-6 months, settling in the months of April-September but peaking in the months of May-June along the California coast (Wilson et al. 2008). Due to the extended pelagic larval duration in *S. melanops*, it is very likely that settling larvae may be from locations other than where they were spawned, causing lower population genetic structure within species. This pattern of outside recruitment due to longer pelagic larval duration is not unlikely as it has been seen in other marine regimes in the Indo-Pacific (Treml et al. 2012). While in the pelagic phase, larvae may encounter oceanographic barriers and ecological shifts that further restrict gene flow between locations, reinforcing divergence across the species range. Restricted gene flow in *S. flavidus* has been previously observed along the California coast with a strong genetic break at Cape Mendocino where strong upwelling characterizes the area during the summer months (Sivasundar & Palumbi 2010; Hess et al. 2011). Pelagic larvae near these strong upwelling zones may be limited in their ability to disperse, effectively restricting gene flow across these oceanographic barriers. Conversely, *S. melanops* do not display the same pattern as *S. flavidus* across Cape Mendocino (Miller et al. 2004; Sivasundar & Palumbi 2010) which may be attributed to the combination of a longer larval duration and the strength of upwelling during larval settlement allowing for more gene flow.

F_{ST} outliers and concordance of detection methods

The ability to identify the relationship between highly differentiated candidate outlier loci and environmental effects has been contentious due to the assumption that direct selection on a gene may not similarly affect adjacent regions (Bierne et al. 2013). In order to increase my ability to discern between a false positive and a true outlier, I implemented two outlier detection methods and chose those outliers that were identified with both methods as the best candidates for being influenced by selection (Pérez-Figueroa et al. 2010). It is important to note that those outliers detected in only one method could still potentially be true outlier loci. In *S. flavidus* there was complete concordance between both outlier detection methods, identifying eleven total candidate outlier loci. The seven loci identified as diverging by directional selection suggests the influence local adaptation has on a high dispersing group of marine fish with moderate genetic structure is detectable. Three of the four loci identified to be influenced by balancing selection were significant only at the lowest confidence intervals in each method. The current methods I used to detect outliers are more sensitive to identifying candidate outlier loci influenced by balancing selection when compared to initial methods used to determine the influence of divergence on outliers. My results are in contrast to what was observed by Mäkinen et al. (2008) in three-spined stickleback, where they found a higher proportion of candidate outlier loci under balancing selection between marine and freshwater populations.

The use of highly polymorphic loci in genome scans for signals of selection may bias the ability to discern between directional and balancing selection, but I observed a higher number of outlier loci with EST-linked microsatellite although these results were not statistically significant. The four outliers identified by both methods to be influenced by balancing selection in *S. flavidus* were categorized in the lowest confidence category

("substantial 0.5 ≤BF≤1) and should be analyzed with caution due to the inherent difficulty in identifying true balancing selection with outlier detection methods. In the *S. melanops* data set, SEB46 is identified as influenced by balancing selection and categorized in the highest confidence intervals by both detection methods ("decisive"; BF≥2 & 99% CI) making it the best candidate for further investigation on the maintenance of balancing selection in a highly dispersing marine organism. Overall, the general trend observed in my study suggests that directional selection influencing outlier loci are possibly responding to subtle environmental differences across the range of a marine population.

The influence of higher gene flow in *S. melanops* was seen in the number of loci identified with the outlier detection methods. In *S. melanops,* when the Kodiak AK population was included, there were only four outliers confirmed by both methods. BayesFst, identified more outliers with eight while BayeScan identified four. This could be that BayesFst is more likely to include false positives (Beaumont & Balding 2004). When the Kodiak AK population was not included, no outliers were confirmed by both methods. Instead, each identified a "decisive" outlier, SEB46 (Identified by BayeScan), as being influenced by balancing selection and Sras309791 (Identified by BayesFst) as diverging due to directional selection. Conversely, a study by Neilsen et al. (2009) reported no evidence for balancing selection in Atlantic cod (*Gadus morhua*) with 98 SNPs developed from an EST library, six of which were expected to be involved in local adaptation. This suggests that elevated gene flow could lead to a smaller proportion of detectable outlier loci in *S. melanops* (Slatkin 1987; Lenormand 2002)

Differences in gene flow due to PLD in each species may also have an impact on the degree of local adaptation as revealed by the number of identified candidate outlier loci. Although I did not detect a significant difference in the number of identified outliers between species, it does suggest that gene flow due to PLD may not be robust enough to wash out signals of selection across a species range. Limitations to gene flow, aside from differences in PLD between the two species, may be influenced by physical barriers in the marine environment restricting recruitment between sampled locations. Comparisons between species revealed two patterns consistent with higher degree of local adaptation in *S. flavidus* as noted by a higher number of candidate outlier loci, while in *S. melanops* fewer identified candidate outliers are a result of increased gene flow preventing any opportunity of local adaptation across the sampled range of this study.

There was also no significant difference between anonymous and EST-linked candidate outliers in either data set. Other studies have reported a higher number of EST-linked microsatellite loci versus anonymous (Li et al. 2004; Scotti-Saintagne et al. 2004; Vasemägi et al. 2005; Martinez et al. 2011) but were not statistically different. Differences in the proportion of candidate outliers detected between anonymous and EST-linked loci is not surprising since the development of "anonymous" microsatellites does not suggest independent sampling from expressed genes and may experience varying degrees of selective pressures. Although sampling the genome of non-model organisms should include "neutral" and expressed regions, due the lack of genomic information about these markers it is important to sample a larger number of genomic regions to increase the probability of ultimately identifying genomic regions responding to selection.

Environmental factors and effects on genetic structure

I was able to detect an association between outliers and environmental factors in *S. flavidus* but no association could be made in the *S. melanops* dataset. Out of the four environmental factors included in the GESTE analysis with eleven candidate outlier loci in *S. flavidus*, three factors showed an association with genetic structure. The one environmental factor that did not associate with any outlier was Jun-SST. The three candidate outliers that were strongly associated with an environmental factor were: KSs6 (Nov-SST, Jun-Chl-a, Nov-Chl-a); Kss16 (Nov-Chl-a); Sgoo_703_1 (Nov-SST). In the California Current system, selective pressures influencing spawning, parturition, and recruitment on organisms with long pelagic life phases increase the probability of successful settlement to suitable habitat (Shanks 2009; Shanks & Eckert 2005). Due to the elevated mortality during the pelagic larval life phases, species in *Sebastosomus* have evolved reproductive strategies to reduce the number of spawned larvae lost in strong offshore currents and maximize the proportion of settling recruits. The relationship between SST and Chl-a concentrations are two major oceanographic features that characterize seasonality in the California Current system and indicate the extent of primary productivity. During times of increased upwelling and higher primary productivity, *Sebastosomus* larval recruitment in *S. flavidus* and *S. melanops* dramatically increase (Lenarz et al. 1995; Ammann 2004; Carr & Syms 2006) while another species (*S. entomelas*) has higher recruitment during upwelling relaxation events (Norton 1987). Furthermore, instances of lower sea level anomalies and temperature differences suggest lower upwelling resulting in weaker larval recruitment in *S. flavidus*, *S. melanops*, and *S. mystinus* (Laidig et al. 2007). Sea surface temperature influences multiple biological processes along the California Current system by providing nutrient-rich deep water to primary producing phytoplankton establishing the basis for many marine regimes along continental shelves (Frank et al. 2007), providing prey for recruiting larvae. Other factors may be used to explain correlation with population structure and candidate outlier loci that have more influence than the environmental factors I used in this study. Those factors may be present during the pelagic larval phase that influence larval behavior and maintain pelagic larvae in low-energy water parcels situated between seaward traveling surface water and shoreward traveling upwelling water (McManus et al. 2005).

Alternatively, the interaction between historical demographic events and oceanographic processes may increase the variance at neutral loci producing increased false positives and forcing an association with unrelated environmental effects (Bierne et al. 2013). While I attempted to reduce the number of false positives prior to environment-outlier analyses by comparing candidate outlier loci identified by both methods, it is still likely that some candidate outliers were identified as significantly diverging due to increases in neutral variance. In a coastal latitudinal marine environment correlations in co-ancestry may increase neutral variation in F_{ST} causing an increase in false positives when implementing a genome scan approach (Bierne et al. 2013). Previous outlier detection methods did not allow for unequal divergence between populations in a data set as implemented by the original methods by Lewontin and Krakauer (1973) but those used in this study adjust for such drawbacks yet can still produce an elevated number of false positives. Such issues will not be completely dismissed when using methods that account for hierarchical structuring like those

implemented by Excoffier et al. (2009), therefore I chose to use a comparison of different methods and only considered candidate outlier loci if they were confirmed by both outlier detection methods. The additional processes that may lead to candidate outliers regardless of the analysis method are related to variance in recombination rates near sampled loci, species-wide selective sweeps, hybrid zones, and edge effects of an expanding population experiencing novel selective pressures. Lastly, the ability to detect selective pressures on the genomes via the identification of outliers is highly dependent on the assumption that I am able to accurately measure neutral variation at loci that are not influenced by selection and therefore I report these outliers as candidates for selection diverging due to environmental factors. Further analysis would be required to confirm the degree at which these loci are linked to genes responding to selective pressures.

Conclusions

My results show that differences in PLD between *S. flavidus* and *S. melanops* influence the degree of gene flow across their ranges as displayed by the higher number of outlier loci in *S. flavidus* leading to local adaptation in their respective habitats. Patterns of population structure vary between the two closely related rockfish species and this difference may be due to differences in PLD although measurements of divergence were not statistically significant F_{ST} and G'_{ST} were higher in *S. flavidus* than *S. melanops*. The shorter PLD in *S. flavidus* may drive signatures of local adaptation, leading to the higher number of identified outlier loci that are related to ecological factors. *S. melanops* has a longer PLD (higher gene flow) that leads to fewer number of identified outlier loci. This study empirically demonstrates that increased gene flow leads to lower levels of genomic adaptation in marine organisms. . Finally, although similar in their distribution and behavior *S.flavidus* and *S. melanops* have evolved two different yet successful life history strategies that optimize larval recruitment. This information coupled with genetic and oceanographic data will provide a better understanding of marine population dynamics in addition to highlighting initial genomic regions possibly linked to genes responding to selection in the marine environment.

Table 3.1 Candidate outlier loci identified by BayeScan and bayesFst in each species. Bold loci were detected by both methods. Loci identified to diverge to balancing selection are denoted with an asterisk while other are diverged due to directional selection.

Species	BayeScan Substantial (0.5 ≤BF≤1)	BayeScan Strong (1 ≤BF≤1.5)	BayeScan Very strong (1.5 ≤BF≤2)	BayeScan Decisive (BF≥2)	BayesFst 95% transformed CI	BayesFst 99% transformed CI
S. flavidus	***Sras413**** ***Sgoo930_1**** ***Sgoo2408_1****	***Sgoo8_703_1*** ***Scau357863*** ***Sras112*** ***Spi7****	-	***KSs6*** ***Sgoo4874_1*** ***KSs16*** ***Sras311217***	***Sgoo8_703_1*** ***Sgoo2408_1**** ***Sgoo930_1**** ***Sras413****	***Sras311217*** ***KSs16*** ***Sgoo4874_1*** ***KSs6*** ***Sras112*** ***Scau357863*** ***Spi7****
S.melanops without Kodiak AK population	-	-	-	*SEB46**	*Sras309791*	-
S.melanops	***Sgoo3574_1***	-	***Sgoo1801a****	***Sras309791*** ***SEB46****	KSs11B Sras311560 KSs16 Sgoo4874_1 ***Sgoo1801a****	***Sras309791*** ***Sgoo3574_1*** ***SEB46****

Table 3.2 Probabilities of environmental variables associated with genetic structure. "Constant" (α0) is the model which only considers F_{ST} as best associating the measured distributed variance. α1 and α2 are Chl-a concentrations for the months of November and June respectively. α3 and α4 are SST for the same months. Only outliers that had more than the "Constant" model associated with environmental variables are shown

Posterior probability	Outlier & regression coefficient	Factor	Mean	Mode	95%HPDI
	KSs6				
0.0763	α_0	Constant	-1.84	-1.84	-8.08; 4.1
	α_3	Nov-SST	-1.55	-1.64	-7.23; 4.78
	α_2	Jun-Chl-a	-0.956	-1.45	-6.99; 5.15
	α_1	Nov-Chl-a	-1.48	-1.16	-6.91; 4.45
	σ^2	-	1.32E+03	877	260; 2.90e+03
	Kss16				
0.258	α_0	Constant	-0.916	-0.846	-2.16; 0.386
	α_1	Nov-Chl-a	1.06	1.04	-0.147; 2.29
	σ^2	-	1.13	0.647	0.115; 3.20
	Sgoo_703_1				
0.105	α_0	Constant	-1.79	-1.46	-7.93; 4.30
	α_3	Nov-SST	-2.52	-2.51	-6.65; 1.41
	σ^2	-	1.33E+03	859	292; 3.00e+03

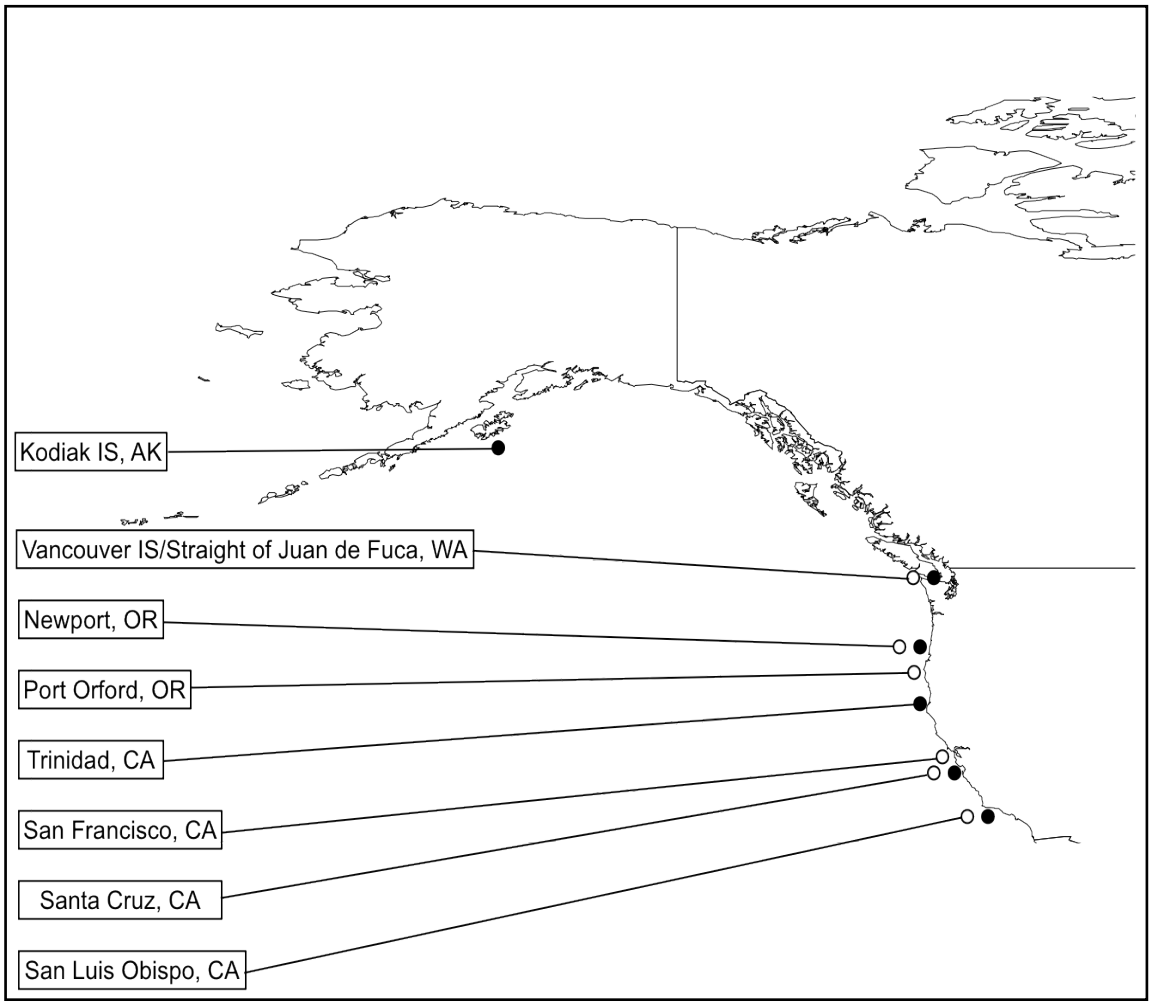


Figure 3.1 Sampling locations for *S. flavidus* (open circles) and *S. melanops* (black circles) used in this study

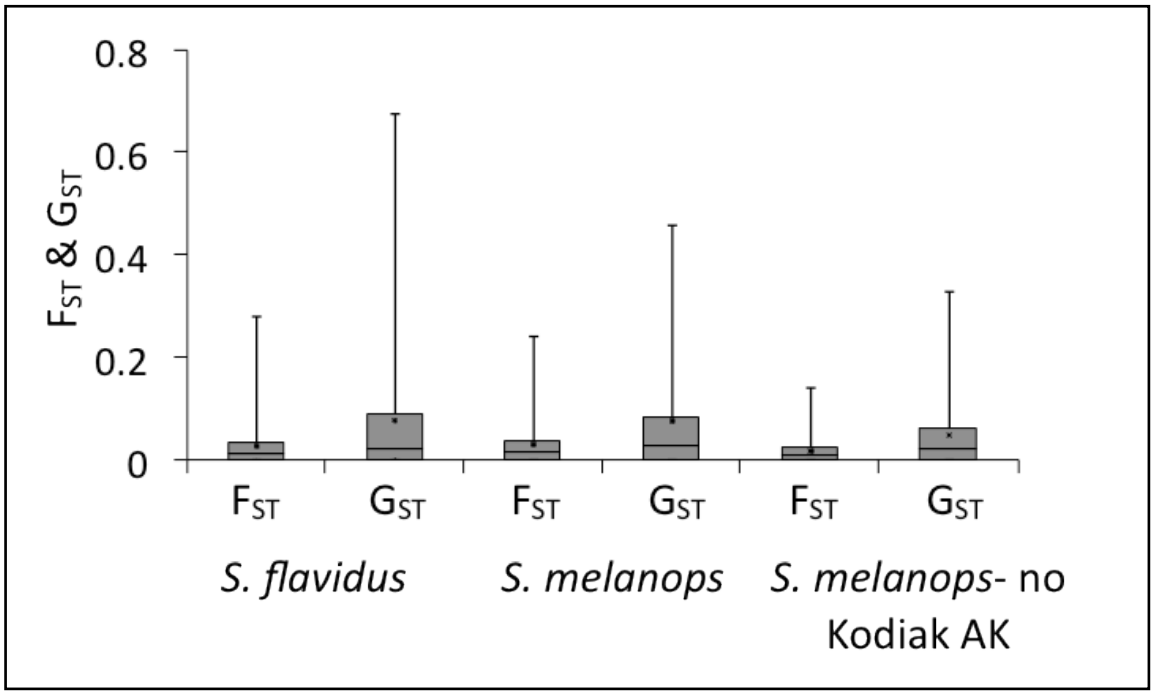


Figure 3.2 Box and whisker plots of F_{ST} and G'_{ST} for S. flavidus and S. melanops. A Student's t-test produced no significant p-values between species datasets across sixty-three shared microsatellite loci

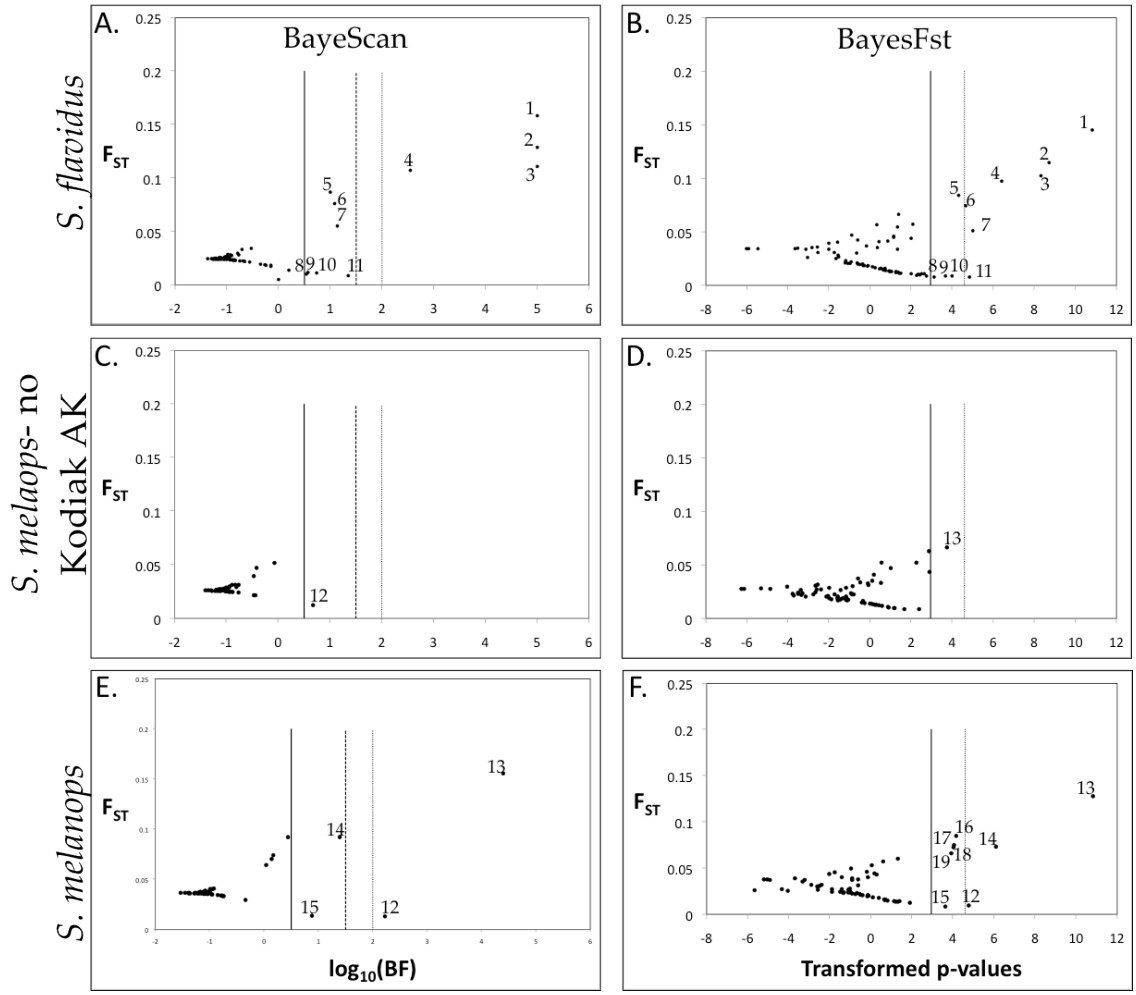


Figure 3.3 Results from the outlier detection methods BayeScan (A,C,E) and Bayesfst (B,D,F) in each species. The solid, dashed, and dotted lines in the BayeScan results designate outliers as "substantial" (0.5 ≤BF≤1), "very strong" (1.5 ≤BF≤2), and "decisive" (BF≤2) respectively. The solid and dotted lines in the BayesFst results correspond to the 95% and 99% transformed confidence intervals respectively. Numbered marker identifiers correspond to the following marker names: 1-Sras311217, 2-KSs16, 3-Sgoo4874_1, 4-KSs6, 5-Sgoo8_703_1, 6-Scau357863, 7-Sras112, 8-Sras413, 9-Sgoo930_1, 10-Sgoo2408_1, 11-Spi7, 12-SEB46, 13-Sras309791, 14-Sgoo3574_1, 15-Sgoo1801a, 16-KSs11b, 17-Sras311560, 18-KSs16, 19-Sgoo4874_1

Chapter 4: Population genomics of two morphologically distinct incipient species of rockfish (Sebastes mystinus) using a 2bRAD sequencing approach

Introduction

Studies on the evolutionary mechanisms that influence reproductive isolation and ultimately lead to speciation of isolated groups of natural populations analyzed on a genome-wide level are key to identifying genes involved in the speciation process (Turner & Hahn 2007; 2010; Michel et al. 2010; Garrigain et al. 2012). Identifying these genomic regions responding to selection from the environment can be attributed to physical, behavioral or ecological pressures producing areas of elevated divergence while other genomic regions evolve in a comparatively neutral manner (Via & West 2008; Via 2009; Feder & Nosil 2010; Via 2012). Characterizing the regions involved in reproductive isolation is not a trivial task and involves the construction of high-density linkage maps from known crosses (Slate et al. 2009; Amores et al. 2011), quantitative trait loci (QTL) mapping (Johnston et al. 2011; Hecht et al. 2012; Miller et al. 2012), and population genomic analyses (Luikart et al. 2003; Storz 2005; Hohenlohe et al. 2010). Despite having these resources, identifying the genomic regions and ultimately the genes themselves that are responding to selection studies have focused on utilizing a small number of sampled loci many of which the location or linkage to a gene is unknown.

In early speciation, populations diverge via reproductive isolation and particular regions in the genome that are under selection diverge much more rapidly than those that are not experiencing the same degree of selection. This mechanism of speciation is straightforward and can be easily applied to populations that are physically separated (allopatric speciation) because the divergence between populations can be attributed to the restriction to gene flow from the physical barrier between the populations. In the case where gene flow is present between diverging species, it is more difficult to identify reproductively isolating mechanisms since recombination will act to homogenize alleles that have adapted to the respective environments (Felsenstein 1981; Coyne & Orr 2004; Nosil 2012). In the cases where researchers cannot sample the genes involved in divergence directly, a large proportion of the genome may be sampled to identify signals of divergence hitchhiking (Via 2012). Divergence hitchhiking involves genomic regions across several megabases involved in local adaptation reveal evolutionary history of reproductive isolation due to ecological pressures have a higher than expected degree of divergence. This allows for the sampling of hundreds of loci without any previous information about the genes involved in divergence.

In addition to analyzing genomes responding to selection after recent speciation historical and contemporary evolutionary factors influencing genomic divergence can provide insight to the processes of speciation producing current distributions of populations within a species. Both, contemporary and historical relationships between divergent populations can be observed in genomic data (Cavalli-Sforza et al. 1967; Felsenstein 1982; Cann et al. 1987). In these initial studies, population histories were reconstructed using few polymorphisms, these studies progressed to hundreds to thousands of markers to genomic level sampling. One method to represent the historical and contemporary relationships between populations is through the reconstruction of a tree. The disadvantage of constructing trees has on representing the relationships

between populations is that they do not represent the contribution of gene flow has on the contemporary relationships between populations. Not having the ability to model gene flow when establishing relationships between populations ignores a possible source of ancestral genomic variation influencing contemporary relationships.

In the process of speciation between diverging populations, the ability to detect those regions under selection are attainable using a genome scan approach with an added advantage of increased genomic sampling from construction of restriction-site-associated DNA (RAD) tagged libraries (Baird et al. 2010; Hohenlohe et al. 2010; Keller et al. 2013; Stölting et al. 2013). The genome scan approach has basis in the original proposed idea by Lewontin & Krakauer (1973) where loci with elevated levels of F_{ST} can be used to identify genomic regions responding to divergent selection. In a genome scan, it is important to sample divergent populations or groups of species since divergence in incipient groups can identify the evolutionary forces that may have shaped current genomic variation. By using populations that have recently diverged across their distribution, the genomic divergence contributing to the reproductive isolation can be identified before any additional genomic differences accumulate to the degree of speciation (Via 2009).

A valuable opportunity for the application of a population genomic approach to identify regions diverging due to recent speciation exits in the marine genus *Sebastes*. Due to the rapid adaptive divergence within this genus, there are over 74 species residing in the Northeastern boundary of the Pacific Ocean (Berntson & Moran 2009; Love et al. 2002). The genus is marked by a species flock driven by an adaptive radiation event that caused the large diversification within the genus (Love et al. 2002). The genus has also been documented with instances of recent speciation events that have caused splits within a species lineage that has led to significant color, morphological, or ecological differences. Studies that have identified two or more phenotypes with distinct genetic divergence between them all associated the groups with habitat preferences within the marine environment (Daníelsdóttir et al. 2008; Hyde et al. 2008). Narum et al. (2004) found that a recently diverged sister-species (*S. carnatus* and *S. chrysomelas*) in the subgenus *Pteropodus* are sympatrically distributed across their ranges but are distinguished by their drastic color morphologies. In the same study, there were only a few individuals that displayed evidence for hybridization between the two species reinforcing the distinction between them but some introgression may still exist in the wild implying post zygotic barriers are still not well established (Buonaccorsi et al. 2011). Furthermore, instances where color or morphology has not developed between diverging groups within a species implying cryptic speciation (Rocha-Olivares et al. 1999; Gharrett et al. 2005; Schmidt 2005; Stefánsson et al. 2009). The *S. inermis* species displayed variable morphological and color variations but these features alone were not sufficient in separating the three genetically distinct types (Kai et al. 2002).

In this study I will focus on a pair of genetically distinct groups of *S. mystinus* previously identified by Cope (2004) and Burford & Bernardi (2008). *S. mystinus* have been reported to have the highest levels of genetic differentiation in the subgenus *Sebastosomus* around Cape Mendocino which has been suggested as incipient speciation as a result of secondary contact of two distinct expanding populations maintained prior to the last glacial maximum (Buford & Bernardi 2008). The two distinct groups of *S.*

mystinus are of particular interest due to their recent divergence and significant genetic break around known oceanographic boundaries. In order to target genomic regions that have participated and responded to reproductively isolating factors, identifying those populations that display incipient speciation characteristics are required. The groups were considered cryptic species and little evidence of hybridization indicates strong reproductive isolation between the two types in the species (Burford et al. 2011). Although both types are found in the same habitat little is known about genomic responses to isolation maintaining the reproductive isolation of these types. The two types can be observed across their ranges from Oregon to the California bight, with a large overlapping zone occurring from southern Oregon to northern California. What has not been described in the two types is that there are two distinct morphological coloration groups and identifying the association with genetic types described by Burford & Bernardi (2008) will provide a unique system to observe the effects of incipient speciation on the genome of two recently diverged groups.

In this study I will perform an initial population genomic analysis of genome-wide patterns of differentiation through outlier detection methods, analyses of population structure, and the historical relationships between populations of a recently diverged groups of marine fishes. I used an alternative restriction-site associated DNA tagged (RAD-tagged) library based on the integration of Illumina sequences to fragments of digested genomic DNA using a type IIB restriction enzyme (Wang et al. 2012) to densely sample genomes for single-nucleotide polymorphisms across six populations of two genetically distinct groups of *S. mystinus*. Because of the recent divergence and previous descriptions of the two genetic groups I expect to identify candidate outlier loci linked to genomic regions responding to selective pressures that caused the isolation between the two genetic groups.

Materials and Methods

Sample Collection and DNA extraction

The two *S. mystinus* morphotypes (Solid & Blotched) were collected using hook and line sampling and salvaged carcasses from recreational vessels from three locations in Newport OR, Albion CA, and San Francisco, CA (Figure 4.1). Twenty four individuals from each location were used (12 of each morphotype) for subsequent analyses. Fish from Newport OR were identified by morphotype based on morphometric and meristic data in addition to confirmation using a subset of eight microsatellite loci with the rest of the sampled individuals. Tissue samples were placed directly in 95% ethanol before DNA extraction with a DNeasy Blood and Tissue DNA extraction kit (Qiagen Inc.).

2bRAD-tagged library preparation and Illumina sequencing

Genomic DNA was isolated and purified for 72 individuals form three locations, quantification of DNA was analyzed using a Nanodrop 1000 to ensure concentrations of at least 250ng/µl. All samples were then prepared to a volume of 4µL with approximately 1µg of DNA for each individual and amplified according to a modified 2bRAD protocol (Wang et al. 2012). I digested the standardized samples in a total

volume of 6µL with the following reagents: 0.78 µL of nuclease free water (NFW), 0.6 µL 10x Buffer (Thermo Scientific), 0.12 µL SAM (500 µM;Thermo Scientific), 4.0µL gDNA (~1µg), and 0.5 µL AlfI (2U/µL-Thermo Scientific). Ligations of the digested product were conducted in 20 µL reactions with the following reagent volumes: 11.25µL NFW, 2.5 µL 10x T4 ligase buffer/ATP (New England BioLabs), 2.5 µL forward adapter (-NY at 2 µM), 2.5 µL reverse adapter (-NY at 2 µM), 1.25µL T4 DNA ligase (New England BioLabs), and 5µL of the digested product. Library amplifications were conducted in 100µL volumes using the first twenty-four six basepair Illumina TruSeq barcodes for each set of individuals from each location. The reagent volumes for one reaction were: 41.9µL NFW, 20.0µL 5x buffer (New England BioLabs), 5.0µL ILL-Mpx (1µM), 5µL ILL-Barcode (1µM), 2.0 µL ILL-Lib1-20 (10µM), 2.0µL ILL-Lib2 (10µM), 3.125µL dNTPs, 20µL ligation product, and 1.0µL high fidelity Phusion TAQ (New England BioLabs). The thermal profile for library amplification began with an initial denaturing step at 98°C for 15 s, followed by 14 cycles of 98°C for 5 s, 60°C for 20 s, and 72°C for 10 s. PCR products were run on using E-Gel SizeSelect gel apparatus and extracted for proper size range at approximately 150bp. Three independent 100 base pair single end Illumina HiSeq runs were performed containing both morphotypes for each location with 24 individuals per lane (12 for each morphotype)

Processing of sequences and SNP detection

Sequences from each of three runs were de-multiplexed using CASAVA v1.8.2 and only those sequences that matched one of twenty-four Illumina TruSeq six base barcodes were kept for subsequent quality control. All sequences that straddled the restriction site sequence were trimmed prior to sequence quality filtering using various scripts from the FASTX-Toolkit (http://hannonlab.cshl.edu/fastx_toolkit/index.html). The program FastQC was also used to confirm the presence of the AlfI restriction site in over 90% of reads in the FASTQ files for all individuals. Sequences shorter than 5 bases, homopolymer repeats, and poor overall quality scores were removed from further analysis discarding any fragments with 18 bases or more with a minimum score of 20.

I constructed a de novo reference "genome" using clustering algorithms provided in the SHRiMP analysis package in addition to custom filtering scripts provided from Dr. Eli Meyer's website (http://people.oregonstate.edu/~meyere/tools.html). I used a subset of four random individuals from each location and type (24 total) to reconstruct the cluster derived reference (CDR). The maximum number of hits per read in the *gmapper* command was reduced from the default 10 to 3 to minimize the number of highly similar sequences and the confidence alignment probability to the CDR was set to a p-value of 0.05. To eliminate short, low quality, or uncertain reads I set the minimum length of alignment to be the complete length of the restriction fragment (32) with a minimum of 30 matching bases. Base calling composition for each alignment was conducted if there were at least ten reads per site. Finally, the genotyping of each SNP site according to predetermined frequency thresholds were set to: maximum minor allele frequency in homozygotes (Max-hom) of 0.01, minimum minor allele frequency in heterozygotes (Min-het) of 0.3, with a minimum coverage of 10 reads.

Population analysis

I calculated observed and expected heterozygosity, and tested for Hardy-Weinberg equilibrium (HWE) for the Solid and Blotched morphotypes from the entire dataset using Genepop v4.0 (Rousset 2008) with 10,000 dememorization steps, 50 batches, and 1000 iterations. The statistical significance for loci out of HWE was evaluated using a 0.05 and 0.01 p-value cutoff rather than the more conservative Bonferroni due to the large number of comparisons and low adjusted p-value as a result. I also estimated a pair wise population divergence parameter (F_{ST}) to determine the degree of gene flow between populations and overall F_{ST} to characterize the overall population divergence between locations and morphotypes.

Outlier analysis

To identify all SNPs from the dataset suggesting evidence for divergent and balancing selection between and among morphotypes, I conducted a genome scan using BayeScan v2.1 (Foll & Gaggiotti 2008) with 21,318 SNP loci. This method is useful in identifying candidate loci responding to selection by comparing a model of neutral expectation versus a model of selection. The program was run with the default parameters but with an increased number of burn-in (100,000), thinning interval (20), number of iterations (300,000), number of pilot runs (20), and length of each pilot run (10,000). These corrected Bayes factors (BF) are established as analogues to p-value cutoffs but the intention is to reduce the false discovery rate in the dataset. The method by which this approach reduces the false discovery rate is by establishing criteria before the analyses and adjusting according to recommendations by the authors I proceeded with the following: $BF=\log_{10}(FDR)$; for FDR below 0.01 and 0.03 which gave cutoff BF of 2 and 1.5 respectively. Loci that exhibited outlier status according to the $\log_{10}$("Bayes factor") were considered substantial ($0.5 \leq BF \leq 1$), strong ($1 \leq BF \leq 1.5$), very strong ($1.5 \leq BF \leq 2$), and decisive ($2 \leq BF$).

Population structure analysis

To investigate the population substructure in my dataset and divergence between morphotypes across all populations, I implemented the program STRUCTURE v2.3.2.1 (Pritchard et al. 2000). I used 10,000 burn-in and 50,000 Markov-chain Monte Carlo steps and assumed an admixture model with no other prior population information. I used the entire dataset that consisted of 21,318 SNP loci which includes SNPs with at least two genotypes across individuals at a locus. This removed any SNPs that did not provide a measurement of variation (i.e.-monomorphic) and accounted for sampling the three theoretical genotypes present at a site (i.e.-two homozygotes and the heterozygote). This filtering reduced the possibility of sampling genotypes with a sequencing error at a homologous site across individuals. Although it is possible to have more than three genotypes at a homologous site, I chose to implement a more conservative criteria in order to reduce type II error in my dataset and provide a reliable representation of genomic variation. The *K* values were varied from 1-6 and optimal *K* value was determined using Structure Harvester (http://taylor0.biology.ucla.edu/structureHarvester/) which implements the Evanno et al. (2005) method.

Admixture and demographic history

To determine the historical demography and admixture events within *S. mystinus* I used the analysis program TreeMix v1.12 (Pickrell & Pritchard 2012). This method models a maximum likelihood tree between populations and places a pre-determined number of migration events on the tree as an attempt to explain the relationship of the genomic data between populations in a species. This method allows me to propose historical admixture events between morphotypes and reveal evolutionary trajectories that gave rise to the current degree of isolation within the species. I assembled a subset of SNP loci that represented the highest number of variable sites sampled across all individuals. This dataset solely consisted of bi-allelic loci with the alternate allele present in at least one individual across the sampled sites. I modeled the maximum likelihood tree without admixture events and sequentially increased the migration events to six in each subsequent simulation until I achieved a graph with the smallest non-zero residuals. I ran TreeMix with bootstrapping across blocks of thirty SNPs (-k 30 option) and did not set any populations as an out group since a distant population was not sequenced.

Results

Raw data and processing

The average number of reads produced from sequencing after de-multiplexing was 17.9 million reads per individual. The minimum number of reads in one individual was 43,971 while the maximum number of reads was 60.2 million. After processing through the FASTX-toolkit to remove unwanted primer sequences and removal short or homopolymer sequences, the average number of reads per individual dropped to 8.9 million reads with a minimum of 8,762 and a maximum of 25.8 million reads. The total number of reads produced by the CDR *de novo* 2bRAD-taged genome assembly was 131,893 from 24 individuals represented by four individuals form each type and site. The initial SNP file contained 1,712,187 loci but not all loci met the minimum requirement of the minimum number of genotypes in order to consider a locus polymorphic. The final working dataset consisted of 21,318 SNP loci with at least two genotypes present in order to remove any monomorphic loci or loci with missing data. In later assemblies, I will use the entire dataset for a more complete representation of 2bRAD-tagged reads that may have been missed in the analysis due to low representation in the dataset in a fraction of the individuals from sampled sites.

Population analysis

Given the large number of tests for significance, I will report the proportion of loci out of HWE for p-value significance levels under 0.05 and 0.01. In the Oregon-blotched collection, 4,253 loci produced a viable HW exact test such that 3.9% of loci were significantly out of HWE under the 0.05 p-value and 0.9% were significant under 0.01. In the Oregon-solid collection, 4.095 loci produced viable tests which produced 5.5% of loci out of HWE while 0.8% were significant under the 0.01 p-value. In the Albion-solid collection 4,152 loci produced p-values which identified 4.8% of the loci under 0.05, and 1.2% under the 0.01 p-value. In the Albion-blotched collection, 4,664 loci produced successful tests which identified 4.3% of loci under the 0.05 and 1% under the 0.01 p-value significance level cutoff. In the San Francisco-solid collection 4,878

loci produced viable tests identifying 6.8% of loci under the 0.05 and 2% under the 0.01 p-value cutoff. Finally, the San Francisco-blotched collection produced 4,163 tests that identified 13% of loci under the 0.05 and 2.7% of loci under the p-value significance level (Table 4.1).

The overall population differentiation among six sampled collections and 21,318 SNP loci was estimated to be 0.187, and ranged from 0 to 1 (Figure 4.2) while F_{ST} for all blotched morphotypes was estimated as 0.049 while all solid morphotypes had an F_{ST} estimated as 0.086 . The overall pairwise F_{ST} values for the blotched collection were 0.027 between Albion and Oregon, 0.099 between San Francisco and Oregon, and 0.059 between San Francisco and Albion (Table 4.2). The pairwise F_{ST} values for the solid collections were 0.012 between Albion and Oregon, 0.15 between San Francisco and Oregon, and 0.137 between San Francisco and Albion (Table 4.2). Overall F_{ST} values between all locations and morphotypes displayed a general trend towards having higher differentiation between morphotypes than between locations (Table 4.2). The highest values of F_{ST} observed between locations and morphotypes was observed between the San Francisco blotched and Oregon solid collections with a value of 0.321 while the smallest value was estimated to be 0.095 between the San Francisco blotched and solid collections (Table 4.2).

Outlier analysis results from BayeScan

Out of 21,318 loci between all collections across locations, there were 209 loci identified as outliers with a minimum corrected bayes factor (BF) of 1.5. The proportion of 209 candidate outliers corresponds to one percent of all sampled loci and in order to reduce the degree of false positives in the dataset, I implemented two cutoff BF values: $1.5 \geq BF \geq 2$ (very strong) and $BF \geq 2$ (decisive). BayeScan identified 42 candidate outlier loci under the "very strong" cutoff and 167 candidate outlier loci were identified under the "decisive" cutoff value (Figure 4.3). All candidate outlier loci were identified to be diverging due to diversifying selection, the BF was lowered to the minimum cutoff for significance (0.5) no additional loci were identified to be diverging due to balancing selection. Of the 209 candidate outlier loci identified, six pair of loci were from the same 2bRAD-tagged sequence fragment with a mean distance of 13.8bp (range: 1-27bp).

Population structure among populations and morphotypes

The population structure analyses across six locations and both morphotypes based on 21,318 SNP loci clustered the data into two clusters regardless of preset K values (Figure 4.4). The optimal K value as reported by Structure Harvester which uses the second order rate of change of the likelihood distribution between values of K was 2 for all simulations and predetermined values for K. One individual (IND18) from the Newport OR-solid collection appeared to have genomic ancestry from both morphotypic lineages but this individual was missing more than 99% of identified variable sites. The average proportion of missing data among individuals was 37% when IND18 was removed. All other individuals had probabilities of assignment >92% to their respective morphotypic collections. In the San Francisco CA-solid collection, there were two individuals (IND53, IND55) that were assigned to have ancestries most like that of the

blotched collections. The probable ancestry range was 0.598-0.885 with a mean of 0.769 for the solid morphotype (Figure 4.4).

Admixture and demographic history

To determine the signal of admixture between locations and historical gene flow between morphotypes, I used TreeMix to find an admixture graph that best fits the connections between each collection among sites. I first analyzed the maximum likelihood trees of all six collections followed by an analysis of the residuals to detect population pairs that display higher gene flow than what was represented by the tree. The initial analysis that did not allow for any migration events among collections and sites produced a maximum likelihood tree produced relationships among sites regardless of morphotype (Figure 4.5). The corresponding residual (SE=2) for this tree suggested higher admixture between Albion_Blotched-Oregon_Blotched, San Francisco_Blotched-Oregon_Blotched, Albion_Solid-Oregon_Solid, San Francisco_Solid-Albion_Solid, and San Francisco_Blotched-Albion_Blotched (Figure 4.5). The maximum likelihood tree produced with one migration event produced a shallow tree with the Oregon collections highly divergent from the other four with the most likely migration event occurring between Albion_Solid-San Francisco_Solid (Figure 4.6). The residual graph for the one migration event (SE=2.3) suggested migration events between Albion_Blotched-Oregon_Blotched, San Francisco_Solid-Oregon_Solid, San Francisco_Blotched-Oregon_Blotched, San Francisco_Solid-Albion_Solid, San Francisco_Blotched-San Francisco_Solid (Figure 4.6). The maximum likelihood tree for two migration events was similar to that of one migration event with less divergence between the San Francisco populations (Figure 4.7). The residual graph (SE=1.4) identified eight admixture events Oregon_Solid-Oregon_Blotched, Albion_Blotched-Oregon_Blotched, San Francisco_Blotched-Oregon_Blotched, Albion_Solid-Oregon_Solid, Albion_Blotched-Albion_Solid, San Francisco_Solid-Albion_Solid, San Francisco_Blotched-Albion_Blotched, San Francisco_Blotched-San Francisco_Solid (Figure 4.7). The maximum likelihood tree that allowed for three migration events produced a tree with strong divergence in the Oregon collections from the other four and pairing of the two San Francisco collections (Figure 4.8). The residual (SE=1) identified eight admixture events: Oregon_Solid-Oregon_Blotched, Albion_Blotched-Oregon_Blotched, San Francisco_Blotched-Albion_Blotched, Albion_Solid-Oregon_Solid, San Francisco_Solid-Oregon_Solid, Albion_Blotched-Albion_Solid, San Francisco_Solid-Albion_Solid, San Francisco_Solid-Albion_Blotched (Figure 4.8). The maximum likelihood tree that allowed for four migration events placed the Oregon_Solid collection as the most divergent than the other five collections which more closely reflected the morphotypic separation in the species (Figure 4.9). The residual was the lowest for all analyses (SE=0.1) with seven admixture events: Albion_Blotched-Oregon_Blotched, San Francisco_Solid-Oregon_Blotched, Albion_Blotched-Oregon_Solid, Albion_Solid-Oregon_Solid, San Francisco_Solid-Albion_Solid, San Francisco_Blotched-Albion_Solid, San Francisco_Blotched-Albion_Blotched (Figure 4.9). The weights for each of the four migration events were 34% (Albion_Solid to San Francisco_Solid), 36% (Basal location split from Oregon_Solid and rest of the tree to Oregon_Blotched, 62% (Albion_Solid to San Francisco_Solid), 43% (Albion_Blotched

to San Francisco_Blotched). The maximum likelihood tree for five migration events similar to the tree for four migrations isolated the Oregon collections from the other four (Figure 4.10). The residual for the tree was the second lowest (SE=0.5) with seven migration events not reflected in the tree: Oregon_Solid-Oregon_Blotched, San Francisco_Blotched-Oregon_Blotched, San Francisco_Solid-Oregon_Solid, Albion_Blotched-Albion_Solid, San Francisco_Solid-Albion_Solid, San Francisco_Blotched-Albion_Blotched, San Francisco_Blotched-San Francisco_Solid (Figure 4.10). The final maximum likelihood tree with six allowed migration events produced a tree with very deep branches from the San Francisco_Solid collection and the other five collections which had very shallow branches between them (Figure 4.11). The corresponding residual was the first to have a standard error of zero which indicated the highest number of migration events the data supported to produce a viable tree (Figure 4.11).

Discussion

My results show the utility of assembling a reduced representation genomic library with type IIB restriction enzymes coupled with current next generation sequencing platforms for non-model organisms for the production genome-wide sampling of SNPs. I also demonstrated the ability to construct a clustered assembly of genomic fragments of average uniform length as a reference for the identification of SNP loci across all populations. Until recently, the ability of researchers to identify a genome-wide collection of SNPs for population genomic studies require an extended processes of identification and confirmation across multiple populations and individuals. The 2bRAD technique demonstrates the ability for researchers to produce large genome-wide samples of variability of non-model organisms for the purpose of fine-scale divergence of populations

Population structure

Although my dataset sampled 21,318 SNPs from the the *S. mystinus* genome, not all of these sites are independent since some are on the same locus (restriction site fragment) and lead to a reduction in the overall sampling of the genome. Linked sites should not be a problem in estimating population divergence and structure considering that tens-of-thousands of loci were used to measure these parameters (Pritchard & Wen 2004). Using this set of 21,318 SNPs I was able to provide clear evidence for distinct genetic groups that correspond to the two morphotypes seen in *S. mystinus* regardless of proposed subdivision values.

At each location, I observed strong divergence between each morphotype confirming genomic divergence between groups. Specifically I found that the F_{ST} values between the two morphotypes were higher than those among morphotypes (Table 4.2). The divergence among morphotype collections across the three sampled locations have a similar trend where the furthest locations between Newport OR and San Francisco CA had the highest divergence values. The next closest populations also had a similar trend but the divergence among them was the lowest between Albion CA and Newport OR. The closest collections between Albion CA and San Francisco CA between morphotypes had the median value of divergence. These trends along with the F_{ST} values observed in

each morphotypic collection, suggest that the solid morphotype may have a limitation to dispersal through a shorter pelagic larval duration or timing of parturition has been selected to match specific physical oceanographic events that prevents similar gene flow patterns to that of the blotched morphotypes. A detailed in-field analysis would be required to determine whether differences in reproductive events, timing of parturition, or pelagic larval duration are influencing differences in divergence between morphotypes across their ranges. In a study on F1 hybrid analyses on *S. mystinus*, Burford et al. (2011) did not observe substantial evidence for the introgression between morphotypes where adults occur in roughly equal numbers, reinforcing the idea that the two morphotypes have diverged to the degree of speciation albeit not complete as seen in the San Francisco solid collection. The breakdown of reproductive isolating factors in the San Francisco solid collection correlates to the unidirectional hybridization of the blotched to the solid group on a genome-wide level as seen by the dense sampling of SNPs collected from this study. Identifying the portions of the genome that are introgressed into the solid morphotype genome are particularly interesting from the perspective of analyzing the degree of "porosity" and the identification of any new adaptive variation introduced by the hybridization.

Identification of candidate outlier loci

In my study, the 209 candidate outlier SNP loci represent genomic regions responding to divergent selection or regions tightly linked to genomic regions responding to strong selection. Although these candidate outlier loci are preliminary and the independent identification of each outlier using different methods would provide confirmation of the outlier status, additional methods are prone to sensitivity identifying a larger number of outliers that are most likely false positives. Candidate outlier loci detected in this study most likely represent genomic regions that are highly differentiated, but since this preliminary analyses identifies outliers in six populations between a pair of morphotypes results provide a set of genomic regions potentially involved in divergence influencing or maintaining reproductive isolation between the morphotypes. The fact that I did not detect any outliers diverging due to balancing selection is not surprising since the sampled genomes across morphotypes most likely represent two groups that are diverged to the level of species and balancing selection would be more appropriate to identify within each morphotype across multiple populations. Since I used SNP loci to sample variation from the genome balancing selection due to heterozygote advantage or ongoing frequency-dependent selection would have to be extremely exaggerated or directly linked to pathways related to genes coding for antigens (Amambua-Ngwa et al. 2012; Gokcumen et al 2013). Future analyses between morphotypes focused on identifying genomic regions linked to divergence and speciation should focus on collections of the two types across their ranges and confirm whether the identified loci in this study are still considered outliers.

Considering the number of candidate outlier loci identified from the 21,318 SNP loci, I identified that approximately 1% of sampled loci were candidate outlier loci. The ability to detect candidate outlier loci involved in divergence during speciation while using a RAD-tagged library presents unique circumstances that may leave actual regions that are involved in divergence unaccounted for (Keller et al. 2013). First, the restriction

site must be near the actual target of selection or close enough such that recombination does not break up these combinations. Second, SNPs must be present within the restriction fragment at a frequency detectable by the thresholds established prior to genotyping. Third, if a SNP is present and tightly linked to a target of selection then sequencing of that fragment must be present in numbers high enough to provide sufficient read depth based on the researcher's criteria.

Admixture and demographic history

My results from the TreeMix analysis gave mixed results but the optimal tree with imposed migration events confirmed gene flow between types with little to no admixture between morphotypes. Considering the Structure analyses I suspected a signal of migration from the blotched morphotype into the San Francisco solid collection, but this was not the case given the migration events plotted on the tree. When analyzing the corresponding residual graph, the San Francisco solid population appeared to be too distant on the tree with highest probability suggesting that the relationship between the collections maintained a small signal of gene flow between them. To fully determine the historical patterns of population splits and admixture events from genome-wide data in *S. mystinus* I would require a close outgroup (e.g. *S. entomelas* Hyde & Vetter 2007). Adding an outgroup and additional populations from the *S. mystinus* distribution would assist in resolving the relationships between these recently diverged species as the currently sampled collections does not have sufficient resolution to identify admixture events between morphotypes. This would provide a more accurate prediction of the historical relationships between the different morphotypes. Additionally, the current program only accepts genotypes from SNPs with two alleles present in the sampling of populations. This limitation in data restricts more accurate measurements of distribution of genomic variation between these diverged groups by removing potentially informative SNP loci.

Conclusions

I have developed the first RAD-tagged DNA sequencing dataset in the *Sebastes* genus to investigate the genomic divergence between the two recently diverged morphotypes of *S. mystinus*. This dataset allowed me to identify candidate genomic regions linked to genes involved in the speciation process. The rapid speciation in the genus has been described as a species flock and suggests intricate evolutionary processes that most likely include a combination of initial divergence in allopatry with ecological differences influencing localized adaptive divergence reinforcing pre-zygotic barriers. My results also identify the incomplete isolation between morphotypes in the San Francisco solid collection. A comparative genomic approach to the densely annotated genome of the closely related *S. aleutinanus* will provide starting points in the genome to associate the divergence of outliers to molecular pathways influencing behaviors or adaptations that maintain reproductive isolation between the morphotypes.

Table 4.1 Measured values of heterozygosity and proportion of SNP loci out of HWE from 21,318 SNP loci

	Oregon-solid	***Oregon-blotched***	***Albion-solid***	***Albion-blotched***	***San Francisco-solid***	***San Francisco-blotched***
Population-observed heterozygosity	0.131	0.146	0.123	0.131	0.213	0.195
Range-observed heterozygosity	0-1	0-1	0-1	0-1	0-1	0-1
Range-expected heterozygosity	0.086-0.625	0.079-0.611	0.079-0.625	0.079-0.625	0.079-0.625	0.079-0.625
Proportion of loci out of HWE ($p<0.01$)	0.80%	0.90%	1.20%	1%	2%	2.70%
Proportion of loci out of HWE ($p<0.05$)	5.50%	3.90%	4.80%	4.30%	6.80%	13%

Table 4.2 Pairwise F_{ST} values for all collections used in this study

	Oregon solid	Oregon blotched	Albion solid	Albion blotched	San Fran. solid	San Fran. blotched
Oregon solid	-					
Oregon blotched	0.282	-				
Albion solid	0.012	0.286	-			
Albion blotched	0.3	0.027	0.283	-		
San Fran. solid	0.15	0.172	0.137	0.152	-	
San Fran. blotched	0.321	0.099	0.31	0.059	0.095	-

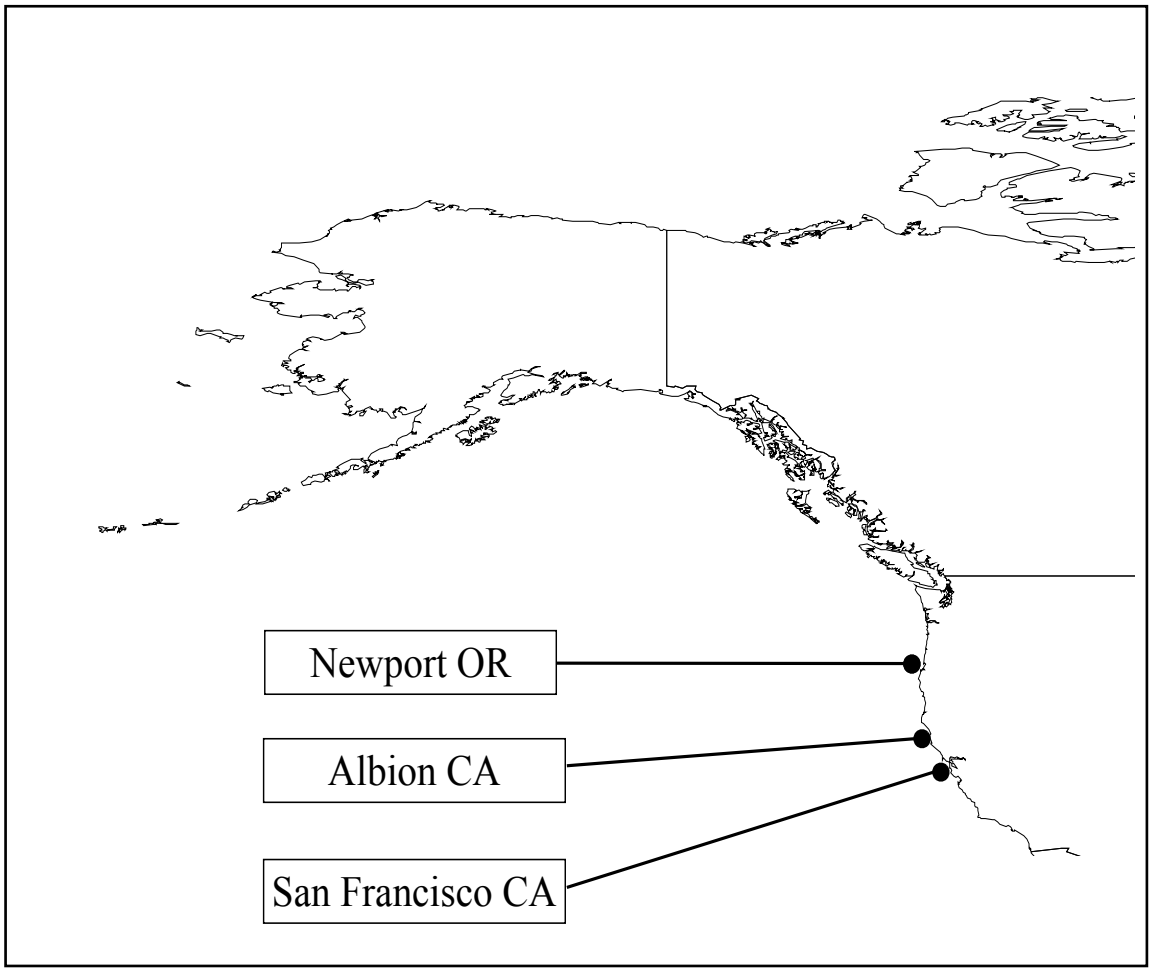


Figure 4.1 Sampled locations for *S. mytinus*. Both morphotypes were collected at each site

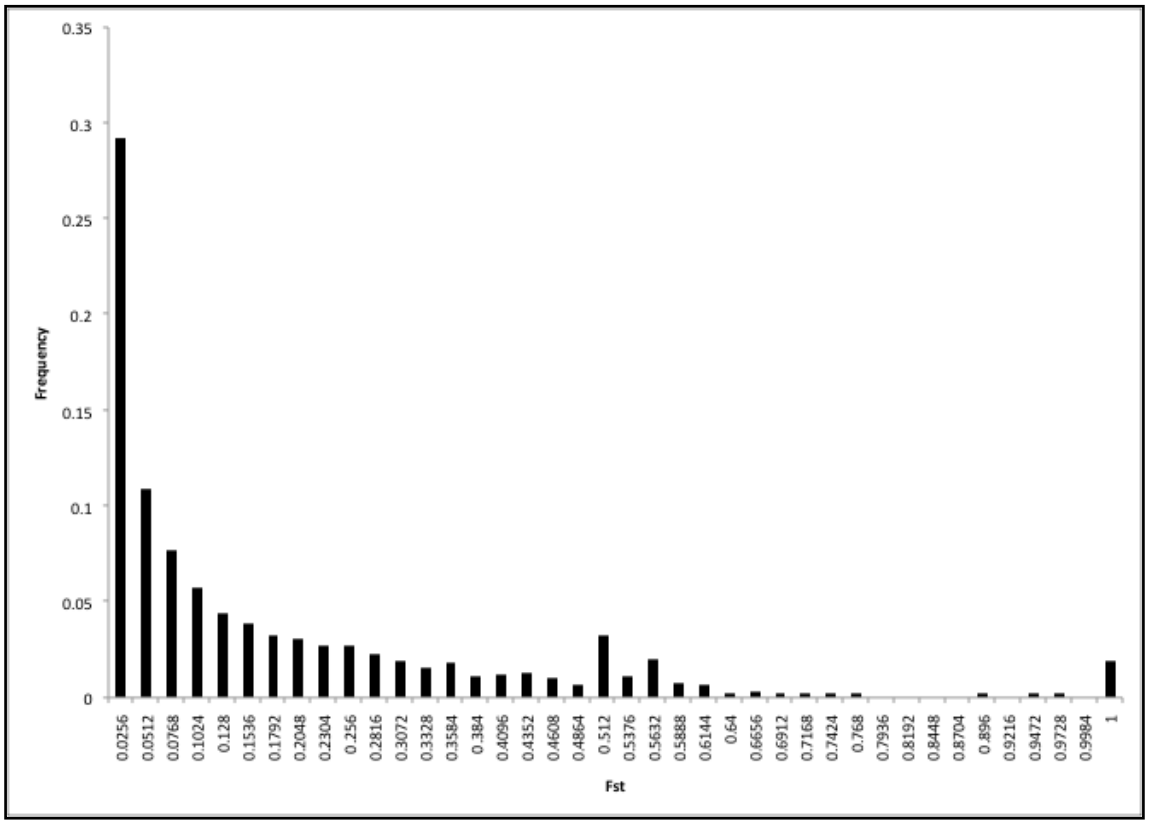


Figure 4.2 Distribution of F_{ST} values between all locations and morphotypes for all 21,318 SNP loci used in population analyses.

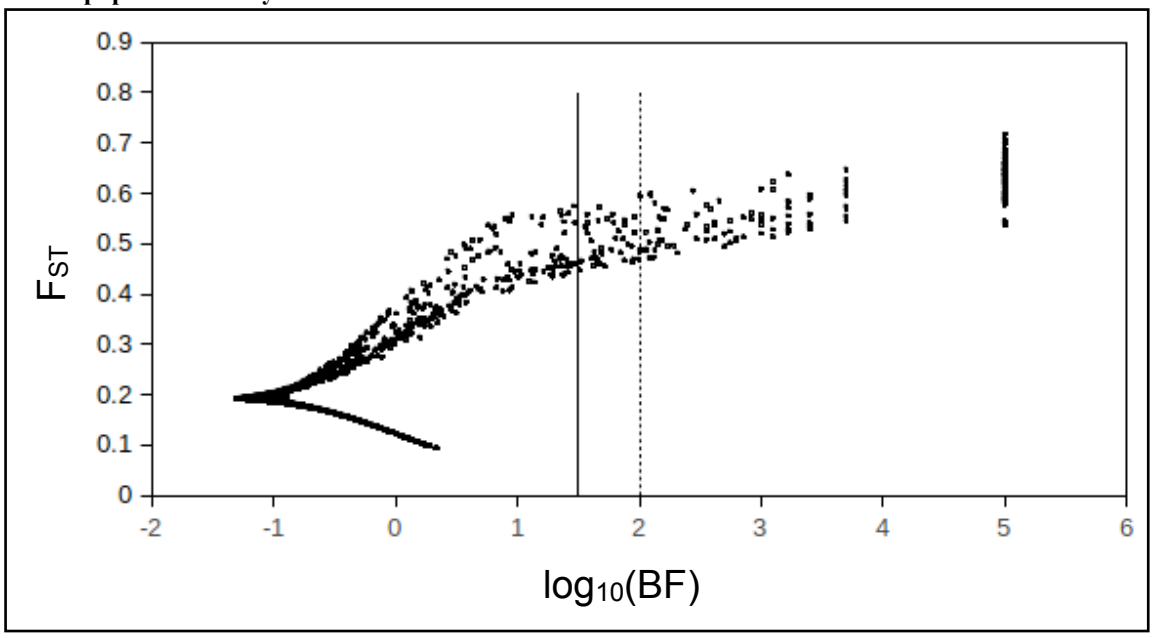


Figure 4.3 Outlier detection method BayeScan results from all collections and locations. The solid and dashed line corresponds to "very strong" (1.5 ≤BF≤2), and "decisive" (BF≤2) outlier cutoff values respectively.

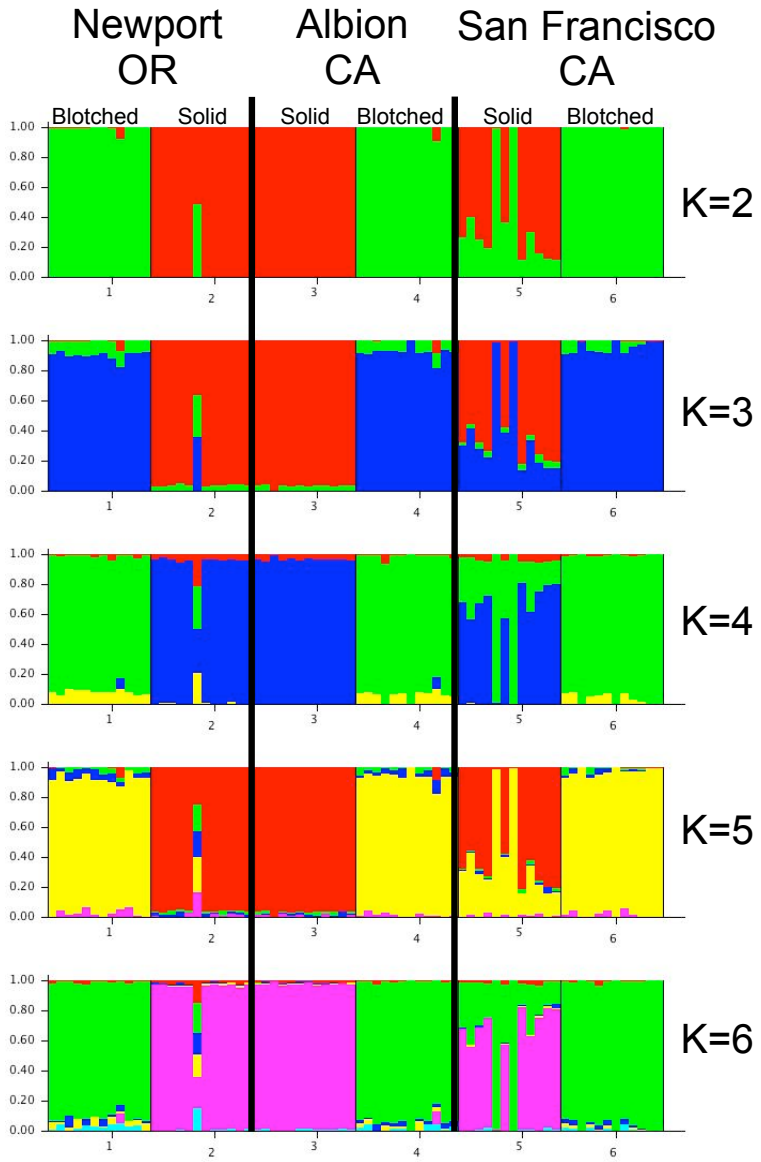


Figure 4.4 Structure plots for K's between 2-6 for all individuals collected using 21,318 SNP loci. Each vertical bar represents an individual from a collection. The segregation of colors within each vertical bar corresponds to the probability of assignment to each respective genetic cluster.

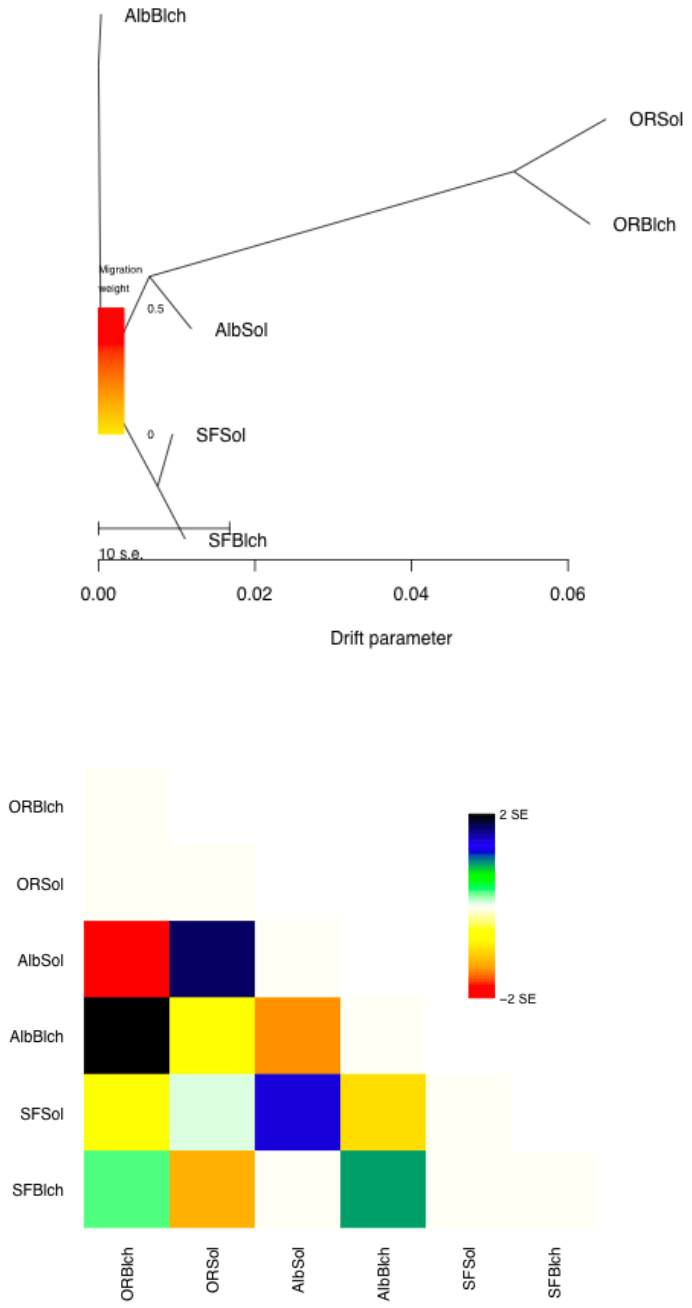


Figure 4.5 Results from the TreeMix analysis for zero migration events. (Top) Maximum-likelihood tree of six collections from the two morphotypes in *S. mystinus*. Abbreviations of each collection are as follows: Oregon: OR, Albion: Alb, San Francisco: SF; Abbreviations for each morphotype are as follows: Solid: Sol, Blotched: Blch. (Bottom) Residual fit for the corresponding tree, residuals above zero are pair-wise candidates for admixture.

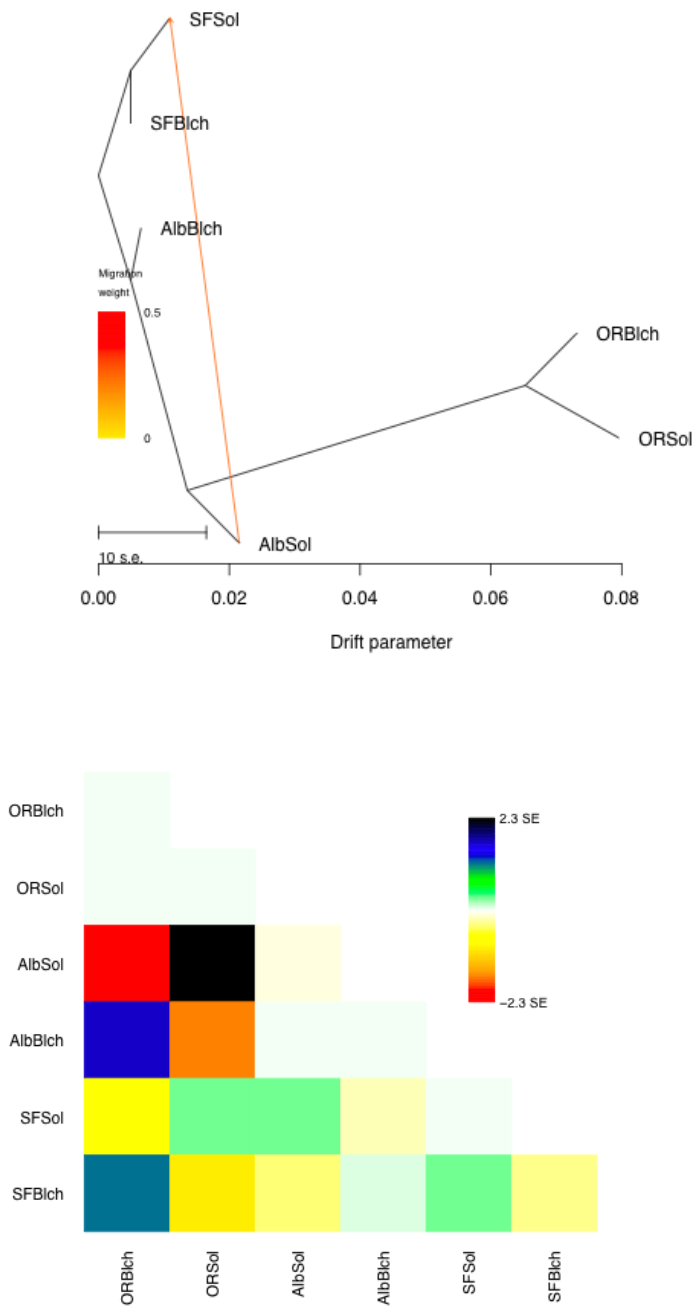


Figure 4.6 Results from the TreeMix analysis for one migration event. (Top) Maximum-likelihood tree of six collections from the two morphotypes in *S. mystinus*. Abbreviations of each collection are as follows: Oregon: OR, Albion: Alb, San Francisco: SF; Abbreviations for each morphotype are as follows: Solid: Sol, Blotched: Blch. (Bottom) Residual fit for the corresponding tree, residuals above zero are pair-wise candidates for admixture.

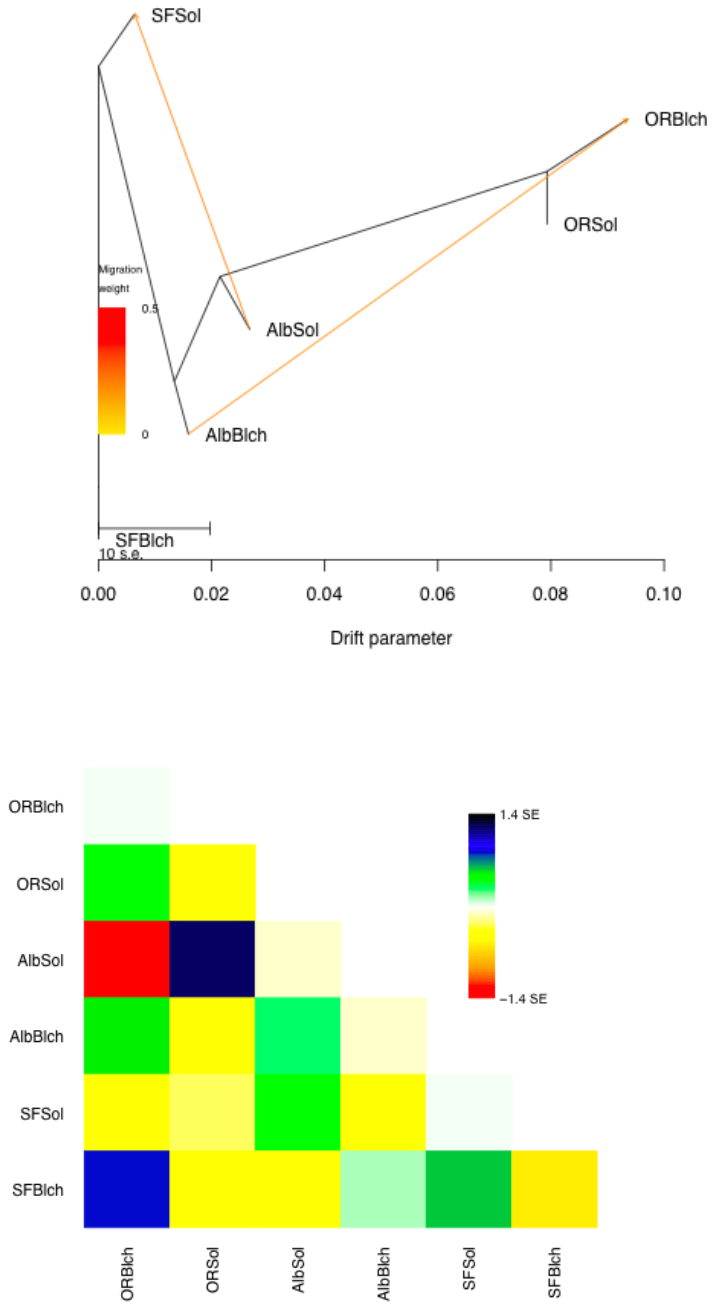


Figure 4.7 Results from the TreeMix analysis for two migration events. (Top) Maximum-likelihood tree of six collections from the two morphotypes in *S. mystinus*. Abbreviations of each collection are as follows: Oregon: OR, Albion: Alb, San Francisco: SF; Abbreviations for each morphotype are as follows: Solid: Sol, Blotched: Blch. (Bottom) Residual fit for the corresponding tree, residuals above zero are pair-wise candidates for admixture.

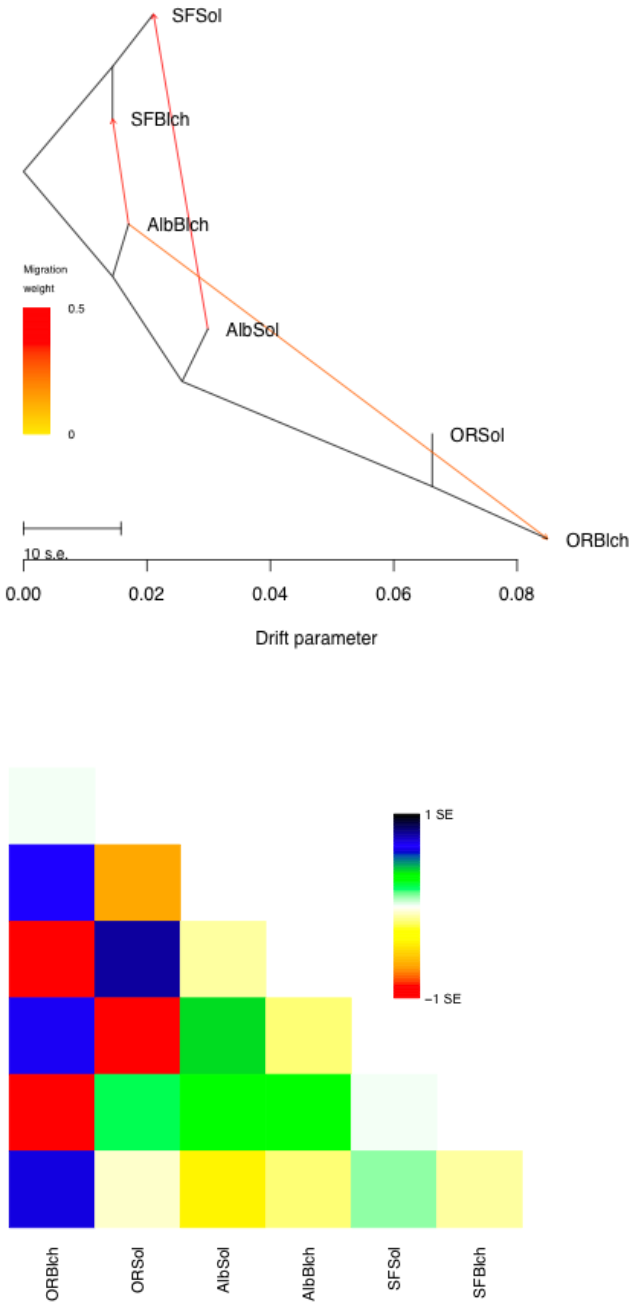
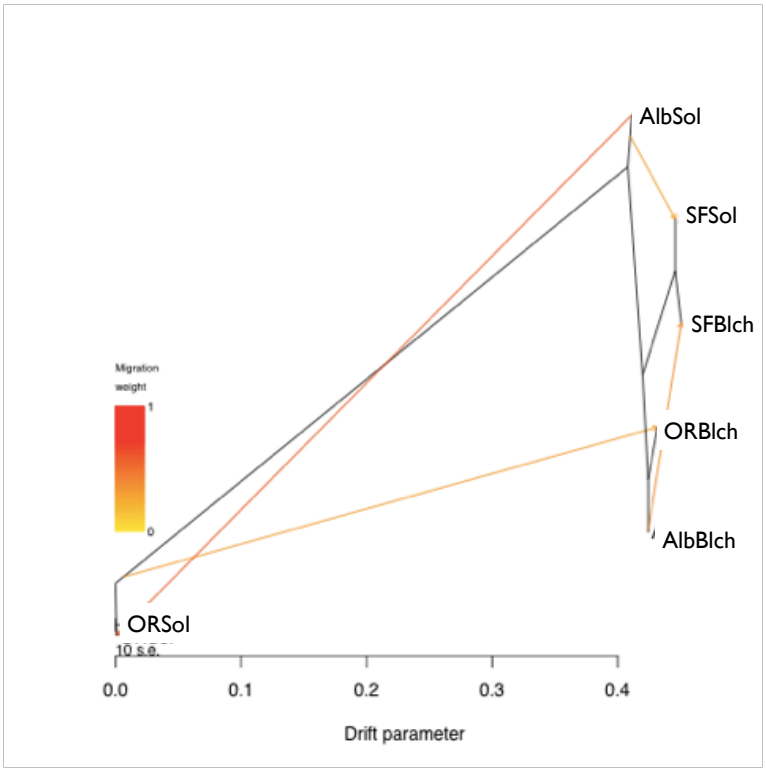


Figure 4.8 Results from the TreeMix analysis for three migration events. (Top) Maximum-likelihood tree of six collections from the two morphotypes in *S. mystinus*. Abbreviations of each collection are as follows: Oregon: OR, Albion: Alb, San Francisco: SF; Abbreviations for each morphotype are as follows: Solid: Sol, Blotched: Blch. (Bottom) Residual fit for the corresponding tree, residuals above zero are pair-wise candidates for admixture.



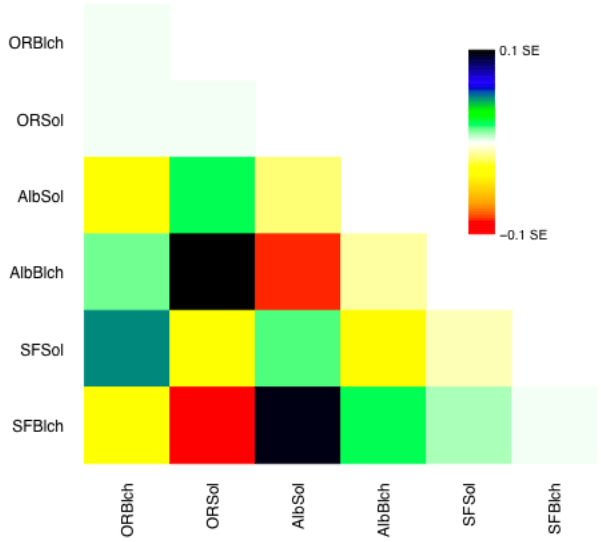


Figure 4.9 Results from the TreeMix analysis for four migration events and the best fit based on the smallest residuals. (Top) Maximum-likelihood tree of six collections from the two morphotypes in *S. mystinus*. Abbreviations of each collection are as follows: Oregon: OR, Albion: Alb, San Francisco: SF; Abbreviations for each morphotype are as follows: Solid: Sol, Blotched: Blch. (Bottom) Residual fit for the corresponding tree, residuals above zero are pair-wise candidates for admixture.

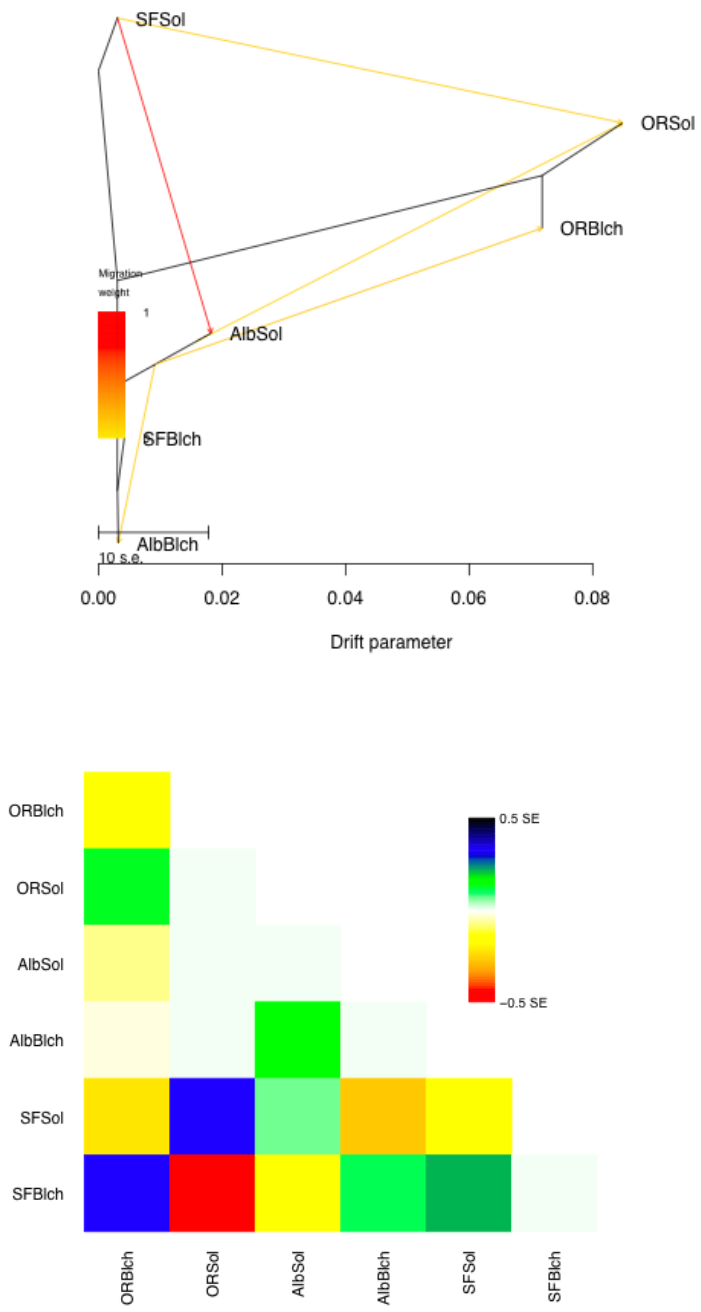
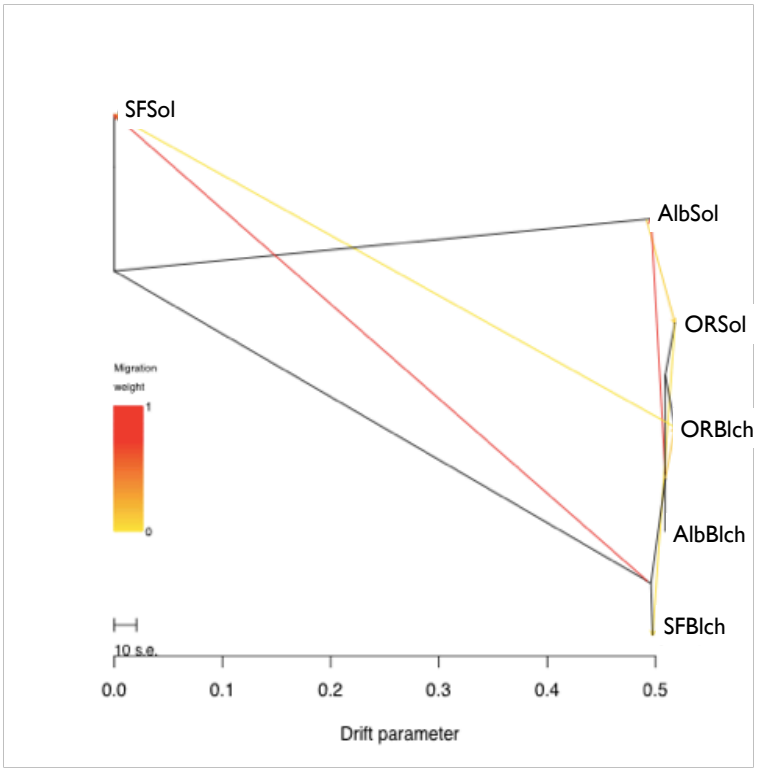


Figure 4.10 Results from the TreeMix analysis for five migration events. (Top) Maximum-likelihood tree of six collections from the two morphotypes in *S. mystinus*. Abbreviations of each collection are as follows: Oregon: OR, Albion: Alb, San Francisco: SF; Abbreviations for each morphotype are as follows: Solid: Sol, Blotched: Blch. (Bottom) Residual fit for the corresponding tree, residuals above zero are pair-wise candidates for admixture.



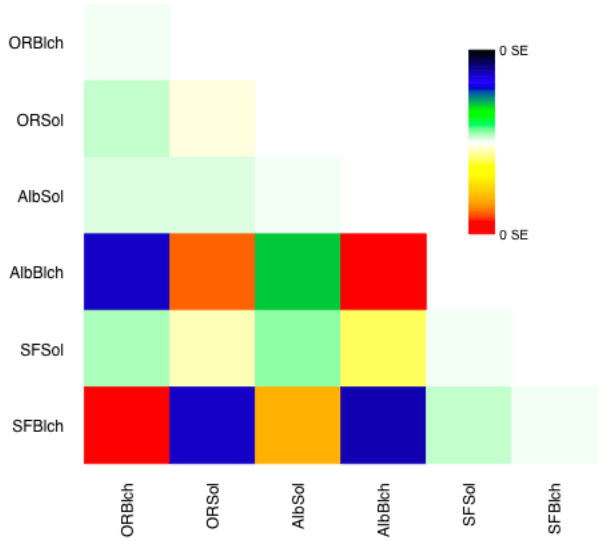


Figure 4.11 Results from the TreeMix analysis for six migration events. (Top) Maximum-likelihood tree of six collections from the two morphotypes in *S. mystinus*. Abbreviations of each collection are as follows: Oregon: OR, Albion: Alb, San Francisco: SF; Abbreviations for each morphotype are as follows: Solid: Sol, Blotched: Blch. (Bottom) Residual fit for the corresponding tree, residuals above zero are pair-wise candidates for admixture.

Chapter 5: Conclusions

I performed the first genome scan within the genus *Sebastes* to identify the effects of natural selection on a highly diverse group of marine fish characterized by a rapid adaptive radiation from the Atlantic to the Pacific Ocean and subsequent adaptation to multiple novel environments within the Pacific. The mechanisms involved in adaptation and ultimately speciation can be observed along the boundary between the ecology of the environment and the genomic responses to selective pressures (Hohenlohe et al. 2010; Nosil & Feder 2012; Via 2012; Keller et al. 2013). Ecological adaptation to the local environment is a key initial step in the evolution of reproductive isolation between populations and can have a large degree of influence on genomic regions involved in functional genes (Hemmer-Hansen 2007; Nielsen et al. 2009; Via 2009; Pespeni & Palumbi 2013). The ultimate goal of studying such a system is identifying the ecological pressures that interact with functional genomic regions and their effects on phenotypic variation that may lead to reproductive isolation.

Identifying the interaction between environment and genome for the purpose of locating genomic regions responsible for rapid adaptation in novel or varying environments is a complicated process, even with model organisms. Fortunately, the rapid advances in genome sequencing technologies provides researchers with a wealth of genomic sequences that can be applied to mapping of genomic regions involved in adaptation and differentiation of populations. The highly differentiated candidate markers identified in my dissertation between populations and across species within a subgenus provide key regions for further fine scale mapping using next-generation sequencing and comparative genomics with closely related teleost species. The use of genome scans to identify candidate genomic regions responding to selection are not meant as ultimate goals of studies, due to the assumptions of adaptive evolution on the target species and historical evolutionary events that may have shaped the genomic diversity of derived populations. Mutation rates and recombination of homologous regions of a population may produce patterns similar to those produced by selection and must be approached with caution before absolute statements of divergence can be concluded (Hohenlohe et al. 2010). It is more appropriate to consider candidate outlier loci as preliminary genomic regions warranting further exploration through targeted sequencing and using candidate outliers as testable hypothesis to the observed divergence between populations or species collections.

The evolutionary processes leading to increased divergence between populations in response to local adaptation may not always associate to ecological features in the local environment, but may be a result of genetic incompatibilities between populations coinciding with barriers due to ecological differences (Bierne et al. 2011). The decoupling of genetic incompatibilities and environmental responses have been generally ignored in genome scan approaches, but this can produce spurious outliers believed to be diverging due to selection when in fact they are diverging due to incompatibilities between loci adapted to local habitats. Much of the literature describing this phenomenon has been observed in the in the hybrid literature along hybrid zones between species and can be implemented on my current analyses of divergence due to the potential of local adaptation across a species' distribution. Although the degree of divergence between hybrids is much greater than what would be detected between

populations, it is important to acknowledge that divergence via local adaptation should not be the deferred explanation of outlier loci. Other explanations ignored in outlier detection analyses should incorporate whether selective pressures producing the divergence are current or historical, the direct linkage to the selected gene or genes is a result of environmental differences, and "hybrid" fitness depression between populations (Bierne et a. 2011). What may explain the genetic structure previously attributed to local adaptation by ecological differences is that many loci exhibiting genetic structure may be due to many loci experiencing small but significant degrees of incompatibilities coinciding with an environmental boundary (Bierne et al. 2011). In my study where marine boundaries vary spatially and temporally at different times of the year, leading to ecological differences influencing genes directly and other linked genomic regions, the direct causation of elevated divergence of each outlier would require further investigation to separate the genetic and ecological influences. Alternative hypothesis have been proposed to explain the causation of outliers include the historical demographics of the populations, the genetic interactions of barriers to gene flow, the interactions of selected loci in different populations, and the interaction between the directly selected gene and the neutral variation flanking it (Bierne et al. 2011).

The evolutionary question addressing speciation and the genomic responses to reproductive isolation, in terms of the number of regions involved in the speciation process, is still inconclusive and variable (Nosil et al. 2012). Genomic studies that have aimed to identify genomic regions differentiating due to speciation in divergent lineages of *Anopheles* identified a very low number of regions while later sequencing revealed many more than originally identified (Turner et al. 2005; Lawniczak et al. 2010). This finding highlights the descriptions of the genomic "continent" versus "island" model describing initially in the case for *Anopheles* where very few loci were identified to have diverged due to selection and little to no signal of divergence would be identified around these strongly selected regions. In the latter case, selection on the regions around the divergent genomic region many more loci than what is expected in the island model are affected beyond neutral expectations producing an "island" of the genome diverging due to selection. In my dissertation, the number of loci identified in the between species comparison of *S. flavidus* and *S. melanops* identified 13% and 6% candidate outlier loci from each respective dataset and approximately 1% of SNP loci from a much larger dataset producing a much larger sampling of the genome. Ideally the candidate outlier loci identified from each dataset would have gone through sufficiently stringent criteria but further analysis would be required to determine with certainty whether the candidate outlier loci are located near genes or other targets of selection. In addition to the number of candidate loci identified with outlier detection methods, the imprint or the size of the "continent" displaying the degree of divergence may span a few hundred kilobases (Buonaccorsi et al. 2011; Nadeau 2012), but instances where the nearest known quantitative trait locus is much further have been documented (Renaut et al 2012; Via et al. 2012).

The development of genomic information for *Sebastes* is still in its early stages, which makes the identification of the genes diverging in response to selection or the genetic mechanisms producing the outlier remains a difficult task. The size and location of genomic regions diverging due to selection will reveal the relative location of the gene

to the candidate outlier, additional genes within the chromosomal vicinity, and the strength of selection on the gene itself. As a future approach to identify the targets of selection linked to candidate outlier loci would involve increasing the sampled genomic region and comparative genomic mapping to fully mapped or annotated teleost genomes. I will utilize a preliminarily annotated *Sebastes* genome developed from *S. aleutianus* (Evgrafov unpublished data) to first increase the sequence length of the regions around the candidate outlier loci to analyze a larger portion from the *Sebastes* genome and identify functional genomic regions in other teleost genomes.

To determine the degree of introgression between candidate outlier loci relative to neutrally evolving genomic regions I would implement coalescent based simulations that would highlight the maintenance of divergence at candidate outlier regions in the face of gene flow. Such methods can model population divergences, gene flow, and the outcomes of founder effects which will have marked genetic variation signatures in the genome. Programs that model historical and contemporary demographic effects, such as IMa2 (Hey 2011), can model these demographic features of multiple diverging populations between selected and neutral genomic regions. Additionally, approximate Bayesian computation (ABC) can also model demographic histories through time to assess differences in genetic variation between selected and neutral genomic regions by summarizing population statistics and fitting the model that best fits the variation in the data (Beaumont et al. 2002).

The environmental and ecological variables associated with the genetic divergence identified in *S. flavidus* and *S. melanops* is a first approach to understand the interaction between the genomic responses to selection. If environmental factors have influenced the candidate outlier loci, it is likely that other factors aside from those selected influence their divergence. Although the sampled environmental variables are crude (sea surface temperature and chlorophyll concentrations), they provide an initial glimpse into environmental factors influencing genetic structure. Another difficulty in correlating environmental variables to genetic structure is that when comparing multiple populations the environmental factor associated with the candidate outlier may not produce the same patterns in other locations. For example, in one location temperature may associate the best with an outlier locus while in another location habitat preferences may influence a different outlier. Fine scale differences such as timing of larval release and selection during the larval phase may also contribute to the association with environmental variables, but not directly explain the relationship. In order to substantiate the effects of environmental factors influencing divergence in the marine environment, I would include environmental variables from smaller local scales such as larval release ranges, suitable habitat for recruitment or larval behavior increasing successful settlement.

A major improvement on these studies in respect to the effect of pelagic larval duration on population connectivity in a widely dispersing marine organism would be to collect an increased number of individuals from currently sampled locations and increase the number of sampled locations. Sampling an increased number of individuals and locations would provide additional representation of genomic variation as well as an increase in possible selective forces contributing to local adaptation within the respective environments. Given that I used two species with different pelagic larval durations that

differ by one month and spawn at different times of the year may not provide a sufficiently convincing relationship between pelagic larval duration, gene flow, and local adaptation. Each species has the potential to cross large distances and with such long pelagic larval durations and the role ecological features have in each of their respective niches, the relationship between pelagic larval duration and local adaptation is too weak to confidently state with a limited dataset that increased gene flow significantly influences the number of identified outliers. This may lead to the false identification of loci as outliers that are not related to local adaptation as a result of gene flow due to the similarities in pelagic larval duration between the two species. A more appropriate analysis would consist of sampling multiple populations of different species within the genus with different pelagic larval durations and ecological niches. This approach would specifically strengthen the suggestion that longer pelagic larval duration (higher migration) limits local adaptation, which reduces the number of loci identified with outlier detection methods. As a supplement to the anonymous and EST microsatellite markers used in this study, a RAD-tagged sequencing approach across multiple species with variable pelagic larval durations would provide a sufficient dataset to clarify the role gene flow has on local adaptation in the marine environment.

Genome-wide sampling is now possible with the advantage of reduced representation libraries through RAD-tagged sequencing methods for non-model organisms. The dense sampling of the genome has allowed me to sample a larger portion of the *S. mysintus* genome to identify regions of the genome possibly involved in speciation and begin to describe historical demographic events that produced the current patterns of reproductive isolation between the recently diverged morphotypes. The data I produced with this study serves as a preliminary analysis into the understanding of how the genome responds during speciation in the marine environment. The RAD-tagged sequencing method sampled thousands of potentially informative variable sites from a short fragment at adequate coverage rates across collections. The fragments produced from the construction of the library and the subsequent assemblies into reference loci provide a short but practical sequence that I will use to align against a preliminarily annotated genome from *S. aleutianus* (Evgrafov unpublished data). Currently the reference genome does not contain detailed information about chromosomal location or known interactions with other genomic regions, but it does serve as an initial source of genomic information that I can mine and extract for further comparative genomic analyses with other more densely mapped and annotated genomes available. Those candidate outliers identified in both studies presented here provide an initial set of sequences for comparative analyses. Ideally the markers will assist in identifying functional genes and relative locations in other teleost genomes. In addition to locating the gene in another genome, identification will aid in future development of fine-scale mapping of genes that code for morphological or life history traits that may play a critical role in the initial genes that experience strong divergent selection in a novel habitat. Identification of these genes can provide possible explanations to local adaptation and possible speciation genes, which can possibly reveal genomic mechanisms that explain the rapid adaptive radiation in *Sebastes*. The identification of the genes can also provide evolutionary biologists studying genome evolution with insights on how the organization, duplication and conservation of particular regions are arranged.

The data and conclusions from my dissertation have added to the understanding of population divergence and speciation in the marine environment. Additionally, the number of outlier loci identified between *S. melanops* and *S. flavidus* highlight possible genomic regions that may be directly involved in adaptive divergence, although these loci would require subsequent genomic analyses to determine their roles in divergence within each species. Although many of the tests attempting to relate PLD and population divergence where not significant, there are signals of population divergence in the interspecies comparison as suggested by F_{ST} and G'_{ST} values. Alternate analyses would investigate the degree of population structure in these species to associate the relationship between gene flow via PLD and the degree of population structure. The exact genomic mechanisms involved in the observed population divergence is not clear, but my data provides a set of loci associated with distinct ecological features across the distribution of these two species. In order to accurately associate the relationship between PLD and local adaptation I would sample a larger portion of the genus with population-level sampling of species with varying degrees of PLD across similar ranges, which would provide stronger evidence for the relationship rather than the two species analyzed.

The data collected from the incipient species pair of *S. mystinus* provides another step in identifying the fine-scale genomic interactions involved in speciation within the marine environment. Addressing the genomic processes driving divergence towards speciation in natural populations should assist in identifying the effects selection has on the genome. Increased sampling of the two morphotypes across their ranges will identify how genomic divergence across recently diverged species. The genomic data set represents another step in identifying the degree of gene flow between incipient species at particular genes which do not appear to be influenced by the divergence influencing the rest of the genome.

Natural or artificial ecological impacts have profound impacts on the genomic structure and gene frequencies of populations. With the advances in genomic information, detecting selection on a genomic level allows one the opportunity to identify specific genomic regions responding to selection. The *Sebastes* genus is a unique system for evolutionary biologists to study questions in evolution that range from adaptation to speciation. Due to many species in the genus characterized by long life spans, the effects of fishery-induced selection on the genome would not be seen for some time and my have permanent results on the genome. The identification of genomic regions that are sensitive to fishery-induced selection is an important aspect towards understanding outcomes of intense selection on marine populations. This information can be used to determine how a population will respond to selective pressure by fisheries or extended periods of habitat alteration. A better assessment of population health should take a genomic approach, as proposed here, using a larger sample of genomic data to provide an accurate representation of response to selective pressure across multiple genes. Undoubtedly, it is a system that remains to be observed through a genomic perspective and will serve as an example of a systems biology approach incorporating the fields of evolutionary biology, genomics, management, and ocean sciences.

References

Addison JA, Ort BS, Mesa KA, Pogson GH (2008) Range-wide genetic homogeneity in the California sea mussel (Mytilus californianus):a comparison of allozymes, nuclear DNA markers, and mitochondrialDNA sequences. Mol Ecol 17:4222–4232.

Amambua-Ngwa A, Tetteh KK, Manske M, Gomez-Escobar N, Stewart LB, Deerhake ME, ... Conway DJ (2012). Population genomic scan for candidate signatures of balancing selection to guide antigen characterization in malaria parasites. PLoS genetics, 8(11), e1002992.

Ammann AJ (2004). SMURFs: standard monitoring units for the recruitment of temperate reef fishes. Journal of Experimental Marine Biology and Ecology,299(2), 135-154.

Amores A, Catchen J, Ferrara A, Fontenot Q, Postlethwait JH (2011) Genome evolution and meiotic maps by massively parallel DNA sequencing: spotted gar, an outgroup for the teleost genome duplication. Genetics, 188, 799–808.

An HS, Park JY, Kim MJ, Lee EY, Kim KK (2009) Isolation and characterization of microsatellite markers for the heavily exploited rockfish, and cross-species amplification in four related spp. Conservation Genetics,10(6), 1969-1972.

Baird NA, Etter PD, Atwood TS, Currey MC, Shiver AL, Lewis ZA, ... Johnson EA (2008). Rapid SNP discovery and genetic mapping using sequenced RAD markers. PloS one, 3(10), e3376.

Bakun A (1996) Patterns in the ocean: ocean processes and marine population dynamics. California Sea Grant College System, National Oceanic and Atmospheric Administration, University of California at San Diego, La Jolla, Ca.

Beaumont MA, Balding DJ (2004) Identifying adaptive genetic divergence among populations from genome scans. Molecular Ecology 13: 969-980.

Beaumont MA, Nichols RA (1996) Evaluating loci for use in the genetic analysis of population structure. Proceedings of the Royal Society London 263: 1619-1626

Beaumont, MA, Zhang W, Balding D J (2002) Approximate Bayesian computation in population genetics. Genetics, 162(4), 2025-2035.

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

Bernatchez L, Renaut S, Whiteley AR, Derome N, Jeukens J, Landry L, ... & St-Cyr J (2010). On the origin of species: insights from the ecological genomics of lake whitefish. Philosophical Transactions of the Royal Society B: Biological Sciences, 365(1547), 1783-1800

Bernson EA, Moran P (2009) The utility and limitations of genetic data for stock identification and management of North Pacific rockfish (Sebastes spp.) Rev. Fish Bio. Fisheries 19:233-247.

Benson G. (1999) Tandem repeats finder: a program to analyze DNA sequences. Nucleic Acids Res. 27(2): 573-580.

Bierne N, Welch J, Loire E, Bonhomme F, David, P (2011). The coupling hypothesis: why genome scans may fail to map local adaptation genes. Molecular Ecology, 20(10), 2044-2072.

Bierne N, Roze D, Welch JJ. (2013) Pervasive selection or is it…? why are Fst outliers sometimes so frequent? Mol Ecol 22 2061-2064.

Bobko, SJ, Berkeley, SA (2004) Maturity, ovarian cycle, fecundity, and age-specific parturition of black rockfish (Sebastes melanops). Fishery Bulletin,102(3), 418-429.

Bonin A, Taberlet P, Miaud C, Pompanon F (2006) Explorative genome scan to detect condidate loci for adaptation along a gradient of altitude in the common frog (Rana temporaria). Molecular Biology and Evolution 23(4): 773-783.

Bonin A, Taberlet P, Miaud C, Pompanon F (2006) Explorative genome scan to detect condidate loci for adaptation along a gradient of altitude in the common frog (Rana temporaria). Molecular Biology and Evolution 23(4): 773-783.

Buonaccorsi VP, Kimbrell CA, Lynn EA, Vetter RD (2002) Population structure of copper rockfish (Sebastes caurinus) reflects postglacial colonization and contemporary patterns of larval dispersal. Can J Fish Aquat Sci 59:1374–1384

Buonaccorsi VP, Narum SR, Karkoska KA, Gregory S, Deptola T, Weimer AB (2011). Characterization of a genomic divergence island between black-and-yellow and gopher Sebastes rockfishes. Molecular ecology, 20(12), 2603-2618.

Buonaccorsi VP, Westerman M, Stannard J, Kimbrell CA, Lynn EA, Vetter RD (2004) Molecular genetic structure suggests limited larval dispersal in grass rockfish, Sebastes rastrelliger. Mar. Biol. 145, 779–788.

Burford MO (2009) Demographic history, geographical distribution and reproductive isolation of distinct lineages of blue rockfish (Sebastes mystinus), a marine fish with a high dispersal potential. J. Evol. Biol. 22, 1471–1486.

Burford MO, Bernardi G (2008) Incipient speciation within a subgenus of rockfish (Sebastosomus) provides evidence of 2 recent radiations within an ancient species flock. Mar Biol (Berl) 154(4):701–717.

Burford MO, Bernardi G, Carr MH (2011). Analysis of individual year-classes of a marine fish reveals little evidence of first-generation hybrids between cryptic species in sympatric regions. Marine biology, 158(8), 1815-1827.

Burford MO, Larson RJ (2007) Genetic heterogeneity in a single year-class from a panmictic population of adult blue rockfish (Sebastes mystinus). Mar Biol (Berl) 151:451–465.

Burton RS (1998) Intraspecific phylogegraphy across the Point Conception biogeographic boundary. Evolution 52: 734-745.

Butlin RK (2008) Population genomics and speciation. Genetica online Sep 6.

Cann RL, Stoneking M, Wilson AC (1987) Mitochondrial DNA and human evolution. Nature 325: 31–6.

Carr M, Syms C (2006) Recruitment. In: Allen LG, Pondella DJ, Horn MH (eds) The ecology of marine Wshes: California and adjacent waters. University of California Press, Berkeley, pp 411–427

Cavalli-Sforza LL, Edwards AW (1967) Phylogenetic analysis. Models and estimation proce-dures. Am J Hum Genet 19: 233–57.

Checkley Jr DM, Alheit J, Oozeki Y, Roy C 2009a Climate Change and Small Pelagic Fish. Cambridge University Press, Cambridge. pp. 372.

Checkley Jr DM, Barth JA 2009b Patterns and processes in the California Current System. Progress in Oceanography, 83(1), 49-64.

Conesa A, Götz S, García-Gómez JM, Terol J, Talón M, Robles M (2005) Blast2GO: A universal tool for annotation, visualization and analysis in functional genomics research. Bioinformatics 21: 3674-3676

Cope JM (2004). Population genetics and phylogeography of the blue rockfish (Sebastes mystinus) from Washington to California. Canadian Journal of Fisheries and Aquatic Sciences, 61(3), 332-342.

Cowen RK, Lwiza, KMM, Sponaugle S, Paris CB, Olson DB (2000) Connectivity of marine populations: Open or closed? Science 287 (5454) 857-859.

Coyne JA, Orr HA (2004) Speciation. Sunderland, MA: Sinauer Associates.

Daníelsdóttir AK, Gíslason D, Kristinsson K, Stefánsson M, Johansen T, Pampoulie C (2008). Population structure of deep-sea and oceanic phenotypes of deepwater redfish in the Irminger Sea and Icelandic continental slope: are they cryptic species?. Transactions of the American Fisheries Society, 137(6), 1723-1740.

Dassanayake M, Haas JS, Bohnert HJ, Cheeseman JM (2009). Shedding light on an extremophile lifestyle through transcriptomics. New Phytologist 183: 764–775

Dawson MN (2001) Phylogeography in coastal marine animals: a solution from California? Journal of Biogeography, 28, 723–736

De Wit P, Palumbi SR (2012) Transcriptiome-wide polymorphisms of red abalone (Haliotis rufescens) reveal patterns of gene flow and local adaptation. Molecular Ecology 22: 2884-2897.

Doherty PJ, Planes S, Mather P (1995) Gene flow and larval duration in seven species of fish from the Great Barrier Reef. Ecology 76 2373-2391.

Echeverria, TW (1987). Thirty-four species of California rockfishes: Maturity and seasonality of reproduction. Fishery Bulletin, 85(2), 229-250.

Ellegren, H (2008) Sequencing goes 454 and takes large-scale genomics into the wild. Mol. Ecol. 17: 1629–1631.

Emelianov I, Frantisek M, Mallet J (2004) Genomic evidence for divergence with gene flow in host races of the larch budmoth. Proceedings of the Royal Society London. 271: 97-105.

Evanno G, Regnaut S, Goudet J (2005) Detecting the number of clusters of individuals using the software STRUCTURE: a simulation study. Mol Ecol 14:2611–2620

Excoffier L, Hofer T, Foll M (2009). Detecting loci under selection in a hierarchically structured population. Heredity, 103(4), 285-298.

Feder JL, Nosil P (2010) The efficacy of divergence hitchhiking in generating genomic islands during ecological speciation. Evolution 64, 1729-1747.

Felsenstein, J (1981) Skepticism toward Santa Rosalia, or why are there so few kinds of animals? Evolution 35, 124–138.

Felsenstein J (1982) How can we infer geography and history from gene frequencies? J Theor Biol 96: 9–20.

Ferguson L, Lee SF, Chamberlain N, Nadeau N, Joron, M, Baxter S, ... & Jiggins C (2010). Characterization of a hotspot for mimicry: assembly of a butterfly wing transcriptome to genomic sequence at the HmYb/Sb locus.Molecular ecology, 19(s1), 240-254.

Foll M, Gaggiotti O (2006). Identifying the environmental factors that determine the genetic structure of populations. *Genetics*, *174*(2), 875-891.

Foll M, Gaggiotti O (2008) A Genome-Scan Method to identify Selected Loci Appropiate for Both Dominant and Codominant Markers: A Bayesian Perspective. Genetics 180: 977-993.

Frank KT, Petrie B, Shackell NL (2007) The ups and downs of trophic control in continental shelf ecosystems. Trends in Ecology & Evolution, 22(5), 236-242.

Gaggiotti OE, Bekkevold D, Jorgensen HB, Foll M, Carvalho GR, et al. (2009) Disentangling the effects of evolutionary, demographic, and environmental factors influencing genetic structure of natural populations: Atlantic herring as a case study. Evolution 63: 2939–2951

Galindo HM, Olson DB, Palumbi SR (2006) Seascape genetics: a coupled oceanographic-genetic model predicts population structure of Caribbean corals. Current Biology, 16, 1622–1626.

Galindo J, Grahame JW, Butlin RK (2010) An EST-based genome scan using 454 sequencing in the marine snail LIttorina saxatillis. J. Evol Bio. 23: 2004-2016

Garrigan D, Kingan SB, Geneva AJ, et al. (2012) Genome sequencing reveals complex speciation in the Drosophila simulans clade. Genome Research 22, 1499-1511.

Gharrett AJ, Matala AP, Peterson EL, Gray AK, Li Z, Heifetz J (2005) Two genetically distinct forms of rougheye rockfish are different species. Trans Am Fish Soc 134:242–260.

Gharrett AJ, Mecklenburg CW, Seeb LW, Li Z, Matala AP, Gray AK, Heifetz J (2006) Do genetically distinct rougheye rockfish sibling species differ phenotypically? Trans Am Fish Soc 135:792–800.

Gilbert-Horvath, EA, Larson, RJ, Garza JC (2006) Temporal recruitment patterns and gene flow in kelp rockfish (Sebastes atrovirens). Molecular Ecology 15, 3801–3815.

Gilg MR, Hilbish TJ (2003) The geography of marine larval dispersal: Coupling genetics with fine-scale physical oceanography. Ecology 84(11): 2989-2998.

Gokcumen O, Zhu Q, Mulder LC, Iskow RC, Austermann C, Scharer CD, ... & Lee C (2013). Balancing selection on a regulatory region exhibiting ancient variation that predates human–Neandertal divergence. PLoS genetics,9(4), e1003404.

Gomez-Urchida D, Hoffman EA, Ardren EA, Banks MA. (2003) Microsatellite markers for the heavily exploited canary (Sebastes pinniger) and other rockfish species. Mol Ecol 3: 387-389.

Haldorson L, Stanley RD, Leaman, BM, O'Connell VM (1994) Movements of tagged adult yellowtail rockfish, Sebastes flavidus, off the west coast of North America. Fishery Bulletin, 92, 655-663.

Hansen MM, Hemmer-Hansen J (2007) Landscape genetics goes to sea. Journal of Biology, 6, 6.

Hecht BC, Thrower FP, Hale MC, Miller MR, Nichols KM (2012) The genetic architecture of migration-related traits in rainbow and steelhead trout, Oncorhynchus mykiss. G3: Genes, Genomes, Genetics, 2, 1113–1127.

Hedrick PW (2005). A standardized genetic differentiation measure..Evolution, 59(8), 1633-1638.

Hedrick PW in Genetics, Demography, and Viability of Fragmented Populations (eds Young, A. & Clarke, G.) 113–125 (Cambridge Univ. Press, Cambridge, UK, 2000).

Hess JE, Vetter RD, Moran P (2011) A steep genetic cline in yellowtail rockfish Sebastes flavidus, suggests regional isolation across the Cape Mendocino faunal break. Can J Fish Aquat Sci 68: 89-104.

Hemmer-Hansen J, Nielsen EE, Frydenberg J, Loeschcke V (2007) Adaptive divergence in a high gene flow environment: Hsc70 variation in the European flounder (Platichthys flesus L.). Heredity, 99, 592–600.

Hey J (2011) Documentation for IMa2.

Hohenlohe PA, Bassham S, Etter PD, Stiffler N, Johnson EA, Cresko, WA (2010). Population

genomics of parallel adaptation in threespine stickleback using sequenced RAD tags. PLoS genetics, 6(2), e1000862.

Holm S (1979) A simple sequentially rejective multiple test procedure. Scand J Stat 6:65–70

Hudson ME (2008) Sequencing breakthroughs for genomic ecology and evolutionary biology. Mol. Ecol. Resour. 8: 3–17.

Huang X, Madan A (1999) CAP3: A DNA sequence assembly program. Genome Res. 9: 868-877.

Hyde JR (2007) The origin, evolution, and diversification of rockfishes of the genus Sebastes (Cuvier): insights into speciation and biogeography of temperate reef fishes. University of California, San Diego

Hyde JR, Kimbrell, CA, Budrick, JE, Lynn EA, & Vetter RD (2008). Cryptic speciation in the vermilion rockfish (Sebastes miniatus) and the role of bathymetry in the speciation process. Molecular Ecology, 17(4), 1122-1136.

Hyde JR, Vetter RD (2007) The origin, evolution, and diversification of rockfishes of the genus Sebastes (Cuvier). Molecular Phylogenetics and Evolution 44: 790-811.

Johansson, ML, Banks, MA, Glunt, KD, Hassel-Finnegan, HM, Buonaccorsi, VP (2008) Influence of habitat discontinuity, geographical distance, and oceanography on fine-scale population genetic structure of copper rockfish (Sebastes caurinus). Mol. Ecol. 17(13): 3051–3061.

Johnston SE, McEwan JC, Pickering NK, et al. (2011) Genome- wide association mapping identifies the genetic basis of discrete and quantitative variation in sexual weaponry in a wild sheep population. Molecular Ecology, 20, 2555– 2566

Jump AS, Hunt LM, Martínez-Izquierdo JA, Peñuelas J (2006) Natural selection and climate change: temperature-linked spatial and temporal trends in gene frequency in *Fagus sylvatica*. Molecular Ecology 15(11): 3469-3480.

Kai Y, Nakayama K, Nakabo T (2002). Genetic differences among three colour morphotypes of the black rockfish, Sebastes inermis, inferred from mtDNA and AFLP analyses. Molecular ecology, 11(12), 2591-2598.

Kalinowski ST (2005) HP-Rare: a computer program for performing rarefaction on measures of allelic diversity. Molecular Ecology Notes 5:187-189.

Kaplan NL, Hudson RR, Langley CH (1989) The 'hitchhiking effect' revisited. Genetics 123: 887-899.

Kawecki T, Ebert D (2004) Conceptual issues in local adaptation. Ecol Lett 7:1225–41.

Keller I, Wagner CE, Greuter L, Mwaiko S, Selz OM, Sivasundar A, ... & Seehausen O (2013). Population genomic signatures of divergent adaptation, gene flow and hybrid speciation in the rapid radiation of Lake Victoria cichlid fishes. Molecular Ecology, 22(11), 2848-2863.

Knutsen H, Jorde PE, André C, Stenseth NC (2003) Fine-scaled geographical population structuring in a highly mobile marine species: the Atlantic cod. Mol. Ecol. 12 385-394.

Kudela RM, Banas NS, Barth JA, Frame ER, Jay DA, Largier JL, Lessard E.J, Peterson TD, Woude AJV 2008 New insights into the controls and mechanisms of plankton productivity in coastal upwelling waters of the northern California current system.

Oceanography 21 (4), 46–59.

Laidig TE, Chess JR, Howard DF (2007). Relationship between abundance of juvenile rockfishes (Sebastes spp.) and environmental variables documented off northern California and potential mechanisms for the covariation. Fishery Bulletin, 105: 39–48

Laroche WA, Richardson SL. (1980) Development and occurrence of larvae and juveniles of the rockfish *Sebastes flavidus* and *Sebastes melanops* (Scorpaenidae) Oregon. Fish Bull. (Wash. , D.C.) 77:901-924)

Lawniczak MKN et al. (2010) Widespread divergence between incipient Anopheles gambiae species revealed by whole genome sequences. Science 330, 512–514.

Lenarz WH, Ventresca DA, Graham WM, Schwing FB, Chavez, F (1995) Explorations of El Niño events and associated biological population dynamics off central California. California Cooperative Oceanic Fisheries Investigations Reports, 36, 106-119.

Lenormand, T (2002) Gene flow and the limits to natural selection. Trends in Ecology & Evolution, 17(4), 183-189.

Lewontin RC, Krakauer J (1973) Distribution of gene frequency as a test of the theory of the selective neutrality of polymorphisms. Genetics, 74(1), 175-195.

Li, YC, Korol, AB, Fahima T, Nevo, E (2004) Microsatellites within genes: structure, function, and evolution. Molecular biology and evolution,21(6), 991-1007.

Love MS (1996) Probably more than you want to know about the fishes of the Pacific coast. Really Big Press, Santa Barbara, California. 215 p.

Love MS, Yoklavich MM, Thosrsteinson L (2002) The rockfishes of the Northeast Pacific. University of California Press, Berkeley, Ca.

Luikart G, England PR, Tallmon D, Jordan S, Taberlet P (2003) The power and promise of population genomics: from genotyping to genome typing. Nature Reviews Genetics, 4, 981–994.

McManus MA, Cheriton OM, Drake PJ, Holliday DV, Storlazzi CD, Donaghay PL, Greenlaw CF (2005) Effects of physical processes on structure and transport of thin zooplankton layers in the coastal ocean. Mar Ecol Prog Ser 301:199–215.

Mäkinen HS, Cano J M, Merilä, J (2008) Identifying footprints of directional and balancing selection in marine and freshwater three-spined stickleback (Gasterosteus aculeatus) populations. Molecular Ecology, 17(15), 3565-3582.

Manel S, Gaggiotti OE, Waples RS (2005) Assignment methods: matching biological questions with appropriate techniques. Trends in Ecology & Evolution, 20, 136–142.

Marchesiello P, McWilliams JC, Shchepetkin A. (2003). Equilibrium structure and dynamics of the California Current System. Journal of Physical Oceanography, 33(4), 753-783.

Martinez AN, Aguilar A (2011). Development of fifty-one novel EST-SSR loci for use in rockfish (genus Sebastes). Conservation Genetics Resources,3(2), 335-340.

Martinez A, Garza JC, Pearse DE (2011) A microsatellite genome screen identifies chromosomal regions under differential selection in steelhead and rainbow trout. Trans. Am. Fish. Soc. 140: 829–842

Maynard Smith, J. and J. Haigh. (1974) The hitch-hiking effect of a favourable gene. Genet. Res. 23: 23-35.

Mayr E (1963) Animal species and evolution. Cambridge, MA: Harvard University Press.

Miller JA, Shanks AL (2004) Evidence for limited larval dispersal in black rockfish (Sebastes melanops): implications for population structure and marine-reserve design. Can J Fish Aquat Aci 61: 1723-1735.

Michel AP, Sim S, Powell THQ, et al. (2010) Widespread genomic divergence during sympatric speciation. Proceedings of the National Academy of Sciences 107, 9724-9729.

Miller KM, Schultze AD, Withler RE (1999) Characterization of microsatellite loci in Sebastes alutus and their conservation in congeneric rockfish species. Mol. Ecol. 9(2): 240-242.

Mitamura H, Arai N, Sakamoto W, Mitsunaga Y, Maruo T, Mukai, Y, Yoneda Y (2002) Evidence of homing of black rockfish Sebastes inermis using biotelemetry. Fisheries Science, 68(6), 1189-1196.

Moser HG (1996) Scorpaenidae: scorpionfishes and rockfishes. *In* The Early Stages of Fishes in the California Current region. *Edited by* H.G. Moser. California Cooperative Oceanic Fisheries Investigation, Atlas 33. Allen Press, Lawrence, Kansas.

Nadeau NJ et al. (2012) Genomic islands of divergence in hybridizing Heliconius butterflies identified by largescale targeted sequencing. Phil. Trans. R. Soc. B 367, 343–353.

Narum, SW, Buonaccorsi VP, Kimbrell CA, Vetter RD (2004) Genetic divergence between gopher rockfish (Sebastes carnatus) and black and yellow rockfish (Sebastes chrysomelas). Copeia 4, 926–931

Nielsen EE, Hemmer-Hansen J, Larsen PF, Bekkevold D (2009) Population genomics of marine fishes: identifying adaptive variation in space and time. Molecular Ecology, 18, 3128–3150.

Nielsen EE, Hemmer-Hansen J, Poulsen NA, Loeschcke V, Moen T, Johansen T, Carvalho GR (2009) Genomic signatures of local directional selection in a high gene flow marine organism; the Atlantic cod (Gadus morhua). BMC Evolutionary Biology, 9(1), 276

Norton J (1987) Ocean climate influences on groundfish recruitment in the California Current. In Proceedings of the International Rockfish Symposium, Anchorage, Alaska (AK Sea Grant Rep 87-2). University of Alaska, Anchorage(pp. 73-99).

Nosil P. Ecological Speciation (Oxford Univ. Press, 2012)

Nosil P, Feder JL (2012). Genomic divergence during speciation: causes and consequences. Philosophical Transactions of the Royal Society B: Biological Sciences, 367(1587), 332-342.

Oetjen K, Reusch TBH (2007) Genome scans detect consistent divergent selection among subtidal vs intertidal population of the marine angiosperm *Zostera marina*. Molecol 16: 5156-5157

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

Palumbi SR, Wilson AC (1990) Mitochondrial DNA diversity in the sea urchins *Strongylocentrotus purpuratus* and *S. droebachiensis*. Evolution 44:403–415.

Parker SJ, Berkeley SA, Golden JT, Gunderson DR, Heifetz J, Hixon MA, Larson R, Leaman BM, Love MS, Misick JA, O'Connell VM, Ralston S, Weeks HJ, Yoklavich MM (2000) Management of Pacific Rockfish Fisheries 25:22-30.

Pérez-Figueroa A, García-Pereira MJ, Saura M, Rolán-Alvarez E, Caballero A (2010) Comparing three different methods to detect selective loci using dominant markers. Journal of Evolutionary Biology, 23(10), 2267-2276.

Pespeni MH, Garfield DA, Manier MK, Palumbi SR (2012) Genome-wide polymorphisms show unexpected targets of natural selection. Proceedings of the Royal Society B: Biological Sciences 279: 1412-1420.

Pespeni MH, Palumbi SR (2013). Signals of selection in outlier loci in a widely dispersing species across an environmental mosaic. Molecular Eclology 22: 3580-3597.

Petersen CH, Drake PT, Edwards CA, Ralston S (2010) A numerical study of inferred rockfish (Sebastes spp.) larval dispersal along the central California coast. Fisheries Oceanography, 19(1), 21-41.

Pickrell, JK, Pritchard JK (2012). Inference of population splits and mixtures from genome-wide allele frequency data. PLoS genetics, 8(11), e1002967.

Primmer CR (2009). From conservation genetics to conservation genomics. Ann NY Acad Sci 1162, 357–368.

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

Pritchard JK, Wen W (2004). Documentation for STRUCTURE Software: Version 2. Department of Human Genetics, University of Chicago, Chicago.

Renaut S, Maillet N, Normandeau E, Sauvage C, Derome N, Rogers SM, Bernatchez L (2012) Genome-wide patterns of divergence during speciation: the lake whitefish case study. Phil. Trans. R. Soc. B 367, 354–363.

Rexroad III CE, Palti Y, Gahr SA, Vallejo RL.(2008) A second generation genetic map for rainbow trout (Oncorhynchus mykiss). BMC Genetics, v.9, doi: 10.1186/1471-2156-9-74.

Rice WR (1989) Analyzing tables of statistical tests. Evolution, 43, 223–225.

Rocha-Olivares A, Kimbrell CA, Eitner BJ, Vetter RD (1999) Evolution of a mitochondrial cytochrome b gene sequence in the species-rich genus Sebastes (Teleostei, Scorpaenidae) and its utility in testing the monophyly of the subgenus Sebastomus. Mol Phylogenet Evol 11:426–440

Rocha-Olivares A, Leal-Navarro RA, Kimbrell C, Lynn EA, Vetter RD (2003) Microsatellite variation in the Mexican rockfish Sebastes macdonaldi. Scientia Marina 67, 451–460.

Rocha-Olivares A, Rosenblatt RH, Vetter RD (1999) Cryptic species of rockfishes (Sebastes: Scorpaenidae) in the Southern Hemisphere inferred from mitochondrial lineages. Journal of Heredity, 90(3), 404-411.

Rogers SM, Bernatchez L (2007) The Genetic Architecture of Ecological Speciation and the Association Signatures of Selection in Natural Lake Whitefish (Coregonus sp. Salmonidae) Species Pairs. Mol Biol. Evol 24(6) 1423-1438.

Roques S, Pallotta D, Sévingy J, Bernatchez L (1999) Isolation and characterization of polymorphic microsatellite makers in the North Atlantic redfish (Teleostei: Scorpaenidae, genus Sebastes). Mol Ecol 8: 685-702.

Rousset F (2008). genepop'007: a complete re-implementation of the genepop software for Windows and Linux. Molecular ecology resources, 8(1), 103-106.

Rozen S, Skaletsky HJ (2000) Primer3 on the WWW for general users and for biologist programers. Bioinformatics Methods and Protocols: Methods in Molecular Biology. Humana Press, Totowa , NJ, 365-386.

Sanford E, Kelly MW (2010) Local Adaptation in Marine Invertebrates. Annual Review of Marine Science 3, 509–535

Schemske DW (2000) Understanding the origin of species. Evolution. *54, 1069–1073.*

Schluter, D (2000) The ecology of adaptive radiation. Oxford, UK: Oxford University Press.

Schluter D (2001) Ecology and the origin of species. Trends in Ecol. Evol. 16: 372.

Schmidt C (2005) Molecular genetic studies on species and population structure of North Atlantic redfish (genus Sebastes; Cuvier 1829).

Schuelke M. (2000). An economic method for the fluorescent labeling of PCR fragments. Nature biotechnology, 18(2).

Schlötterer, C. (2003) Hitchhiking mapping- functional genomics from the population genetics perspective. Trends in Genetics 19: 32-38.

Schwarz D, Robertson H, Feder J, Varala K, Hudson M, Ragland G et al. (2009) Sympatric ecological speciation meets pyrosequencing: sampling the transcriptome of the apple maggot Rhagoletis pomonella. BMC Genomics 10: 633.

Scotti-Saintagne C, Mariette S, Porth I, Goicoechea PG, Barreneche T, Bodénes C, Kremer A (2004) Genome scanning for interspecific differentiation between two closely

related oak species [Quercus robur L. and Q. petraea (Matt.) Liebl.]. Genetics, 168(3), 1615-1626.

Sekino M, Takagi N, Hara M, Takahashi H. (1999) Microsatellites in rockfish Sebastes thompsoni (Scorpaenidae). Mol Ecol. 9: 629-644.

Selkoe KA, Henzler CM, Gaines SD (2008) Seascape genetics and the spatial ecology of marine populations. Fish and Fisheries, 9, 363–377.

Shanks AL (2009) Pelagic larval duration and dispersal distance revisited. Biol. Bull. 216: 373-385.

Shanks AL, Eckert GL (2005) Population persistence of California Current fishes and benthic crustaceans: a marine drift paradox. Ecological Monographs, 75(4), 505-524.

Sivasundar A, Palumbi, SR (2010) Life history, ecology and biogeography of strong genetic breaks among 15 species of Pacific rockfish Sebastes.Marine Biology, 157, 1433–1452.

Slate J, Gratten J, Beraldi D, Stapley J, Hale M, Pemberton JM (2009) Gene mapping in the wild with SNPs: guidelines and future directions. Genetica, 136, 97–107.

Slatkin M (1987) Gene flow and the geographic structure of natural populations. Science 236: 787-792

Sotka EE (2005) Local adaptation in host use among marine invertebrates. Ecol Lett ;8:448-59.

Sotka EE, Wares JP, Barth JA, Grosberg RK, Palumbi SR (2004) Strong genetic clines and geographical variation in gene flow in the rocky intertidal barnacle Balanus glandula. Mol Ecol 13:2143-2156.

Stefánsson MÖ, Sigurdsson T, Pampoulie C, Daníelsdóttir AK, Thorgilsson B, Ragnarsdóttir A, Bernatchez L (2009). Pleistocene genetic legacy suggests incipient species of Sebastes mentella in the Irminger Sea. Heredity, 102(5), 514-524.

Stefánsson MÖ, Reinert J, Sigurðsson Þ, Kristinsson K, Nedreaas K, Pampoulie C (2009). Depth as a potential driver of genetic structure of Sebastes mentella across the North Atlantic Ocean. ICES Journal of Marine Science: Journal du Conseil, 66(4), 680-690.

Stephan W, Wiehe T, Lenz MW (1992) The effect of strongly selected substitutions on neutral polymorphism: analytical results based on diffusion theory. Theoretical Population Biology 41: 237-254.

Stinchcombe JR, Hoekestra HE (2007) Combining population genomics and quantitative genetics: finding the genes underlying ecologically important traits. Heredity 1-13.

Stölting KN, Nipper R, Lindtke D, Caseys C, Waeber S, Castiglione S, Lexer C (2013). Genomic scan for single nucleotide polymorphisms reveals patterns of divergence and gene flow between ecologically divergent species. Molecular ecology, 22(3), 842-855.

Storz JF (2005) Using genome scans of DNA polymorphism to infer adaptive population divergence. Molecular Ecology, 14, 671–688.

Treml EA, Roberts J, Chao Y, Halpin P, Possingham HP, Riginos C (2012) Reproductive output and duration of the pelagic larval stage determine seascape-wide connectivity of marine populations. Integrative and Comparative Biology 52(4): 525-537

Tagart J (1991) Population dynamics of yellowtail rockfish (Sebastes flavidus) stocks in the northern California to southwest Vancouver Island region(Doctoral dissertation, University of Washington).

Toth AL, Varala K, Newman TC, Miguez FE, Hutchison SK, Willoughby DA et al. (2007). Wasp gene expression supports an evolutionary link between maternal behavior and eusociality. Science 318: 441–444.

Turner TL, Hahn MW (2007) Locus- and population-specific selection and differentiation between incipient species of Anopheles gambiae. Molecular Biology and Evolution 24, 2132-2138.

Turner TL, Hahn MW (2010) Genomic islands of speciation or genomic islands and speciation? Molecular Ecology 19, 848-850.

Turner TL, Hahn MW, Nuzhdin SV (2005) Genomic islands of speciation in Anopheles gambiae. PLoS Biology, 3, e285, DOI: 10.1371/journal.pbio.0030285.

Vasemägi A, Nilsson J, McGinnity P, et al. (2012) Screen for Footprints of Selection during domestication/Captive Breeding of Atlantic Salmon. Comparative and Functional Genomics 628204

Vasemägi A, Nilsson J, Primmer CR (2005) Expressed sequence tag-linked microsatellites as a source of gene-associated polymorphisms for detecting signatures of divergent selection in Atlantic salmon (Salmo salar L.). Molecular Biology and Evolution 22: 1067-1076.

Via S (2001). Sympatric speciation in animals: the ugly duckling grows up. Trends Ecol. Evol. 16, 381-390.

Via S (2009) Natural selection in action during speciation. Proceedings of the National Academy of Sciences 106, 9939-9946.

Via S (2012) Divergence hitchhiking and the spread of genomic isolation during ecological speciation-with-gene-flow. Philosophical Transactions of the Royal Society B: BiologicalSciences 367, 451-460.

Via S, West J (2008) The genetic mosaic suggests a new role for hitchhiking in ecological speciation. Mol. Ecol. 17:4334-4345.

Vitalis R, Dawson K, Boursot P (2001) Interpretation of variation across marker loci as evidence of selection. Genetics 158:1811–1823.

Wang S, Meyer E, McKay JK, Matz MV (2012). 2b-RAD: a simple and flexible method for genome-wide genotyping. Nature Methods, 9(8), 808-810.

Weir BS, Cockerham CC (1978) Testing hypotheses about linkage disequilibrium with multiple alleles. Genetics, 88(3), 633-642.

Westerman ME, Bounaccorsi VP, Stannard A, Galver C, Taylor C, Lynn EA, Kimbrell

CA, Vetter RD. (2005) Cloning and characterization of novel microsatellite DNA markers for the grass rockfish (*Sebastes rastrelliger*), and cross-species amplification in 10 related *Sebastes* spp. Mol Ecol. 5: 74-76.

Westrheim SJ (1975) Reproduction, maturation, and identification of larvae of some Sebastes (Scorpaenidae) species in the northeast Pacific Ocean.Journal of the Fisheries Board of Canada, 32(12), 2399-2411.

Wiehe T (1998) The effect of selective sweeps on the variance of the allele distribution of a linked multiallele locus: hitchhiking of microsatllites. Theoretical Population Biology 53: 272-283.

Wiehe T, Stephan W (1993) Analysis of a genetic hitchhiking model and its application to DNA polymorphism data from *Drosophila melanogaster*. Molecular Biology and Evolution 10: 842-854.

Wilson JR, Broitman BR, Caselle JE, Wendt DE (2008) Recruitment of coastal fishes and oceanographic variability in central California.Estuarine, Coastal and Shelf Science, 79(3), 483-490.

Wimberger P, Burr J, Gray A, Lopez A, Bentzen P (1999) Isolation and characterization of twelve microsatellite loci for rockfish (Sebastes). Mar. Biotech. 1: 311-315.

Withler RE, Beacham TD, Schulze AD, Richards LJ, Miller KM (2001) Co-existing populations of Pacific ocean perch, *Sebastes alutus*, in Queen Charlotte Sound, British Columbia. Mar Biol (Berl) 139:1–12.

White TA, Stamford J, Hoelzel AR (2010) Local selection and population structure in a deep-sea fish, the roundnose grenadier (Coryphaenoides rupestris). Mol. Ecol. 19: 216–226.

Wood HM, Grahame JW, Humphray S, Rogers J, Butlin RK (2008) Sequence differentiation in regions identified by a genome scan for local adaptation. Mol Ecol 17(13):3123–3135.

Yoshida K, Nakagawa M, Wada S (2005) Multiplex PCR system applied for analyzing microsatellite loci of Schlegel's black rockfish, Sebastes schlegeli. Mol Ecol. 5: 416-418.