ICES Journal of Marine Science, 2024, Vol. 0, Issue 0, 1–15
https://doi.org/10.1093/icesjms/fsae070
Received: 26 January 2024; **revised:** 19 April 2024; **accepted:** 8 May 2024
Original Article

Whole genome resequencing of sablefish at the northern end of their range reveals genetic panmixia and large putative inversions

Laura E. Timm [1,2,*,‡], Wesley A. Larson[1,‡], Andrew J. Jasonowicz[3,5], Krista M. Nichols[4]

[1]Genetics Program, Auke Bay Laboratories, Alaska Fisheries Science Center, Juneau, AK 99801, United States
[2]College of Fisheries and Ocean Sciences, University of Alaska Fairbanks, Fairbanks, AK 99775, United States
[3]School of Aquatic and Fishery Science, University of Washington, Box 355020, Seattle, WA 98195-5020, United States
[4]Conservation Biology Division, Northwest Fisheries Science Center, NOAA, National Marine Fisheries Service, Seattle, WA 98112, United States

[*]Corresponding author. Auke Bay Laboratories - Genetics Program 17109 Point Lena Loop Road Juneau, AK 99801. E-mail: laura.timm@noaa.gov
[5]Present address: International Pacific Halibut Commission, 2320 West Commodore Way, Suite 300, Seattle, WA 98199-1287, USA.
[‡]Authors contributed equally.

Abstract

Sablefish (*Anoplopoma fimbria*) are a highly mobile species that support important commercial fisheries in the North Pacific Ocean. Information on the genetic stock structure of sablefish is vital for constructing management strategies that ensure the long-term viability of the species. Most previous genetic studies on sablefish have found panmixia throughout the majority of their range, but a recent study suggested that a population structure may exist. Here, we use low-coverage whole genome resequencing to investigate genetic structure in the northern end of the species' range (from Washington State, USA to the Bering Sea and Aleutian Islands, AK, USA). Additionally, we reanalyzed an existing genomic dataset containing 2661 markers to test specific hypotheses about genetic structure by sex. Genome resequencing data from 119 individuals screened at 7 110 228 markers revealed no evidence of population structure, and reanalysis of the existing genomic dataset supported the same conclusion. Differentiation across the genome was largely driven by variation at two putative inversions located ~1 megabase apart, which did not display any signals of geographic differentiation. Our study further supports the conclusion of genetic panmixia in sablefish throughout its northern range.

Keywords: population genomics; low coverage whole genome resequencing; *Anoplopoma fimbria*; inversion; black cod; DAPC; fisheries management

Introduction

Understanding the geographic distribution of genetic variation in exploited populations is critical for informing management because it provides the information necessary to construct spatially discrete management units (Allendorf et al. 1987). These units represent genetically, and therefore demographically distinct population segments that should be managed separately to ensure that they are not overexploited (Carvalho & Hauser 1994, Moritz 1994). However, defining genetically distinct management units can be difficult, especially in marine fish, which are often characterized by low genetic structure due to large population sizes, long-distance dispersal, and migration in larvae and adults, respectively, and a lack of physical barriers to prevent or reduce migration (Waples 1998, Nielsen & Kenchington 2001, Cowen & Sponaugle 2009). Genomic tools represent a potential way to improve delineation of management units in marine fish species because these tools increase resolution to detect subtle genetic structure and may identify putatively adaptive loci that can be leveraged to conserve important adaptive variation (Nielsen et al. 2009, Funk et al. 2012, Hemmer-Hansen et al. 2014). Current genomic tools, like low-coverage whole genome resequencing (lcWGS), represent a substantial advance over earlier genomic approaches such as restriction site-associated DNA (RAD) sequencing (Lou et al. 2021). lcWGS facilitates the genotyping of orders of magnitude more markers than RAD (millions vs thousands), increasing resolution for detecting population structure and loci that may be under selection (Lou et al. 2021). lcWGS can also be used to identify important features of an organism's genome, such as sex-determining regions and structural variants (Mérot et al. 2020, Lou et al. 2021, Hansen et al. 2022).

Here, we use lcWGS to investigate the genetic structure of sablefish (*Anoplopoma fimbria*; also referred to as black cod) across the northern end of its range and characterize structural genomic variation. Sablefish inhabit mesopelagic waters (200–1500 m) across the continental shelf of the Pacific Ocean, from central Baja California to the Aleutian Islands in the eastern Pacific, northwards and across the Bering Sea to the Kamchatka Peninsula, extending to southern Japan in the western Pacific (Allen & Smith 1988). Sablefish are prized for their delicate meat and support high-value fisheries, especially in Alaska, which supports the largest fishery for sablefish with an ex-vessel value often exceeding US$100 million (Huppert & Best 2004, Hanselman et al. 2015). Adult sablefish spawn along the continental shelf and fertilized eggs remain at depth (200–800 m; Thompson 1941, Kodolov 1968, Moser et al. 1994) before juveniles migrate to the surface during the larval phase (McFarlane & Beamish 1992). Juveniles then settle in nearshore benthic habitats and move offshore as they mature

(Krieger et al. 2020). Sablefish reach maturity at ~6 years of age and can live more than 90 years (Echave et al. 2012). A defining feature of sablefish biology is their high propensity to move long distances at all life stages. For example, extensive tagging studies have documented that adult sablefish can move >2000 km over their lifetime and that, in general, adult sablefish are highly mobile (Hanselman et al. 2015).

Sablefish have traditionally been considered two stocks: a northern stock, which inhabits waters from northern British Columbia north, and a southern stock, which is found from southern British Columbia south (Kimura et al. 1998, Head et al. 2014). These putative stocks show differences in growth and maturity, with fish from north of the northern border of British Columbia growing to larger sizes (Head et al. 2014). More recently, Kapur et al. (2020) analyzed spatial growth patterns and found evidence of additional zonation associated with oceanographic features. Sablefish are currently managed based on four primary spatial management zones that largely correspond with political boundaries: (i) Alaska Federal (offshore) Waters, (ii) Alaska State (inshore) Waters, (iii) British Columbia, and (iv) West Coast United States. Genetic structure in sablefish has been analyzed with multiple genetic markers, including allozymes (Tsuyuki & Roberts 1969, Gharrett et al. 1982), mitochondrial DNA (mtDNA) (Orlova et al. 2019, Orozco-Ruiz et al. 2023), microsatellites (Tripp-Valdez et al. 2012, Orozco-Ruiz et al. 2023), and single-nucleotide polymorphisms (SNPs) derived from RAD sequencing (Jasonowicz et al. 2017). Nearly all previous studies have documented either no population structure or weak population structure between samples taken from the extreme southern end of the range (Mexico) and collections taken from primary areas of the species distribution. However, the two most recent and most comprehensive genetic studies in sablefish (Jasonowicz et al. 2017, Orozco-Ruiz et al. 2023) documented different patterns of population structure, with some differences in the geographic ranges covered.

The only study that has used genomic approaches to investigate population structure in sablefish to date was conducted by Jasonowicz et al. (2017), who genotyped 2661 SNPs with RAD sequencing in 404 individuals distributed nearly continuously from southern California to the Aleutian Islands (no samples from British Columbia). Jasonowicz et al. (2017) found low and insignificant levels of population structure and concluded that sablefish are likely panmictic across this range. Contrastingly, Orozco-Ruiz et al. (2023) genotyped individuals from Mexico to the Kamchatka Peninsula using microsatellites (11 microsatellites in 252 individuals) and mtDNA (D-loop in 121 individuals) and concluded that significant population structure exists in the species. Orozco-Ruiz et al. (2023) found evidence of genetic structure between their southernmost site (Mexico) and the rest of the species range with microsatellites and also found evidence of genetic structure between some northern sites (Russian waters off the Kamchatka Peninsula) and the rest of the geographic range with mtDNA. Additionally, Orozco-Ruiz et al. (2023) analyzed patterns of genetic structure in the microsatellite dataset with discriminant analysis of principal components (DAPC) (Jombart et al. 2010) for males, females, and both sexes combined, in the Bering Sea (Russian waters off Kamchatka not included) and Gulf of Alaska. This analysis suggested genetic structure among some sites in this region when both sexes were combined and in the male-specific dataset, but not the female-specific dataset. However, this genetic structure was not detected with the program STRUCTURE (Pritchard et al. 2000).

Table 1. Information associated with each geographic region in the analyzed dataset, including the geographic region (and code) from which samples were collected, the sample size (*N*), and the years in which samples were collected.

Region (code)	*N*	Collection years
Bering Sea and Aleutian Islands (BSAI)	25	2006–2018
Western Gulf of Alaska (wGOA)	18	2011–2013
Eastern Gulf of Alaska (eGOA)	29	2011–2013
Coast of Washington State, USA (south)	49	2017–2018

The primary goal of our study was to use a high-resolution genetic approach to enhance understanding of the genetic structure of sablefish in Alaska. Our study represents a methodological improvement over past studies in two important ways: (i) we use lcWGS to genotype orders of magnitude more genetic markers than past studies, and (ii) we focus our analysis on juvenile samples, which have likely had much less opportunity to move than adults. All previous genetic studies in sablefish focused on non-spawning adults because spawning aggregations are difficult to locate and few genetic samples have been collected during the spawning season. However, analyzing individuals that are not actively spawning can obscure signals of population structure if individuals from multiple populations mix outside the spawning season. Juvenile sablefish can still move relatively large distances through larval dispersal, but these distances should be smaller than the 1000 s of km movements documented in adults (Hanselman et al. 2015). We therefore hypothesize that the juvenile samples analyzed in our study represent the closest approximation to individuals associated with spawning events to date. In addition to the lcWGS data, including separate analysis of the nuclear and mitochondrial genomes, we also reanalyze data from Jasonowicz et al. (2017), grouped by sex and with similar methods to Orozco-Ruiz et al. (2023), to determine if we can recover the same patterns of population structure found in Orozco-Ruiz et al. (2023). Finally, we leverage the recently published sablefish genome (Flores et al. 2023) to identify large putative structural variants.

Methods

Sample collection

Low-coverage whole genome resequencing (lcWGS) was conducted on 123 sablefish collected between 2006 and 2019. Collection sites were grouped into four geographic regions that contained sufficient sample sizes for analysis: the Bering Sea and Aleutian Islands (BSAI, $n = 27$), the western Gulf of Alaska (wGOA, $n = 18$), the eastern Gulf of Alaska (eGOA, $n = 29$), and the coast of Washington State, USA (south, $n = 49$) (Table 1; Fig. 1). Samples from all areas except Washington State were young-of-year and were <10 g. These samples were either collected from larval surveys conducted by the Ecosystems and Fisheries–Oceanography Coordinated Investigations (ECO–FOCI) group at the National Oceanographic and Atmospheric Administration (NOAA) Alaska Fisheries Science Center or collected from beak loads from bird surveys conducted by the US Fish and Wildlife Service on Buldir Island, AK, USA. Samples from Washington State were adults (~3.8 kg) that were as close to maturity

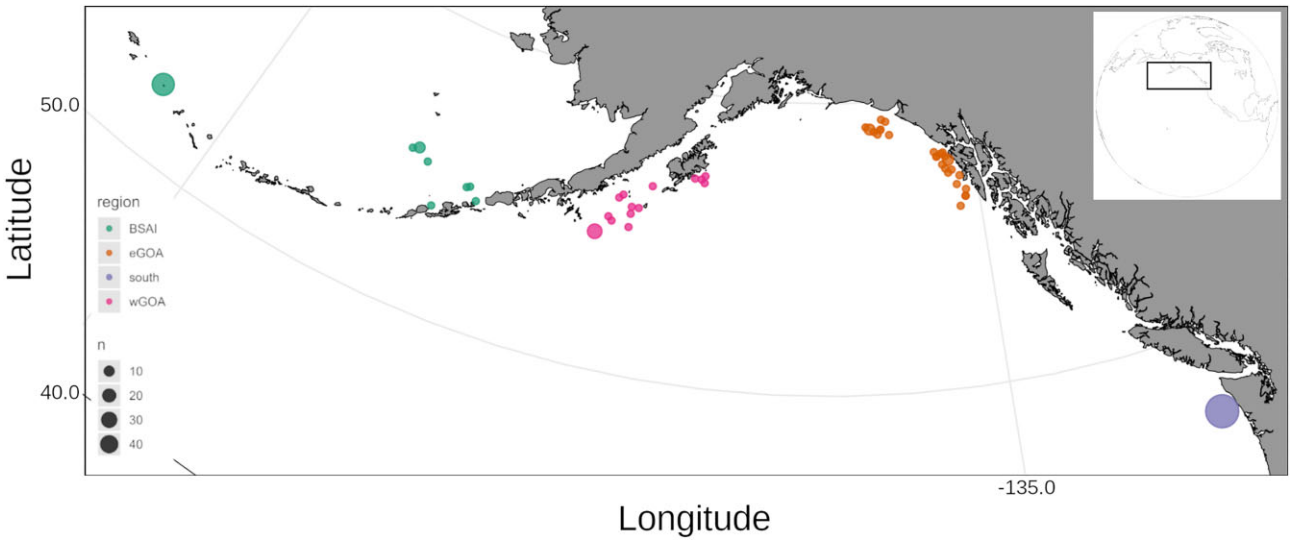

Figure 1. Map of sampling locations for individuals included in the final dataset ($n = 119$). Geographic region is denoted by color: the Bering Sea and Aleutian Islands (BSAI) in green, the western Gulf of Alaska (wGOA) in pink, the eastern Gulf of Alaska (eGOA) in orange, and the coast of Washington State, USA (south) in purple. Sample size (*n*) is represented by marker size.

as possible, collected by the NOAA Northwest Fisheries Science Center as broodstock for sablefish aquaculture experiments. Phenotypic sex was identified in these adult samples by examining gametes prior to spawning for aquaculture. All samples were preserved in ethanol and DNA was extracted with Qiagen DNeasy blood and tissue kits (Hilden, Germany) using manufacturer's protocols. Collection metadata for every sequenced individual is reported in Supplementary Table S1.

Some analyses and all plotting were conducted in R v4.2.0 (R Core Team 2022), utilizing a number of libraries: *dplyr* v1.1.2 (Wickham et al. 2023), *ggOceanMaps* v1.3.4 (Vihtakari 2023), *ggplot2* v3.4.4 (Wickham 2016), *ggpubr* v0.4.0 (Kassambara 2020), *gtools* v3.9.2 (Warnes et al. 2021), *RColorBrewer* v1.1.3 (Neuwirth 2022), *readxl* v1.4.3 (Wickham & Bryan 2023), *reshape2* v1.4.4 (Wickham 2007), *runner* v0.4.3 (Kałędkowski 2023), *stringr* v1.5.0 (Wickham 2022), and *tidyverse* v2.0.0 (Wickham et al. 2019). All figures were finalized in GIMP (GNU Image Manipulation Program). For detailed information regarding data assembly, analysis, and visualization, see https://github.com/letimm/sablefish_lcWGS.

Whole genome library preparation and low coverage sequencing

Whole genome resequencing library preparation was conducted using methods similar to Baym et al. (2015), Therkildsen & Palumbi (2017), and Euclide et al. (2023). For each individual, genomic DNA input was normalized to 10 ng. As in Euclide et al. (2023), sample purification, product normalization, and pooling were conducted with SequalPrep plates (ThermoFisher Scientific: Waltham, MA, USA) in lieu of AMPure XP beads (Beckman Coulter: Brea, CA, USA). Following Therkildsen & Palumbi (2017), AMPure XP beads were used for a 0.6× size selection, clean-up, and concentration of the normalized, pooled library. Prior to sending for sequencing, the final pooled library was visualized on a 2% agarose E-gel (ThermoFisher Scientific) and quantified with the Qubit HS dsDNA Assay Kit (ThermoFisher Scientific). Sequencing occurred at Novogene (Sacramento, CA, USA) on $\sim\frac{2}{3}$ of a NovaSeq S4 lane.

Data filtering, alignment, and genotype likelihood estimation

Our workflow for quality filtering and obtaining genotype likelihoods was similar to that of Clucas et al. (2019). We used FastQC (Andrews 2010) and multiQC (Ewels et al. 2016) to aggregate sequence QC results. The *ILLUMINACLIP* tool in TRIMMOMATIC v0.39 (Bolger et al. 2014) was used to cut adapters and Illumina-specific sequences. Two seed mismatches were allowed, the palindrome and simple clip thresholds were set to 30 and 10, respectively; the minimum adapter length was set to 1 bp; and both reads were kept (*2:30:10:1: true*). Finally, we discarded reads <40 bp after trimming (*MINLEN:40*). Trimmed, paired reads were aligned to the *Anoplopoma fimbria* genome (GCF_027596085.1; Flores et al. 2023) with the *mem* algorithm in BWA v0.7.17 (Li & Durbin 2009). To enable compatibility with downstream filtering programs, reads that mapped to two different genomic regions were marked as duplicates (*-M*). After alignment, reads were sorted by coordinates with the *sort* tool in Samtools v1.11 (Danecek et al. 2021). We removed PCR duplicated reads with MarkDuplicates in Picard v2.23.9 (http://broadinstitute.github.io/picard/), allowing variable length reads to remain coded (*VALIDATION_STRINGENCY SILENT*), clipped overlapping sequence termini for each mapped read pair with *clipOverlap* in bamutil v1.0.5 (https://genome.sph.umich.edu/wiki/BamUtil), and calculated coverage statistics for each individual with the depth tool in samtools. Individuals with a mean depth of <1× were removed from downstream analyses.

Single nucleotide polymorphisms (SNPs) were called and genotype likelihoods were estimated in ANGSD v0.933 (Korneliussen et al. 2014) with the Samtools model (*-GL 1*). To call a SNP, the minimum and maximum total depth thresholds were 119 (the number of samples in the dataset; *-setMinDepth 119*) and 2380 (20× the number of samples in the dataset; *-setMaxDepth 2380*), respectively. The minimum

depth threshold reflected the 1× coverage filter applied prior to estimating genotype likelihoods and the maximum depth filter was set to remove repetitive regions, as it was much higher than the average mean individual coverage observed in the dataset. Sites were excluded from consideration as SNPs if their mapping quality score was <15 (*-minMapQ 15*) or if the *P*-value for the site being polymorphic was $>10^{-10}$ (*-SNP_pval 1e-10*). Finally, SNPs were only retained if the minor global allele frequency was ≥5% (*-minMaf 0.05*). Genotype likelihoods across all individuals were used to infer major and minor alleles for polymorphic sites (*-doMajorMinor 1*).

Population structure with whole genome data

Genetic distances between individuals were explored with principal components analysis (PCA) in PCAngsd v0.99 (Meisner & Albrechtsen 2018). A covariance matrix was calculated among individuals in PCAngsd, and we decomposed this matrix to eigenvectors with the *eigen* function, as implemented in R. The number of eigenvalues retained (-e parameter) was determined with a minimum average partial (MAP) test, the default method in PCAnsgd. Principal component analyses were conducted for SNPs across the full genome and for the full genome without chromosome 22, which contains two large putative inversions (see results). The latter dataset was also used to estimate individual admixture proportions with NGSadmix v32 (Skotte et al. 2013), testing $K = 1$–4 with 3 replicates. Optimal K was determined as having the highest log likelihood.

The weighted population differentiation metric, F_{ST}, was estimated between all pairs of sampling regions in ANGSD. Site allele frequency (SAF) likelihood for each region was estimated (*-doSaf 1*) prior to calculating the 2D site frequency spectrum for each region pair (*-realSFS*) and the weighted F_{ST} (*realSFS fst*). The SAF calculation utilized the same thresholds as set in the genotype likelihood calculation for SNPs, where the minimum depth and maximum depth were set to 1× and 20× the number of individuals representing the region. Additionally, the sablefish reference genome was supplied as an ancestral genome (*-anc* was the same as *-ref*), as no ancestral genome was available.

To determine whether weighted F_{ST} values were statistically significant, we conducted an individual-based permutation test. Population number and sample sizes matched empirical data and every individual was randomly assigned to a population, without replacement, for each permutation. Once all individuals had been sampled, a weighted F_{ST} was calculated for all population pairs. A distribution of F_{ST} values was built from 50 permutations for every population pair. We then calculated the mean of each distribution and used it to estimate the cumulative distribution function (CDF) of the F_{ST} values for the population pair under an exponential distribution (Elhaik 2012), where *P*-value = 1–CDF.

To investigate the genomic landscape of differentiation among regions, pairwise F_{ST} for each regional comparison was estimated for every polymorphic site in the genome with ANGSD (*realSFS fst stats2*) and Manhattan plots were constructed.

Population structure with mitochondrial data

We analyzed the mitochondrial genome from the lcWGS dataset separately to facilitate comparisons between our data and Orozco-Ruiz et al. (2023). Raw reads were trimmed, aligned to the reference genome (GCF_027596085.1) and the mitochondrial genome (NC_018119.1; Rondeau et al. 2013), and filtered. Genotype likelihoods were calculated at polymorphic sites across the mitochondrial genome as described above, and the resulting beagle file was converted to a fasta file with a custom Python script that coded heterozygous sites as *N*s (https://github.com/letimm/sablefish_lcWGS/blob/main/scripts/beagle2fasta.py). The multiple sequence alignment was edited manually to remove any sites where all individuals were genotyped as *N*.

Analysis of the Fasta file occurred in R. A haplotype network was built from the fasta with *haploNet* in *pegas* v1.1 (Paradis 2010). We conducted an Analysis of Molecular Variance (AMOVA) with *poppr* v2.9.3 (Kamvar et al. 2014, 2015). Finally, we calculated pairwise F_{ST} with *hierfstat* v0.5–10 (Goudet 2005), testing for significance with an individual-based permutation test similar to that described above: individuals were randomly assigned to populations for each permutation, maintaining original population sizes, and pairwise F_{ST} values were calculated. The observed F_{ST} values were compared to distributions constructed from randomly sampled individuals with a permutation test ($N = 1000$ permutations) in *ade4* v1.7.19 (Chessel et al. 2004, Dray & Dufour 2007, Dray et al. 2007, Bougeard & Dray 2018, Thioulouse et al. 2018).

Population structure by sex with published RAD dataset

We used genotypes from the RAD dataset published in Jasonowicz et al. (2017) (Genbank Bioproject PRJNA279314) to investigate whether we could recover the same patterns of population structure documented in Orozco-Ruiz et al. (2023). This dataset includes 404 individuals (198 males and 206 females) genotyped at 2661 SNPs. Individuals were partitioned into groups denoted in Fig. 1 of Jasonowicz et al. (2017): northern Bering Sea, southern Bering Sea, Aleutian Islands, western Gulf of Alaska, central Gulf of Alaska, West Yakutat, East Yakuat, Washington, Oregon, northern California, and southern California. These groups correspond closely to those in Orozco-Ruiz et al. (2023), with the important exception that Jasonowicz et al. (2017) did not include any samples from Mexico or Russia. Therefore, we focus on comparing the dataset from Jasonowicz et al. (2017), both with sexes combined and separated, to the Alaska-wide datasets presented in Orozco-Ruiz et al. (2023). This Alaska-wide dataset contains samples from the northern Bering Sea, southern Bering Sea, Aleutian Islands, western Gulf of Alaska, central Gulf of Alaska, and eastern Gulf of Alaska. Jasonowicz et al. (2017) conducted a PCA that indicated no genetic structure across their full dataset; however, they did not analyze the data by sex. We therefore conducted separate PCAs for each sex using the *dudi.pca* function in the *ade4* v1.7.22 package in R (v4.3.1). This is the same PCA function that *adegenet* uses for the current implementation of DAPC.

As DAPC results were the primary evidence for population structure in sablefish in Orozco-Ruiz et al. (2023), we also conducted DAPC using *adegenet* (v2.1.10). DAPC is a frequently applied analysis for investigating population structure, but recent papers have suggested that incorrect conclusions can be drawn if caution is not taken when conducting this analysis and interpreting the results (Miller et al. 2020, Thia 2023). In particular, the number of principal components (PCs) retained can substantially influence DAPC results,

and retaining too many PCs can lead to overfitting (Miller et al. 2020, Thia 2023). Recently, Thia (2023) used simulations to demonstrate that no more than *K*-1 PCs should be retained, where *K* is the number of effective populations in a dataset. However, many frequently used methods for choosing the number of PCs suggest retaining many more than *K* effective populations (Thia 2023). This includes cross-validation, a popular method for choosing the number of PCs to retain, which employs a training-holdout approach to compare DAPC results with different subsets of a dataset and select the number of PCs that maximize re-assignment accuracy (Jombart et al. 2010).

It is important to note that Orozco-Ruiz et al. (2023) used sampling locations as *a priori* groups and determined the optimal number of PCs to retain with a cross-validation approach implemented in *poppr* (Kamvar et al. 2014, 2015), resulting in the retention of 60 PCs for the male and combined sexes datasets and 10 for the female dataset. We used the following three approaches to choose the number of PCs to retain for DAPCs conducted with the data from Jasonowicz et al. (2017): the *K*-1 approach described in Thia (2023), retaining 60 PCs similar to Orozco-Ruiz et al. (2023), and cross-validation. We believe $K = 1$ is the most appropriate value for sablefish in the (Jasonowicz et al. 2017) dataset given their finding of panmixia. However, this *K* cannot be used for DAPC, so we decided to use 2 *K*-values: $K = 11$, the number of fine-scale regions defined in Jasonowicz et al. (2017) resulting in 10 retained PCs; and $K = 3$, the number of broad-scale groups defined in Jasonowicz et al. (2017) (Bering Sea, Gulf of Alaska, and West Coast) resulting in 2 retained PCs. We also conducted an additional DAPC, retaining a higher number of PCs (60 PCs), similar to Orozco-Ruiz et al. (2023), in order to illustrate the sensitivity of DAPC to the number of PCs retained. Finally, we implemented cross-validation to select PCs. For this approach, the *xvalDapc* function in *adegenet* was used with default argument values under the assumption that this is how Orozco-Ruiz et al. (2023) conducted their DAPC cross-validation since specific argument values were not detailed in their methods.

Identification of putative structural variants

Genotype patterns at large structural variants such as chromosomal inversions can drive patterns in genome-wide PCA analyses, especially when population structure is low (Nowling et al. 2020, Kess et al. 2021). This was the case in our lcWGS data, where the initial PCA we constructed contained six discrete clusters unassociated with specific populations, which we labeled A–F from left-to-right along PC1, indicating potential structural variation. Manhattan plots of F_{ST} across the genome with each discrete PCA cluster coded as a population revealed that differentiation among groups was found on chromosome 22 in two discrete regions separated by a small (~1 MB) area of low differentiation. To investigate this variation further, we constructed separate PCAs for each region containing loci with $F_{ST} > 0.3$. We also constructed genotype heatmaps in R using a custom script that summarized each set of three genotype likelihood values into a single estimate of allele dosage for each individual at each SNP. Finally, linkage disequilibrium (LD) was estimated with ngsLD (Fox et al. 2019) for each region, including 500 bp upstream and downstream, with SNPs subsampled by 200× for computational tractability.

Based on these analyses, we were able to assign genotypes for individuals in each region. We then calculated observed genotype frequencies and expected genotype frequencies for each region based on Hardy–Weinberg equilibrium. We calculated these frequencies in three ways: (i) assuming the two inversions were unlinked; (ii) assuming the inversions were fully linked; and (iii) assuming that the inversions were unlinked but a certain haplotype that was not found in the observed data was deleterious. We tested for significant deviations from Hardy–Weinberg and linkage equilibrium using the inferred genotypes for each region with exact tests in Genepop V4 (Rousset 2008). We also calculated r^2 for inferred genotypes to measure LD using NeEstimator v2.1 (Do et al. 2014).

Results

Sequencing and genotype likelihood calculation

Sequencing 123 individuals yielded 1 722 328 206 raw paired sequences—~14 002 668 paired sequences per individual (Timm and Larson 2024). After aligning reads to the reference genome, three individuals were removed for insufficient coverage (<1×). Preliminary PCAs revealed two samples were duplicates, so one was removed. The resulting 119 individuals had a mean depth per individual of 4.1 ($SD = 1.20$) and the final dataset contained 7 110 228 single nucleotide polymorphisms (SNPs).

Population structure with whole genome data

Initially, PCA revealed six clusters separating along PC1, representing 1.95% variance (PC2 represented 1.03% variance). These clusters were not associated with the geographic region from which samples were collected (Fig. 2a). Manhattan plots of F_{ST} among PCA clusters localized the signal to chromosome 22. When this chromosome was removed from the dataset and PCA was rerun on the remaining 23 chromosomes, no population genetic structure was evident (Fig. 2b). Specifically, most individuals grouped in a single cluster containing samples from all collections, while a few individuals from the southernmost population in Washington State (south) were separated slightly on PC2. Analysis of individual admixture proportions produced similar results to the PCA, with no signal of genetic structure associated with geography (Supplementary Fig. S1) and an optimal *K* value of one. Log likelihood values ranged from 16775.82 ($K = 4$) to nearly infinite ($K = 1$, optimal), increasing as *K* decreased.

Pairwise F_{ST} values were extremely low, ranging from 0.004 (eGOA vs south) to 0.007 (wGOA vs eGOA; wGOA vs south) and none were statistically significant (Table 2). Weighted F_{ST} values never differed by >0.001 from the average of the distribution built from 50 permutations (wGOA vs eGOA distribution average was 0.006) and *P*-values ranged from 0.888 (wGOA vs eGOA) to 1.0 (BSAI vs wGOA). Manhattan plots of genetic differentiation across the genome did not reveal any observable high-F_{ST} regions or conspicuous signals of putative adaptive divergence among regions (Fig. 3).

Population structure with mitochondrial data

The lcWGS data produced genotypes for 99 SNPs across the mitochondrial genome. The haplotype network visualized from the mitochondrial data revealed extreme haplotypic diversity, with haplotypes grouped into three central hubs, as well as a separate branch leading to seven individual

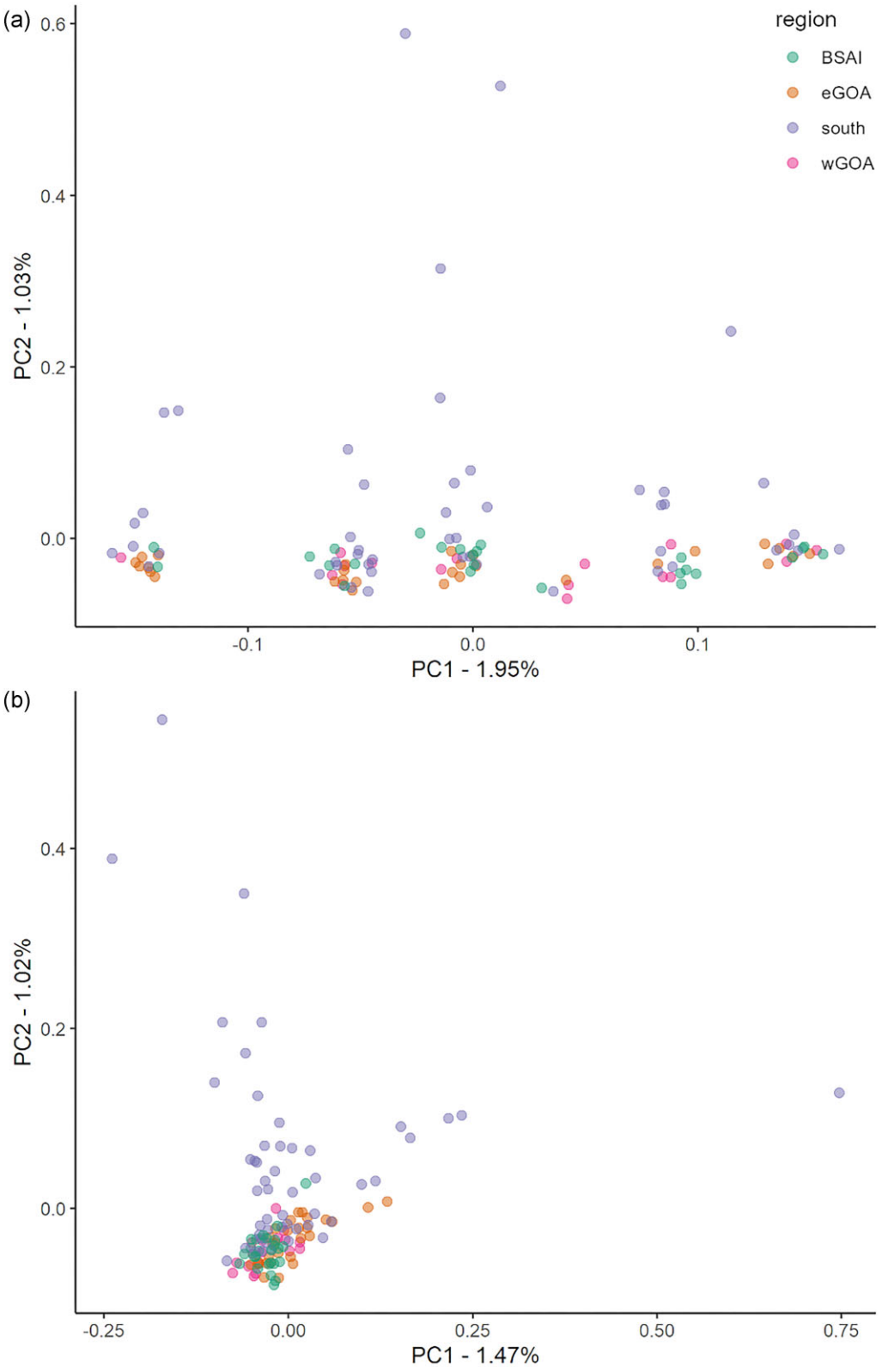

Figure 2. Principal component analyses (PCAs) of 119 sablefish from the Bering Sea and Aleutian Islands (BSAI), the western Gulf of Alaska (wGOA), the eastern Gulf of Alaska (eGOA), and the coast of Washington State, USA (south). Colors representing regions are consistent with those in the sampling map (Fig. 1). (a) PCA includes variant sites from across the whole genome. The six clusters separated along PC1 are hereafter referred to as clusters A–F, from left to right. These clusters correspond to genotypes of two putative inversions identified on chromosome 22 (see Fig. 7). (b) PCA includes variant sites from across the whole genome, with the exception of chromosome 22.

Table 2. Pairwise F_{ST} values for every pair of geographic regions [the Bering Sea and Aleutian Islands (BSAI), the western Gulf of Alaska (wGOA), the eastern Gulf of Alaska (eGOA), and the coast of Washington State, USA (south)].

	BSAI	wGOA	eGOA	South
BSAI		0.006 | 0.006	0.005 | 0.000	0.005 | 0.001
wGOA	1.0 | 1.0		0.007 | 0.008	0.007 | 0.008
eGOA	0.9 | 1.0	0.9 | 1.0		0.004 | 0.001
South	1.0 | 1.0	0.9 | 1.0	1.0 | 1.0	

F_{ST} values are reported above the diagonal; *P*-values are reported below. The values represent those calculated from the nuclear genome and the mitochondrial genome (nuclear | mitochondrial), respectively.

haplotypes (Fig. 4). Each major hub contained haplotypes shared between geographic regions and the seven satellite haplotypes included individuals from all four regions. We observed 105 unique haplotypes and no haplotype was found in more than three individuals. AMOVA estimated Φ_{ST} to be 0.003, suggesting little molecular variance could be attributed to differences between geographic regions. This was further supported by analysis of pairwise F_{ST}. Values ranged from 0 (eGOA vs BSAI) to 0.008 (eGOA vs wGOA and wGOA vs south; Table 2). Permutation testing revealed none of the pairwise F_{ST} values were statistically significant (all simulated *P*-values were =1).

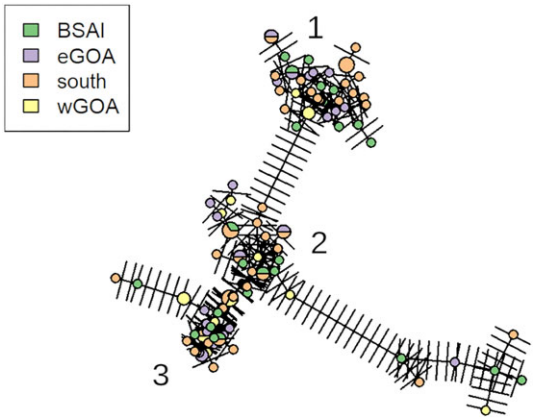

Figure 4. Haplotype network of the full mitochondrial genome constructed using data from 99 SNP loci. The network illustrates extreme haplotypic diversity, primarily centered around three major hubs (labeled 1–3).

Population structure by sex with published RAD dataset

Visual inspection of the top two PCs did not reveal any clear patterns of spatial population structure when both sexes were analyzed together or when males and females were analyzed separately (Fig. 5). The various PC selection approaches

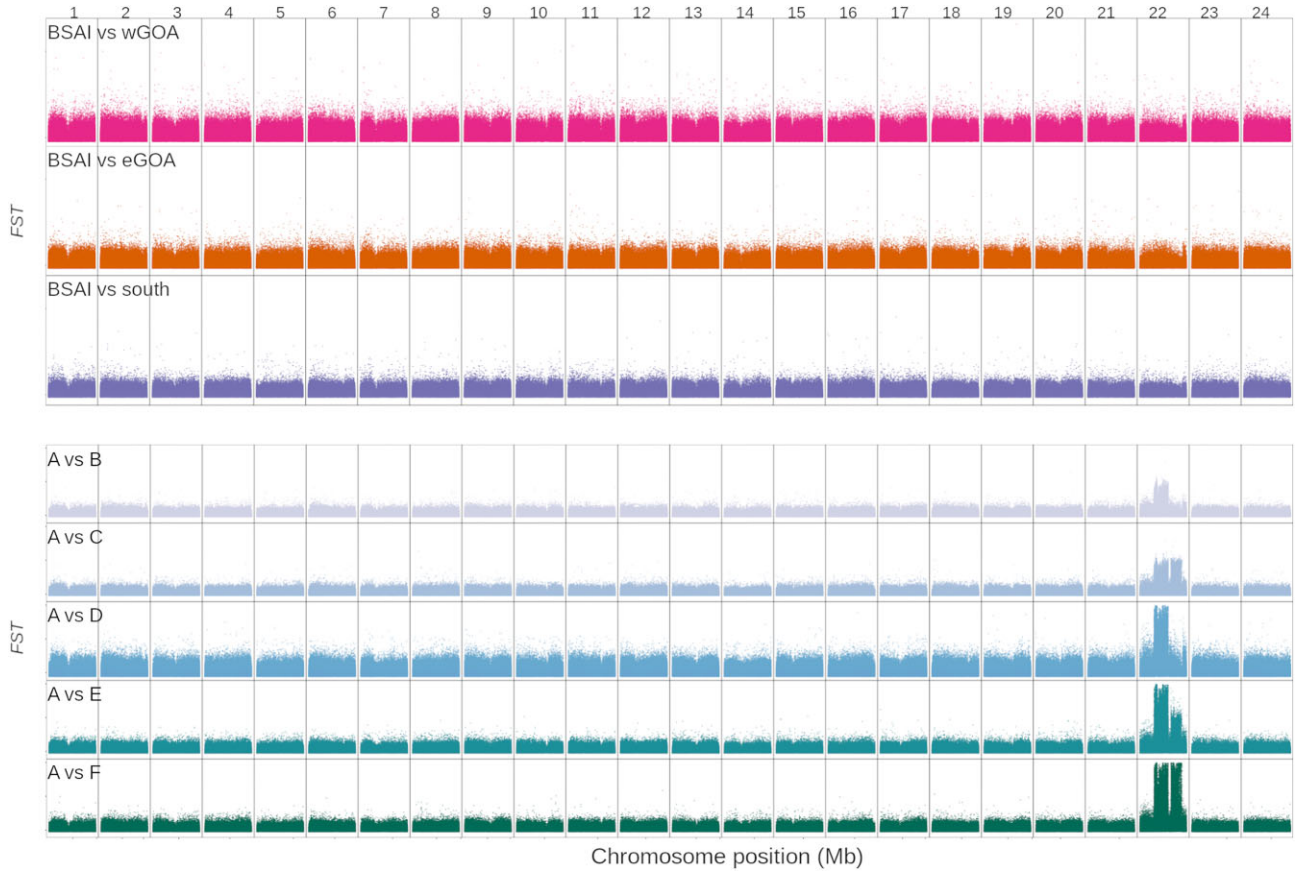

Figure 3. Manhattan plots of FST values when comparing pairs of geographic regions (upper): the Bering Sea and Aleutian Islands to the western Gulf of Alaska (BSAI vs wGOA), the eastern Gulf of Alaska (BSAI vs eGOA), and the coast of Washington State, USA (BSAI vs south); and clusters identified with the whole genome PCA (lower): A vs each of the other clusters. The 15 Mb position is marked along the bottom of each chromosome. Each chromosome is labeled at the top of the plot.

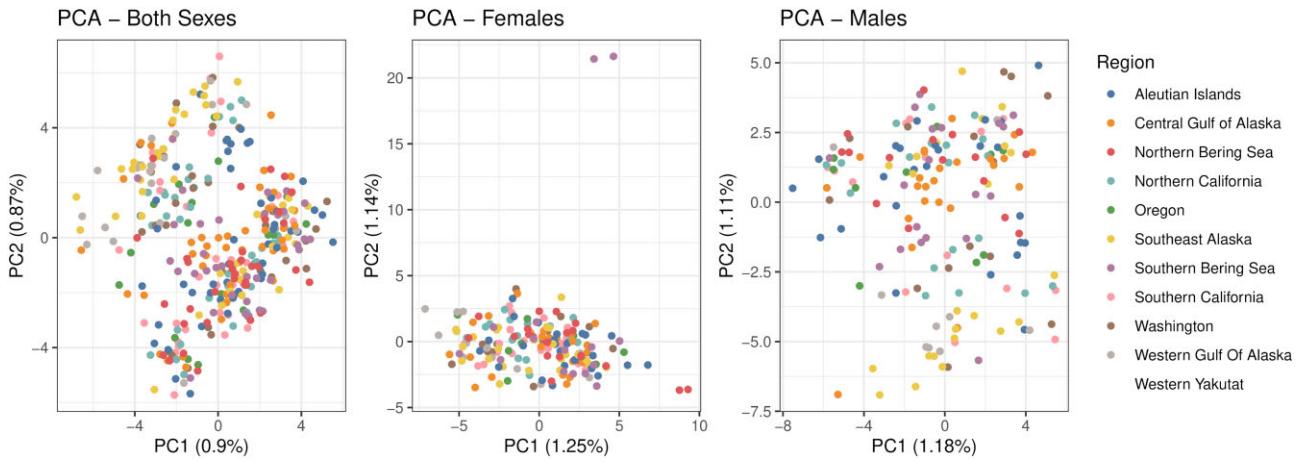

Figure 5. Principal components analysis of 2661 RADseq-derived SNP loci from sablefish collected in the northeast Pacific Ocean (originally published in Jasonowicz et al. 2017).

Table 3. Summary of the % variance explained when different numbers of PCs are retained for discriminant analysis of principal components runs using 2661 RADseq-derived SNP loci in both sexes, females-only and males-only.

	2 PCs	10 PCs	60 PCs	Cross-validation
Both sexes	1.77%	6.14%	27.43%	79.38% (250)
Females	2.39%	9.49%	42.64%	87.7% (160)
Males	2.30%	8.90%	42.80%	64.13% (100)

We included analyses when 2, 10, and 60 PCs were retained. When cross-validation was used to determine the number of PCs to retain, the number of PCs retained is reported in parentheses.

resulted in a wide range of % variance of the dataset being retained for the discriminant analysis portion of DAPC (Table 3). This is especially dramatic in the case where cross-validation was used to select the number of PCs: 250 PCs accounted for 79.38% of the total variance when both sexes were analyzed together; however, when males and females were analyzed separately, 160 PCs were retained for the female-only dataset and 100 for the male-only dataset, resulting in 87.7 and 64.13% of the total variance retained in each, respectively. Not surprisingly, as the number of PCs retained increased (and % variance explained by the PCs retained), separation among the *a priori* specified groupings increased as well (Fig. 6).

Identification of putative structural variants

Manhattan plots of pairwise F_{ST} calculated among the six clusters found in the PCA containing all loci across the genome (Fig. 2b) revealed two blocks of elevated divergence on chromosome 22 (Fig. 3): the first was ~4Mbs in length and spanned positions 4.13e6-8.19e6 [hereafter referred to as inversion 1 (inv1)]; the second was ~3Mbs in length and spanned positions 8.97e6-11.90e6 [hereafter referred to as inversion 2 (inv2)] (Fig. 7a). Sites within these regions with $F_{ST} > 0.3$ were targeted to generate a dataset for each block (16 802 and 7248 SNPs, respectively). PCAs of each block revealed that each contained three clusters, which likely correspond to alternate homozygotes with a heterozygote cluster in the middle (Fig. 7b and c). The variation explained by PC 1 was much higher than PC2 in both PCAs (inv1 PCA PC1 38%, PC2 1%; inv2 PCA PC1 56%, PC2 1%). There was evidence of a small amount of additional sequence variation in the PCAs based on the presence of multiple clusters on PC2 for homozygote (hom) 0 in inv1 and PC2 for hom2 in inv2. Heatmaps of LD revealed elevated LD within putative inversion regions; average Pearson correlation coefficients within blocks were 0.056 for inv1 and 0.103 for inv2 (Fig. 7d and e).

Observed genotype frequencies within each inversion were similar to expected frequencies (Table 4) and no evidence of deviations from Hardy–Weinberg equilibrium was detected (P-value = 0.85 for inv1 and 0.16 for inv2). However, observed genotype frequencies differed substantially from expected frequencies when inversions were treated as unlinked. Specifically, genotypes containing the haplotypes inv1 allele (al) 0 inv2 al0 and inv1 al1 inv2 al1 were generally more common than expected, while genotypes containing the inv1 al1 inv2 al0 haplotype were absent, even though genotypes containing these haplotypes were expected at a frequency of 0.16. Additionally, the expected proportion of the inv1 hom0 inv2 hom1 genotype was found at a frequency 0.11 less than expected. When the loci were treated as fully linked, observed genotype frequencies closely matched expected frequencies. Finally, when the inv1 al1 inv2 al0 haplotype is treated as deleterious and any individual with this haplotype is not viable, observed genotypes deviate substantially from expectations. Deviations are generally similar to those observed when the inversions were treated as unlinked, with the inv1 hom0 inv2 hom1 genotype showing the largest deviation from expectations (observed frequency = 0.16, expected = 0.06). Significant LD between the putative inversions was detected based on analysis of inversion genotypes (P-value = 0, R2 = 0.42), providing further evidence that the inversions are linked.

Discussion

Our study represents the first to use whole genome resequencing to investigate population structure in sablefish and provides strong evidence for panmixia in a large portion of the northern end of their range. We also reanalyzed a RAD-seq dataset, which provided further evidence of panmixia in this region. Finally, we identified two putative chromosomal inversions found near each other on chromosome 22 that appear to be in strong LD. Our findings suggest that, since sablefish are

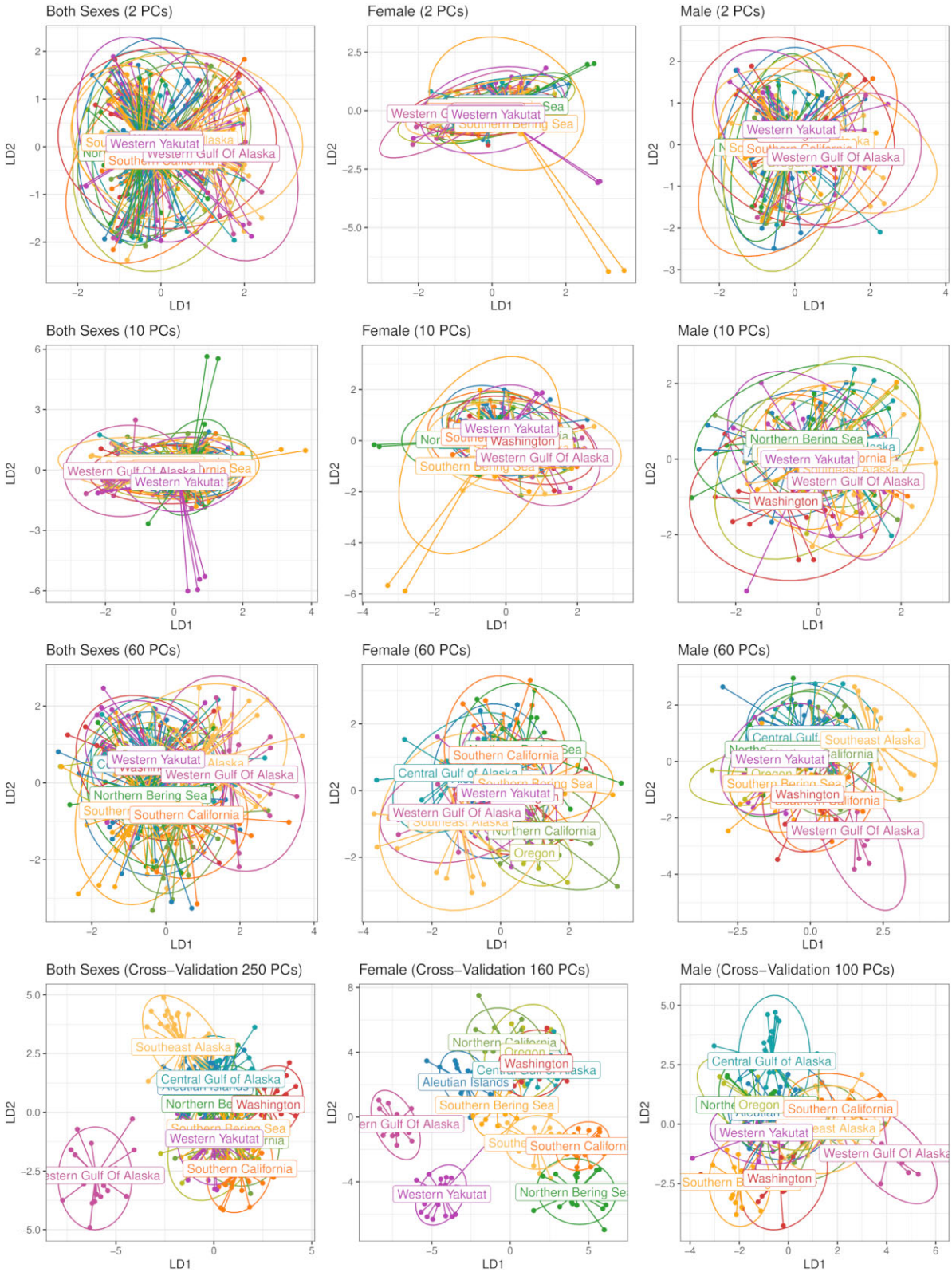

Figure 6. Discriminant analysis of principal components of 2661 RADseq-derived SNP loci (originally published in Jasonowicz et al. 2017). Each panel represents a unique combination of sex (from left-to-right: both sexes, female-only and male-only) and number of retained PCs [from top-to-bottom: 2, 10, 60, and a number inferred from cross-validation (number of retained PCs given in parentheses)].

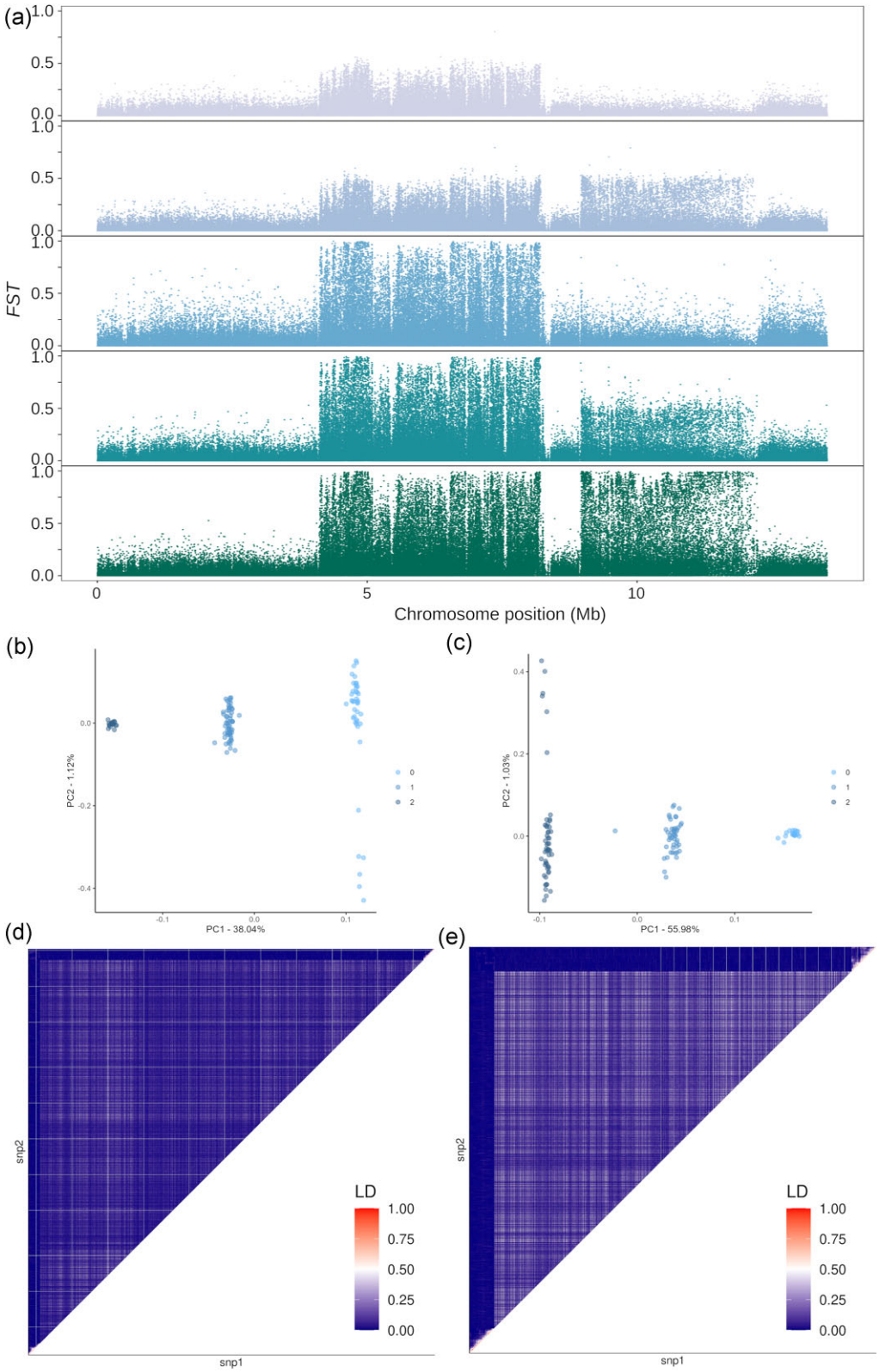

Figure 7. (a) A scan of *FST* values across sites on chromosome 22 identifies two putative inversions. The first is ∼4 Mbs (4.13e6-8.19e6) and the second is ∼3 Mbs (8.97e6-11.90e6). (b) PCA of the first block reveals three genotypes: homozygous for the minor allele (clusters D, E, and F in the whole genome PCA), heterozygous (clusters B and C), and homozygous for the major allele (cluster A). (c) PCA of the second block also reveals three genotypes: homozygous for the minor allele (clusters A, B, and D in the whole genome PCA), heterozygous (clusters C and D), and homozygous for the major allele (cluster F). In both PCAs, individuals from cluster D in the whole genome PCA spread away from the larger genotype cluster on PC2, indicating some sub-variation. (d) Linkage disequilibrium values, expressed as Pearson correlation coefficients for all SNP pairs in the first block. (e) Linkage disequilibrium values, expressed as Pearson correlation coefficients for all SNP pairs in the second block.

Table 4. Inverted genotype frequencies from chromosome 22.

Cluster	inv1_genotype	inv2_genotype	Observed proportion	Unlinked expected	Linked expected	Deleterious expected
A	1/1	1/1	0.15	0.07	0.16	0.09
None	*1/1*	*0/1*	*0.00*	*0.06*	*0.00*	*0.00*
None	*1/1*	*0/0*	*0.00*	*0.02*	*0.00*	*0.00*
B	0/1	1/1	0.24	0.22	0.20	0.27
C	0/1	0/1	0.25	0.20	0.29	0.15
None	*0/1*	*0/0*	*0.00*	*0.08*	*0.00*	*0.00*
D	0/0	1/1	0.05	0.16	0.06	0.20
E	0/0	0/1	0.14	0.14	0.17	0.23
F	0/0	0/0	0.16	0.06	0.13	0.06

The cluster representing the combination, as well as the observed and expected frequencies and the observed and expected sample sizes ("obs" and "exp," respectively) for each inverted genotype pair under three situations: assuming the two inversions were unlinked ("unlinked"), assuming the inversions were fully linked ("linked"), and assuming that the inversions were unlinked but absent haplotypes are deleterious ("deleterious"). When a genotype combination is not seen in the data, the cluster is labeled "none."

highly mobile and panmictic in their northern range, management boundaries in this region do not need to consider genetic structure.

Genetic evidence for panmixia from Washington State to the Bering Sea

We did not document any evidence of genetic structure in sablefish using data from millions of SNPs genotyped in individuals from Washington State, USA to the Bering Sea and Aleutian Islands, AK, USA. Our study is also the first to analyze mostly juvenile samples, which should be more representative of spawning groups than adults because they have had less time to disperse, and we still did not document evidence of genetic structure. In general, our findings are similar to Jasonowicz et al. (2017), who used 2661 markers genotyped with RAD sequencing to investigate the genetic structure of adult sablefish from California to the Aleutian Islands and Bering Sea and also did not find any evidence of genetic structure. Sablefish are extremely mobile based on tagging studies (Hanselman et al. 2015). Thus, not surprisingly, Jasonowicz et al. (2017) found that sablefish display similar values of genetic differentiation as two species that are known to be panmictic, European eels (*Anguilla anguilla*) and American eels (*Anguilla rostrata*) (Côté et al. 2013, Pujolar et al. 2014). Notably, Ulmo-Diaz et al. (2023) recently found panmixia in the American eel extended into the tropical part of its range and, similar to sablefish, identified two putative inversions, which occurred across the geographic sampling region.

The conclusion of panmixia in our study and Jasonowicz et al. (2017) is not consistent with a recent study by Orozco-Ruiz et al. (2023), who analyzed microsatellite data from Mexico to Russia and concluded that genetic structure exists in sablefish throughout this range as well as within the range of our study. In particular, they documented genetic structure within Alaska with DAPC analysis in a male-only dataset and a dataset with both sexes combined, but not in a female-only dataset. Structure was detected among nearly all of the six collections analyzed with the male-only dataset and between the Gulf of Alaska and the Bering Sea/Aleutian Islands with the full dataset. The number of PCs retained for these datasets were 60, 60, and 10, respectively. We reanalyzed the RAD dataset presented in Jasonowicz et al. (2017), which consisted of adult individuals with known sex, to determine if we could recreate these patterns. PCAs of male-specific and female-specific data showed no clustering by region, supporting our conclusion of panmixia. We then constructed DAPCs using various methods to select the number of PCs retained and found that when the number of PCs retained was relatively small ($\leq$10) and likely close to the *K*-1 threshold suggested by Thia (2023), no population structure was observed. However, when the number of PCs retained was larger (>100), populations began to separate, but no specific geographic pattern was observable. Interestingly, population structure was still not visible with 60 PCs retained in the RAD data, even though structure was observable when 60 PCs were retained in Orozco-Ruiz et al. (2023). We hypothesize that this was because the RAD dataset was more robust to variation in retained PCs due to increased marker numbers. It is also possible that the characteristics of microsatellites (i.e. many alleles) may make microsatellite datasets more susceptible to finding artifacts with DAPC. Finally, it is notable that the female-only microsatellite dataset in Orozco-Ruiz et al. (2023) was analyzed with 10 retained PCs and did not show any structure, suggesting that retaining a number of PCs closer to *K*-1 led to results in agreement with our own, namely no population structure.

The DAPC analyses described above illustrate the potential pitfalls associated with conducting DAPC in high gene flow species. Two recent simulation studies address this issue and both highlight the fact that DAPC can create artifacts of population structure in high-gene flow systems (Miller et al. 2020, Thia 2023). DAPC is designed to maximize among-group differentiation, and when groups are defined *a priori* as in the analyses conducted above, DAPC will often return a result that suggests the groups are part of discrete clusters (Miller et al. 2020). This problem is substantially exacerbated when the number of retained PCs is high, as is often supported by many methods used to choose the number of PCs to retain (Thia 2023). If these results are taken as facts, they can lead to incorrect conclusions that could be integrated into management. In our case, postulating that population structure exists in sablefish throughout Alaska could lead to a re-evaluation of the management paradigm for this species. We therefore encourage researchers to be extremely cautious about implementing DAPC in high-gene flow species and also recommend that conclusions of population structure be based on more than one analysis.

Additionally, we did not detect genetic structure using data from the full mitochondrial genome. Diversity was extremely high, with no haplotype shared in more than three individuals, and no spatial clustering was present. Orozco-Ruiz et al. (2023) also found extremely high diversity at the D-loop region but concluded that structure did exist, with the most structure observed between the western Bering Sea and the

remaining samples. They also identified genetic discontinuities between locations in British Columbia and the Gulf of Alaska using genetic landscape shape interpolation analysis (Miller 2005), but did not detect significant structure in this region based on Φ_{ST}. Our study did not include samples from the western Bering Sea, but our finding of no significant structure based on Φ_{ST} in the North Pacific and Gulf of Alaska is in accordance with that of Orozco-Ruiz et al. (2023). We are unsure why the genetic landscape shape interpolation analysis detected potential structure in the North Pacific in Orozco-Ruiz et al. (2023), but it is possible that this approach is not appropriate when structure is weak or non-existent. This discordance between analysis methods again stresses the importance of using multiple lines of evidence to support conclusions of genetic structure, especially when structure is low (Waples 1998).

Putative structural variation on chromosome 22

Our analyses identified two putative structural variants on chromosome 22 that were relatively large (3–4 Mb) and separated by ~1 Mb. We hypothesize that these are inversions identified by clustering patterns in local PCAs, high divergence between homozygous types, and high linkage (Huang et al. 2020, Mérot et al. 2020). Additional evidence for these inversions in sablefish was documented by Flores et al. (2023), who found that recombination in both males and females was highly suppressed in this region, including across the 1 Mb region spanning the inversions. Proximate inversions have also been documented in Atlantic cod (*Gadus morhua*, Kirubakaran et al. 2016). The inversions in cod appear to be adjacent, with no gap between them, whereas the inversions we documented contained a small 1 Mb gap.

We did not document regional differences in genotype frequencies at inversions, suggesting that these regions are not undergoing divergent selection. Our finding of inversions not involved in spatial divergence dominating overall patterns of genetic structure is similar to Kess et al. (2021), who documented a large inversion in another high gene flow marine species, Atlantic halibut (*Hippoglossus hippoglossus*). Kess et al. (2021) discuss the possibility that the inversion they discovered may be related to phenotypic variation within genetically similar populations, and recommended that future studies investigate potential connections between inversion genotypes, behavior, and other phenotypes. We suggest a similar approach for sablefish. It is also important to note that our study did not include the full species range of sablefish, so it is possible the putative inversions we discovered are involved in adaptive divergence in other geographic regions.

Genotype patterns at the two putative inversions that we documented indicated that both inversions are in Hardy–Weinberg equilibrium when analyzed independently, and analysis of LD indicated that the two inversions are highly linked. When both inversions were assumed to be independent, observed genotype frequency differed substantially from expectations. Similar deviations were observed when the loci were assumed to be independent and the inv1 hom1 inv2 hom0 haplotype (which was not observed) was assumed to be deleterious. However, when the inversions were assumed to be in LD, observed genotype patterns closely matched expectations. We hypothesize that linkage likely explains the patterns we observed and the absence of the inv1 hom1 inv2 hom0 haplotype. For example, if the second inversion arose on a DNA strand containing the first inversion, and no recombination occurred, then the inv1 hom1 inv2 hom0 haplotype would never be present. Recent research suggests that recombination can be suppressed outside of inversion breakpoints due to subtle changes in chromosomal pairing during meiosis (Li et al. 2023). It is possible that this or a similar mechanism has reduced recombination to the point where these inversions do not act independently even on long timescales, potentially explaining the absence of a specific haplotype. While we cannot rule out the possibility that the combination of inv1 hom1 inv2 hom0 is deleterious and is being selected against, this hypothesis seems unlikely. Very little research has been conducted to understand whether combinations of inversions are the targets of selection rather than each inversion separately, and the only paper that we could find investigating this topic found no evidence of selection on inversion combinations (Zivanovic et al. 2016). One potential avenue for future research to understand these inversions could be to genotype thousands of individuals at the inversions using amplicon sequencing to determine if the absent haplotype is indeed absent, which would suggest that recombination does not occur in this region.

Conclusions and management implications

Both the genome resequencing and RAD-seq data analyzed in this paper provide strong evidence of panmixia in sablefish in their northern range from the West Coast US to the Bering Sea and Aleutian Islands. This result is generally consistent with past genetic studies and with the life history of sablefish, which involves frequent and long-distance movements. Our results provide additional evidence that the current management strategy for sablefish in the northern part of their range, which is based on high movement rates and genetic panmixia, is appropriate. It is important to note that our study did not include some areas of the sablefish range, such as Mexico, Russia, and Japan. Previous studies (Tripp-Valdez et al. 2012, Orozco-Ruiz et al. 2023) have suggested that genetic structure exists in these areas, and reevaluating those results with whole genome sequencing could confirm their findings and shed light on important adaptive variation in the species.

Acknowledgments

The authors would like to thank Katie D'Amelio, Kirby Karpan, and Diana Baetscher for assistance in sample processing and preparation for sequencing, as well as Sara Schaal and Nicolas Lou for their help interpreting inversions and consultation on lcWGS data. Gratitude is also due to Kris Christensen, Adam Luckenbach, and Eric Rondeau for feedback on an earlier version of this work, as well as to Eric Anderson for his advice on significance testing. We also would like to thank Rick Goetz, Ali Deary, and other researchers on ECO-FOCI cruises, and Alaska Maritime National Wildlife field crews for collecting samples.

Author contributions

W.A.L. and K.M.N. designed the study, with input from all authors. L.E.T. analyzed the whole genome data and A.J.J. analyzed the RAD data. L.E.T. and W.A.L. drafted the manuscript, with assistance from all coauthors. All authors commented on the manuscript and gave final approval for publication.

Supplementary data

Supplementary data is available at *ICES Journal of Marine Science* online.

Conflict of interest: The authors have no conflicts of interest to declare.

Funding

This work is funded in part by the Cooperative Institute for Climate, Ocean, and Ecosystem Studies (CICOES) under NOAA Cooperative Agreement NA20OAR4320271, contribution no. 2024–1365.

Data availability

The data underlying this article are available in the NCBI SRA repository at https://www.ncbi.nlm.nih.gov/sra and can be accessed with PRJNA1099646.

References

Allen MJ, Smith GB. Atlas and zoogeography of common fishes in the Bering Sea and northeastern Pacific. *US Dep Comm NOAA Tech Rep NMFS* 1988;**66**: p. 151.

Allendorf F, Ryman N, Utter F. Genetics and fishery management: past, present, and future. In: N. Ryman, F. Utter (eds), *Population Genetics and Fishery Management*. Seattle, WA: University of Washington Press, 1987, pp. 1–19.

Andrews S. FastQC: a quality control tool for high throughput sequence data. 2010. https://github.com/s-andrews/FastQC (1 May 2023, date last accessed).

Baym M, Kryazhimskiy S, Lieberman TD *et al.* Inexpensive multiplexed library preparation for megabase-sized genomes. *PLoS One* 2015;**10**:e0128036. https://doi.org/10.1371/journal.pone.0128036

Bolger AM, Lohse M, Usadel B. Trimmomatic: a flexible trimmer for Illumina sequence data. *Bioinformatics* 2014;**30**:2114–20. https://doi.org/10.1093/bioinformatics/btu170

Bougeard S, Dray S. Supervised multiblock analysis in R with the ade4 package. *J Stat Softw* 2018;**86**:1–17. https://doi.org/10.18637/jss.v086.i01

Carvalho GR, Hauser L. Molecular genetics and the stock concept in fisheries. *Rev Fish Biol Fish* 1994;**4**:326–50. https://doi.org/10.1007/BF00042908

Chessel D, Dufour A, Thioulouse J. The ade4 package—I: one-table methods. *R News* 2004;**4**:5–10. https://cran.r-project.org/doc/Rnews/

Clucas GV, Lou RN, Therkildsen NO *et al.* Novel signals of adaptive genetic variation in northwestern Atlantic cod revealed by whole-genome sequencing. *Evol Appl* 2019;**12**:1971–87. 10.1111/eva.12861

Côté CL, Gagnaire PA, Bourret V *et al.* Population genetics of the American eel (*Anguilla rostrata*): $F_{ST} = 0$ and North Atlantic oscillation effects on demographic fluctuations of a panmictic species. *Mol Ecol* 2013;**22**:1763–76. 10.1111/mec.12142

Cowen RK, Sponaugle S. Larval dispersal and marine population connectivity. *Annl Rev Mar Sci* 2009;**1**:443–66. https://doi.org/10.1146/annurev.marine.010908.163757

Danecek P, Bonfield JK, Liddle J *et al.* Twelve years of SAMtools and BCFtools. *Gigascience* 2021;**10**:giab008.

Do C, Waples RS, Peel D *et al.* NeEstimator v2: re-implementation of software for the estimation of contemporary effective population size (Ne) from genetic data. *Mol Ecol Resour* 2014;**14**:209–14. https://doi.org/10.1111/1755-0998.12157

Dray S, Dufour A. The ade4 package: implementing the duality diagram for ecologists. *J Stat Softw* 2007;**22**:1–20. 10.18637/jss.v022.i04

Dray S, Dufour A, Chessel D. The ade4 package—II: two-table and K-table methods. *R News* 2007;**7**:47–52. https://cran.r-project.org/doc/Rnews/

Echave KB, Hanselman DH, Adkison MD *et al.* Interdecadal change in growth of sablefish (*Anoplopoma fimbria*) in the northeast Pacific Ocean. *Fish Bull* 2012;**110**:361–74.

Elhaik E. Empirical distributions of F_{ST} from large-scale human polymorphism data. *PLoS One* 2012;**7**:e49837. https://doi.org/10.1371/journal.pone.0049837

Euclide PT, Larson WA, Shi Y *et al.* Conserved islands of divergence associated with adaptive variation in sockeye salmon are maintained by multiple mechanisms. *Mol Ecol* 2023;**00**:1–21. https://doi.org/10.1111/mec.17126

Ewels P, Magnusson M, Lundin S *et al.* MultiQC: summarize analysis results for multiple tools and samples in a single report. *Bioinformatics* 2016;**32**:3047–8. https://doi.org/10.1093/bioinformatics/btw354

Flores AM, Christensen KA, Campbell B *et al.* Sablefish (*Anoplopoma fimbria*) chromosome-level genome assembly. *G3* 2023;**13**:jkad089. https://doi.org/10.1093/g3journal/jkad089

Fox EA, Wright AE, Fumagalli M *et al.* ngsLD: evaluating linkage disequilibrium using genotype likelihoods. *Bioinformatics* 2019;**35**:3855–6. https://doi.org/10.1093/bioinformatics/btz200

Funk WC, McKay JK, Hohenlohe PA *et al.* Harnessing genomics for delineating conservation units. *Trends Ecol Evol* 2012;**27**:489–96.

Gharrett A, Thomason M, Wishard L. Biochemical genetics of sablefish. *NWAFC Process* 1982; Report 82-05. https://repository.library.noaa.gov/view/noaa/9009 (15 January 2024, date last accessed).

Goudet J. Hierfstat, a package for R to compute and test hierarchical *F*-statistics. *Mol Ecol Notes* 2005;**5**:184–6. https://doi.org/10.1111/j.1471-8286.2004.00828.x

Hanselman DH, Heifetz J, Echave KB *et al.* Move it or lose it: movement and mortality of sablefish tagged in Alaska. *Can J Fish Aquat Sci* 2015;**72**:238–51. https://doi.org/10.1139/cjfas-2014-0251

Hansen CCR, Westfall KM, Pálsson S. Evaluation of four methods to identify the homozygotic sex chromosome in small populations. *BMC Genom* 2022;**23**:160. https://doi.org/10.1186/s12864-022-08393-z

Head MA, Keller AA, Bradburn M. Maturity and growth of sablefish, *Anoplopoma fimbria*, along the US west coast. *Fish Res* 2014;**159**:56–67. https://doi.org/10.1016/j.fishres.2014.05.007

Hemmer-Hansen J, Therkildsen NO, Pujolar JM. Population genomics of marine fishes: next-generation prospects and challenges. *Biol Bull* 2014;**227**:117–32. https://doi.org/10.1086/BBLv227n2p117

Huang K, Andrew RL, Owens GL *et al.* Multiple chromosomal inversions contribute to adaptive divergence of a dune sunflower ecotype. *Mol Ecol* 2020;**29**:2535–49. https://doi.org/10.1111/mec.15428

Huppert DD, Best B. *Study of Supply Effects on Sablefish Market Price*. Washington, WA: School of Marine Affairs, University of Washington, 2004, 46.

Jasonowicz AJ, Goetz FW, Goetz GW *et al.* Love the one you're with: genomic evidence of panmixia in the sablefish (*Anoplopoma fimbria*). *Can J Fish Aquat Sci* 2017;**74**:377–87. https://doi.org/10.1139/cjfas-2016-0012

Jombart T, Devillard S, Balloux F. Discriminant analysis of principal components: a new method for the analysis of genetically structured populations. *BMC Genet* 2010;**11**:1–15. https://doi.org/10.1186/1471-2156-11-94

Kałędkowski D. runner: running operations for vectors. R package version 0.4.3, 2023. https://CRAN.R-project.org/package=runner (19 December 2023, date last accessed).

Kamvar ZN, Brooks JC, Grünwald NJ. Novel R tools for analysis of genome-wide population genetic data with emphasis on clonality. *Front Genet* 2015;**6**:208. https://doi.org/10.3389/fgene.2015.00208

Kamvar ZN, Tabima JF, Grünwald NJ. Poppr: an R package for genetic analysis of populations with clonal, partially clonal, and/or sexual reproduction. *PeerJ* 2014;**2**:e281. doi:10.7717/peerj.281

Kapur M, Haltuch M, Connors B *et al.* Oceanographic features delineate growth zonation in northeast Pacific sablefish. *Fish Res* 2020;**222**:105414. https://doi.org/10.1016/j.fishres.2019.105414

Kassambara A. ggpubr: 'ggplot2' based publication ready plots. R package version 0.4.0, 2020. https://CRAN.R-project.org/package=ggpubr (19 December 2023, date last accessed).

Kess T, Einfeldt AL, Wringe B *et al.* A putative structural variant and environmental variation associated with genomic divergence across the northwest Atlantic in Atlantic Halibut. *ICES J Mar Sci* 2021;**78**:2371–84. https://doi.org/10.1093/icesjms/fsab061

Kimura DK, Shimada AM, Shaw FR. Stock structure and movement of tagged sablefish, *Anoplopoma fimbria*, in offshore northeast Pacific waters and the effects of El Nino southern oscillation on migration and growth. *Fish Bull* 1998;**96**:462–81.

Kirubakaran TG, Grove H, Kent MP *et al.* Two adjacent inversions maintain genomic differentiation between migratory and stationary ecotypes of Atlantic cod. *Mol Ecol* 2016;**25**:2130–43. https://doi.org/10.1111/mec.13592

Kodolov LS. Reproduction of the sablefish (*Anoplopoma fimbria*). *Probl Ichthyol* 1968;**8**:531–5.

Korneliussen TS, Albrechtsen A, Nielsen R. ANGSD: analysis of next generation sequencing data. *BMC Bioinf* 2014;**15**:1–13. https://doi.org/10.1186/s12859-014-0356-4

Krieger JR, Beaudreau AH, Heintz RA *et al.* Growth of young-of-year sablefish (*Anoplopoma fimbria*) in response to temperature and prey quality: insights from a life stage specific bioenergetics model. *J Exp Mar Biol Ecol* 2020;**526**:151340. https://doi.org/10.1016/j.jembe.2020.151340

Li H, Berent E, Hadjipanteli S *et al.* Heterozygous inversion breakpoints suppress meiotic crossovers by altering recombination repair outcomes. *PLoS Genet* 2023;**19**:e1010702. https://doi.org/10.1371/journal.pgen.1010702

Li H, Durbin R. Fast and accurate short read alignment with Burrows–Wheeler transform. *Bioinformatics* 2009;**25**:1754–60. https://doi.org/10.1093/bioinformatics/btp324

Lou RN, Jacobs A, Wilder AP *et al.* A beginner's guide to low-coverage whole genome sequencing for population genomics. *Mol Ecol* 2021;**30**:5966–93. https://doi.org/10.1111/mec.16077

McFarlane GA, Beamish RJ. Climatic influence linking copepod production with strong year-classes in sablefish, *Anoplopoma fimbria*. *Can J Fish Aquat Sci* 1992;**49**:743–53. https://doi.org/10.1139/f92-083

Meisner J, Albrechtsen A. Inferring population structure and admixture proportions in low-depth NGS data. *Genetics* 2018;**210**:719–31. https://doi.org/10.1534/genetics.118.301336

Mérot C, Oomen RA, Tigano A *et al.* A roadmap for understanding the evolutionary significance of structural genomic variation. *Trends Ecol Evol* 2020;**35**:561–72.

Miller JM, Cullingham CI, Peery RM. The influence of *a priori* grouping on inference of genetic clusters: simulation study and literature review of the DAPC method. *Heredity* 2020;**125**:269–80. https://doi.org/10.1038/s41437-020-0348-2

Miller MP. Alleles in space (AIS): computer software for the joint analysis of interindividual spatial and genetic information. *J Hered* 2005;**96**:722–4. https://doi.org/10.1093/jhered/esi119

Moritz C. Defining 'evolutionarily significant units' for conservation. *Trends Ecol Evol* 1994;**9**:373–5.

Moser HG, Charter RL, Smith PE *et al.* Early life history of sablefish, *Anoplopoma fimbria*, off Washington, Oregon, and California, with application to biomass estimation. *Cal Coop Ocean Fisheries Invest Rep* 1994;**35**:144–59.

Neuwirth E. RColorBrewer: ColorBrewer Palettes. R package version 1.1-3, 2022. https://CRAN.R-project.org/package=RColorBrewer (19 December 2023, date last accessed).

Nielsen EE, Hemmer-Hansen JAKOB, Larsen PF *et al.* Population genomics of marine fishes: identifying adaptive variation in space and time. *Mol Ecol* 2009;**18**:3128–50. https://doi.org/10.1111/j.1365-294X.2009.04272.x

Nielsen EE, Kenchington E. A new approach to prioritizing marine fish and shellfish populations for conservation. *Fish Fish* 2001;**2**:328–43. https://doi.org/10.1046/j.1467-2960.2001.00055.x

Nowling RJ, Manke KR, Emrich SJ. Detecting inversions with PCA in the presence of population structure. *PLoS One* 2020;**15**:e0240429. https://doi.org/10.1371/journal.pone.0240429

Orlova SY, Schepetov DM, Mugue NS *et al.* Evolutionary history told by mitochondrial markers of large teleost deep-sea predators of family Anoplopomatidae Jordan & Gilbert 1883, endemic to the North Pacific. *J Mar Biol Assoc UK* 2019;**99**:1683–91.

Orozco-Ruiz AM, Galvan-Tirado C, Orlov AM *et al.* Genetic analyses reveal a non-panmictic genetic structure in the sablefish *Anoplopoma fimbria* in the northern Pacific. *ICES J Mar Sci* 2023;**80**:1319–28.

Paradis E. pegas: an R package for population genetics with an integrated–modular approach. *Bioinformatics* 2010;**26**:419–20. https://doi.org/10.1093/bioinformatics/btp696

Pritchard JK, Stephens M, Donnelly P. Inference of population structure using multilocus genotype data. *Genetics* 2000;**155**:945–59. https://doi.org/10.1093/genetics/155.2.945

Pujolar JM, Jacobsen MW, Als TD *et al.* Genome-wide single-generation signatures of local selection in the panmictic European eel. *Mol Ecol* 2014;**23**:2514–28. https://doi.org/10.1111/mec.12753

R Core Team. *R: A Language and Environment for Statistical Computing*. Vienna: R Foundation for Statistical Computing, 2022. https://www.R-project.org/

Rondeau EB, Messmer AM, Sanderson DS *et al.* Genomics of sablefish (*Anoplopoma fimbria*): expressed genes, mitochondrial phylogeny, linkage map and identification of a putative sex gene. *BMC Genom* 2013;**14**:1–19.

Rousset F. Genepop'007: a complete reimplementation of the Genepop software for Windows and Linux. *Mol Ecol Resour* 2008;**8**:103–6. https://doi.org/10.1111/j.1471-8286.2007.01931.x

Skotte L, Korneliussen TS, Albrechtsen A. Estimating individual admixture proportions from next generation sequencing data. *Genetics* 2013;**195**:693–702. https://doi.org/10.1534/genetics.113.154138

Therkildsen NO, Palumbi SR. Practical low-coverage genomewide sequencing of hundreds of individually barcoded samples for population and evolutionary genomics in nonmodel species. *Mol Ecol Resour* 2017;**17**:194–208. https://doi.org/10.1111/1755-0998.12593

Thia JA. Guidelines for standardizing the application of discriminant analysis of principal components to genotype data. *Mol Ecol Resour* 2023;**23**:523–38. https://doi.org/10.1111/1755-0998.13706

Thioulouse J, Dray S, Dufour A *et al. Multivariate Analysis of Ecological Data with ade4*. New York, NY: Springer, 2018. https://doi.org/10.1007/978-1-4939-8850-1

Thompson WF. A note on the spawning of the black cod. *Copeia* 1941;**1941**:270. https://doi.org/10.2307/1437492

Timm LE, Larson WA. Whole genome resequencing of sablefish at the northern end of their range. 2024. PRJNA1099646. https://www.ncbi.nlm.nih.gov/sra/PRJNA1099646 (12 April 2024, date last accessed).

Tripp-Valdez MA, García-de-León FJ, Espinosa-Pérez H *et al.* Population structure of sablefish *Anoplopoma fimbria* using genetic variability and geometric morphometric analysis. *J Appl Ichthyol* 2012;**28**:516–23. https://doi.org/10.1111/j.1439-0426.2012.01942.x

Tsuyuki H, Roberts E. Muscle protein polymorphism of sablefish from the eastern Pacific Ocean. *J Fish Res Board Can* 1969;**26**:2633–41. https://doi.org/10.1139/f69-255

Ulmo-Diaz G, Engman A, McLarney WO *et al.* Panmixia in the American eel extends to its tropical range of distribution: biological implications and policymaking challenges. *Evol Appl* 2023;**16**:1872–88. https://doi.org/10.1111/eva.13599

Vihtakari M. ggOceanMaps: plot data on oceanographic maps using 'ggplot2'. R package version 2.1.12, 2023. https://mikkovihtakari.github.io/ggOceanMaps (19 December 2023, date last accessed).

Waples RS. Separating the wheat from the chaff: patterns of genetic differentiation in high gene flow species. *J Hered* 1998;**89**:438–50. https://doi.org/10.1093/jhered/89.5.438

Warnes GR, Bolker B, Lumley T. gtools: various R programming tools. R package version 3.9.2, 2021. https://CRAN.R-project.org/package=gtools (19 December 2023, date last accessed).

Wickham H. Reshaping data with the reshape package. *J Stat Softw* 2007;**21**:1–20. http://www.jstatsoft.org/v21/i12/

Wickham H. *ggplot2: Elegant Graphics for Data Analysis*. New York, NY: Springer-Verlag, 2016. ISBN 978-3-319-24277-4 https://ggplot2.tidyverse.org (19 December 2023, date last accessed).

Wickham H. stringr: simple, consistent wrappers for common string operations. R package version 1.5.0. 2022. https://CRAN.R-project.org/package=stringr (19 December 2023, date last accessed).

Wickham H, Averick M, Bryan J *et al.* Welcome to the tidyverse. *J Open Source Software* 2019;**4**:1686. https://doi.org/10.21105/joss.01686

Wickham H, Bryan J. readxl: read excel files. R package version 1.4.3, 2023. https://CRAN.R-project.org/package=readxl (19 December 2023, date last accessed).

Wickham H, François R, Henry L *et al.* dplyr: a grammar of data manipulation. R package version 1.1.4, 2023. https://github.com/tidyverse/dplyr, https://dplyr.tidyverse.org (19 December 2023, date last accessed).

Zivanovic G, Arenas C, Mestres F. Individual inversions or their combinations: which is the main selective target in a natural population of *Drosophila subobscura*? *J Evol Biol* 2016;**29**:657–64. https://doi.org/10.1111/jeb.12800

Handling Editor: Kristi Miller-Saunders