

Genomic signatures of natural selection and population structure in West Coast and Alaskan sablefish (*Anoplopoma fimbria*)

Andrew J. Jasonowicz

A thesis

submitted in partial fulfillment of the

requirements for the degree of

Master of Science

University of Washington

2015

Committee:

Steven Roberts

Frederick Goetz

Lorenz Hauser

Krista Nichols

Program Authorized to Offer Degree:

School of Aquatic and Fishery Sciences

©Copyright 2015
Andrew J. Jasonowicz

University of Washington

Abstract

Genomic signatures of natural selection and population structure in West Coast and Alaskan sablefish (*Anoplopoma fimbria*)

Andrew J. Jasonowicz

Chair of Supervisory Committee:

Steven Roberts, Associate Professor

School of Aquatic and Fishery Sciences

Previous population genetic studies have not been able to find clear population genetic structure across the range of the sablefish (*Anoplopoma fimbria*) in North America. They have focused on using a relatively small number of genetic markers that are presumed to be neutral to the effects of natural selection. Genomic approaches enable researchers to examine variation at thousands of genetic markers throughout the genome simultaneously. This allows for the detection of markers influenced by natural selection. Many marine species have large population sizes leading to low levels of genetic drift that often cause little or no differentiation to be observed at neutral genetic markers. On the other hand, there is potential for significant adaptive variation to be present between populations even when levels of genetic differentiation at neutral markers are low. In marine fish species, studies of adaptive variation may help resolve genetic structure that is weak at neutral loci. In the present study, traditional population genetic analyses were used to examine differentiation among sablefish collected in the Bering Sea, Gulf of Alaska and off the West Coast of the United States. Individual based methods were used to investigate the possibility of cryptic population structure, and a landscape genomics

approach was used to examine adaptive variation in the sablefish genome by a testing for associations between environmental conditions and SNP genotypes. Restriction site associated DNA sequencing (RAD-seq) was used to identify and genotype single nucleotide polymorphism (SNP) markers throughout the sablefish genome. Low and insignificant levels of population differentiation among survey areas were observed and the individual based methods support a single genetic cluster of sablefish across their North American range. The results are suggestive of high rates of gene flow or movement of sablefish in the northeast Pacific Ocean. The landscape genomic analyses found two SNPs to be associated with depth.

Table of Contents

List of Figures	i
List of Tables	ii
List of Supplemental Files	iii
Acknowledgements	iv
Chapter 1: General Introduction	1
1) Molecular Genetics in Fisheries	2
a) The Stock Concept.....	2
b) Population Genetic Structure of Marine Fish.....	3
c) Molecular Markers and Their Application in Fisheries.....	5
2) The Sablefish	11
a) Sablefish Biology.....	12
b) Sablefish Stock Structure.....	13
i) Seamount Sablefish	14
ii) Population Genetics	16
iii) Evidence of Fine Scale Population Differences.....	18
iv) Ecotypic Variation	19
3) Statement of Purpose	20
Chapter 2: Genomic Evidence for Large Scale Panmixia in the Sablefish	22
1) Introduction	22
2) Methods	25
a) Sample Collection.....	25
b) Genomic Library Construction.....	27
c) Processing Sequence Reads	29
d) SNP Quality Filtering.....	30
e) Population Structure	31
f) Landscape Genomics.....	33
i) Environmental Variables	33
ii) Association Analyses	34
iii) Annotation	34
3) Results	35
a) Sequencing.....	35
b) Population Structure	35
c) Environmental Associations	40
i) Annotation.....	40
4) Discussion.....	42
a) Population Structure	42
b) Environmental Associations	50
c) Relation to Fine Scale Biological Differences and Other Population Structure...	51
d) Conclusion	53
5) Data Accessibility.....	54
References	55
Appendix:	69

List of Figures

Figure 1. Preparation of digested genomic DNA for RAD sequencing	8
Figure 2. Map of sablefish samples used for RAD sequencing in this study	28
Figure 3. Patterns of isolation by distance for sablefish in the northeast Pacific Ocean..	38
Figure 4. Plot of the first two principal component axes in a principal component analysis of 2661 RAD-seq derived SNPs.	41
Figure S 1. Estimated length at 50% probability of maturity for female and male sablefish samples collected for this study	69
Figure S 2. Histogram of F_{IS} calculated by locus for the population groupings used for the population genetic analyses as well as globally.....	70

List of Tables

Table 1. Table of F_{ST} estimates for each pair of survey groups. F_{ST} below diagonal and p-values above.....	37
Table 2. P-values for tests of allelic differentiation between each population pair and statistical power of each comparison	37
Table 3. P-values for tests of genotypic differentiation between each population pair....	37
Table 4. Mantel correlation coefficients used to test the association between geographic and genetic distance matrices.....	39
Table 5. Significant SNP – environment associations found using LFMM.	43
Table 6. Some life history characteristics of marine fish in which panmixia has been proposed.....	48
Table S 1. Summary of F_{IS} for the SNP loci used in this study before and after filtering SNPs on the basis of Hardy - Weinberg equilibrium (HWE).....	71
Table S 2. Summary of BLASTN alignments of sablefish contig 152341 (gi:559591484, gb:AWGY01151885.1) to the stickleback BORAD S1 genome	72
Table S 3. Summary of BLASTN alignments of sablefish contig 133848 (gi: 559609892, gb: AWGY01133477.1) to the stickleback BORAD S1 genome.....	74

List of Supplemental Files

Supplemental file 1. Sample information for individual sablefish included in this study (File S 1)

Supplemental file 2. Fasta file containing sequences for each RAD tag included in this study (File S 2)

Supplemental file 3. Variant call format (vcf) file containing quality filtered genotypes of sablefish included in this study (File S 3)

Acknowledgements

There are many people that have worked to make this research possible. I would like to thank my UW advisor Steven Roberts and committee member Lorenz Hauser for their support and input at all stages of this thesis, their advice has been invaluable throughout this project. I would also like to extend a sincere thank you to Rick Goetz and Krista Nichols, my supervisors at the Northwest Fisheries Science Center, for providing me with this opportunity and experience, it has been a privilege to study under their guidance.

I would also like to thank the staff of the Conservation Biology Division's Molecular Genetics Team at NOAA's Northwest Fisheries Science Center for the generous use of their laboratory and equipment. Their willingness to give advice and answer numerous questions was invaluable while working in their lab. Thank you to Giles Goetz for sharing his bioinformatic expertise and help on many computer related issues.

A large group of people were involved in the sampling effort of this project and it could not have been done without their help. I would like to thank the Fisheries Resource and Monitoring Division (FRAM) at the Northwest Fisheries Science Center, both the Resource Conservation and Engineering Division (RACE) and the Marine Ecology and Stock Assessment Group (MESA) at the Alaska Fisheries Science Center, as well as the crews of the F/V Vesterallen, F/V Alaska Leader, F/V Ms. Julie, F/V Noah's Ark and F/V Excalibur for their hard work and expertise during sample collection.

This work was made possible with funding from NOAA's aquaculture program. This work was facilitated through the use of advanced computational, storage, and networking infrastructure provided by the Hyak supercomputer system at the University of Washington. Nils Ryman provided a version of the computer program POWSIM that was compiled specifically to accommodate the size of this dataset.

Lastly, I would like to thank my friends and family for their constant support and encouragement throughout this endeavor.

Chapter 1: General Introduction

Typical approaches to fisheries management often focus on the causes of short-term demographic changes. However, with the integration of genetic data into a management plan, possible genetic effects can be considered (Hauser & Carvalho 2008; Waples *et al.* 2008). For example, plans may be implemented that mitigate the effects of genetic drift which can lead to the loss of genetic diversity, ultimately compromising the adaptive potential and long-term viability of a population (Allendorf & Luikart 2007; Hauser & Carvalho 2008). Understanding the genetic structure of a fishery can provide managers with a tool to make decisions that ensure a lasting resource under continually changing environmental conditions (Hauser & Carvalho 2008). Recent advances in DNA sequencing technology and methods has made the rapid development of large numbers of genetic markers accessible for many investigators (Baird *et al.* 2008). With large genomic datasets, the investigation of adaptive or non-neutral genetic variation has become popular for many species. With the potential for significant adaptive variation to be present between populations even when gene flow is high, studies of adaptive variation may help resolve genetic structure that is weak at neutral loci (Conover *et al.* 2006; Hauser & Carvalho 2008).

The sablefish (*Anoplopoma fimbria*) is a deep-water groundfish species widely distributed throughout the northern Pacific Ocean. Previous studies have used molecular markers to characterize the population structure of this species (Tsuyuki & Roberts 1969; Gharrett *et al.* 1982; Tripp-Valdez *et al.* 2012). These studies suggest that there is little or no genetic structure across large geographic scales, however they have only examined variation at markers presumed to be neutral to the effects of natural selection. Newer

methods using genomic technologies that simultaneously investigate a large number of markers throughout the genome enable researchers to investigate the effects of natural selection on specific loci or genomic regions. This type of approach may prove useful in identifying adaptive variation in marine species like the sablefish that typically have low levels of background differentiation. The use of molecular genetics in fisheries and how genomic tools may help identify adaptive variation in marine fish is discussed below along with sablefish biology and stock structure.

1) Molecular Genetics in Fisheries

a) The Stock Concept

The concept of a stock is central to fisheries management and while genetic data can be a powerful tool in fisheries management, their use in stock delineation can be contentious. In part, this is because there are several definitions for a stock within a management context (Carvalho & Hauser 1994; Begg & Waldman 1999). One definition and perhaps the most comprehensive would be that of Ihssen *et al.* (1981), where a stock is defined as “an interspecific group of randomly mating individuals with temporal and spatial integrity”. This definition characterizes an idealized view of a stock, because it implies that gene flow between groups is reduced to the point where the groups are genetically distinct from one another (Ihssen *et al.* 1981; Shaklee & Bentzen 1998; Ward 2002). While this serves as a baseline to define a stock, it may not be realistic in practice because there are many different patterns of population structure that may arise that do not fit this model definition (see Hellberg *et al.* 2002). Under the stock concept, managers define semi-discrete boundaries to fish populations in order to make decisions regarding those management units (Begg *et al.* 1999). These semi-discrete boundaries

may or may not relate directly to Ihssen's definition of spatial and temporal integrity (Ihssen *et al.* 1981). When thinking about biological populations in the wild, it is easy to envision situations where little data on population structure is available; where a species' stock structure may not conform to a clear population genetic model, or where single populations occupy a geographic area managed by multiple governments. In these instances, defining stocks and management units becomes more complex.

b) Population Genetic Structure of Marine Fish

Many commercially harvested marine finfish could be considered "classical marine species", characterized by wide distributions, large population sizes, high fecundity and pelagic larval periods (Nielsen & Kenchington 2001). Low levels of genetic differentiation typical of many marine species (Ward *et al.* 1994) are thought to result from a combination of relatively large effective population sizes and few obvious barriers to migration in the sea (Nielsen *et al.* 2009a). In some cases sufficient time for migration-drift equilibrium has not passed since postglacial recolonization events. As a result, many classical marine species have relatively shallow population histories resulting in low levels of genetic differentiation (Nielsen *et al.* 2009a). The degree that marine populations are structured can vary greatly, ranging from completely open populations (panmixia) to completely closed, demographically independent populations (Hellberg *et al.* 2002). The development of these genetic patterns depends on the interplay between gene flow (migration), drift (effective population size) and natural selection. Hellberg *et al.* (2002) defines six types of genetic patterns that may arise in marine populations:

1. **Highly structured/Isolation:** When migration is low and effective population sizes are small, it is likely that closed, genetically isolated populations will develop.
2. **Weak structuring/Panmixia:** If migration rate is high and effective population sizes are large, very little or no genetic structure will result. This could also occur if populations have not reached migration-drift equilibrium after a postglacial expansion or colonization or event.
3. **Stepping stone:** Under scenarios where migration only occurs between neighboring populations a stepping-stone or isolation by distance pattern is likely to develop.
4. **Abrupt genetic differences across a geographic barrier:** This type of pattern may develop when genetic drift overwhelms migration across the barrier. Relatively small population sizes and low migration rates would be ideal conditions under which this type of pattern would develop rapidly.
5. **Cline:** A gradual change in genetic diversity over space may result from secondary contact between two previously isolated populations or different patterns of selection at the ends of a species' range.
6. **Chaotic genetic patchiness:** This type of pattern may develop under a number of scenarios. Oceanographic conditions may enable successful recruitment of larvae from different sources to a population. In species with pelagic larvae it is possible that some natural selection occurs during recruitment, or a species may have high reproductive output and low survival rates of larvae and a small number of adults may contribute offspring to the next generation. In these cases, unpredictable variable structure is likely to arise. This pattern highlights the dynamic processes at play during the recruitment process. Sweepstakes reproductive success, differential

selection at early life stages, and unstable oceanographic conditions may all play a role in the development of this pattern.

While this list does not necessarily consider all possible scenarios, it provides a general framework for interpreting how demographic and evolutionary processes might influence genetic population structure in marine fish.

c) Molecular Markers and Their Application in Fisheries

The use of molecular markers in studies of population structure can shed light on many of the issues central to resource management and overcome many of the technical obstacles associated with the direct observation of individuals in large populations. For example, genetic markers can be used to infer movement or interaction between populations without directly observing and tracking individuals with tagging or telemetry-based methods (Brown & Epifanio 2003). Molecular approaches and methods have been used for a number of applications in fisheries management; species and stock identification, assessing the effects of harvest on genetic diversity, introgression from hatchery stocks, estimation of effective dispersal distance, and local adaptation (Shaklee & Bentzen 1998; Ward 2002; Nielsen *et al.* 2009b; Cunningham *et al.* 2009; Pinsky *et al.* 2010; Bradbury *et al.* 2010). In addition, the use of genetic markers can be useful in aquaculture programs where populations with desirable traits are often selected as sources for broodstock (Hedgecock *et al.* 1976; Sonesson & Meuwissen 2009).

Despite typically low levels of genetic differentiation, marine species may be well suited to studies of natural selection. They may respond to relatively weak selection because their large population sizes will reduce the effects of genetic drift and increase

the likelihood that adaptive alleles will sweep through the population (Nielsen *et al.* 2009a). Also, the number of migrants needed to effectively counteract the effects of natural selection is more than the number required to maintain genetic homogeneity in the face of genetic drift (Conover *et al.* 2006). Thus, for species with relatively large effective population sizes, in which very small migration rates may be enough to homogenize already weak population genetic structure, it is possible for significant adaptive variation to remain (Larsen *et al.* 2007; Hauser & Carvalho 2008). With recent developments in sequencing technology and bioinformatics, it has become economically feasible for researchers to rapidly develop large sets of genetic markers that span the entire genome in non-model organisms. As a result, studies on adaptive variation have become quite popular. Investigating adaptive variation in marine fish holds potential to resolve genetic structure even in instances where neutral loci may fail to do so.

While genetic markers can be utilized for defining populations and their interactions with each other, it is not easily done if adequate markers are not available for the species of interest or the questions at hand. Recent advances in sequencing technology have made the discovery of large numbers of new markers increasingly efficient for non-model organisms (Baird *et al.* 2008; Elshire *et al.* 2011; Kumar *et al.* 2012; Peterson *et al.* 2012). These new high throughput sequencing methods enable the rapid development of large marker sets that can be used to simultaneously investigate natural selection and neutral genetic variation in wild populations (Luikart *et al.* 2003). Restriction site associated DNA sequencing (RAD-seq) is one such method. RAD-seq provides a platform for simultaneous single nucleotide polymorphism (SNP) discovery and genotyping in non-model organisms for which sequenced genomes are unavailable

(Baird *et al.* 2008). Briefly, RAD-seq uses restriction enzymes to digest the genomic DNA of an individual (Figure 1). The resulting fragments (RAD tags) are then ligated to a nucleotide sequence containing a sequencing primer and a nucleotide barcode that is used to identify individuals (Baird *et al.* 2008). Samples from many individuals can then be pooled and sequenced using massively parallel next generation sequencing technologies (Baird *et al.* 2008). As their name suggests, SNPs are point mutations dispersed throughout the genome and as such, they are sometimes found in genes that code for functional proteins. In species with large geographic ranges, the environment is often heterogeneous and different genes can be associated with local adaptation under varying environmental conditions (Eckert *et al.* 2010; Bradbury *et al.* 2010). SNPs may be found in these genomic regions under selection and may show patterns of differentiation that are different than patterns observed at SNPs found that are not influenced by natural selection.

The Atlantic cod (*Gadus morhua*) is perhaps the best-studied species when it comes to using molecular tools to assess the stock structure of a marine fish. It is a widespread commercially important species and with major declines in biomass since the 1950's (Myers *et al.* 1996). With the collapse of many fisheries (see Myers *et al.* 1996), a substantial effort has gone into identifying relevant stock structure in cod. A number of studies have used neutral genetic markers to investigate cod stock structure and have uncovered population structure at varying scales (see Ruzzante *et al.* 1999). Using adaptive variation to address the question of stock structure has proven useful in instances where neutral population structure is found to be weak. For example, Nielsen *et al.* (2009b) found signatures of directional selection in cod populations with low levels

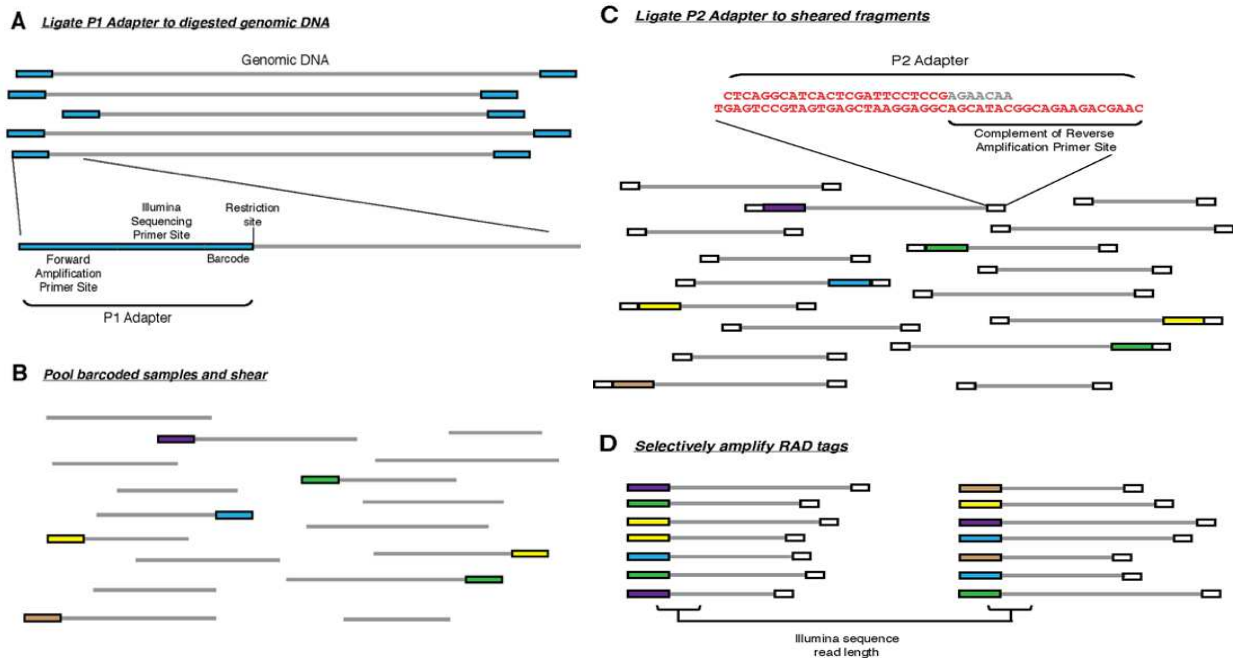


Figure 1. Preparation of digested genomic DNA for RAD sequencing; A) Ligate adapter with Illumina primer and nucleotide barcode. B) Pool barcoded fragments and shear. C) Ligate primer for reverse amplification to sheared fragments. D) Amplify and sequence. From: Baird et al. 2008

of background divergence, showing that signatures of selection can be detected in large populations with high rates of gene flow. It also emphasizes the notion that selection can be an important force shaping the stock structure of marine fish, even when population sizes are large and gene flow is high among groups. Markers associated with adaptive divergence have also been used to resolve fine-scale population structure across the geographic range of Atlantic cod that neutral SNPs alone could not (Bradbury *et al.* 2013). Genetic markers presumed to be under selection can also be quite powerful in assignment tests and make it possible to assign individuals back to their population of origin with very high accuracy. Nielsen *et al.* (2012) ranked SNP loci under selection according to F_{ST} and used the top eight to accurately assign fish from 21 populations of Eastern Atlantic cod back to their population of origin except for a single individual identified as a migrant. Genes have also been found to be under selection in response to environmental variables in Atlantic cod. Similar sets of genes have been shown to be associated with temperature on both sides of the Atlantic suggesting parallel adaptive responses to environmental conditions (Bradbury *et al.* 2010). Population genomic analyses have also been used to detect genomic regions of differentiation that separate stationary and migratory cod ecotypes (Hemmer-Hansen *et al.* 2013; Karlsen *et al.* 2013).

While studies on natural selection in Atlantic cod comprise a large portion of the literature on selection in marine fishes, they are certainly not alone. Signatures of natural selection have also been observed in the genomes of other marine species. The investigation of genetic structure at candidate genes thought to be under selection in European flounder (*Platichys flesus*) (Hemmer-Hansen *et al.* 2007) and walleye pollock (*Theragra chalcogramma*) (Canino *et al.* 2005) has shown that variation at non-neutral

loci can be different than patterns observed at neutral loci even in populations with high rates of gene flow. The recent growth in genomic sequencing technology has enabled the use of non-targeted approaches designed to detect outlier genes that are considered to be under selection. For example, including a small number of outlier loci in analyses of population structure can sometimes resolve population differences that a set of neutral loci cannot and has been observed in Atlantic herring (*Clupea harengus*) (Limborg *et al.* 2012) and European hake (*Merluccius merluccius*) (Milano *et al.* 2013). Increased ability to resolve population differences with outlier loci is further support for the notion that natural selection can lead to locally adapted populations even in the face of high rates of gene flow (Limborg *et al.* 2012). Adaptive variation in marine fish may become an important characteristic to consider since it implies that, for some fish stocks, local recruitment may be more important and vital to their persistence than immigration from neighboring populations (Hauser & Carvalho 2008).

Many genetic methods require individuals to be grouped into pre-defined populations or groupings prior to analysis, hence the name population genetics. While these methods are powerful and have been shown to benefit studies of stock structure, there are inherent problems with this type of analysis when samples are continuously distributed or sampling locations do not correspond to reproductive groups, as is the case for many marine species. In situations like this individual-based analyses that do not rely on *a priori* population groupings may be better suited for analyzing genetic structure (Jones *et al.* 2013). When combined with genotyping by sequencing methods such as RAD-seq, individual-based analyses provide a potentially powerful platform by which both neutral and non-neutral genetic variation can be investigated in non-model species

(Luikart *et al.* 2003; Baird *et al.* 2008) and the emerging field of landscape genomics provides a framework for an individual-based approach. Schwartz *et al.* (2010) refers to landscape genomics as "...the simultaneous study of tens-to-hundreds of markers, ideally including markers in candidate adaptive genes (genes under selection), with georeferenced samples collected across a landscape". While approaches like this are not particularly abundant in the literature on classical marine fishes, they have been successfully applied to analyses of local adaptation and population structure in other fishes such as salmon (Zueva *et al.* 2014), and other organisms such as trees (Sork *et al.* 2013; Frichot *et al.* 2013), plants (Bothwell *et al.* 2013; Jones *et al.* 2013) goats (Pariset *et al.* 2009) and humans (Frichot *et al.* 2013). When compared directly, landscape genomic analyses can identify significant and interesting relationships between genes and the environment that population based approaches may miss (Frichot *et al.* 2013).

2) The Sablefish

The sablefish (*Anoplopoma fimbria*) is a deep-water groundfish species widely distributed throughout the northern Pacific Ocean. In Asia, they are found in the waters off of Japan, and Kamchatka, Russia. In North America, they inhabit waters from the Bering Sea south to Baja California (Sasaki 1985). Sablefish are commercially important throughout the North American part of their range. In 2012, sablefish landings totaled 11.58 million pounds valued at 28.1 million dollars in waters off the West Coast of the United States, and in Alaska harvest totals were 29.7 million pounds worth 113 million dollars (NMFS 2013). The sablefish resource in the northeast Pacific is shared between Canada and the United States. Accordingly, sablefish are assessed and managed

separately in the waters off the West Coast of the U.S., Alaska (Stewart *et al.* 2011; Hanselman *et al.* 2012b), and Canada (DFO 2011).

a) Sablefish Biology

The sablefish is a long lived species, and it is common for ages in excess of 40 years to be recorded (Kimura *et al.* 1993; Beamish & McFarlane 2000; Head *et al.* 2014). The sablefish is a highly mobile species and occupies a variety of habitat types throughout development. Adults are found along the continental slope, where it is thought that spawning occurs at depths exceeding 300 m (Mason *et al.* 1983). The eggs are pelagic and incubate at depths greater than 200 m (Mason *et al.* 1983; Kendall & Matarese 1987; Alderdice *et al.* 1988; McFarlane & Beamish 1992). While empirical catch data in the literature only documents larvae to 700 m it is thought that the eggs and larvae sink to depths exceeding 1000 m prior to hatching (McFarlane & Beamish 1992). This conclusion is based on specific gravity calculations made during laboratory experiments that would suggest the eggs sink to approximately 1200 m prior to hatching (Alderdice *et al.* 1988). Also, the prevalence of newly hatched larvae and a lack of late stage eggs in samples taken down to 500 m has been interpreted as evidence of this phenomenon (Mason *et al.* 1983; Kendall & Matarese 1987). The larvae then ascend to the surface near the end of the yolk sac stage and are a dominant species in neuston assemblages until they reach about 80 mm (Kendall & Matarese 1987; Alderdice *et al.* 1988; Doyle 1992; McFarlane & Beamish 1992; Wing & Kamikawa 1995). It is estimated that larval development takes approximately 2 – 3 months in the wild (Boehlert

& Yoklavich 1985; Kendall & Matarese 1987). Juveniles then utilize nearshore areas for 1 – 2 years before moving out to the slope as they begin to mature (Maloney & Sigler 2008).

b) Sablefish Stock Structure

While managed separately in Alaska, Canada and off the West Coast of the U.S., the current perception of sablefish stock structure suggests that there are two biological stocks of sablefish. There is a northern stock that inhabits Alaskan waters and northern British Columbia and a southern stock that ranges from southern British Columbia to Baja California (Sumaila *et al.* 2005; Hanselman *et al.* 2012b). Differences in growth, size at maturity and tagging support the division of the stock into two groups (Hanselman *et al.* 2012b). However, these stocks are not discrete and, based on tagging studies, some mixing occurs (Kimura *et al.* 1998).

Kimura *et al.* (1993) estimated von Bertalanffy growth parameters using both mark-recapture and length at age data. Different parameter estimates for sablefish off the West Coast of the U.S. and the Gulf of Alaska were obtained, suggesting that growth between these two stocks is different (Kimura *et al.* 1993). A number of studies have estimated the size at maturity for southern sablefish (Fujiwara & Hankin 1988; Hunter *et al.* 1989; Macewicz & Hunter 1994; Head *et al.* 2014) and it appears that they mature at a smaller size than the northern stock (Sasaki 1985; Kimura *et al.* 1998). While few studies have explicitly examined the age at 50% probability of maturity (A_{50}) in the northern stock, it is estimated to be 6.6 years for females (Hanselman *et al.* 2012b). In contrast, a number of studies have examined maturity status of the southern stock and estimates of A_{50} range from ages 3.8 to 7 years old (Mason *et al.* 1983; Fujiwara &

Hankin 1988; McFarlane & Beamish 1990; Head *et al.* 2014). The estimate of A_{50} used by the Alaskan stock assessment is within the range of values found for the southern stock. While the size at maturity differs between the two stocks, it is unclear if age at maturity differs between them.

Tagging studies also support the current division of these stocks. While there is some exchange between the northern and southern stocks, there is far more movement within these areas (Kimura *et al.* 1998). Movements within the northern stock have been described more extensively than movements within the southern stock. The general pattern of movement in the Gulf of Alaska has been described as counterclockwise (Maloney & Sigler 2008). Tagging studies have shown that young sablefish predominantly move from the eastern Gulf of Alaska and westward toward the Aleutian Islands as they mature (Heifetz & Fujioka 1991; Maloney & Sigler 2008; Echave *et al.* 2013). They then return to the Eastern Gulf as adults (Heifetz & Fujioka 1991; Maloney & Sigler 2008; Echave *et al.* 2013). Sex-biased movement has also been documented for sablefish; female sablefish have been observed to move longer distances than males (Morita *et al.* 2012; Hanselman *et al.* 2015). A relationship between growth rate and movement is also observed; however, sablefish exhibit a female-biased sexually dimorphic growth rate so the difference in movement distance could be attributed to differences in growth rather than sex (Morita *et al.* 2012).

i) Seamount Sablefish

Sablefish are also known to inhabit offshore seamounts throughout their range and in many cases, form a major component of the ichthyofauna present on the seamounts (Hughes 1981; Maloney 2004). They are found on seamounts of both the

United States and Canada and there are seamount fisheries that occur in Canada (Maloney 2004; Clark 2009; DFO 2013). Once sablefish reach maturity, there is some rate of migration to offshore seamounts and these migrations are assumed to be the source of recruitment for the seamount populations (Maloney 2004; DFO 2013). Juvenile sablefish are absent on seamounts and catches are comprised exclusively of mature adult fish (Shaw & Parks 1997; Maloney 2004). While sablefish have been observed moving from the slope to seamounts and vice-versa, there are no data that support movement between seamounts; all tag recoveries from fish tagged on a seamount have come from either the slope or the same seamount that the individual was tagged and released (Maloney 2004; DFO 2013).

Fish generally move to seamounts between the ages of 5 and 15, and as a result the populations are primarily made up of older fish (Maloney 2004). The growth of sablefish on seamounts may also differ from fish on the slope. For example, the otoliths of seamount fish typically show a compressed pattern in the annuli, indicating a reduction in growth rate and the age at which the individual moved to the seamount (Maloney 2004). This pattern is very common to seamount sablefish and while it is relatively rare in otoliths sampled from fish collected on the slope (Maloney 2004) and it may indicate that an individual has been to the seamounts and migrated back to the slope. Similar to coastal populations, the sex ratio of mature fish on Canadian and Alaskan seamounts is skewed toward males (Hughes 1981; Murie & McFarlane 1996; Maloney 2004; DFO 2013).

ii) Population Genetics

Even though a number of studies have supported the paradigm that two sablefish stocks exist, there have also been studies yielding results that might challenge this position. To date, there have been only a few studies that have examined the population genetics of sablefish and few conclusions have been drawn from the results. An early study examined polymorphism of a single muscle protein in fish collected from Unimak Pass, AK to Cape Blanco, OR and found no evidence of discrete northern and southern stocks (Tsuyuki & Roberts 1969). Another study examined eleven allozyme loci in collections of fish made from Baja California to the Bering Sea (Gharrett *et al.* 1982). In discussing the results, the authors stated that their observations are “consistent with the existence of a number of somewhat discrete populations between which some gene flow exists” (Gharrett *et al.* 1982). The authors also mentioned that more genetic heterogeneity was observed near the center of the sampling area compared to the ends and suggested that more breeding stocks are found near the center of the range (Gharrett *et al.* 1982). They also stated that the similarities in allele frequencies among collections would be consistent with very large effective population sizes and some degree of gene flow throughout the North American range (Gharrett *et al.* 1982), this could also be interpreted that very little or no population structure is present.

More recently, a study using four microsatellite loci and a single mitochondrial gene, suggested that there may be subtle differentiation between sablefish populations in the northeastern Pacific (Tripp-Valdez *et al.* 2012). Significant F_{ST} values were observed between a collection of fish made near the southern limit of the sablefish range, San Quentin Mexico and collections in the Gulf of Alaska and the Bering Sea (Tripp-Valdez

et al. 2012). Two genetic clusters were also inferred using the program *STRUCTURE* (Pritchard *et al.* 2000) but no clear geographic pattern to the clusters was observed (Tripp-Valdez *et al.* 2012). This analysis is somewhat more sophisticated than other studies done to date in the sense that *STRUCTURE* employs a model based method designed to cluster individuals into genetic populations without any prior information on an individual's location of origin (Pritchard *et al.* 2000). However, caution should be applied when interpreting these results for a number of reasons. First, null alleles were found at all loci and while a correction was applied in this study, the authors mention that the bias in F_{ST} is greater when gene flow is low and a null allele correction is used (Tripp-Valdez *et al.* 2012). Population genetic phenomena can sometimes be responsible for the false impression that null alleles are present in a data set (Dakin & Avise 2004), in which case the conclusion of genetic differentiation is lost. Reported p-values were not significant for any F_{ST} calculations that were not adjusted for the presence of null alleles (Tripp-Valdez *et al.* 2012). Second, population-based sampling schemes may be problematic for species that are continuously distributed, as is the case for sablefish. The four collections made in this study were all at least 1000 km apart, leading to large areas of the species range left un-sampled. Under such scenarios it is possible to infer population structure when in fact a cline or isolation by distance pattern may better describe the data (Serre & Pääbo 2004). Lastly, the values of pairwise F_{ST} that were observed between groups are roughly an order of magnitude lower than the point ($F_{ST} = 0.02$) at which *STRUCTURE* begins to perform poorly and may provide a false sense of the true number of clusters in the data (see Latch *et al.* 2006). Also, while the best *STRUCTURE* model suggested two clusters, the statistic used to select the best supported

number of clusters cannot actually identify the correct number of clusters when it is one (Evanno *et al.* 2005).

iii) Evidence of Fine Scale Population Differences

While the current paradigm suggests that there are two biological stocks of sablefish in the northeast Pacific Ocean and population genetic studies suggest that there may be a single genetic group, studies using other approaches have found evidence of biological differences on a finer geographic scale. For example, stable isotope analysis of otoliths from juvenile sablefish suggests that three spawning stocks may exist from Oregon to Vancouver Island, British Columbia (Gao *et al.* 2004). It is possible that stable isotope signatures may show higher levels of structuring than genetic markers do. As otoliths are deposited in isotopic equilibrium with the environment, stable isotope signatures can serve as a natural tag that can be used to trace individuals back to a site of origin even when genetic markers may fail to do so (Thorrold *et al.* 2001). This can occur if enough time has not passed in order for genetic drift to cause genetic differentiation at neutral genes (Peterson & Fry 1987; Thorrold *et al.* 2001; Gao *et al.* 2004).

Head *et al.* (2014) assessed the size and age at first sexual maturity of sablefish off the West Coast of the U.S. and found differences in life history parameters between regions that might suggest population differences occur within the southern stock. Fish collected south of Cape Mendocino, CA mature at a smaller size and at an older age than fish collected north of this line. Von Bertalanffy growth parameters were also different between fish collected north and south of Cape Mendocino, suggesting that growth is different between fish in these areas.

iv) Ecotypic Variation

It has been suggested that genetic differences could be responsible for the so called “soft flesh problem” observed in sablefish (Norris *et al.* 1987; Karinen *et al.* 2010). Fish caught in deeper water generally have softer flesh, less protein content, increased lipid levels and have a less desirable flesh quality when compared to fish harvested from shallower water (Norris *et al.* 1987; Karinen *et al.* 2010). In addition to flesh condition and desirability, observations by Norris *et al.* (1987) noted that sablefish captured at depths greater than 400 fathoms had darker skin coloration, lighter visceral lining, whiter fillets, increased scale loss, thinner belly walls, and a different rib cage shape when compared to the firm fleshed fish caught at shallower depths. Soft flesh has been observed in other marine species as well, such as dover sole (*Microstomus pacificus*), American plaice (*Hippoglossoides platessoides*) and giant grenadier (*Albatrossia pectoralis*) (Patashnik & Groninger 1964; Hunter *et al.* 1990). In some species the incidence of soft flesh is related to depth but much like sablefish, it is not clear if these individuals represent a genetically distinct deep-water ecotype.

Deep-water ecotypes have been documented for some freshwater fishes, and in particular the lake trout (*Salvelinus namaycush*), where increased lipid levels are presumably an adaptation to living at greater depths (Eshenroder 2008). Interestingly, much like the deep-water sablefish, thinner ventral walls are also a distinctive characteristic of some deep-water lake trout ecotypes (Eshenroder 2008). A genetic basis has been demonstrated for phenotypic differences observed between deep and shallow-water forms of lake trout (Goetz *et al.* 2010, 2014). It has been proposed that there may be some bathymetric separation of sablefish stocks, where adult fish move to a preferred

depth and remain there throughout their life (Norris 1993). If this is the case there may be some adaptive significance to the phenotypic characteristics associated with deep-water sablefish (Norris 1993). If the phenotypic differences observed between shallow and deep-water sablefish are under genetic control, then deep-water sablefish could possibly represent a genetically distinct ecotype.

3) Statement of Purpose

With recent advances in sequencing technology, genomic analyses of non-model organisms have become increasingly cost effective. While there are genetic markers available for sablefish (McCraney *et al.* 2011), microsatellites are generally not found in protein coding regions and are not particularly suited for identifying genes that may be under selection, especially in the absence of a reference genome or linkage map. RAD-seq presents an economical way to gather genetic data from sablefish and generate a large number of SNP markers throughout the genome that can be used to answer questions regarding stock structure and local adaptation in sablefish throughout their range in the Northeast Pacific Ocean. With some genomic resources available (Rondeau *et al.* 2013) genomic regions under selection in sablefish stocks could be identified.

To date, only a few studies have examined the genetic variation and structure of sablefish populations. Results of these studies suggest that current stocks may be composed of mixtures of individuals from many different genetic populations, or that no structure exists (Tsuyuki & Roberts 1969; Gharrett *et al.* 1982; Tripp-Valdez *et al.* 2012). These studies have only evaluated differentiation using neutral genetic markers such as microsatellites that are generally not influenced by the effects of natural selection. While previous studies have attempted to investigate the genetic structure of this species, a

large-scale study using samples collected continuously throughout the extent of their North American range utilizing a large number of markers remains to be done. Combined with a draft genome sequence available for the sablefish (Rondeau *et al.* 2013), the use of RAD-seq would enable the characterization of a large number of SNP markers that can be used for the investigation of population structure. This type of approach is designed to have increased power over studies using a smaller number of loci, enhancing the ability to resolve fine scale population structure. In addition, with a large number of markers, signatures of natural selection can also be identified in the sablefish genome, providing evidence of local adaptation and insight on the biological differences observed across their range. Not only will this type of information aid in further describing the biological differences seen in wild stocks of sablefish but it can also be used in aquaculture programs where certain traits such as fast growth and flesh quality are desired in order to maximize efficiency and marketability (Hedgecock *et al.* 1976; Sonesson & Meuwissen 2009).

This study aims to use RAD-seq to:

- Develop large numbers of SNP markers for sablefish
- Assess neutral population genetic structure using RAD-seq derived SNPs
- Test for signatures of selection in the sablefish genome using RAD-seq derived SNPs

Chapter 2: Genomic Evidence for Large Scale Panmixia in the Sablefish

1) Introduction

The concept of a stock is central to the conservation and management of marine fisheries (Begg & Waldman 1999). Under the stock concept, managers define semi-discrete boundaries to fish populations in order to make decisions regarding those management units (Begg *et al.* 1999). Knowledge of genetic structure or an understanding of localized adaptation may help delineate fish stocks and can be extremely relevant in the management of a fishery (Hauser & Carvalho 2008). Many commercially important marine species fall into the category of “classical marine species”, characterized by large geographic distributions and population sizes, high fecundity, and pelagic larvae (Nielsen & Kenchington 2001). Many of these species typically display relatively low levels of genetic differentiation among population groupings, however significant genetic differentiation has been observed in a number of these species, even at relatively small spatial scales (Nielsen & Kenchington 2001; Nielsen *et al.* 2009a).

The sablefish (*Anoplopoma fimbria*) could be considered a “classical marine species”. It is a long-lived (max age 113 years), demersal species found in the northern Pacific Ocean, where it inhabits deep water (up to 1830 m) along the continental slope (Sasaki 1985; Beamish & McFarlane 1988; Beamish *et al.* 2006). It is found along the west coast of North America from Baja California to the Bering Sea, throughout the Aleutian Islands into waters off the Kamchatka Peninsula, Russia and northern Japan (Sasaki 1985). Based primarily on tagging studies and differences in growth and maturity, it is generally accepted that two sablefish stocks exist in the Northeast Pacific

Ocean (Kimura *et al.* 1998; Hanselman *et al.* 2012b). A northern stock inhabits waters off of the Alaskan Coast and northern British Columbia while a southern stock is found off southern British Columbia and the West Coast of the lower United States (Hanselman *et al.* 2012b).

The sablefish is a highly mobile species, and very little is known about the timing or location of spawning. It is thought that spawning occurs at depth along the continental slope (Mason *et al.* 1983) though no spawning sites have been documented in the literature. The eggs are pelagic and believed to incubate at depths of about 400 m and descend to depths of 1000 m or more after hatching (Alderdice *et al.* 1988; McFarlane & Beamish 1992). During yolk sac absorption the larvae ascend to the surface where they are very abundant in the neuston (Doyle 1992) and the development of eggs to 40 mm larvae is estimated to take about 10 weeks in the wild (Boehlert & Yoklavich 1985; Kendall & Matarese 1987). Juvenile sablefish utilize nearshore areas (1-2 years) before they begin to gradually move back out to the slope as they reach maturity (Maloney & Sigler 2008).

Previous studies have used molecular markers to characterize the population structure of this species (Tsuyuki & Roberts 1969; Gharrett *et al.* 1982; Tripp-Valdez *et al.* 2012). These studies suggest that there is little or no genetic structure across large geographic scales, however they have only examined variation at markers presumed to be neutral to the effects of natural selection. While useful in examining demographic processes, care should be taken when interpreting low or insignificant values of genetic differentiation derived from neutral markers in marine populations, as these estimates may not accurately represent contemporary patterns of stock structure in such large

populations (Allendorf & Phelps 1981; Conover *et al.* 2006). Large population sizes typical of many marine species may make them less susceptible to the effects of genetic drift; therefore apparent genetic homogeneity may persist at neutral genes even at low levels of migration (Allendorf & Phelps 1981; Nielsen *et al.* 2009a). Given that large populations are less susceptible to drift, beneficial alleles are less likely to be lost through drift and may respond to relatively weak selection pressure in marine fish species (Nielsen *et al.* 2009a). Also, many more migrants are needed to effectively counteract the effects of natural selection than are required to maintain genetic homogeneity in the face of genetic drift (Conover *et al.* 2006). So, while neutral markers may fail to show any patterns of significant differentiation, significant adaptive variation may persist (Conover *et al.* 2006; Larsen *et al.* 2007; Hauser & Carvalho 2008). The morphological differences observed by Tripp-Valdez and colleagues (2012) might suggest that natural selection is influencing the genetic stock structure of sablefish. In species such as the sablefish that have large geographic ranges and large population sizes, it may be more appropriate to examine the variation of genes subject to natural selection rather than loci that are presumed to be selectively neutral. Genetic drift alone may not be strong enough to cause genetic differences between stocks and by incorporating non-neutral genes it may be possible to more accurately assess the genetic structure of sablefish stocks. The emerging field of landscape genomics provides a framework to examine the association between environmental variables and genomic variation in wild populations. Schwartz *et al.* (2010) refers to landscape genomics as "...the simultaneous study of tens-to-hundreds of markers, ideally including markers in candidate adaptive genes (genes under selection), with georeferenced samples collected across a landscape". While approaches

like this are not particularly abundant in the literature on classical marine fishes, they have been successfully applied to analyses of local adaptation and population structure in other fishes such as salmon (Zueva *et al.* 2014), and other organisms such as trees (Sork *et al.* 2013; Frichot *et al.* 2013), plants (Bothwell *et al.* 2013; Jones *et al.* 2013) goats (Pariset *et al.* 2009) and humans (Frichot *et al.* 2013).

Here, we used SNP data derived from restriction site associated DNA sequencing (RAD-seq) to 1) evaluate population structure in sablefish collections from the Bering Sea, Gulf of Alaska, and the West Coast of the U.S. using SNP markers throughout the genome, and 2) test for signatures of natural selection in the sablefish genome by associating SNP genotypes with environmental conditions.

2) Methods

a) Sample Collection

Fin clips and were collected and stored in 95% ethanol as a part of the National Oceanic and Atmospheric Administration's National Marine Fisheries Service (NOAA, NMFS) surveys in the summer of 2012. The Fishery Resource Analysis and Monitoring (FRAM) Division at NOAA's Northwest Fisheries Science Center (NWFSC) collected fin clips (n=600) during their annual U.S West Coast Groundfish Bottom Trawl Survey. As part of their annual longline survey, the Marine Ecology and Stock Assessment (MESA) Division of NOAA's Alaska Fisheries Science Center (AFSC) collected fin clips (n=1929) from the Gulf of Alaska. The Resource Assessment and Conservation Engineering Division (RACE) at AFSC collected fin clips (n=140) from sablefish opportunistically encountered as part of their trawl survey of the Bering Sea Slope. In addition to fish clips, additional biological data were collected for each sample as well;

length, weight, sex, and maturity (FRAM and MESA). The three major regions (West Coast, Gulf of Alaska, Bering Sea) covered by the surveys, were further divided into eleven sub-regions, to evenly choose samples collected across the study area (Figure 2).

In order to restrict the genetic analysis to adult fish, maturity data were used to select a subset of samples. The data from the macroscopic examination of maturity at the time of sampling could not be used because maturity data were not collected on fish caught in the Bering Sea. The maturity data for the fish collected in the Gulf of Alaska and West Coast were pooled and logistic regressions were fit separately for males and females to determine the length at 50% probability of maturity (L_{50}) (Figure S 1). Only fish above the size thresholds set by L_{50} were considered for genetic analysis. A logistic regression model (GLM, R version 3.1.2) was fit to the data to estimate the parameters of the equation where the probability of being mature (P) was calculated by:

$$P = \frac{e^{\beta_0 + \beta_1 X}}{1 + e^{\beta_0 + \beta_1 X}}$$

where X is length and β_0 and β_1 are regression coefficients (Quinn & Deriso 1999). The inflection point of the L_{50} was calculated by:

$$L_{50} = \frac{-\beta_0}{\beta_1}$$

Males and females above the size thresholds (females: 614.8 mm, males: 573.7). were then randomly selected from each sub-area to help ensure an even sample distribution across the study area. In areas where a small number of individuals were above those size thresholds, the largest individuals were selected. Out of the total number of fin clips

(n=2669), 441 (West Coast: 144, Gulf of Alaska: 224, Bering Sea: 73) were selected for RAD-seq.

b) Genomic Library Construction

Genomic DNA was extracted using Qiagen's DNeasy Blood and Tissue Kit (Qiagen, Valencia, CA). DNA was quantified using the Quant-iT PicoGreen dsDNA Assay Kit (Invitrogen, Carlsbad, CA), and concentrations were adjusted to 12.5 ng/μl. Libraries were prepared as described in Miller *et al.* (2007) with modifications described below. DNA from each sample (500 ng) was digested with either *SbfI* or *PstI* restriction endonucleases (New England Biolabs, Ipswich, MA). Since the *PstI* recognition site is part of the *SbfI* recognition site, the samples digested with *PstI* should contain all of the *SbfI* loci making them suitable for inclusion in this analysis. Barcoded P1 adapters unique to each individual and complimentary to the restriction enzyme cut site were ligated onto the digested DNA of each individual. For single end sequencing runs, barcoded, digested DNA was then pooled into libraries containing DNA from 10 (*PstI* digestions) or 12 (*SbfI* digestions) individuals for single end sequencing and a library of four individuals was prepared for paired end sequencing. DNA was then randomly sheared with a Q-sonica Q800R DNA Sonicator (QSonica, LLC, Newtown, CT). Gel electrophoresis was used to identify and cut out DNA fragments that were 300-500 bp in length for single end libraries and 200-400, 400-600, 600-900 bp for paired end. Illumina P2 adapters were then blunt-end ligated to the other end of the DNA fragments. Libraries were quantified via qPCR using Kapa Biosystems, Kapa Illumina library quantification kit (Kapa Biosystems Inc, Woburn, MA, USA) on an ABI 7900 real time PCR system

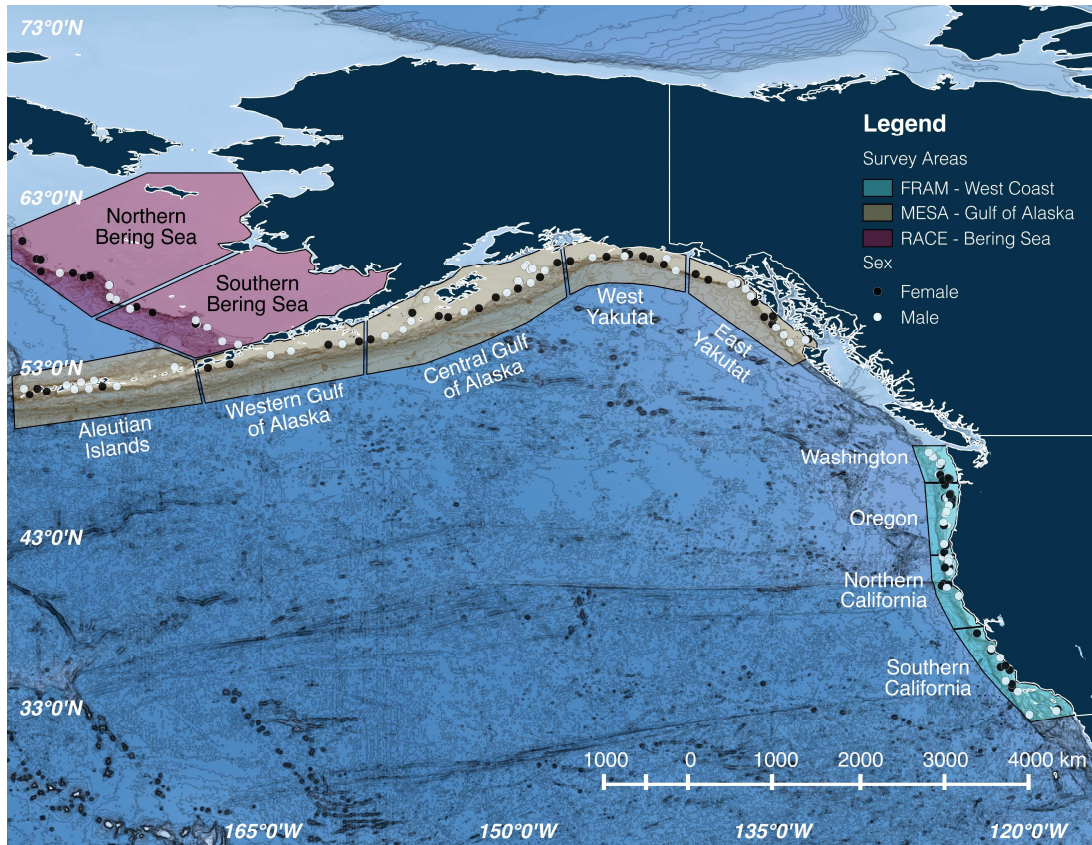


Figure 2. Map of sablefish samples used for RAD sequencing in this study. Survey areas are shaded (red = RACE/Bering Sea, yellow = MESA/Gulf of Alaska, green = FRAM/West Coast of the U.S.). Black circles represent female fish and white circles are males.

(Life Technologies, Carlsbad, CA). Four libraries were pooled in equimolar concentrations prior to sequencing.

Single end sequencing libraries consisting of samples digested with the restriction endonuclease *SbfI* were sequenced on an Illumina HiSeq 2500 (Illumina, San Diego, CA) for read lengths of 100 bp. From preliminary sequencing runs we determined that four *SbfI* digested libraries of 12 individuals could be pooled into a single lane of 48 samples for the single end runs. Paired end sequencing was done on a library of 4 individuals digested with *PstI* on the Illumina HiSeq 2500 for read lengths of 100 bp in a single lane and a single library of 10 individuals digested with *PstI* on an Illumina MiSeq for read lengths of 250 bp in a single lane.

c) Processing Sequence Reads

FastQC (Andrews 2012) was used to assess the quality of each base position in the sequence reads for each Illumina sequencing run. Based on that, the raw sequence reads were de-multiplexed, quality filtered and trimmed to a length of 75 bp using *process_radtags* in Stacks (version 1.21) (Catchen *et al.* 2011, 2013). Any read with an uncalled base was removed as well as reads with low quality scores. Bowtie (version 1.0.0) (Langmead *et al.* 2009) was used to align sequence reads from each individual to an index built from both the sablefish draft genome (GenBank: AWGY00000000.1) and mitogenome (NCBI RefSeq: NC_018119.1) (Rondeau *et al.* 2013). Up to three mismatches were allowed when aligning reads back to the genome and reads with more than one valid alignment were discarded. The Stacks program *pstacks* was used to genotype individuals. A minimum of five reads was required to form a stack at a locus. The bounded SNP calling model implemented in *pstacks* was used (upper bound: 0.1,

lower bound: 0.001, α : 0.05). Eleven individuals were then selected at random from each of the 11 geographic sub-regions (Figure 2) to create a catalog of reference loci in the Stacks program *cstacks*. The Stacks program *sstacks* was then used to match loci in each individual back to the catalog of reference loci. The program *populations* was used to export a Variant Call Format (VCF) file (Danecek *et al.* 2011) containing the genotype information for the set of individuals. A minimum stack depth of eight was required for a locus to be processed by *populations*.

d) SNP Quality Filtering

First, to restrict the analyses to nuclear SNPs, all RAD loci in the mitochondrial genome were removed from further analysis. Next, RAD tags were culled from further filtering if they contained more than three SNPs. The program *VCFTools* (v0.1.12) (Danecek *et al.* 2011) was then used to filter out loci with a global minor allele frequency (MAF) of 0.05 and more than 20% missing data. Individuals were then removed from further analyses if they were missing genotypes at more than 40% of the filtered loci. For tags that contained more than one SNP, only the SNP with the highest minor allele frequency was retained. Linkage disequilibrium was assessed by calculating the r^2 statistic between each pair of SNPs using the program *Plink* (v1.07) (Purcell *et al.* 2007). In cases where $r^2 > 0.8$, the SNP with more missing data was culled from the dataset. Exact tests for Hardy-Weinberg equilibrium (HWE) were conducted in *VCFTools* and SNPs that were out of HWE in any one population (West Coast, Bering Sea, Gulf of Alaska) were removed in all populations prior to conducting analyses that required population groups to be defined *a priori*.

e) Population Structure

To assess the genetic structure of sablefish across their sampled range, we combined the use of traditional population genetic techniques with individual based analyses that may be better suited when individuals are sampled across a landscape (Jones *et al.* 2013). We pooled individuals into population groups defined by the survey on which they were collected (West Coast: FRAM, Gulf of Alaska: MESA, Bering Sea: RACE). Levels of expected heterozygosity were calculated in *VCFTools* for each population and F_{IS} was calculated for each SNP both globally and in each survey area. Fisher's exact test as implemented in *Genepop* v4.2.2 (Rousset 2008) was used to test for differences in allele frequency between survey areas. Markov chain parameters of 10,000 dememorization steps, 100 batches and 5000 iterations per batch were used for the exact test in *Genepop*. *POWSIM* (Ryman & Palm 2006) was used to assess the statistical power of this dataset to detect differences in allele frequency between survey areas at known levels of genetic divergence or F_{ST} for both the global and pairwise tests. Twenty five replicate runs were used with the same Markov chain parameters as *Genepop* and the number of samples selected from each population at the end of the drift simulations was defined by the sample sizes of each area in the filtered SNP dataset used for this study. A version of *POWSIM* was specifically compiled to accommodate the size of this dataset (Ryman & Palm 2006). *Genepop* was also used to test for differences in genotype frequency between the survey areas using the same Markov chain parameters used to test for differences in allele frequency. F_{ST} (Weir & Cockerham 1984) was estimated globally and between all group pairs using *Genepop*. The R package *StAMPP* (Pembleton *et al.* 2013) was used to test for significance of the pairwise F_{ST} values obtained in *Genepop*. *StAMPP* was chosen because it also estimates F_{ST} using the

method of Weir & Cockerham (1984) and is comparable to *Genepop*. Individuals were randomized among groups for 1000 permutations and a p-value was obtained by converting the simulated values to an object of the class *randtest* in the R package *ade4* (Chessel *et al.* 2004).

Isolation by distance (IBD) was evaluated by calculating Rousset's \hat{a} (Rousset 2000) between individuals in *SPAGeDi* v1.4 (Hardy & Vekemans 2002). Geographic distance between individuals was calculated as great-circle distance in the *fields* package in R (Nychka *et al.* 2014). Distance between individuals along the 500 m isobath was also estimated using the R package *gdistance* to route individuals between sample locations along the 500 m isobath (van Etten 2014). Significance of IBD relationships were assessed by Mantel's test with 1000 permutations implemented in R using the *vegan* package (Oksanen *et al.* 2013). Values of great circle distance were log transformed prior to analysis to account for the possibility of dispersal in two dimensions, while distances along the 500 m isobath were not because under that dispersal scenario, it is assumed to take place in one dimension (Rousset 1997).

To investigate the possibility of cryptic population genetic structure we used a combination of methods that do not rely on predefined population groupings. First, the program *SmartPCA* in the software package *Eigensoft* v5.0.2 was used to compute the principal components (Patterson *et al.* 2006). Default outlier removal parameters were used in *SmartPCA* ($> |6\sigma|$ on one top 10 PCs, 10 iterations). Principal component analysis (PCA) was used because it does not rely on underlying population genetic models and does not attempt to group individuals into populations (Patterson *et al.* 2006). Individuals are placed along axes of variation (principal components) and once plotted, a

representation of the genetic structure in the sample can be visualized (Patterson *et al.* 2006). The spatial clustering algorithm implemented in *BAPS6* was also used to examine population genetic structure (Corander *et al.* 2008). Unlike PCA, *BAPS6* uses a model that attempts to minimize linkage and Hardy-Weinberg disequilibrium within groups of samples while also incorporating spatial information to improve the partitioning of individuals into biologically meaningful groups (Latch *et al.* 2006; Corander *et al.* 2008). Partitions of one to 30 groups were tested in *BAPS6*.

f) Landscape Genomics

i) Environmental Variables

The samples obtained for this study were collected using bottom contact fishing methods so depth of capture for each individual was extracted from the *gebco_08* 0.5 arc second grid bathymetry data (IOC, IHO & BDOC 2003) using the GPS location of each individual in *QGIS* (QGIS Development Team 2009). Environmental conditions at the sampling location of each individual were estimated using salinity and temperature profiles collected by the Argo profiling float network. These data were collected and made freely available by the International Argo Program and the national programs that contribute to it, (<http://www.argo.ucsd.edu>, <http://argo.jcommops.org>). The Argo Program is part of the Global Ocean Observing System. Argo data from the time period that the samples were collected (03/01/2012 – 10/31/2012) were obtained from the Coriolis project's webserver (<http://www.coriolis.eu.org>). Temperature and salinity were estimated using the sampling location and bottom depth for each individual using the computer program *Ocean Data View* v4.6.2 (Schlitzer 2014). To reduce collinearity among the environmental variables, Pearson's correlation coefficient (r) was calculated

between each pair of variables in *R* (R Development Core Team 2013) and if a pair of variables were correlated with an $|r| > 0.8$, one was removed from the association analyses.

ii) Association Analyses

Latent factor mixed models implemented in *LFMM* (Frichot *et al.* 2013) were used to test for genotype – environment associations. *LFMM* has been shown to reduce the number of false positives in association studies by accounting for past population history including patterns of isolation by distance (Jones *et al.* 2013). Population structure is modeled using unobserved variables or “latent factors” (Frichot *et al.* 2013). We chose to use a single latent factor (*k*) because our analyses suggested the existence of a single cluster. *LFMM*’s Gibbs sampler was set to estimate model parameters using 10,000 sampling iterations with 1,000 iterations discarded as burn-in. The *LFMM* results did not change over replicate runs for the same variable so we determined that the Gibbs sampler was run long enough. P-values obtained in *LFMM* were adjusted for a false discovery rate (FDR) of 0.05 by calculating q-values (Storey & Tibshirani 2003) in the *R* package *q-value* (Dabney & Storey 2014).

iii) Annotation

Single nucleotide polymorphisms with a significant association to an environmental variable were annotated by using Ensembl’s BLASTN search tool (http://www.ensembl.org/Gasterosteus_aculeatus/Tools/Blast?db=core) to align the full contig sequence from the sablefish genome assembly (Rondeau *et al.* 2013) containing the SNP/RAD tag of interest to release 78 of the stickleback (*Gasterosteus aculeatus*) BROAD S1 genome (Kersey *et al.* 2014). An E-value of $< 1e^{-03}$ was used as a cutoff to

identify valid alignments. The alignments were then viewed using the BROAD S1 genome browser in Ensembl (Kersey *et al.* 2014). The stickleback was chosen because it is the most closely related species to the sablefish with a complete genome (Rondeau *et al.* 2013). It is thought that the divergence of sablefish and stickleback lineages occurred less than 150 mya and possibly as recent as 50 mya (Rondeau *et al.* 2013) A high degree of synteny between stickleback and sablefish chromosomes is also observed (Rondeau *et al.* 2013).

3) Results

a) Sequencing

There were on average, 145,960,310 reads per lane for the Illumina HiSeq runs. On average we were able to recover 80% of the barcoded reads per lane with 3,274,890 reads per individual. On average, 65% of the reads for each individual aligned to the draft genome unambiguously, 16% of the reads had more than 1 valid alignment and were discarded, and 19% failed to align altogether. The *Stacks* pipeline produced 100,317 SNPs in 29,242 tags and after SNP quality filtering in *VCFtools* and *Plink*, the final dataset consisted of 2,661 SNPs in 404 individuals (West Coast: 117, Gulf of Alaska: 217, Bering Sea: 70) (File S 1, File S 2, File S 3). While 441 individuals were submitted for sequencing, 37 were removed due to missing data (> 40%), leaving 404 individuals for subsequent analyses (File S 1).

b) Population Structure

For analyses that required *a priori* population groupings (West Coast, Gulf of Alaska, Bering Sea), a number of SNP loci (624) were found to be out of HWE and were filtered prior to any population genetic analyses, for a total of 2,037 SNPs. The removal

of SNPs from the data set out of HWE brought the mean F_{IS} across loci closer to 0 both globally and for all population groups (Table S 1, Figure S 2). Differentiation among groups was very low (global $F_{ST} = 0.0002$). Pairwise estimates of F_{ST} between the three survey areas were all 0.0002 (Table 1) and non-significant. No significant differences in allele frequencies (Table 2) or genotype frequencies (Table 3) were detected between population pairs. The simulations conducted in *POWSIM* revealed that power was not sufficient at the levels of differentiation (F_{ST}) observed to identify differences in allele frequency between population pairs using Fisher's exact test (Table 2).

A small number of individual pairs had very low estimates of \hat{a} (< -0.4) between them (Figure 3 a & b). These were considered to be outliers and all comparisons with those samples were removed from the analysis. A number of samples were collected within 1 km from another sample. The log transformation of great circle distance values less than 1 km produced negative values (Figure 3 a) and, as a result, the analysis was restricted to comparisons with a distance greater than 1 km between samples (Figure 3 c & d). The relationship between genetic (\hat{a}) and geographic distance between individuals was non-significant for all tests prior to outlier removal and after, suggesting that there is no pattern of isolation by distance in sablefish (Table 4, Figure 3).

Results from the Bayesian clustering and PCA did not validate the existence of genetic groups defined by geographic region or reveal any cryptic patterns of population genetic structure. Seven samples were more than six standard deviations along one of the first 10 principal components and flagged as outliers by *SmartPCA*. These samples were excluded from the PCA as a result. The first principal component explains a very small

Table 1. Table of F_{ST} estimates for each pair of survey groups. F_{ST} below diagonal and p-values above.

	Bering Sea	Gulf of AK	West Coast
Bering Sea	-	0.1199	0.0889
Gulf of AK	0.0002	-	0.0589
West Coast	0.0002	0.0002	-

Table 2. P-values for tests of allelic differentiation between each population pair and statistical power of each comparison calculated in POWSIM for the level of F_{ST} observed between each population pair (Table 1). P-values below diagonal and power above.

	Bering Sea	Gulf of AK	West Coast
Bering Sea	-	0.0000	0.0000
Gulf of AK	0.9997	-	0.0000
West Coast	0.9970	0.8230	-

Table 3. P-values for tests of genotypic differentiation between each population pair.

	Bering Sea	Gulf of AK	West Coast
Bering Sea	-		
Gulf of AK	0.9998	-	
West Coast	0.9992	0.8886	-

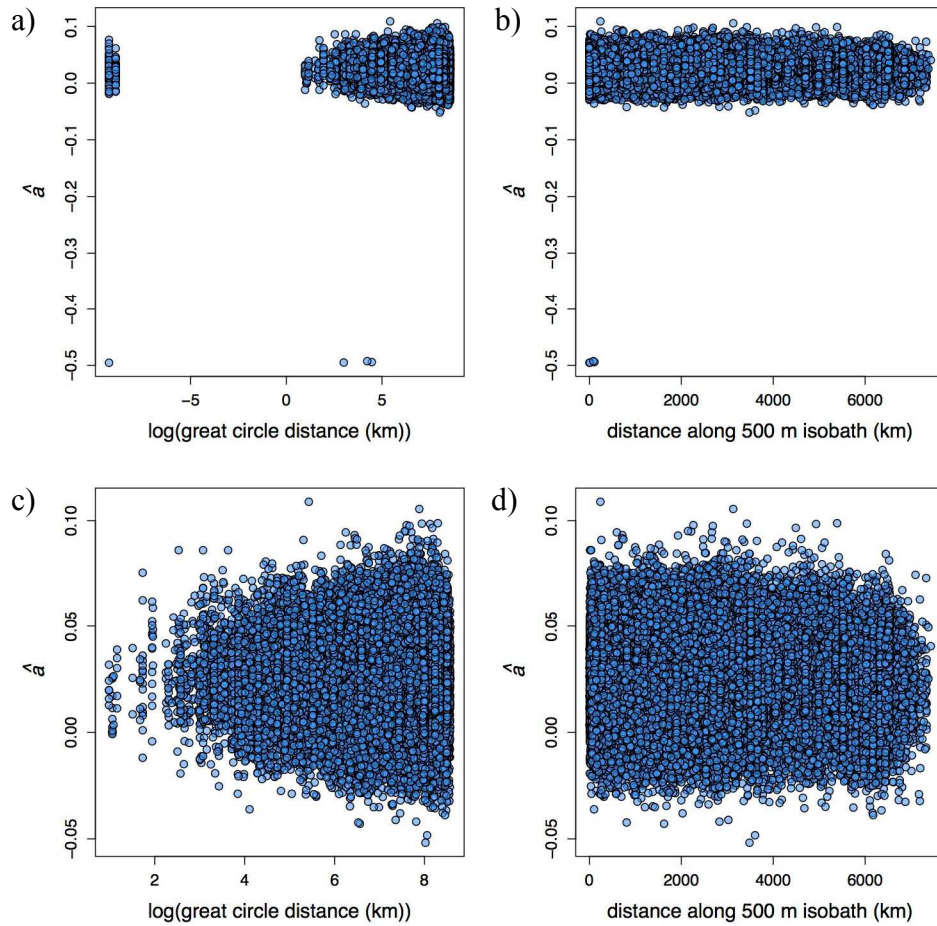


Figure 3. Patterns of isolation by distance for sablefish in the northeast Pacific Ocean. Pairwise genetic distance between individuals (\hat{a}) is plotted against geographic distance calculated as log of great circle distance (a) and distance along the 500 m isobath (b) and the before removing outliers and comparisons where great circle distance is less than 1 km. Relationship between log of great circle distance (c) and distance along the 500 m isobath (d) after outliers were removed.

Table 4. Mantel correlation coefficients used to test the association between geographic and genetic distance matrices.

	Distance Measure	Mantel's r	p-value
w/outliers	log(great circle distance (km))	0.0086	0.1788
	distance along 500 m isobath (km)	-0.0160	0.8482
outliers removed	log(great circle distance (km))	0.0000	0.5095
	distance along 500 m isobath (km)	-0.0164	0.8392

proportion of the total variation (0.73 %). When individuals were plotted on the first two principal component axes calculated from the SNP dataset in *SmartPCA*, a single group of individuals became apparent (Figure 4). There was no clear geographic pattern to the placement of individuals along the PCA axes. Similarly, a single group was also identified by *BAPS6*.

c) Environmental Associations

Out of the five environmental variables selected for this analysis, salinity and depth were correlated with one another (Pearson's $r > |0.8|$). Salinity was excluded from further analyses as a result. After p-values were adjusted for multiple testing, the latent factor mixed models implemented in *LFMM* detected two significant SNP associations with depth (Table 5). These SNPs were contained in RAD tags 9124 and 12048 (Table 5, File S 2, File S 3).

i) Annotation

Each of the sablefish genome contigs containing the SNPs associated with depth in this study were successfully aligned to the stickleback genome. The sablefish contig 152341 (GenBank:AWGY01151885.1) containing the SNP in RAD tag 9124, aligned to the stickleback linkage groupXX (Table S 2). The contig did not align continuously to this region but many fragments aligned in order with gaps between them. While the portion of the contig containing the SNP (bp 11333) did not align directly to the stickleback genome, based on the order of the contig segments that did align, the SNP is presumably part of an intron of the ETS translocation variant 1 (*etv1*) gene (Ensembl ID: ENSGACG00000012802.1). The alignment was similar for sablefish contig 133848

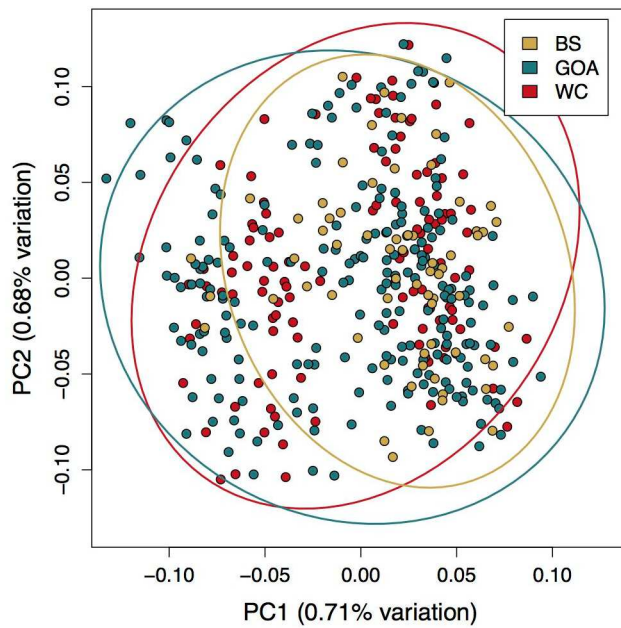


Figure 4. Plot of the first two principal component axes in a principal component analysis of 2661 RAD-seq derived SNPs. Points are colored by the area (BS: Bering Sea, GOA: Gulf of Alaska, WC: West Coast) that each individual was collected. Lines represent 95% confidence ellipses calculated using the R package, *vegan* (Oksanen *et al.* 2013) for each group.

(GenBank: AWGY01133477.1). A number of fragments aligned in order to stickleback linkage group VI (Table S 3). Based on the alignment order of the fragments, this SNP (bp 4995) occurs in an intron of the gene that codes for the FAM196ab protein (*fam196ab-201*) (Ensembl ID: ENSGACG00000009611.1).

4) Discussion

a) Population Structure

Using thousands of SNP markers distributed throughout the sablefish genome, we were unable to uncover evidence that would suggest that these stocks are genetically distinct. While the current paradigm of sablefish stock structure suggests the existence of northern and southern biological stocks, results from population genetic and individual based analyses are suggestive of a single genetic group across the North American range of the sablefish. Our results agree with earlier population genetics studies that also were unable to identify distinct population genetic structure in the sablefish (Tsuyuki & Roberts 1969; Gharrett *et al.* 1982; Tripp-Valdez *et al.* 2012). Allele frequencies were identical for all survey areas, suggesting that these regions are part of a single gene pool. The F_{ST} estimates from these data were non-significant and very small suggesting that considerable gene flow occurs between these areas. This would serve to homogenize allele frequencies and may be why we could not detect any significant difference in allele frequency between the survey areas. A very minor proportion of the genetic variability observed could be explained by the survey areas incorporated in the study. The lack of genetic structure is not surprising when viewed in light of tagging data that suggests these fish move frequently and over long distances (Kimura *et al.* 1998; Maloney & Sigler 2008; Morita *et al.* 2012; Hanselman *et al.* 2015).

Table 5. Significant SNP – environment associations found using LFMM. Q-values were calculated from p-values using an FDR threshold of 0.05.

RAD Tag	Contig	BP Position	Variable	p-value	q-value
9124	gi 559591484 gb AWGY01151885.1	11333	Depth	6.35E-06	0.0338
12048	gi 559609892 gb AWGY01133477.1	4995	Depth	2.81E-08	0.0003

Both the PCA and Bayesian clustering analyses yielded results that may support the notion of panmixia. However, Bayesian clustering methods like *BAPS6* may have difficulty resolving the population structure when levels of differentiation are low among groups or individuals are continuously distributed throughout the sampling area (Latch *et al.* 2006; Frantz *et al.* 2009). Both of those conditions were met in this study. Principal component analysis on the other hand, does not necessarily suffer from reduced performance under these circumstances. Principal component analysis does not attempt to classify individuals into groups, rather it places individuals along axes of variation (Patterson *et al.* 2006). So while PCA can depict clusters in the data when they are present it can also represent genetic structure along a continuum and is useful in situations where groups overlap. This is attractive in scenarios where model based clustering methods may not be able to accurately characterize groups. In the current analysis, both the spatial clustering method implemented in *BAPS6* and the PCA suggested the existence of a single genetic group and results from *BAPS6* made sense in light of the results from the population genetic analyses as well. While *BAPS6* may have identified a single group as a result of reduced performance, the PCA and population genetic analyses confirm this result.

Our analysis of isolation by distance is also supportive of a single panmictic population of sablefish. A non-significant relationship was observed between genetic and geographic distance, suggesting that dispersal occurs over large spatial scales. Sablefish undergo a pelagic larval period and utilize nearshore areas as juveniles before they move back to deep water habitat on the continental slope (Sasaki 1985; Maloney & Sigler

2008). To our knowledge there are no studies that have examined the capacity for natal homing in this species, so the location where an individual migrates back to the slope could be very far from where it was spawned. That, combined with the capacity for frequent long distance movements as adults (Kimura *et al.* 1998; Maloney & Sigler 2008; Morita *et al.* 2012; Hanselman *et al.* 2015), would not promote the development of a pattern of genetic isolation by distance. The pattern of panmixia or genetic homogeneity we present here is not unique to sablefish; it has been proposed in the literature for a number of other marine fish species (Table 6). Microsatellite data from orange roughy (*Hoplostethus atlanticus*) samples collected from seamounts throughout the North Atlantic Ocean suggests a pattern of panmixia over geographic scales comparable to those presented here (White *et al.* 2009). Panmixia has also been suggested for Greenland halibut (*Reinhardtius hippoglossoides*) in the Northwest Atlantic (Roy *et al.* 2014). Low levels of differentiation were observed between sample collections and a clear pattern of isolation by distance could not be identified using both straight line distance and distance calculated along ocean currents (Roy *et al.* 2014). Off the coast of California, microsatellite analyses have also uncovered patterns of genetic homogeneity in the kelp rockfish (*Sebastes atrovirens*) (Gilbert-Horvath *et al.* 2006). Multiple studies using both microsatellite and mitochondrial DNA have considered panmixia as a possibility for red snapper (*Lutjanus campechanus*) in the Gulf of Mexico (Gold & Richardson 1998; Heist & Gold 2000; Gold *et al.* 2001; Garber *et al.* 2004).

While these studies report low non-significant estimates of F_{ST} , the values are generally higher than what we observe here for sablefish. Interestingly, the values of F_{ST} we present here have been reported for two species that are known to be panmictic, the

European eel (*Anguilla anguilla*) ($F_{ST} \sim 0.0007$ Als *et al.* 2011; Pujolar *et al.* 2014) and American eel (*Anguilla rostrata*) ($F_{ST} \sim 0.00009$ Côté *et al.* 2013). Lowe & Allendorf (2010) suggest that it requires approximately 10 migrants per generation ($F_{ST} < 0.02$) for gene flow to maintain similar allele frequencies among populations, the estimates of population differentiation (F_{ST}) presented here are much smaller than what is required to prevent genetic differentiation due to genetic drift.

While this study does not address all of the biological characteristics that may drive population genetic structure in marine fish species, it is interesting to compare some characteristics of species (including sablefish) for which panmixia has been suggested (Table 6). While there are some commonalities between these species, there seems to be a range of biological characteristics represented. The main unifying characteristic of these species is that they all undergo a pelagic larval period, but it is important to indicate that a pelagic larval period is not necessarily unique to panmictic species. The species in Table 6 can be further divided into two groups; deep (> 100 m) and shallow (< 100 m) water species. In this case, the life history characteristics of sablefish appear to be similar to other deep-water panmictic species that have large adult migration distances, are relatively long lived, and have long pelagic larval periods. While the pelagic period for sablefish may be shorter than that of other species (Table 6), juveniles spend a considerable amount of time nearshore ($\sim 1-2$ years) before migrating back to the slope (Sasaki 1985; Maloney & Sigler 2008). So it is reasonable to think that an individual may move back to the slope very far from where they were originated during spawning.

Adult movements of sablefish seem to be on scales similar to that observed for other deep-water panmictic marine fish (Table 6) and mark recapture studies have shown

that sablefish have the capacity for very complex long distance movements as adults (Beamish & McFarlane 1988; McFarlane & Saunders 1993; Kimura *et al.* 1998; Maloney & Sigler 2008; Morita *et al.* 2012; Hanselman *et al.* 2015). For the northern stock, annual movement probability among management areas is quite high (Hanselman *et al.* 2015). It is common for individual sablefish to be recaptured over 200 km from where they were tagged and released (McFarlane & Saunders 1993; Kimura *et al.* 1998; Morita *et al.* 2012). A recent ongoing tagging study estimates that the average distance traveled in a single year by sablefish in the northern stock is 191 km and this number is much larger for the total time between tagging and recapture (Hanselman *et al.* 2015). There may be a density dependent nature to these movement patterns in the northern stock. Large fish tend to move more when spawning biomass is low while small fish have more erratic movement patterns and may move more in years when recruitment is weak (Hanselman *et al.* 2015). For a long lived species like the sablefish where spawning biomass varies (Stewart *et al.* 2011; Hanselman *et al.* 2012b) and recruitment tends to be erratic (King *et al.* 2000; Beamish *et al.* 2006), movement patterns may not be consistent from year to year. Considering the long life span of sablefish and these complex movement patterns, there are ample opportunities for gene flow to occur.

While the life history characteristics of sablefish are one possible explanation for the apparent genetic homogeneity we observe, the study of differentiation among known

Table 6. Some life history characteristics of marine fish in which panmixia has been proposed.

Species	Marker Type	Adult migration distance	Time to hatch (days)	Pelagic Larval Duration	Depth Range (m)	Lifespan (Years)	Fecundity (eggs/female)
Kelp rockfish (<i>Sebastes atrovirens</i>)	Microsatellite ($F_{ST} = 0.0014$) ¹	Sedentary ²	-	2-3 months ¹	0 – 45 ³	20-25 ¹	-
Orange roughy (<i>Hoplostethus atlanticus</i>)	Microsatellite ($F_{ST} = -0.001$) ⁴	100's km ⁵	7-10 ⁶	7-10 months ⁷	450 – 1800 ^{8,9}	> 84 ⁹	27,180-101,440 ¹⁰
Greenland halibut (<i>Reinhardtius hippoglossoides</i>)	Microsatellite ($F_{ST} = 0.002$) ¹¹	50 - 2500 km ¹²	-	8-9 months ¹³	400 - 1500 ¹⁴	30 – 35 ^{15, 16}	32,500-277,000 ¹⁷
Red Sanpper (<i>Lutjanus campechanus</i>)	Microsatellite, mtDNA ($F_{ST} = -0.004 - 0.012$) ^{18, 19, 20, 21}	2-29 km ^{22, 23}	-	28-30 days ²⁰	0 – 110 ²⁴	> 20 ^{26, 27}	-
Sablefish	RAD SNPs ($F_{ST} = 0.0004$)	200 - 600 km ^{28, 29, 30, 31}	15 ^{32, 33}	2-3 months ^{33,}	100 – 1800 ³⁵	> 40 ^{36, 37, 38}	51,000-556,000 ^{39, 40}

48

Citations as superscripts: 1. Gilbert-Horvath *et al.* (2006), 2. Van Dykhuzien (1983), 3. Miller & Lea (1972), 4. White *et al.* (2009), 5. Francis & Clark (1998), 6. Bulman & Koslow (1995; Zeldis *et al.* (1995), 7. Mace *et al.* (1990), 8. Shephard *et al.* (2007), 9. Branch (2001), 10. Minto & Nolan (2006), 11. Roy *et al.* (2014), 12. Boje (2002), 13. Ådlandsvik *et al.* (2004), 14. Bowering & Nedreaas (2000), 15. Treble *et al.* (2008), 16. Gregg *et al.* (2006), 17. Gundersen *et al.* (2001), 18. Gold *et al.* (2001), 19. Heist & Gold (2000), 20. Gold & Richardson (1998), 21. Garber *et al.* (2004). 22. Strelcheck *et al.* (2007), 23. Patterson *et al.* (2001), 24. Rummer & Bennett (2005), 26 Wilson & Nieland (2001), 27. Baker & Wilson (2001), 28. Hanselman *et al.* (2015), 29. Kimura *et al.* (1998), 30. Morita *et al.* (2012), 31. McFarlane & Saunders (1993), 32. Alderdice *et al.* (1988), 33. Boehlert & Yoklavich (1985), 34. Kendall & Matarese (1987), 35. Sasaki (1985), 36. Kimura *et al.* (1993), 37. Beamish & McFarlane (2000), 38. Head *et al.* (2014), 39. Hunter *et al.* (1989), 40. Mason *et al.* (1983)

spawning populations of sablefish is difficult because there is little data available on the specific location or time of spawning for the sablefish. As a result, the sample collections made for this study were unable to target spawning groups of sablefish that has the potential to influence the results of population genetic analyses in particular. For example, if sablefish undergo migration to specific spawning locations, ideally those groups would be targeted during spawning for sample collection and used to define populations. If sablefish populations are mixed outside of the breeding season then low levels of differentiation and apparent genetic homogeneity could be observed as a result of sampling design (Allendorf & Luikart 2007). While specific spawning aggregations may exist, it is unlikely that these fish exhibit philopatry or home site fidelity because one might expect discrete groups to be identified by the clustering analyses or a significant relationship between genetic and geographic distance if migrations outside of the spawning season are limited.

Another possible explanation for the patterns we see here relates to the large population sizes typical of many ‘classical’ marine fishes. In certain circumstances it is possible for low levels of differentiation to be observed between populations that have been recently isolated from one another (Nielsen *et al.* 2009a). If this occurs, it may take many generations for population genetic parameters to reach equilibrium after the isolation event occurs especially in large populations (Hellberg *et al.* 2002). The low levels of genetic differentiation observed here are not likely artifacts of a recent isolation event because long term tagging data clearly demonstrates contemporary movement is occurring at a relatively high rate (Hanselman *et al.* 2015).

b) Environmental Associations

Taking a genomic approach also enabled us to analyze SNP – environment interactions that may be suggestive of adaptation along environmental gradients. While we were able to identify two SNPs associated with depth, some considerations should be kept in mind when interpreting these data. It is important to acknowledge that selection may vary on a temporal scale as well as spatially. Our sample collections were made on adult fish or fish with a high probability of maturity, so tests that associate environmental conditions at the time and location of sampling likely represent a small fraction of the possible environmental conditions and selective forces that the individual has experienced throughout its lifetime. It is possible that selection occurred at a different time or in a different location than the individual previously inhabited; a likely scenario for the sablefish as it is long-lived and highly mobile. Also, significant SNP – environment associations are not necessarily considered direct observations of adaptation (Nielsen *et al.* 2009a). Genetic differences must be tied to trait variation and fitness differences between resident and non-resident individuals in the local area to demonstrate local adaptation (Nielsen *et al.* 2009a). Perhaps incorporating age data or sampling multiple recruiting cohorts of juvenile fish may help identify genomic signatures of selection that vary over time. Beamish *et al.* (2006) proposed that age specific adaptations may help sustain sablefish populations and maintain their ability to respond to climate change. However, disentangling the effects of selection over time with age data may be a challenge for a species such as the sablefish that is long lived and is difficult to age accurately (Hanselman *et al.* 2012a).

In cases where migration rate is large relative to the strength of selection and the strength of selection for alternate alleles is different between two populations, locally adaptive alleles are easily lost. This phenomenon is termed gene-swamping (Lenormand 2002) or patch-swamping (Felsenstein 2015). In order for selection to overcome a high rate of gene flow, it must be very strong. The large distances that sablefish move throughout their life and the high annual movement probabilities (Hanselman *et al.* 2015) likely mask signals of selection on a spatial scale in adult fish populations. For example, if the environment imposes natural selection at an earlier life stage (e.g. during larval development, recruitment, or when juvenile sablefish are in nearshore areas), environmental conditions associated with the sampling location of adult fish may not capture this selective signal, especially if dispersal distances are long. This could be the case for the sablefish since recruitment can be variable from year to year and has been linked to environmental conditions (King *et al.* 2000). Juvenile sablefish spend 1-2 years in nearshore water before they move back to the slope (Sasaki 1985; Maloney & Sigler 2008), so again, it may be interesting to examine earlier life stages for genomic signatures of selection before long distance adult dispersal occurs.

c) Relation to Fine Scale Biological Differences and Other Population Structure

Although the results we present here agree with findings of previous genetic studies that suggest very little or weak population structuring (Tsuyuki & Roberts 1969; Gharrett *et al.* 1982; Tripp-Valdez *et al.* 2012), other studies have documented biological differences that occur on a finer geographic scale. For example, Gao *et al.* (2004) uncovered evidence of three potential stocks between Vancouver, British Columbia and Oregon by examining stable isotope ratios ($^{18}\text{O}/^{16}\text{O}$: $\delta^{18}\text{O}$, and $^{13}\text{C}/^{12}\text{C}$: $\delta^{13}\text{C}$) in juvenile

sablefish otoliths. Otoliths are formed very close to oxygen isotopic equilibrium with the environment and while carbon isotope ratios tend to reflect dietary shifts, examination of these ratios can provide insight to environmental conditions that individuals have experienced earlier in life (Gao *et al.* 2004, 2010). These results would suggest that natal conditions vary in this region. While our results suggest no population genetic structure and little to no influence of natural selection for adult sablefish, the habitat heterogeneity that juvenile fish experience may better describe genomic variation in the sablefish.

There are other areas where sablefish are found that were not sampled in this study such as offshore seamounts (Maloney 2004), the inside waters in southeast Alaska (Baldwin & Stahl 2014) and at the extremes of their range (Orlov & Biryukov 2005). However, it is unclear whether sablefish from these areas represent unique genetic populations. For example, it is well known from tagging data that seamount sablefish populations are supported by the movement of adult fish from the slope to seamounts and back (Maloney 2004). Sablefish have also been shown to move between Alaskan inside waters and the large stock in the eastern Gulf of Alaska (Hanselman *et al.* 2015). Tagging data also shows that while the probability of movement to inside waters from the larger northern stock is generally low, the probability of emigration is higher (Hanselman *et al.* 2015). Hanselman *et al.* (2015) mention that while the probability of movement from the northern stock to inside waters is low, the absolute number of migrants is unknown. The small movement rates may translate to a large number of fish moving from the larger northern stock to smaller stocks in inside Alaskan waters (Hanselman *et al.* 2015) and may be enough to homogenize genetic diversity between these areas. Finally, recent data obtained from mitochondrial DNA suggests the absence of genetic

differentiation between fish collected in Russian waters and those in the northeastern Pacific (Orlova *et al.* 2014).

d) Conclusion

Tagging studies have shown that sablefish have the capacity for frequent long distance movements and this seems to be reflected in the genetic data presented here. (Kimura *et al.* 1998; Maloney & Sigler 2008; Morita *et al.* 2012; Hanselman *et al.* 2015). Our analysis agree with previous genetic studies (Tsuyuki & Roberts 1969; Gharrett *et al.* 1982; Tripp-Valdez *et al.* 2012) finding little population structure across the North American range of the sablefish, and present a signal of panmixia within the sablefish stocks in the northeast Pacific Ocean. Our ability to identify signals of natural selection in the sablefish genome may have been hindered if the low levels of population differentiation observed in this study are a result of high migration rates.

It is important to remember that we cannot prove that panmixia exists by simply failing to reject our null hypothesis of genetic homogeneity (discussed in Gold & Richardson 1998; Heist & Gold 2000; Gold *et al.* 2001). Given that these fish were collected out of the reproductive season, the results we present here are suggestive of considerable mixing of fish between the areas sampled for this study. The low levels of genetic differentiation seen here may result from a number of factors; sampling design, shallow evolutionary population history, high migration rates, or large population sizes typical of many marine species. Defining population groups prior to population genetic analyses is not a trivial issue. A better understanding of the sablefish spawning behavior (timing and location) may help define appropriate groups of sablefish to be included in population genetic analyses. With the potential for natural selection to occur at earlier

life stages, it may also help to examine the genomes of juvenile sablefish collected in nearshore areas to link environmental conditions to genomic variation. While the findings presented here suggest that sablefish populations may be panmictic on an evolutionary timescale, it is still important to consider local population dynamics on shorter ecological timescales for management purposes (Gilbert-Horvath *et al.* 2006).

5) Data Accessibility

Sequence reads for this project have been deposited in the NCBI Sequence Read Archive (BioProject ID: PRJNA279314).

References

- Ådlandsvik B, Gundersen AC, Nedreaas KH, Stene A, Albert OT (2004) Modelling the advection and diffusion of eggs and larvae of Greenland halibut (*Reinhardtius hippoglossoides*) in the north-east Arctic. *Fisheries Oceanography*, **13**, 403–415.
- Alderdice DF, Jensen JOT, Velsen FPJ (1988) Preliminary trials on incubation of sablefish eggs (*Anoplopoma fimbria*). *Aquaculture*, **69**, 271–290.
- Allendorf F., Luikart G (2007) *Conservation and the genetics of populations*. Blackwell.
- Allendorf FW, Phelps SR (1981) Use of allelic frequencies to describe population structure. *Canadian Journal of Fisheries and Aquatic Sciences*, **38**, 1507–1514.
- Als TD, Hansen MM, Maes GE *et al.* (2011) All roads lead to home: panmixia of European eel in the Sargasso Sea. *Molecular Ecology*, **20**, 1333–1346.
- Andrews S (2012) FastQC a quality control tool for high throughput sequence data. *version 0.10.1*.
- Baird NA, Etter PD, Atwood TS *et al.* (2008) Rapid SNP discovery and genetic mapping using sequenced RAD markers (JC Fay, Ed.). *PLoS ONE*, **3**, e3376.
- Baker MS, Wilson C a. (2001) Use of bomb radiocarbon to validate otolith section ages of red snapper *Lutjanus campechanus* from the northern Gulf of Mexico. *Limnology and Oceanography*, **46**, 1819–1824.
- Baldwin A, Stahl J (2014) *2013 NSEI (Northern Southeast Inside Subdistrict) sablefish mark-tag survey*. Alaska Department of Fish and Game, Division of Sport Fish, Research and Technical Services.
- Bashalkhanov S, Eckert AJ, Rajora OP (2013) Genetic signatures of natural selection in response to air pollution in red spruce (*Picea rubens*, Pinaceae). *Molecular Ecology*, **22**, 5877–5889.
- Beamish R, McFarlane C (1988) Resident and dispersal behavior of adult sablefish (*Anaplopoma fimbria*) in the slope waters off Canada's west coast. *Canadian Journal of Fisheries and Aquatic Sciences*, **45**, 152–164.
- Beamish R, McFarlane G (2000) Reevaluation of the interpretation of annuli from otoliths of a long-lived fish, *Anoplopoma fimbria*. *Fisheries Research*, **46**, 105–111.
- Beamish RJ, McFarlane GA, Benson A (2006) Longevity overfishing. *Progress in Oceanography*, **68**, 289–302.

- Begg GA, Friedland KD, Pearce JB (1999) Stock identification and its role in stock assessment and fisheries management: an overview. *Fisheries Research*, **43**, 1–8.
- Begg GA, Waldman JR (1999) An holistic approach to fish stock identification. *Fisheries Research*, **43**, 35–44.
- Boehlert GW, Yoklavich MM (1985) Larval and juvenile growth of sablefish, (*Anoplopoma fimbria*), as determined from otolith increments. *Fishery Bulletin*, **83**, 475.
- Boje J (2002) Intermingling and seasonal migrations of Greenland halibut (*Reinhardtius hippoglossoides*) populations determined from tagging studies. *Fishery Bulletin*, **100**, 414–422.
- Bothwell H, Bisbing S, Therkildsen NO *et al.* (2013) Identifying genetic signatures of selection in a non-model species, alpine gentian (*Gentiana nivalis* L.), using a landscape genetic approach. *Conservation Genetics*, **14**, 467–481.
- Bowering WR, Nedreaas KH (2000) A comparison of Greenland halibut (*Reinhardtius hippoglossoides* (Walbaum)) fisheries and distribution in the Northwest and Northeast Atlantic. *Sarsia*, **85**, 61–76.
- Bradbury IR, Hubert S, Higgins B *et al.* (2010) Parallel adaptive evolution of Atlantic cod on both sides of the Atlantic Ocean in response to temperature. *Proceedings of the Royal Society B: Biological Sciences*, **277**, 3725–3734.
- Bradbury IR, Hubert S, Higgins B *et al.* (2013) Genomic islands of divergence and their consequences for the resolution of spatial structure in an exploited marine fish. *Evolutionary Applications*, **6**, 450–461.
- Branch TA (2001) A review of orange roughy (*Hoplostethus atlanticus*) fisheries, estimation methods, biology and stock structure. *South African Journal of Marine Science*, **23**, 181–203.
- Bulman C, Koslow J (1995) Development and depth distribution of the eggs of orange roughy, *Hoplostethus atlanticus* (Pisces: Trachichthyidae). *Marine and Freshwater Research*, **46**, 697.
- Canino MF, O'Reilly PT, Hauser L, Bentzen P (2005) Genetic differentiation in walleye pollock (*Theragra chalcogramma*) in response to selection at the pantophysin (Pan I) locus. *Canadian Journal of Fisheries and Aquatic Sciences*, **62**, 2519–2529.
- Carvalho G, Hauser L (1994) Molecular genetics and the stock concept in fisheries. *Reviews in Fish Biology and Fisheries*, **4**, 326–350.

- Catchen JM, Amores A, Hohenlohe P, Cresko W, Postlethwait JH (2011) Stacks: building and genotyping goci *de novo* from short-read sequences. *G3: Genes, Genomes, Genetics*, **1**, 171–182.
- Catchen J, Hohenlohe PA, Bassham S, Amores A, Cresko WA (2013) Stacks: an analysis tool set for population genomics. *Molecular Ecology*, **22**, 3124–3140.
- Chessel D, Dufour AB, Thioulouse J (2004) The ade4 package - I: one-table methods. *R News*, **4**, 5–10.
- Clark MR (2009) Deep sea seamount fisheries: a review of global status and future prospects. *Latin American Journal of Aquatic Research*, **37**, 501–512.
- Conover DO, Clarke LM, Munch SB, Wagner GN (2006) Spatial and temporal scales of adaptive divergence in marine fishes and the implications for conservation. *Journal of Fish Biology*, **69**, 21–47.
- Corander J, Marttinen P, Sirén J, Tang J (2008) Enhanced Bayesian modelling in BAPS software for learning genetic structures of populations. *BMC bioinformatics*, **9**, 539.
- Côté CL, Gagnaire PA, Bourret V *et al.* (2013) Population genetics of the American eel (*Anguilla rostrata*): $F_{ST} = 0$ and North Atlantic Oscillation effects on demographic fluctuations of a panmictic species. *Molecular Ecology*, **22**, 1763–1776.
- Cunningham KM, Canino MF, Spies IB, Hauser L (2009) Genetic isolation by distance and localized fjord population structure in Pacific cod (*Gadus macrocephalus*): limited effective dispersal in the northeastern Pacific Ocean. *Canadian Journal of Fisheries and Aquatic Sciences*, **66**, 153–166.
- Dabney A, Storey JD (2014) qvalue: Q-value estimation for false discovery rate control. *R package version 1.40.0*.
- Dakin EE, Avise JC (2004) Microsatellite null alleles in parentage analysis. *Heredity*, **93**, 504–509.
- Danecek P, Auton A, Abecasis G *et al.* (2011) The variant call format and VCFtools. *Bioinformatics*, **27**, 2156–2158.
- DFO (2011) Management procedures for the multi-gear sablefish (*Anoplopoma fimbria*) fishery in British Columbia, Canada. *DFO Can. Sci. Advis. Sec. Sci. Advis. Rep.* 2011/025.
- DFO (2013) A review of sablefish population structure in the northeast Pacific Ocean and implications for Canadian seamount fisheries. *DFO Can. Sci. Advis. Sec. Sci. Rep.* 2013/017.

- Doyle MJ (1992) Neustonic ichthyoplankton in the northern region of the California Current ecosystem. *Calif. Coop. Oceanic Fish. Invest. Rep.*, **33**, 141–161.
- Van Dykhuzien G. (1983) Activity patterns and feeding chronology of the kelp rockfish (*Sebastes atrovirens*) in a central California kelp forest. San Jose State University.
- Echave KB, Hanselman DH, Maloney NE (2013) *Report to industry on the Alaska sablefish tag program, 1972 - 2012.*
- Eckert AJ, Bower AD, González-Martínez SCSC *et al.* (2010) Back to nature: ecological genomics of loblolly pine (*Pinus taeda*, Pinaceae). *Molecular Ecology*, **19**, 3789–3805.
- Elshire RJ, Glaubitz JC, Sun Q *et al.* (2011) A robust, simple genotyping-by-sequencing (GBS) approach for high diversity species (L Orban, Ed.). *PLoS ONE*, **6**, e19379.
- Eshenroder RL (2008) Differentiation of deep-water lake charr *Salvelinus namaycush* in North American lakes. *Environmental Biology of Fishes*, **83**, 77–90.
- Van Etten J (2014) gdistance: distances and routes on geographical grids. *R package version 1.1-5.*
- Evanno G, Regnaut S, Goudet J (2005) Detecting the number of clusters of individuals using the software structure: a simulation study. *Molecular Ecology*, **14**, 2611–2620.
- Felsenstein J (2015) *Theoretical Evolutionary Genetics*. University of Washington, Seattle, WA.
- Francis RICC, Clark MR (1998) Inferring spawning migrations of orange roughy (*Hoplostethus atlanticus*) from spawning ogives. *Marine and Freshwater Research*, **49**, 103.
- Frantz AC, Cellina S, Krier A, Schley L, Burke T (2009) Using spatial Bayesian methods to determine the genetic structure of a continuously distributed population: clusters or isolation by distance? *Journal of Applied Ecology*, **46**, 493–505.
- Frichot E, Schoville SD, Bouchard G, Francois O (2013) Testing for associations between loci and environmental gradients using latent factor mixed models. *Molecular Biology and Evolution*, **30**, 1687–1699.
- Fujiwara S, Hankin D (1988) Sex ratio, spawning period, and size and age at maturity of sablefish *Anoplopoma fimbria* off Northern California. *Bulletin of the Japanese Society of Scientific Fisheries*.

- Gao Y, Dettman DL, Piner KR, Wallace FR (2010) Isotopic correlation ($\delta^{18}\text{O}$ versus $\delta^{13}\text{C}$) of otoliths in identification of groundfish stocks. *Transactions of the American Fisheries Society*, **139**, 491–501.
- Gao Y, Joner SH, Svec RA, Weinberg KL (2004) Stable isotopic comparison in otoliths of juvenile sablefish (*Anoplopoma fimbria*) from waters off the Washington and Oregon coast. *Fisheries research*, **68**, 351–360.
- Garber A, Tringali M, Stuck K (2004) Population structure and variation in red snapper (*Lutjanus campechanus*) from the Gulf of Mexico and Atlantic Coast of Florida as determined from mitochondrial DNA control region sequence. *Marine Biotechnology*, **6**, 175–185.
- Gharrett A, Thomason M, Wishard L (1982) Biochemical genetics of sablefish. *NWAFRC Processed Report*, **82-05**.
- Gilbert-Horvath EA, Larson RJ, Garza JC (2006) Temporal recruitment patterns and gene flow in kelp rockfish (*Sebastes atrovirens*). *Molecular Ecology*, **15**, 3801–3815.
- Goetz F, Jasonowicz A, Johnson R *et al.* (2014) Physiological differences between lean and siscowet lake trout morphotypes: Are these metabolotypes? *Canadian Journal of Fisheries and Aquatic Sciences*, **71**, 427–435.
- Goetz F, Rosauer D, Sitar S *et al.* (2010) A genetic basis for the phenotypic differentiation between siscowet and lean lake trout (*Salvelinus namaycush*). *Molecular Ecology*, **19**, 176–196.
- Gold JR, Pak E, Richardson LR (2001) Microsatellite variation among red snapper (*Lutjanus campechanus*) from the Gulf of Mexico. *Marine Biotechnology*, **3**, 293–304.
- Gold JR, Richardson LR (1998) Genetic homogeneity among geographic samples of snappers and groupers: evidence of continuous gene flow? *Proceedings of the Gulf and Caribbean Fisheries Institute*, 709–726.
- Gregg JL, Anderl DM, Kimura DK (2006) Improving the precision of otolith-based age estimates for Greenland halibut (*Reinhardtius hippoglossoides*) with preparation methods adapted for fragile sagittae. *Fishery Bulletin*, **104**, 643–648.
- Gundersen AC, Ronneberg JE, Boje J (2001) Fecundity of greenland halibut (*Reinhardtius hippoglossoides* Walbaum) in East Greenland waters. *Fisheries Research*, **51**, 229–236.
- Hanselman DH, Clark WG, Heifetz J, Anderl DM (2012a) Statistical distribution of age readings of known-age sablefish (*Anoplopoma fimbria*). *Fisheries Research*, **131–133**, 1–8.

- Hanselman DH, Heifetz J, Echave KB, Dressel SC (2015) Move it or lose it: movement and mortality of sablefish tagged in Alaska. *Canadian Journal of Fisheries and Aquatic Sciences*, **72**, 238–251.
- Hanselman DH, Lunsford CR, Rodgveller CJ (2012b) *Assessment of the sablefish stock in Alaska*. Anchorage, AK.
- Hardy OJ, Vekemans X (2002) SPAGeDi: a versatile computer program to analyse spatial genetic structure at the individual or population levels. *Molecular Ecology Notes*, **2**, 618–620.
- Hauser L, Carvalho GR (2008) Paradigm shifts in marine fisheries genetics: ugly hypotheses slain by beautiful facts. *Fish and Fisheries*, **9**, 333–362.
- Head MA, Keller AA, Bradburn M (2014) Maturity and growth of sablefish, (*Anoplopoma fimbria*), along the U.S. West Coast. *Fisheries Research*, **159**, 56–67.
- Hedgecock D, Shleser RA, Nelson K (1976) Applications of biochemical genetics to aquaculture. *Journal of the Fisheries Research Board of Canada*, **33**, 1108–1119.
- Heifetz J, Fujioka JT (1991) Movement dynamics of tagged sablefish in the northeastern Pacific. *Fisheries Research*, **11**, 355–374.
- Heist EJ, Gold JR (2000) DNA Mmicrosatellite loci and genetic structure of red snapper in the Gulf of Mexico. *Transactions of the American Fisheries Society*, **129**, 469–475.
- Hellberg ME, Burton RS, Neigel JE, Palumbi SR (2002) Genetic assessment of connectivity among marine populations. *Bulletin of Marine Science*, **70**, 273–290.
- 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.
- Hemmer-Hansen J, Nielsen EE, Therkildsen NO *et al.* (2013) A genomic island linked to ecotype divergence in Atlantic cod. *Molecular Ecology*, **22**, 2653–2667.
- Hughes S (1981) Initial US exploration of nine Gulf of Alaska seamounts and their associated fish and shellfish resources. *Marine Fisheries Review*, **41**, 26–33.
- Hunter J, B.J. Macewicz, Kimbrell CA (1989) Fecundity and other aspects of the reproduction of sablefish, *Anoplopoma fimbria*, in central California waters. *CalCOFI Reports*, **30**, 61–72.

- Hunter J, Butler J, Kimbrell C, Lynn E (1990) Bathymetric patterns in size, age, sexual maturity, water content, and caloric density of Dover sole, *Microstomus pacificus*. *CalCOFI Reports*, **31**, 132–144.
- Ihssen PE, Booke HE, Casselman JM *et al.* (1981) Stock identification: materials and methods. *Canadian Journal of Fisheries and Aquatic Sciences*, **38**, 1838 – 1855.
- IOC, IHO & BDOC (2003) GEBCO Digital Atlas CDROM from BODC.
- Jones M, Forester B, Teufel A (2013) Integrating landscape genomics and spatially explicit approaches to detect loci under selection in clinal populations. *Evolution*, **67**, 3455–3468.
- Karinen JF, Barnett HJ, Masuda M (2010) Soft flesh in sablefish, *Anoplopoma fimbria*, of Southeastern Alaska: relationships with depth, season, and biochemistry. *Marine Fisheries Review*, **72**, 25–35.
- Karlsen BO, Klingan K, Emblem Å *et al.* (2013) Genomic divergence between the migratory and stationary ecotypes of Atlantic cod. *Molecular Ecology*, **22**, 5098–5111.
- Kendall AW, Matarese AC (1987) Biology of eggs, larvae, and epipelagic juveniles of sablefish, *Anoplopoma fimbria*, in relation to their potential use in management. *Mar. Fish. Rev.*, **49**, 1–13.
- Kersey PJ, Allen JE, Christensen M *et al.* (2014) Ensembl genomes 2013: scaling up access to genome-wide data. *Nucleic Acids Research*, **42**, 546–552.
- Kimura DK, Shimada AM, Lowe SA (1993) Estimating von Bertalanffy growth parameters of sablefish *Anoplopoma fimbria* and Pacific cod *Gadus macrocephalus* using tag-recapture data. *Fishery Bulletin*, **91**, 271–280.
- Kimura D, Shimada A, Shaw F (1998) Stock structure and movement of tagged sablefish, *Anoplopoma fimbria*, in offshore northeast Pacific waters and the effects of El Niño-Southern Oscillation on migration and growth. *Fishery Bulletin*, **96**, 462–481.
- King JR, Mcfarlane GA, Beamish RJ (2000) Decadal scale patterns in the relative year class success of sablefish (*Anoplopoma fimbria*). *Fisheries Oceanography*, **9**, 62–70.
- Kumar S, Banks TW, Cloutier S (2012) SNP Discovery through next-generation sequencing and Its applications. *International Journal of Plant Genomics*, **2012**, 831460.
- Langmead B, Trapnell C, Pop M, Salzberg SL (2009) Ultrafast and memory-efficient alignment of short DNA sequences to the human genome. *Genome biology*, **10**, R25.

- Larsen PF, Nielsen EE, Williams TD *et al.* (2007) Adaptive differences in gene expression in European flounder (*Platichthys flesus*). *Molecular Ecology*, **16**, 4674–4683.
- Latch EK, Dharmarajan G, Glaubitz JC, Rhodes OE (2006) Relative performance of Bayesian clustering software for inferring population substructure and individual assignment at low levels of population differentiation. *Conservation Genetics*, **7**, 295–302.
- Lenormand T (2002) Gene flow and the limits to natural selection. *Trends in Ecology & Evolution*, **17**, 183–189.
- Limborg MT, Helyar SJ, De Bruyn M *et al.* (2012) Environmental selection on transcriptome-derived SNPs in a high gene flow marine fish, the Atlantic herring (*Clupea harengus*). *Molecular Ecology*, **21**, 3686–3703.
- Lowe WH, Allendorf FW (2010) What can genetics tell us about population connectivity? *Molecular Ecology*, **19**, 3038–51.
- 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**, 981–94.
- Mace PM, Fenaughty JM, Coburn RP, Doonan IJ (1990) Growth and productivity of orange roughy (*Hoplostethus atlanticus*) on the north Chatham Rise. *New Zealand Journal of Marine and Freshwater Research*, **24**, 105–119.
- Macewicz BJ, Hunter JR (1994) Fecundity of sablefish, *Anoplopoma fimbria*, from Oregon coastal waters. *CalCOFI Reports*, **35**, 160–174.
- Maloney NE (2004) Sablefish, *Anoplopoma fimbria*, populations on Gulf of Alaska seamounts. *Marine Fisheries Review*, **66**, 1–12.
- Maloney NE, Sigler MF (2008) Age-specific movement patterns of sablefish (*Anoplopoma fimbria*) in Alaska. *Fishery Bulletin*, **106**, 305–316.
- Mason JC, Beamish RJ, McFarlane GA (1983) Sexual maturity, fecundity, spawning, and early life history of sablefish (*Anoplopoma fimbria*) off the Pacific coast of Canada. *Canadian Journal of Fisheries and Aquatic Sciences*, **40**, 2126–2134.
- McCraney TW, Saski CA, Guyon JR (2011) Isolation and characterization of 12 microsatellites for the commercially important sablefish, *Anoplopoma fimbria*. *Conservation Genetics Resources*.

- McFarlane G, Beamish R (1990) Effect of an external tag on growth of sablefish (*Anoplopoma fimbria*), and consequences to mortality and age at maturity. *Canadian Journal of Fisheries and Aquatic Sciences*, **47**, 1551–1557.
- McFarlane GA, Beamish RJ (1992) Climatic influence linking copepod production with strong year-classes in sablefish, *Anoplopoma fimbria*. *Canadian Journal of Fisheries and Aquatic Sciences*, **49**, 743–753.
- McFarlane G, Saunders M (1993) Dispersion of juvenile sablefish, *Anoplopoma fimbria*, as indicated by tagging in Canadian waters. In: *Biology and management of sablefish, Anoplopoma fimbria* (eds Wilkins ME, Saunders MW), pp. 137–150. U.S. Dep. Commer., NOAA Tech Rep. NMFS 130.
- Milano I, Babbucci M, Cariani A *et al.* (2013) Outlier SNP markers reveal fine-scale genetic structuring across European hake populations (*Merluccius merluccius*). *Molecular Ecology*, **23**, 118–135.
- Miller MR, Dunham JP, Amores A, Cresko WA, Johnson EA (2007) Rapid and cost-effective polymorphism identification and genotyping using restriction site associated DNA (RAD) markers. *Genome Research*, **17**, 240–248.
- Miller DJ, Lea RN (1972) *Guide To The Coastal Marine Fishes of California*. California Department of Fish and Game, Sacramento, CA.
- Minto C, Nolan CP (2006) Fecundity and maturity of orange roughy (*Hoplostethus atlanticus* Collett 1889) on the Porcupine Bank, Northeast Atlantic. *Environmental Biology of Fishes*, **77**, 39–50.
- Morita SH, Morita K, Nishimura A (2012) Sex-biased dispersal and growth in sablefish (*Anoplopoma fimbria*) in the northeastern Pacific Ocean. *Environmental Biology of Fishes*, **94**, 505–511.
- Murie DJ, McFarlane G (1996) Canadian trap-fishery for sablefish on seamounts in the northeastern Pacific Ocean, 1983-1993. *Canadian Manuscript Report Fisheries Aquatic Science*, 107.
- Myers R, Hutchings J, Barrowman N (1996) Hypotheses for the decline of cod in the North Atlantic. *Marine Ecology Progress Series*, **138**, 293–308.
- Nielsen EE, Cariani A, Aoidh E Mac *et al.* (2012) Gene-associated markers provide tools for tackling illegal fishing and false eco-certification. *Nature Communications*, **3**, 851.
- Nielsen EE, Hemmer-Hansen J, Larsen PF, Bekkevold D (2009a) Population genomics of marine fishes: identifying adaptive variation in space and time. *Molecular Ecology*, **18**, 3128–3150.

- Nielsen E, Hemmer-Hansen J, Poulsen N *et al.* (2009b) Genomic signatures of local directional selection in a high gene flow marine organism; the Atlantic cod (*Gadus morhua*). *BMC Evolutionary Biology*, **9**, 276.
- Nielsen EE, Kenchington E (2001) A new approach to prioritizing marine fish and shellfish populations for conservation. *Fish and Fisheries*, **2**, 328–343.
- NMFS (2013) *Fisheries Economics of the United States 2012*. United States.
- Norris J (1993) Adaptive radiation and sablefish, *Anoplopoma fimbria*. In: *Biology and management of sablefish, Anoplopoma fimbria* (eds Wilkins ME, Saunders MW), pp. 99–114. U.S. Dep. Commer., NOAA Tech Rep. NMFS 130.
- Norris J, Rowley J, Mathews S (1987) *Analysis of four factors affecting the sablefish soft flesh problem*.
- Nychka D, Furrer R, Sain S (2014) fields: Tools for spatial data. *R package version 7.1*.
- Oksanen J, Blanchet FG, Kindt R *et al.* (2013) vegan: Community Ecology Package. *R package version 2.0-10*.
- Orlov AM, Biryukov IA (2005) First report of sablefish in spawning condition off the coast of Kamchatka and the Kuril Islands. *Journal of Marine Science*, **62**, 1016–1020.
- Orlova SY, Orlov AM, Volkov AA, Novikov RN (2014) Micro-evolutionary processes in sablefish *Anoplopoma fimbria*, based on polymorphism of the two sites of mitochondrial DNA. *Doklady Biochemistry and Biophysics*, **458**, 172–176.
- Pariset L, Joost S, Marsan P, Valentini A, Econogene Consortium (2009) Landscape genomics and biased FST approaches reveal single nucleotide polymorphisms under selection in goat breeds of North-East Mediterranean. *BMC Genetics*, **10**, 7.
- Patashnik M, Groninger HSJ (1964) Observations on the milky condition in some Pacific Coast fishes. *Journal of the Fisheries Research Board of Canada*, **21**, 335–346.
- Patterson N, Price AL, Reich D (2006) Population structure and eigenanalysis. *PLoS Genetics*, **2**, e190.
- Patterson WF, Watterson JC, Shipp RL, Cowan JH (2001) Movement of tagged red snapper in the northern Gulf of Mexico. *Transactions of the American Fisheries Society*, **130**, 533–545.
- Pembleton LW, Cogan NOI, Forster JW (2013) StAMPP: an R package for calculation of genetic differentiation and structure of mixed-ploidy level populations. *Molecular Ecology Resources*, **13**, 946–952.

- Peterson BJ, Fry B (1987) Stable isotopes in ecosystem studies. *Annual Review of Ecology and Systematics*, **18**, 293–320.
- Peterson BK, Weber JN, Kay EH, Fisher HS, Hoekstra HE (2012) Double digest RADseq: an inexpensive method for de novo SNP discovery and genotyping in model and non-model species. *PLoS ONE*, **7**, e37135.
- Pinsky ML, Montes HR, Palumbi SR (2010) Using isolation by distance and effective density to estimate dispersal scales in anemonefish. *Evolution*, **64**, 2688–700.
- Pritchard JK, Stephens M, Donnelly P (2000) Inference of population structure using multilocus genotype data. *Genetics*, **155**, 945.
- Pujolar JM, Jacobsen MW, Als TD *et al.* (2014) Genome-wide single-generation signatures of local selection in the panmictic European eel. *Molecular Ecology*, **23**, 2514–2528.
- Purcell S, Neale B, Todd-Brown K *et al.* (2007) PLINK: a tool set for whole-genome association and population-based linkage analyses. *The American Journal of Human Genetics*, **81**, 559–575.
- QGIS Development Team (2009) QGIS geographic information system. *Version 2.2.0-Valmiera*.
- Quinn TJ, Deriso RB (1999) *Quantitative fish dynamics*. Oxford University, New York.
- R Development Core Team (2013) R: a language and environment for statistical computing.
- Rondeau EB, Messmer AM, Sanderson DS *et al.* (2013) Genomics of sablefish (*Anoplopoma fimbria*): expressed genes, mitochondrial phylogeny, linkage map and identification of genetic sex markers. *BMC Genomics*, **14**, 452.
- Rousset F (1997) Genetic differentiation and estimation of gene flow from *F*-statistics under isolation by distance. *Genetics*, **145**, 1219–1288.
- Rousset F (2000) Genetic differentiation between individuals. *Journal of Evolutionary Biology*, **13**, 58–62.
- Rousset F (2008) genepop'007: a complete re-implementation of the genepop software for Windows and Linux. *Molecular Ecology Resources*, **8**, 103–106.
- Roy D, Hardie DC, Treble MA, Reist JD, Ruzzante DE (2014) Evidence supporting panmixia in Greenland halibut (*Reinhardtius hippoglossoides*) in the northwest Atlantic. *Canadian Journal of Fisheries and Aquatic Sciences*, **71**, 763–774.

- Rummer JL, Bennett W a. (2005) Physiological effects of swim bladder overexpansion and catastrophic decompression on red snapper. *Transactions of the American Fisheries Society*, **134**, 1457–1470.
- Ruzzante DE, Taggart CT, Cook D (1999) A review of the evidence for genetic structure of cod (*Gadus morhua*) populations in the NW Atlantic and population affinities of larval cod off Newfoundland and the Gulf of St . Lawrence. *Fisheries Oceanography*, **43**, 79–97.
- Ryman N, Palm S (2006) POWSIM: a computer program for assessing statistical power when testing for genetic differentiation. *Molecular Ecology Notes*, **6**, 600–602.
- Sasaki T (1985) Studies on the sablefish resources of the North Pacific. *Bulletin of the Far Seas Fisheries Research Laboratory*, **22**, 1–107.
- Schlitzer R (2014) Ocean data view. *Version 4.6.2*.
- Schwartz MK, Luikart G, Mckelvey KS, Cushman SA (2010) Landscape genomics: a brief perspective. In: *Spatial Complexity, Informatics, and Wildlife Conservation* (eds Cushman SA, Huettmann F), pp. 165–174. Springer Japan, Tokyo.
- Serre D, Pääbo S (2004) Evidence for gradients of human genetic diversity within and among continents. *Genome Research*, **14**, 1679–1685.
- Shaklee JB, Bentzen P (1998) Genetic identification of stocks of marine fish and shellfish. *Bulletin of Marine Science*, **62**, 589–621.
- Shaw F, Parks N (1993) Movement patterns of tagged sablefish, *Anoplopoma fimbria*, recovered on seamounts in the northeast Pacific Ocean and Gulf of Alaska. In: *Biology and management of sablefish, Anoplopoma fimbria* (eds Wilkins ME, Saunders MW), pp. 151–158. U.S. Dep. Commer., NOAA Tech Rep. NMFS 130.
- Shephard S, Trueman C, Rickaby R, Rogan E (2007) Juvenile life history of NE Atlantic orange roughy from otolith stable isotopes. *Deep-Sea Research I*, **54**, 1221–1230.
- Sonesson AK, Meuwissen THE (2009) Testing strategies for genomic selection in aquaculture breeding programs. *Genetics Selection Evolution*, **41**, 9.
- Sork VL, Aitken SN, Dyer RJ *et al.* (2013) Putting the landscape into the genomics of trees: approaches for understanding local adaptation and population responses to changing climate. *Tree Genetics & Genomes*, **9**, 901–911.
- Stewart IJ, Thorson JT, Wetzel C (2011) Status of the US sablefish resource in 2011. *Pacific Fishery Management Council, Portland, OR*, **441**.

- Storey JD, Tibshirani R (2003) Statistical significance for genomewide studies. *Proceedings of the National Academy of Sciences*, **100**, 9440–9445.
- Strelcheck AJ, Cowan JH, Patterson WF (2007) Site fidelity, movement, and growth of red snapper: Implications for artificial reef management. *Red Snapper Ecology and Fisheries in the U.S. Gulf of Mexico*, **60**, 147–162.
- Sumaila R, Volpe J, Liu Y (2005) Ecological and economic impact assessment of sablefish aquaculture in British Columbia. *Fisheries Centre Research Reports*, **13**.
- Thorrold SR, Latkoczy C, Swart PK, Jones CM (2001) Natal homing in a marine fish metapopulation. *Science*, **291**, 297–299.
- Treble MA, Campana SE, Wastle RJ, Jones CM, Boje J (2008) Growth analysis and age validation of a deepwater Arctic fish, the Greenland halibut (*Reinhardtius hippoglossoides*). *Canadian Journal of Fisheries and Aquatic Sciences*, **65**, 1047–1059.
- Tripp-Valdez MA, García-de-León FJ, Espinosa-Pérez H, Ruiz-Campos G (2012) Population structure of sablefish *Anoplopoma fimbria* using genetic variability and geometric morphometric analysis. *Journal of Applied Ichthyology*, 1–8.
- Tsuyuki H, Roberts E (1969) Muscle protein polymorphism of sablefish from the eastern Pacific Ocean. *Journal of the Fisheries Board of Canada*, **26**, 2633–2641.
- Waples RS, Punt AE, Cope JM (2008) Integrating genetic data into management of marine resources: how can we do it better? *Fish and Fisheries*, **9**, 423–449.
- Ward R. (2002) Genetics of fish populations. In: *Handbook of fish biology and fisheries* (eds Hart P, Reynolds J), pp. 200–224. Blackwell Science, Malden, MA.
- Ward RD, Woodwark M, Skibinski DOF (1994) A comparison of genetic diversity levels in marine, freshwater, and anadromous fishes. *Journal of Fish Biology*, **44**, 213–232.
- Weir BS, Cockerham CC (1984) Estimating *F*-statistics for the analysis of population structure. *Evolution*, **38**, 1358–1370.
- White TA, Stefanni S, Stamford J, Hoelzel AR (2009) Unexpected panmixia in a long-lived, deep-sea fish with well-defined spawning habitat and relatively low fecundity. *Molecular Ecology*, **18**, 2563–2573.
- Wilson CA, Nieland DL (2001) Age and growth of red snapper, *Lutjanus campechanus*, from the northern Gulf of Mexico off Louisiana. *Fishery Bulletin*, **99**, 653–664.
- Wing BL, Kamikawa DJ (1995) *Distribution of neustoninc sablefish larvae and associated ichthyoplankton in the eastern Gulf of Alaska.*

Zeldis JR, Grimes PJ, Ingerson JK (1995) Ascent rates, vertical distribution, and a thermal history model of development of orange roughy, *Hoplostethus atlanticus*, eggs in the water column. *Fishery Bulletin*, **93**, 373–385.

Zueva KJ, Lumme J, Veselov AE *et al.* (2014) Footprints of directional selection in wild atlantic salmon populations: evidence for parasite-driven evolution? *PLoS ONE*, **9**, e91672.

Appendix:

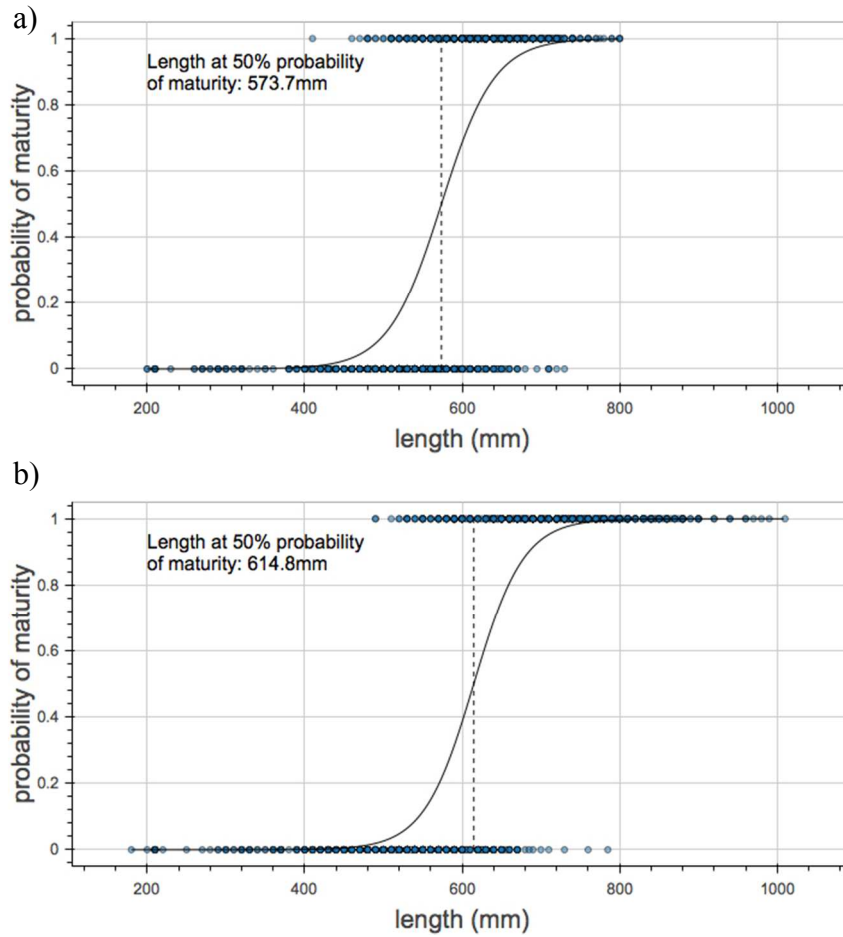


Figure S 1. Estimated length at 50% probability of maturity for female (a) and male (b) sablefish samples collected for this study.

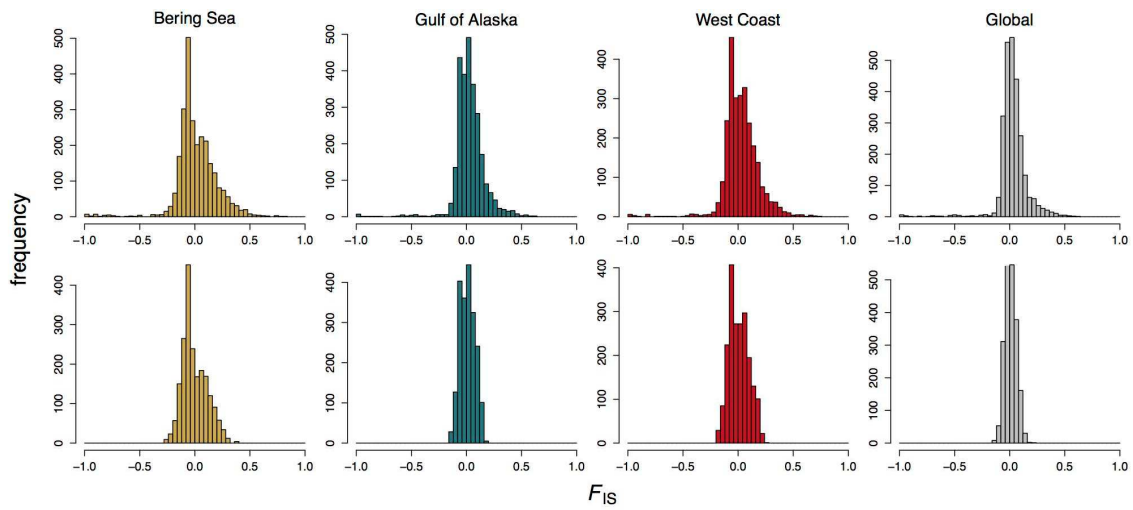


Figure S 2. Histogram of F_{IS} calculated by locus for the population groupings used for the population genetic analyses as well as globally. Top row is prior to removing SNPs on the basis of HWE and the bottom row is after.

Table S 1. Summary of F_{IS} for the SNP loci used in this study before and after filtering SNPs on the basis of Hardy - Weinberg equilibrium (HWE). Calculated for each population (BS: Bering Sea, GOA: Gulf of Alaska, WC: West Coast) and globally.

	Population	Mean	Median	σ
All QC loci	BS	0.0160	-0.0152	0.1824
	GOA	0.0308	0.0234	0.1382
	WC	0.0302	0.0166	0.1591
	Global	0.0311	0.0232	0.1317
Filtered on HWE	BS	-0.0038	-0.0337	0.1106
	GOA	0.0097	0.0099	0.0655
	WC	0.0076	0.0000	0.0886
	Global	0.0096	0.0080	0.0515

Table S 2. Summary of BLASTN alignments of sablefish contig 152341 (gi:559591484, gb:AWGY01151885.1) to the stickleback BORAD S1 genome. Results are ordered by their genomic location in the stickleback genome.

Stickleback Genomic Location	Orientation	Query start	Query end	Query ori	Length	Score	E-val	%ID
groupIII:7438342-7438369	Forward	10269	10296	Forward	28	56	8.00E-05	100.00
groupIII:9945023-9945088	Forward	6492	6555	Forward	66	60	5.00E-06	89.39
groupXX:14624125-14624397	Forward	72	364	Forward	293	248	1.00E-62	85.67
groupXX:14624913-14625548	Forward	1055	1675	Forward	643	414	8.00E-113	84.14
groupXX:14625850-14625961	Forward	2053	2164	Forward	112	190	2.00E-45	96.43
groupXX:14625990-14626081	Forward	2184	2275	Forward	92	79.8	6.00E-12	85.87
groupXX:14626362-14626462	Forward	2588	2688	Forward	101	81.8	1.00E-12	85.15
groupXX:14626513-14626702	Forward	2772	2955	Forward	191	222	6.00E-55	90.58
groupXX:14626779-14626967	Forward	3001	3189	Forward	189	232	6.00E-58	90.48
groupXX:14627240-14627284	Forward	3589	3633	Forward	45	58	2.00E-05	91.11
groupXX:14627550-14627784	Forward	3941	4177	Forward	238	222	6.00E-55	87.39
groupXX:14627954-14628035	Forward	4320	4401	Forward	82	99.6	6.00E-18	90.24
groupXX:14628123-14628198	Forward	4470	4547	Forward	78	77.8	2.00E-11	88.46
groupXX:14628262-14628435	Forward	4628	4803	Forward	176	167	3.00E-38	88.07
groupXX:14628550-14628834	Forward	4897	5181	Forward	285	406	2.00E-110	92.98
groupXX:14628902-14628941	Forward	5277	5316	Forward	40	56	8.00E-05	92.50

Table S 2. continued

Stickleback Genomic Location	Orientation	Query start	Query end	Query ori	Length	Score	E-val	%ID
groupXX:14629253-14629480	Forward	5776	6015	Forward	240	289	3.00E-75	90.00
groupXX:14630385-14630912	Forward	7536	8060	Forward	534	442	4.00E-121	86.14
groupXX:14630969-14631303	Forward	8133	8477	Forward	345	375	7.00E-101	88.70
groupXX:14631484-14631784	Forward	8691	8982	Forward	301	254	2.00E-64	87.38
groupXX:14632072-14632221	Forward	9319	9467	Forward	150	210	2.00E-51	93.33
groupXX:14632278-14632325	Forward	9512	9559	Forward	48	63.9	3.00E-07	91.67
groupXX:14632661-14632803	Forward	10300	10442	Forward	143	244	2.00E-61	96.50
groupXX:14633048-14633097	Forward	10810	10859	Forward	50	67.9	2.00E-08	92.00
groupXX:14633459-14633606	Forward	11118	11265	Forward	148	135	1.00E-28	86.49
groupXX:14633693-14633896	Forward	11357	11560	Forward	204	325	6.00E-86	95.10
groupXX:14634494-14634561	Forward	12751	12818	Forward	68	95.6	1.00E-16	92.65

Table S 3. Summary of BLASTN alignments of sablefish contig 133848 (gi: 559609892, gb: AWGY01133477.1) to the stickleback BORAD S1 genome. Results are ordered by their genomic location in the stickleback genome

Stickleback Genomic Location	Orientation	Query start	Query end	Query ori	Length	Score	E-val	%ID
groupV:10115094-10115124	Forward	1000	1030	Forward	31	61.9	7.00E-07	100.00
groupVI:11485733-11485859	Reverse	6059	6186	Forward	128	151	9.00E-34	90.62
groupVI:11486211-11486456	Reverse	5371	5611	Forward	246	192	3.00E-46	85.37
groupVI:11486517-11486632	Reverse	5166	5285	Forward	122	77.8	1.00E-11	84.43
groupVI:11486658-11486734	Reverse	5051	5127	Forward	77	73.8	2.00E-10	87.01
groupVI:11487005-11487307	Reverse	4417	4721	Forward	305	224	8.00E-56	84.92
groupVI:11487498-11488004	Reverse	3667	4185	Forward	520	392	1.00E-106	84.81
groupVI:11488131-11488370	Reverse	3257	3497	Forward	241	343	1.00E-91	93.36
groupVI:11488704-11489211	Reverse	2403	2925	Forward	533	274	9.00E-71	81.24
groupVI:11489604-11490064	Reverse	1489	1944	Forward	467	283	1.00E-73	83.08
groupVI:11490265-11490335	Reverse	542	612	Forward	71	54	2.00E-04	84.51
groupVI:11490427-11490562	Reverse	308	443	Forward	136	63.9	2.00E-07	80.88
groupXI:16431271-16431298	Forward	1003	1030	Forward	28	56	4.00E-05	100.00
groupXIV:1140798-1140831	Reverse	1006	1039	Forward	34	52	6.00E-04	94.12