

# Strong population differentiation in lingcod (*Ophiodon elongatus*) is driven by a small portion of the genome

Gary C. Longo<sup>1</sup>  | Laurel Lam<sup>2,3</sup> | Bonnie Basnett<sup>3</sup> | Jameal Samhour<sup>4</sup> | Scott Hamilton<sup>3</sup> | Kelly Andrews<sup>4</sup> | Greg Williams<sup>2</sup> | Giles Goetz<sup>5</sup> | Michelle McClure<sup>6,7</sup> | Krista M. Nichols<sup>4</sup> 

<sup>1</sup>NRC Research Associateship Program, Northwest Fisheries Science Center, National Marine Fisheries Service, National Oceanic and Atmospheric Administration, Seattle, WA, USA

<sup>2</sup>Pacific States Marine Fisheries Commission, Under contract to Northwest Fisheries Science Center, National Marine Fisheries Service, National Oceanic and Atmospheric Administration, Seattle, WA, USA

<sup>3</sup>Moss Landing Marine Laboratories, Moss Landing, CA, USA

<sup>4</sup>Conservation Biology Division, Northwest Fisheries Science Center, National Marine Fisheries Service, National Oceanic and Atmospheric Administration, Seattle, WA, USA

<sup>5</sup>UWJISAO, Under contract to Northwest Fisheries Science Center, National Marine Fisheries Service, National Oceanic and Atmospheric Administration, Seattle, WA, USA

<sup>6</sup>Fisheries Resource Analysis and Monitoring Division, Northwest Fisheries Science Center, National Marine Fisheries Service, National Oceanic and Atmospheric Administration, Seattle, WA, USA

<sup>7</sup>Pacific Marine Environmental Laboratory, National Oceanic and Atmospheric Administration, Seattle, WA, USA

## Correspondence

Gary C. Longo, Contracted to Conservation Biology Division, Northwest Fisheries Science Center, National Marine Fisheries Service, National Oceanic and Atmospheric Administration, 2725 Montlake Blvd E Seattle, WA 98112, USA.  
Email: garycharleslongo@gmail.com

## Funding information

NMFS Habitat Assessment Improvement Plan (HAIP), Grant/Award Number: #NA15-NMF4550353

## Abstract

Delimiting intraspecific genetic variation in harvested species is crucial to the assessment of population status for natural resource management and conservation purposes. Here, we evaluated genetic population structure in lingcod (*Ophiodon elongatus*), a commercially and recreationally important fishery species along the west coast of North America. We used 16,749 restriction site-associated DNA sequencing (RADseq) markers, in 611 individuals collected from across the bulk of the species range from Southeast Alaska to Baja California, Mexico. In contrast to previous population genetic work on this species, we found strong evidence for two distinct genetic clusters. These groups separated latitudinally with a break near Point Reyes off Northern California, and there was a high frequency of admixed individuals in close proximity to the break. F-statistics corroborate this genetic break between northern and southern sampling sites, although most loci are characterized by low  $F_{ST}$  values, suggesting high gene flow throughout most of the genome. Outlier analyses identified 182 loci putatively under divergent selection, most of which mapped to a single genomic region. When individuals were grouped by cluster assignment (northern, southern, and admixed), 71 loci were fixed between the northern and southern cluster, all of which were identified in the outlier scans. All individuals identified as admixed exhibited near 50:50 assignment to northern and southern clusters and were

This is an open access article under the terms of the Creative Commons Attribution License, which permits use, distribution and reproduction in any medium, provided the original work is properly cited.

© 2020 The Authors. *Evolutionary Applications* published by John Wiley & Sons Ltd

heterozygous for most fixed loci. Alignments of RADseq loci to a draft lingcod genome assembly and three other teleost genomes with chromosome-level assemblies suggest that outlier and fixed loci are concentrated on a single chromosome. Similar genomic patterns have been attributed to chromosomal inversions in diverse taxonomic groups. Regardless of the evolutionary mechanism, these results represent novel observations of genetic structure in lingcod and designate clear evolutionary units that could be used to inform fisheries management.

#### KEYWORDS

fisheries management, heterogeneous genomic divergence, latitudinal cline, Northeast Pacific Ocean, population genomics, RAD sequencing

## 1 | INTRODUCTION

Delineating accurate population boundaries is a critical component of harvest and protected species management; without biologically robust boundaries, estimates of population abundance, productivity, and other dynamics are very likely to be wrong. To delineate population boundaries, species assessments have typically relied on life history and demographic data (Haltuch et al., 2018). Although genetic studies provide an excellent mechanism for detecting population structure, the incorporation of genetic data into management and stock assessments has historically been slow (Hauser & Seeb, 2008; Reiss, Hoarau, Dickey-Collas, & Wolff, 2009; Waples, Punt, & Cope, 2008). This is partly due to the difficulty commonly associated with detecting and interpreting genetic population structure in marine species with large population sizes, high dispersal potential, and few barriers to gene flow. These characteristics, which result in generally low levels of genetic differentiation, make population genetic tools better suited to define populations in an evolutionary context. In contrast, fishery managers tend to be more interested in population differences on ecologically relevant scales (Waples & Gaggiotti, 2006). Genomic techniques (i.e., approaches that generate thousands to millions of polymorphic loci) have advanced the field of population genetics with increased power for detecting population structure (Bernatchez et al., 2017; Waples & Gaggiotti, 2006) and opened the door to identifying genetic markers associated with adaptive differences (Benestan et al., 2016; Sandoval-Castillo, Robinson, Hart, Strain, & Beheregaray, 2018). Despite these recent advances, the best available information on stock structure for most species harvested in commercial fisheries does not include insights from genomic studies (Bernatchez et al., 2017). Closing this gap in knowledge is a major need (Lynch, Methot, & Link, 2018) for species that have (a) geographic ranges that cross jurisdictional boundaries, (b) a stock status with higher uncertainty or greater management concern, and (c) ecological characteristics that may make them more likely to exhibit spatial population structure.

*Ophiodon elongatus*, Girard 1854 (lingcod) are a commercially and recreationally valuable species of groundfish, with landings and revenues within U.S. waters totaling 1,628 metric tons in 2016,

and an average growth of over 200 metric tons per year since 2010 (Haltuch et al., 2018). Lingcod are distributed across international and state borders in the eastern Pacific Ocean, and as such, stocks are assessed and managed primarily along the political boundaries at the international (Canada–United States–Mexico) and national levels (Oregon–California). Boundaries also exist separating the inside waters of the Salish Sea from the outside waters of the Pacific Ocean; the Salish Sea is further divided into U.S. and Canadian waters. Within the United States, separate northern and southern lingcod stock regions, divided at the Oregon–California border, were adopted by the Pacific Fishery Management Council and federal stock assessors. This delineation was implemented due to observed differences in commercial landings data and history of exploitation, rather than reflecting any distinct breaks in lingcod genetic structure, biology, or distribution (Haltuch et al., 2018).

Lingcod exhibit characteristics that can lead to spatial population structure due to several ecological factors, including (a) reproduction is characterized by migrations to relatively shallow, coastal areas; (b) early life-history stages inhabit nearshore and estuarine habitats; and (c) adult home ranges are relatively small with individuals generally exhibiting strong site fidelity (Hart, 1973; Jagielo, 1990; Low & Beamish, 1978; Phillips & Barraclough, 1977; Starr, O'Connell, & Ralston, 2004; Tolimieri, Andrews, Williams, Katz, & Levin, 2009). Previous population genetic studies on lingcod using allozymes, mtDNA, and microsatellites have yielded equivocal results but have primarily indicated high gene flow and a lack of population structure (Jagiello, Leclair, & Vorderstrasse, 1996; Leclair, Young, & Shaklee, 2006; Marko, Rogers-Bennett, & Dennis, 2007). While the most recent lingcod stock assessment suggests that both northern and southern stocks have been rebuilt following an overfished state in 1999, the southern stock has shown significantly slower recovery and is designated in a precautionary zone for the purpose of U.S. federal fisheries management (Haltuch et al., 2018). These disparate population trajectories have increased scientific and management interest in developing clear insight into lingcod population structure. Using genomic analyses to identify structure could increase the accuracy of future assessments and the effectiveness of fisheries management measures, as evidenced by Andrews et al. (2018).

The objective of this study was to evaluate population structure more comprehensively using modern genomic techniques combined with high-resolution spatial sampling throughout the bulk of the species range in order to provide updated information on population structure to aid fisheries management. To address this objective, we used restriction site-associated DNA sequencing (RADseq) to generate thousands of polymorphic loci that sample widely throughout the genome, increasing our power to detect population structure and to identify signatures of adaptive loci. We present these results in the context of a draft lingcod genome assembly to better understand where highly differentiated RADseq loci occur in the genome.

## 2 | MATERIALS AND METHODS

### 2.1 | Species description

Lingcod are benthic predators endemic to the North Pacific, ranging from the Gulf of Alaska to central Baja California, Mexico (Hart, 1973). Lingcod are sexually dimorphic, with females typically growing faster and attaining larger asymptotic sizes than males. Females also reach maturity at larger sizes, between 3 and 5 years of age, whereas males are smaller but reach maturity earlier, at 2 years of age (Cass, Beamish, & McFarlane, 1990; Miller & Geibel, 1973). Additionally, lingcod exhibit a latitudinal trend in growth, longevity, and size at maturity, where individuals from northern waters generally grow faster, live longer, and mature at larger sizes than individuals from southern regions (Lam, 2019). Iteroparous spawning occurs from late winter to early spring when females move from often deeper, offshore habitat into shallow, rocky habitat where they deposit up to 500,000 eggs in high current areas (Hart, 1973; Low & Beamish, 1978). After fertilization, males will guard clutches, sometimes from multiple females, until the eggs hatch in 6–8 weeks (King & Withler, 2005; Withler et al., 2004). Lingcod have a pelagic larval duration (PLD) of around 90 days, where the first two weeks are spent in surface waters followed by 2.5 months offshore epipelagically before recruiting to a suitable habitat (Hart, 1973; Phillips & Barraclough, 1977). Young typically recruit at around 4–6 cm in length to sandy, low relief habitat adjacent to rocky, high relief substrate, which is the preferred habitat of adults (Bassett, Lindholm, Garza, Kvitek, & Wilson-Vandenberg, 2018). Tagging studies have suggested that adults exhibit a high degree of site fidelity and a relatively small home range, although there are infrequent observations of long-range movement in some individuals (Jagiello, 1990; Starr et al., 2004; Tolimieri et al., 2009). There is some evidence that females make longer spawning migrations than males, moving from deeper, offshore habitat to shallower spawning habitat (Low & Beamish, 1978). This is supported by the relative abundance of large females caught in the offshore trawl fishery; however, the nature and extent of this suspected spawning migration are not well characterized (Jagiello, 1990).

### 2.2 | Sample collections

We chartered recreational fishing boats to collect lingcod using hook-and-line fishing gear on high relief rocky habitat at 25 ports from Southeast Alaska to Southern California. Sample collections out of each port typically occurred across 2–5 days. To ensure a thorough collection of lingcod across a wide range of age and size classes, both shallow (<60 m) and deep (60–170 m) nearshore and offshore reefs were targeted by our fishing efforts. Most of the individual samples analyzed in this study came from fin clips or gill tissues obtained from these collections. Supplemental tissue samples were provided by bottom fish trawl surveys conducted in lower-relief habitat by the Alaska Fisheries Science Center's Resource Assessment and Conservation Engineering (RACE) Division and the Northwest Fisheries Science Center's Fishery Resource Analysis and Monitoring (FRAM) Division, opportunistic port sampling, and via collaborators. In total, this study included samples from 42 sites encompassing most of the species range (Table 1). All tissue samples were stored in 95% ethanol. DNA was preferentially extracted from gill tissue, which consistently yielded higher quality and quantity DNA than fin clippings (data not shown). From a single male collected in Hood Canal (Salish Sea), gill and liver tissues were flash-frozen in liquid nitrogen for high molecular weight DNA extraction for genome sequencing; flash-frozen tissues were stored at  $-80^{\circ}\text{C}$  until DNA extraction. In our hook and line surveys, between 25 and 100 individuals were lethally sampled from each port for a related life history and demographic study (Lam, 2019); otherwise, fish were released after measurements and tissue sampling. Removal of gill tissue for sampling (~100 mg) did not appear to adversely affect individual lingcod in trials at the Seattle Aquarium. Lingcod collected via hook-and-line surveys were sexed externally by the presence (male) or absence (female) of a conical papilla. Additionally, we verified sex through visual inspection of gonads when lethally sampled. Lingcod were aged using the fin-ray method (Chilton & Beamish, 1982), which has been found to have the highest accuracy, readability, and minimal between-reader bias when compared to methods using other aging structures (Beamish & Chilton, 1977; Cass & Beamish, 1983; Claiborne, Rosenfield, Topping, Downs, & Tsou, 2016).

### 2.3 | RADseq library preparation, data filtering, and genotype calling

We extracted genomic DNA from 940 samples using the Qiagen DNAeasy Blood & Tissue 96 extraction kit (Qiagen, Inc) and then quantified extractions using a BioTek FLX800 Microplate Fluorescence Reader. Samples with extractions yielding  $<12.5$  ng/ $\mu\text{l}$  were not sequenced. Individual DNA concentrations for the 841 remaining samples were normalized to 12.5 ng/ $\mu\text{l}$  and 125 ng of starting material was used for RADseq library prep. We used the RADseq protocol of Ali et al. (2016) to construct libraries with the following specific details: Genomic DNA from each sample was digested with the restriction enzyme *Pst*I, libraries were sheared to 300–500 bp

**TABLE 1** Lingcod sampling site information for final analyses

Region	Site #	Site name	n	lat	long
GOA	1	South Montague	7	59.1815	-147.9763
GOA	2	East Kodiak	10	58.4054	-151.0168
GOA	3	West Chirikof	10	55.7728	-156.8757
AK	4	Nearshore Yakutat	41	59.5388	-139.9117
AK	5	ADFG stat area 375832	10	58.6167	-137.7500
AK	6	ADFG stat area 395800	16	58.3333	-139.1667
AK	7	ADFG stat area 385800	18	58.3333	-138.5833
AK	8	ADFG stat area 355702	8	57.1167	-136.0833
AK	9	Sitka	19	57.1614	-135.7697
AK	10	Beta Rock	18	55.6709	-133.7392
AK	11	Dall Island	17	55.1640	-133.3629
AK	12	Wolf Rocks	19	55.0157	-133.4981
CAN	13	Hecate Strait	7	53.6362	-131.0692
CAN	14	Queen Charlotte Sound	12	52.1842	-129.2015
CAN	15	Shellfish Section	14	48.9140	-125.4190
SS	16	San Juan Islands	11	48.4712	-123.0710
SS	17	Hood Canal	7	47.7155	-122.8790
SS	18	North Puget Sound	29	48.1415	-122.7057
SS	19	South Puget Sound	8	47.7041	-122.4821
WA	20	Cape Alava	17	48.1543	-124.7720
WA	21	Offshore La Push	16	48.0056	-125.3120
WA	22	Ocean Park	14	46.5353	-124.4240
OR	23	Cannon Beach	14	45.8400	-124.0219
OR	24	West Garibaldi	24	45.7167	-124.3524
OR	25	Stonewall Bank	27	44.5362	-124.4126
OR	26	Coos Bay	21	43.2535	-124.4637
OR	27	Mack Rock	16	42.2296	-124.4092
OR	28	Pt St George Reef	19	41.8230	-124.3624
N_CA	29	Reading Rock	9	41.3428	-124.1875
N_CA	30	Cape Mendocino	16	40.4731	-124.4932
N_CA	31	MacKerricher	9	39.4739	-123.8281
N_CA	32	Stewarts Pt	19	38.6116	-123.3797
N_CA	33	Offshore Pt Reyes	10	38.0606	-123.2999
C_CA	34	Farallon Islands	11	37.5042	-122.8582
C_CA	35	Big Sur	16	36.2082	-121.8235
C_CA	36	North Pt Conception	6	35.2333	-120.9204
S_CA	37	Santa Barbara Channel	10	34.2014	-119.6268
S_CA	38	Carrington Pt	14	34.0759	-120.0435
S_CA	39	Osbourne Bank	9	33.3636	-119.0547
S_CA	40	San Nicolas Island	18	33.3203	-119.4774
S_CA	41	San Diego	12	32.7735	-117.4687
BA	42	Colonet	3	30.7467	-116.5675

Note: General region of sampling site (GOA, Gulf of Alaska; AK, Alaska; CAN, Canada; SS, Salish Sea; WA, Washington; OR, Oregon; N\_CA, Northern California; C\_CA, Central California; S\_CA, Southern California; and BA, Baja California, Mexico), site number (#) and name, number of samples for each site (n), latitude (lat) and longitude (long) of sampling sites.

using a Qsonica sonicator (Newton, CT), and 100 bp paired-end sequencing was conducted in 24 lanes using the HiSeq 4000 (Illumina, Inc) at the University of Oregon Genomics and Cell Characterization Core Facility (GC3F).

We used *Stacks* (Catchen, Hohenlohe, Bassham, Amores, & Cresko, 2013) to discover and identify single nucleotide polymorphisms (SNPs) from raw sequence data. Briefly, raw sequence data were quality-filtered, demultiplexed, trimmed to 85 bp, and filtered for PCR clones using *process\_radtags* and *clone\_filter* in *Stacks* v1.47, using otherwise default parameters. Subsequent data processing of individual samples was carried out in *Stacks* v2.1. In order to determine the optimal *Stacks* parameters specific to our dataset that minimized erroneous splitting or lumping of loci while simultaneously yielding a high number of polymorphic loci, we followed methods outlined in Paris, Stevens, and Catchen (2017) and Rochette and Catchen (2017) and found 3 to be the optimal value both for the maximum number of bp differences between alleles in a sample (*-M*) and the maximum number of bp mismatches between sample loci (*-n*). After initial trimming and filtering for low-quality reads and PCR clones, loci were identified in each individual using *ustacks* with a minimum allele depth (*-m*) of 3. Individual samples with fewer than 3 million reads were excluded from downstream analyses based on an exploratory examination of genotyping rates versus read depth (data not shown). A catalog of consensus loci was constructed with *cstacks* using 84 individuals from throughout the range with high coverage and *sstacks* was used to match individual sample loci to the catalog. We then used *populations* to remove loci that failed to meet the following criteria: present in  $\geq 80\%$  of individuals, minor allele frequency  $\geq 1\%$ , and maximum observed heterozygosity of 70%. We exported the resulting SNP dataset from *Stacks* and further filtered using *VCftools* v.0.1.13 (Danecek et al., 2011). We then dropped all but the first SNP from each RADseq locus (*--thin* 5000), removed loci in individuals that were below 10x depth of coverage (*--minDP* 10), refiltered for loci found in  $\geq 80\%$  of individuals (*--max-missing* 0.8), and then removed individuals with  $>30\%$  missing loci from the final dataset (*--remove*), which was exported for downstream analyses. Finally, we identified and removed duplicate individuals using a custom R script written by Garrett McKinney (<https://github.com/gjmckinney/IDduplicateSamples>).

## 2.4 | Analyses

Our analyses focused on evaluating range-wide genetic differentiation in *lingcod*. It proceeded in three steps. First, we evaluated geographic patterns in genotypic differentiation using population genetic structure analyses to identify population clusters, with no a priori assumptions about population boundaries. Second, we calculated population genetic statistics for these clusters and for individual sampling site locations to quantify differentiation. Third, to better understand the underlying molecular basis and potential evolutionary mechanisms of the observed genetic differentiation, we conducted an outlier analysis, determined the genomic distribution

of outlier and fixed loci using a draft *lingcod* genome assembly together with three other teleost genomes with chromosome-level assemblies, and assessed mtDNA sequence diversity in a subset of individuals showing strong differentiation. These approaches are described in further detail below.

## 2.5 | Population structure analyses

To identify distinct genetic clusters, we first ran population structure analyses without any a priori assumptions about population boundaries. We used a model-based Bayesian clustering analysis implemented in *STRUCTURE* v2.3.4 (Pritchard, Stephens, & Donnelly, 2000) where ten replicates were run for each number of genetic clusters tested ( $K = 1-10$ ) each with a burn-in of 10,000 iterations and 100,000 MCMC reps with admixture allowed (NOADMIX = 0) and no prior location information (LOCPRIOR = 0). Next, we assessed the most likely number of clusters ( $K$ ) across replicate runs using the Evanno method (Evanno, Regnaut, & Goudet, 2005), which assesses the rate of change in log probability of the data between successive values of  $K$  ( $\Delta K$ ), as implemented in *Structure Harvester* (Earl & vonHoldt, 2012). We also used the mean likelihood of the model [ $L(K)$ ] for evaluating  $K$  as the Evanno method cannot detect a scenario of  $a = 1$ . Finally, *CLUMPP* v1.1.2 (Jakobsson & Rosenberg, 2007) was used to summarize results across replicate *STRUCTURE* runs and final plots were created using *DISTRUCT* v1.1 (Rosenberg, 2004). Average individual membership coefficients ( $Q$  values) to each cluster were taken from *CLUMPP* for cluster assignment. We also used principal component analysis (PCA) and discriminant analysis of principal components (DAPC), which do not rely on population genetic models, to summarize the diversity and variation across RADseq loci using the R package *adegenet* v2.1.1 (Jombart, 2008; Jombart et al., 2010). In the DAPC analysis, the optimal number of clusters ( $k$ ) was identified with *find.clusters*, which employs a Bayesian information criterion method. Preliminary results suggested strong clustering in *lingcod*, which appeared to be driven by outlier loci (outlier detection analyses described below). We re-evaluated structure patterns within initially identified clusters as well as with datasets excluding either neutral or outlier loci using *STRUCTURE* and PCAs. We also evaluated structure patterns based on sex and age when known.

## 2.6 | Population genetic analyses

To evaluate genetic differentiation, we calculated population genetic statistics with individuals grouped by sampling site location (42 sites) and then by the distinct cluster assignments from *STRUCTURE* analyses for a subset of analyses including cluster pairwise  $F_{ST}$  comparisons and locus-specific  $F$ -statistics (excluding admixed individuals). We tested each locus over the entire dataset for being in Hardy-Weinberg proportions using an exact test based on 1,000 Monte Carlo permutations of alleles as implemented in *pegas* v0.12 (Paradis, 2010) and exact test  $p$  values were adjusted for multiple

testing with a Benjamini–Hochberg (Benjamini & Hochberg, 1995) false discovery rate (FDR). Allele frequencies, observed heterozygosities, genetic diversity per locus and population, global and locus-specific  $F_{ST}$  statistics, and sampling sites pairwise  $F_{ST}$  (Weir & Cockerham, 1984) were calculated using the R package *HIERFSTAT* (Goudet, 2005). Using *boot.ppfst* and *boot.vc* in *HIERFSTAT*, we tested significance for pairwise  $F_{ST}$  comparisons and global  $F$ -statistics, respectively, when confidence intervals of 1,000 bootstrap replicates did not overlap with zero. Mean observed heterozygosity, mean genetic diversity within populations, and genetic diversity overall were also calculated using *HIERFSTAT*. Pairwise  $F_{ST}$  estimates were subsequently recalculated following identification and removal of candidate outlier loci. Finally, we estimated contemporary effective population size ( $N_e$ ) for *lingcod* with individuals grouped by *STRUCTURE* identified genetic clusters using *NeEstimator* v2.1 (Do et al., 2014) under the random mating model using the linkage disequilibrium (LD) method (Waples & Do, 2008) with a MAF cutoff of 0.05 and report jackknifed 95% confidence intervals.

## 2.7 | Outlier analyses

To better understand the nature of the observed genetic differentiation across sampling sites, we used two methods to detect loci putatively under divergent selection, *BayeScan* v2.1 (Foll & Gaggiotti, 2008) and the R package *pcadapt* (Luu, Bazin, & Blum, 2017). *BayeScan* uses a Bayesian regression approach to break down locus-population  $F_{ST}$  coefficients into a population-specific component and a global component. Loci are identified as outliers if the locus-specific component is required to explain the observed diversity (Foll & Gaggiotti, 2008). We set prior odds to 100 with 100,000 iterations and used a false discovery rate (FDR) of 1%. *pcadapt* uses a principal component analysis to identify loci as outliers while controlling for population structure (Luu et al., 2017). Here, we identified 1 as the optimal number of principal components to consider—denoted  $K$ —by visualizing the proportion of variance explained by each PC after testing  $K = 1$ –20 and then computed  $p$  values for each SNP. Next, we transformed the computed  $p$  values into  $q$  values with a specified FDR of 0.05, as recommended by Luu et al. (2017), using the R package *qvalue* (Storey, Bass, Dabney, & Robinson, 2018) to identify candidate outlier loci. Loci were considered candidates for divergent selection if they were identified in both *BayeScan* and *pcadapt*.

## 2.8 | Mitochondrial DNA sequencing

Because *STRUCTURE* results showed individuals strongly assigning to one of two clusters or as admixed individuals with near equal assignment to each cluster (and no observed back crossing), mitochondrial sequence divergence was evaluated a posteriori to assess the possibility of two cryptic species. Mitochondrial DNA is expected to accumulate differences relatively quickly due to a high mutation rate and

smaller  $N_e$  than nuclear DNA, making mitochondrial markers excellent for differentiating divergent species or populations. We sequenced a 440-bp segment of mitochondrial D-loop using the primers H16498 5' (CCTGAAGTAGGAACCAGATG) (Meyer, Kocher, Basasibwaki, & Wilson, 1990) and Pro-L (CTACCTCCAACCTCCCAAAGC) (Crow, Powers, & Bernardi, 1997) in 45 individuals total; 15 from each of the three cluster assignment scenarios (northern, southern, and admixed). D-loop sequences were edited and aligned using *Codon Code* software v4. We used *ape* v5.2 (Paradis & Schliep, 2019) to read alignments into R; *pegas* v0.12 (Paradis, 2010) was then used to call haplotypes and compute and construct haplotype networks. We quantified differentiation between clusters and then region sampling sites using Nei's  $G_{ST}$  (Nei & Chesser, 1983) and estimated 95% confidence intervals from 100 bootstrap replicates as implemented in *mmod* v1.3.3 (Winter, 2012).

## 2.9 | Genome sequencing and assembly

High molecular weight DNA was extracted from gill tissue (preserved in ethanol) of a coastal Washington male individual using the Qiagen midi-Prep kit (Qiagen, Inc); DNA from this individual was used for shotgun genome sequencing library prep and HiSeqX10 (150 base paired end) sequencing at the University of Washington Genome Sciences lab. High molecular weight DNA from a second individual (Hood Canal) was used for nanopore sequencing, as tissues from the first individual (stored in ethanol) did not yield the high molecular weight DNA ideal for long contiguous sequencing protocols, like that for nanopore (Oxford Nanopore Technologies, Oxford, UK). Both of these individuals were included in RADseq analyses and identified as northern cluster individuals. The Qiagen Midi-Prep (Qiagen, Inc) and the Circulomics Nanobind Tissue Big DNA (Circulomics, Baltimore, MD) kits were both used to extract high molecular weight DNA from the second individual. Several nanopore libraries were prepared from the same individual for multiple instances of sequencing, and they included (a) three libraries prepared using the LSK-108 library prep protocol (Oxford Nanopore Technologies, Oxford, UK) for DNA extracted with the Qiagen midi-prep kit; two of these used the Covaris G-tube (Woburn, MA) shearing following manufacturers protocol, and one library was prepared without shearing, (b) two libraries prepared using the Circulomics extracted DNA; one with the LSK-109 library prep kit (Oxford Nanopore Technologies), and one using the field sequencing kit (Oxford Nanopore Technologies). All libraries were sequenced on the minION using R9.4 flow cells, and raw data were basecalled using either *Albacore* or *minKnow* (Guppy basecaller) running on miniIT (all Oxford Nanopore Technologies). Genome assembly was first conducted only with the HiSeqX10 data. FastQC was used to evaluate the quality of the HiSeqX10 data, and data were then used for assembly using *MaSuRCA* v3.3.1 (Zimin et al., 2013), specifying an average insert size of 244 bases. A final genome assembly was conducted using both the HiSeqX10 and all nanopore data with *MaSuRCA* v3.3.3 with the Flye assembler (<https://github.com/alekseyzimin/masurca>). Assemblies were

evaluated for completeness of conserved eukaryotic genes using BUSCO v3.1.0 with the odb9 databases (Seppey, Manni, & Zdobnov, 2019), whereby the assembly was compared to the Actinopterygii core set of genes.

Analyses of homology among nonsalmonid teleost genomes with chromosomal-level assemblies have detected few inter-chromosomal rearrangements; however, intra-chromosomal reorganization is common (Pettersson et al., 2019; Rondeau et al., 2014). Here, we took advantage of high inter-chromosomal synteny in teleost to gain insight into how contigs in our draft lingcod genome may arrange in chromosomes. This information may result in clearer insight into whether lingcod RADseq outlier loci showed genomic patterns of colocalization, a more random distribution, or some intermediate scenario. We used *raGOO* (Alonge et al., 2019), a reference-guided contig ordering and orienting tool, to scaffold lingcod chromosome-scale pseudomolecules (CSPs) using closely related teleosts with chromosomal-level assemblies. We increased *-i*, the minimum grouping confidence score for localization, from the default of 0.2 to 0.5 in order to prevent low confidence contigs from localizing to CSPs (M. Alonge, personal communication). In consideration of potential reference bias on scaffolding results, we generated three sets of lingcod CSP assemblies using the following teleost reference genomes: *Sebastes schlegelii*, Korean rockfish (CNSA accession: CNP0000222, He et al., 2019); *Gasterosteus aculeatus*, three-spined stickleback (Dryad: <https://doi.org/10.5061/dryad.h7h32>, Peichel, Sullivan, Liachko, & White, 2017); and *Larimichthys crocea*, large yellow croaker (NCBI GenBank assembly accession: GCA\_000972845.2, Ao et al., 2015).

## 2.10 | Manhattan plots, LD heatmaps, and analysis of sex bias in markers

The RADseq contigs from the *Stacks* catalog were aligned to each of three the lingcod CSP assemblies, as well as to each reference genome, with *Bowtie2* v2.3.0 (Langmead & Salzberg, 2012) using the *--very-sensitive-local* parameter. For loci that mapped more than once, the position with highest alignment score was kept. Finally, locus-specific  $F_{ST}$  values generated from *HIERFSTAT* computations using 42 sampling sites were plotted to their corresponding alignment position on each lingcod CSP assembly and on each reference genome assembly using the R package *qqman* (Turner, 2018). For the lingcod CSP that contained most of the outlier loci in each of the three reference-guided assemblies, we estimated linkage disequilibrium (LD) among all loci using  $r^2$  as implemented in *PLINK* v1.9 (Chang et al., 2015) and plotted LD heat maps in R. To assess whether strong selection acting on the sex-determining chromosome could be driving the differentiation pattern, we searched for sex-biased RADseq alleles (i.e., alleles found in significantly different proportions between males and females) and aligned the sequences to the *Sebastes*-guided lingcod CSP assembly with the software *RADsex* (<https://github.com/RomainFeron/RadSex>) with the minimum coverage for a marker to be considered present in an individual (*-min\_cov*) set to

one. We chose to use this CSP assembly because of all the reference genomes used, it yielded the highest alignment rates of lingcod RADseq markers. We visualized *RADsex* results with the R package *radsex-vis* (<https://github.com/RomainFeron/RADSex-vis>).

## 3 | RESULTS

*Stacks* filtering parameters resulted in 443,489 initial SNPs. Using *VCFtools*, we retained only the first SNP on each marker, which reduced the dataset to 228,426 SNPs. Subsequent filtering for  $\geq 10\times$  coverage in each locus of every individual followed by filtering for loci found  $\geq 80\%$  of individuals resulted in 16,749 loci. After individuals with more than 30% missing loci and duplicate individuals were removed, 611 individuals remained (Table S1).

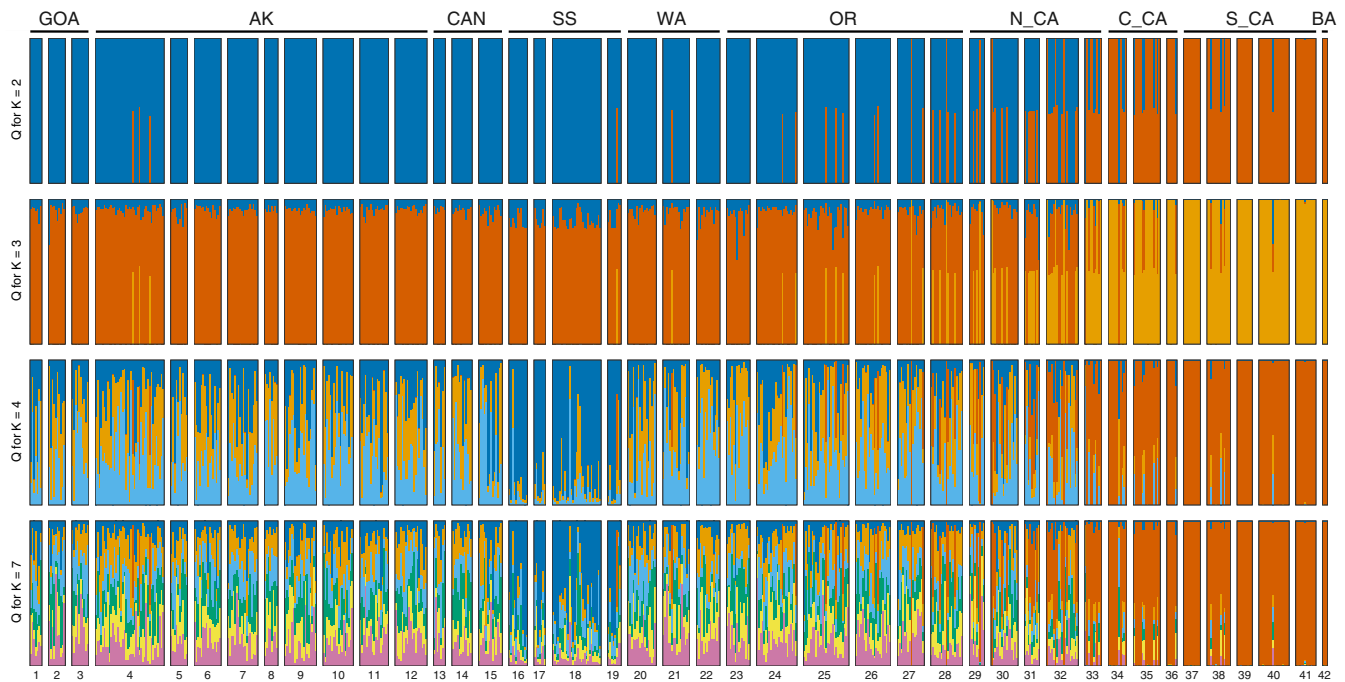
### 3.1 | Outlier analyses

We ran outlier analyses to detect loci potentially under selection and to understand how they may affect the observed genetic differentiation across sampling sites. We report these results first, as most downstream analyses relied on the identification of these loci. Out of 16,749 polymorphic SNPs, *BayeScan* and *pcadapt* identified 243 and 226 candidate outlier loci, respectively. Between detection methods, 182 outlier loci were shared and considered candidate loci under divergent selection.

### 3.2 | Population structure

Across replicate *STRUCTURE* runs with the full dataset, ( $\Delta K$ ) and [ $L(K)$ ] both found  $a = 2$  was the most likely scenario with results showing two distinct genetic clusters. Individuals either definitively assigned to one of the two clusters ( $Q > 0.99$ ), which segregated strongly latitudinally, or showed near equal assignment to both clusters ( $Q = 0.5 \pm 0.05$ ; Figure 1). As we detail below, cluster membership assignment is largely determined by nuclear haplotypes characterized by strongly linked, outlier loci; hereafter, we often refer to cluster haplotype assignments. The location at which the majority of individual memberships switch from one genetic cluster to another, or the break, appears to be centered off Northern California, where the frequency of northern cluster assignment goes from 0.605 at Stewarts Point to 0.3 offshore of Point Reyes (Figure 2; Table S2).

Individuals from the Gulf of Alaska to the northern coast of Oregon almost exclusively assign to the northern cluster haplotype except for 5 admixed individuals; three near Yakutat, AK and one individual each from south Puget Sound, WA, and off the coast of La Push, WA (Figures 1, 2; Table S2). The frequency of pure northern cluster individuals drops off precipitously in Northern California with the southernmost occurrence at the Farallon Islands west of San Francisco Bay (Figure 1; Table S2). Moving south from the Farallon



**FIGURE 1** Bayesian clustering results from *STRUCTURE* for all samples and loci. Plots are shown for the four most likely  $K$  clusters based on ( $\Delta K$ ):  $K_2 = 232.2$ ,  $K_3 = 3.4$ ,  $K_4 = 9.7$ , and  $K_7 = 4.5$ . Sampling sites (numbers below each plot; see Table 1 or Table S2 for site details) are arranged from the western Gulf of Alaska following the coast east and then south down to Baja California, Mexico: GOA Gulf of Alaska, AK Alaska, CAN Canada, SS Salish Sea, WA Washington, OR Oregon, N\_CA Northern California, C\_CA Central California, S\_CA Southern California, and BA Baja California, Mexico

Islands, the frequency of admixed individuals generally decreases, while the frequency of pure southern cluster individuals increases and reaches 100% at some sites. Pure southern cluster individuals are rare north of Point Reyes, CA, with the last observation at Mack Rock, OR. Admixed individuals are the most commonly observed genotype at two sampling sites in Northern CA: MacKerricher & Stewarts Point (Figure 1; Table S2).

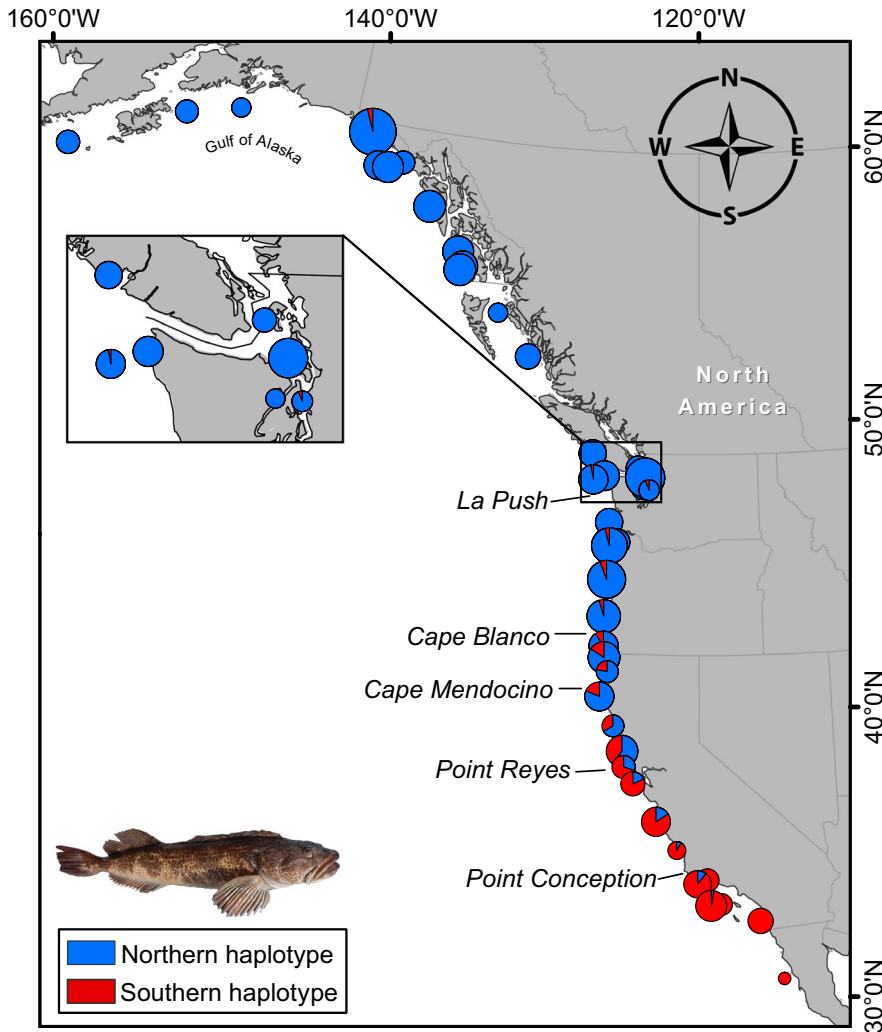
We repeated *STRUCTURE* analyses excluding outlier loci to determine whether these loci were responsible for the clustering we observed using the full dataset. This analysis resulted in the highest  $[L(K)]$  being  $K = 1$ —no structure—while ( $\Delta K$ ), which cannot detect a scenario of  $K = 1$ , suggested  $K = 3$  was most likely (Figure S1). This result implies that nonoutlier loci are not responsible for the strong clustering observed using the full dataset. In contrast, runs containing only outlier loci yielded ( $\Delta K$ ) and  $[L(K)]$  values that strongly suggested  $K = 2$  was the most likely number of genetic clusters (Figure S2), corroborating that these loci were indeed driving the pattern.

To investigate hierarchical population structure within the northern and southern clusters, we ran *STRUCTURE* analyses with only individuals belonging to each of the single clusters from the full dataset. Neither analysis suggested a scenario of  $K > 1$  based on  $[L(K)]$ . For the northern cluster, the highest likelihood ( $\Delta K$ ) was  $K = 3$ ; however, Salish Sea individuals showed signs of slight differentiation for all  $K$  values (Figure S3). In fact, the three *STRUCTURE* analyses that included all samples (i.e., all loci, excluding outlier loci, and exclusively based on outlier loci) showed Salish Sea individuals differentiated from other regions in most  $K > 2$  plots (Figure 1; Figures S1, S2). These

results imply that both neutral and outlier loci may differentiate Salish Sea individuals from other sites. For southern cluster individuals,  $\Delta K$  suggested  $K = 2$ ; however, there was no clear geographic pattern of differentiation among sampling sites or regions (Figure S4).

In the full dataset, PCA using all loci corroborated *STRUCTURE*  $K = 2$  clustering results by revealing three distinct groups separated along PC1 (explaining 0.91% of the variation; Figure 3a). The three groups perfectly match cluster assignment from *STRUCTURE* with northern and southern cluster individuals grouping separately and with admixed individuals falling out in the intermediate grouping. PCA without the outlier loci still shows some separation between individuals based on cluster assignment but the distinct grouping boundaries were lost (Figure 3b). A separate analysis with only the 182 outlier loci mirrored results using all loci, with PC1 explaining 61.7% of variation (Figure 3c). Within the northern and southern clusters, respectively, PCA also corroborated *STRUCTURE* analyses, where little separation was observed among regions except in the northern cluster where Salish Seas individuals showed some distinction along PC1 (Figure S5). In the DAPC analysis of population structure, ( $k = 2$ ) yielded the lowest BIC score (4,305.553), suggesting it was the most likely number of clusters. For this DAPC analysis, we retained 50 PCA axes and one discriminant function. Although the analysis groups admixed individuals with southern cluster individuals as opposed to northern cluster individuals, admixed individuals fall out between the northern and southern cluster groups along discriminant function 1 (Figure S6) as observed in the full dataset PCA along PC1 (Figure 3a). The population structure pattern of





**FIGURE 2** Distribution of northern (blue) and southern (red) lingcod cluster haplotype assignments based on  $K = 2$  genetic clustering results from *STRUCTURE* analyses with all loci. Transition from predominantly northern to southern cluster occurs near latitude  $\sim 38.3^\circ$ . The inset shows the inland waters of the Salish Sea

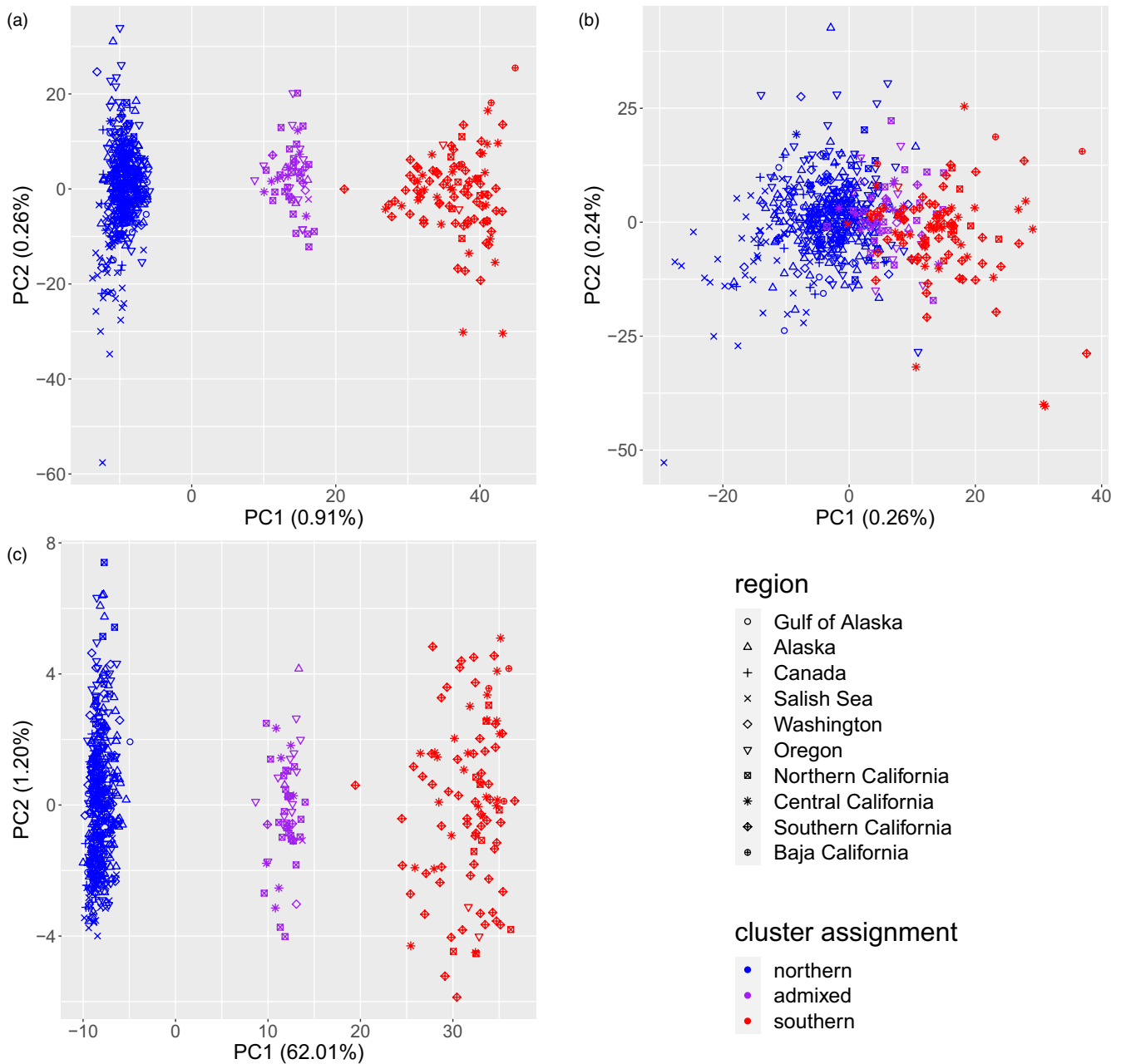
differentiation observed in the full dataset does not appear to be sex linked or related to age stratification, as northern, southern, and admixed individuals assigned roughly evenly to both sexes (Figure S7) and across age classes (Figure S8).

### 3.3 | Population genetics

Although we detected 509 loci out of Hardy-Weinberg proportions using an FDR-adjusted  $p$  value of  $<.05$ , we retained all loci for downstream analyses as these flagged markers contained most of the identified candidate outlier loci (135 out of 182). In order to estimate the strength of population differentiation, we calculated  $F$ -statistics for different groupings of samples with and without outlier loci.  $F$ -statistics analyses with individuals grouped by the 42 sampling sites yielded global  $F_{ST}$ ,  $F_{IS}$ , and  $F_{IT}$  values (with 95% confidence intervals) of 0.0115 (0.0098–0.0132), 0.0024 (0.0006–0.0043), and 0.0139 (0.0112–0.0165), respectively. Locus-specific  $F_{ST}$  varied from  $-0.0575$  to 0.8896, but most loci exhibited low values with a small subset exhibiting higher values (Figure S9; Table S3). Significant pairwise  $F_{ST}$  comparisons ranged from 0.0006 to 0.051 with the most differentiated comparison between north Puget Sound, WA, and Colonet, Baja

California (Table S4). Generally, pairwise  $F_{ST}$  comparisons yielded values much lower between sites on the same side of the latitudinal break ( $\sim 38.3^\circ N$ ) and relatively high values when comparing between sites on opposite sides of the break. When candidate outlier loci were excluded, global  $F_{ST}$ ,  $F_{IS}$ , and  $F_{IT}$  values (with 95% confidence intervals) were 0.0009 (0.0007–0.0011), 0.0021 (0.0002–0.0043), and 0.003 (0.0012–0.0052), respectively. Notably, global  $F_{ST}$  was over an order of magnitude lower, implying that most of the genome is characterized by low differentiation. Similarly, pairwise  $F_{ST}$  comparisons excluding outliers frequently showed values an order of magnitude lower between sites on the opposite side of the latitudinal break, while comparisons on the same side of the break decreased slightly, but generally remained similar to values calculated with outlier loci included (Table S5). However, pairwise comparisons between north Puget Sound and other northern sampling sites outside of the Salish Sea (both including and excluding candidate outlier loci) consistently produced low, but statistically significant  $F_{ST}$  values (Tables S4, S5). These results corroborate patterns of differentiation observed between Salish Sea sites and other regions in some *STRUCTURE* and PCA results (Figures 1, 3; Figures S1, S2, S3, S5a).

When  $F$ -statistics were estimated with individuals grouped by *STRUCTURE* identified clusters—excluding admixed individuals—as



**FIGURE 3** Principal components analysis for lingcod RADseq markers (a) including all loci (b) excluding outlier loci and (c) including only outlier loci with percent variation explained for PC1 and PC2. Distinct groupings seen in (a) and (c) coincide with cluster assignment results observed in *STRUCTURE*  $K = 2$

opposed to by sampling site, most locus-specific  $F_{ST}$  values were again low with the same subset of loci exhibiting even higher values, suggesting strong differentiation between the two groups (Table S6). Indeed, 71 loci were fixed between northern and southern clusters ( $F_{ST} = 1$ ) with admixed individuals being heterozygous at 69 of the 71 fixed sites (Table S7). All 71 fixed loci were included in the 182 outlier loci. Pairwise  $F_{ST}$  comparisons between cluster groups were orders of magnitude higher when calculated with all loci versus without outlier loci, respectively; northern and southern clusters (0.0446 & 0.0007), northern cluster and admixed individuals (0.0119 & 0.0002), and southern cluster and admixed individuals (0.0111 & 0/nonsignificant).

$N_e$  estimates were infinite or undefined for the southern cluster (95% confidence interval: 959 - infinite) and 29,237 (95% confidence interval: 7,643 - infinite) for the northern cluster. Although high, these point estimates presumably are downwardly biased to some extent by physical linkage, which depends on the taxa-specific number of chromosomes (Waples, Larson, & Waples, 2016). The admixed individuals likely represent an artificial population, as they appear to be heterozygous for nonrecombining haplotypes, and  $N_e$  estimates for this group were not considered meaningful (R. Waples, personal communication).

Mitochondrial D-loop haplotypes were shared across clusters and sampling locations (Figure 4) with no signs of haplotype fixation

expected with long-term reproductive isolation. Pairwise Nei's  $G_{ST}$  fixation indices were low between clusters (southern/northern 0.01178, southern/admixed 0.01464, and northern/admixed 0.02029), corroborating the low differentiation among cluster haplotypes visualized in the haplotype networks.

### 3.4 | Genome sequencing and assembly

From the HiSeqX10 data, ~979 million reads were obtained and assembled with *MaSuRCA*. This assembly estimated the genome size of lingcod to be ~900 Mb assembled into 345,190 contigs, with a resulting contig N50 of 4.1 kbp. Using *BUSCO* to evaluate the completeness of the initial assembly against the Actinopterygii core gene set, 75% of the gene set was represented in either complete or fragmented form, and 25% of the gene set was missing. From the Nanopore data, a total of ~5.4 million reads were obtained from multiple runs (see Table S8). Using *MaSuRCA* for a hybrid genome assembly with the HiSeqX10 and nanopore data resulted in 2,224 contigs and 2,175 scaffolds (bridged by the 17 $\times$  coverage nanopore data). The hybrid assembly had a contig N50 of 1,747 kbp. This final genome assembly, which was subsequently scaffolded with three related teleost genomes (using *raGOO*), contained 98% of the Actinopterygii conserved genes.

### 3.5 | Manhattan plots, LD heat maps, and analysis of sex bias in markers

To better understand the genomic distribution patterns of the outlier loci driving the strong differentiation patterns, we aligned all RADseq loci to three lingcod CSP assemblies, as well as to the respective teleost reference genomes. Alignment rates of RADseq

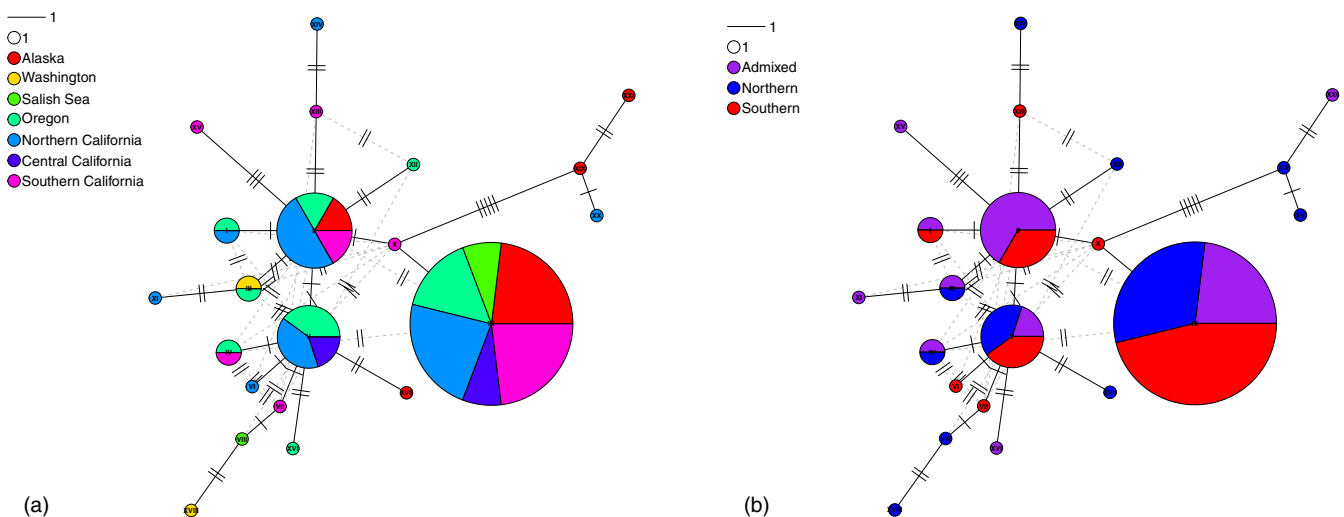
loci to lingcod CSP assemblies and the corresponding reference genomes (i.e., percentage of loci that aligned to lingcod CSPs and reference teleost chromosomes), respectively, were 97.4% and 66.9% for the Korean rockfish, *Sebastes schlegelii*, 93.3% and 50.4% for the three-spined stickleback, *Gasterosteus aculeatus*, and 91.7% and 50.4% for the large yellow croaker, *Larimichthys crocea*. In all three lingcod CSP assemblies, 176/182 (96.7%) outlier and 71/71 (100%) fixed loci aligned to a single CSP (Figure 5; Figure S10). Although alignment rate was expectedly lower, we observed similar patterns of outlier loci predominantly aligning to a single chromosome in the reference genomes as well (Figure S11, S12).

Linkage disequilibrium (LD) heat maps for lingcod loci on each CSP harboring the majority of outlier loci showed strong patterns of linkage in concentrated blocks corroborating outlier patterns seen in the Manhattan plots (Figure S13). Visual assessments of LD heat maps estimate the size of high LD to be ~15 Mb in length in each of the three reference-guided lingcod CSPs, suggesting the area of localized differentiation is relatively large.

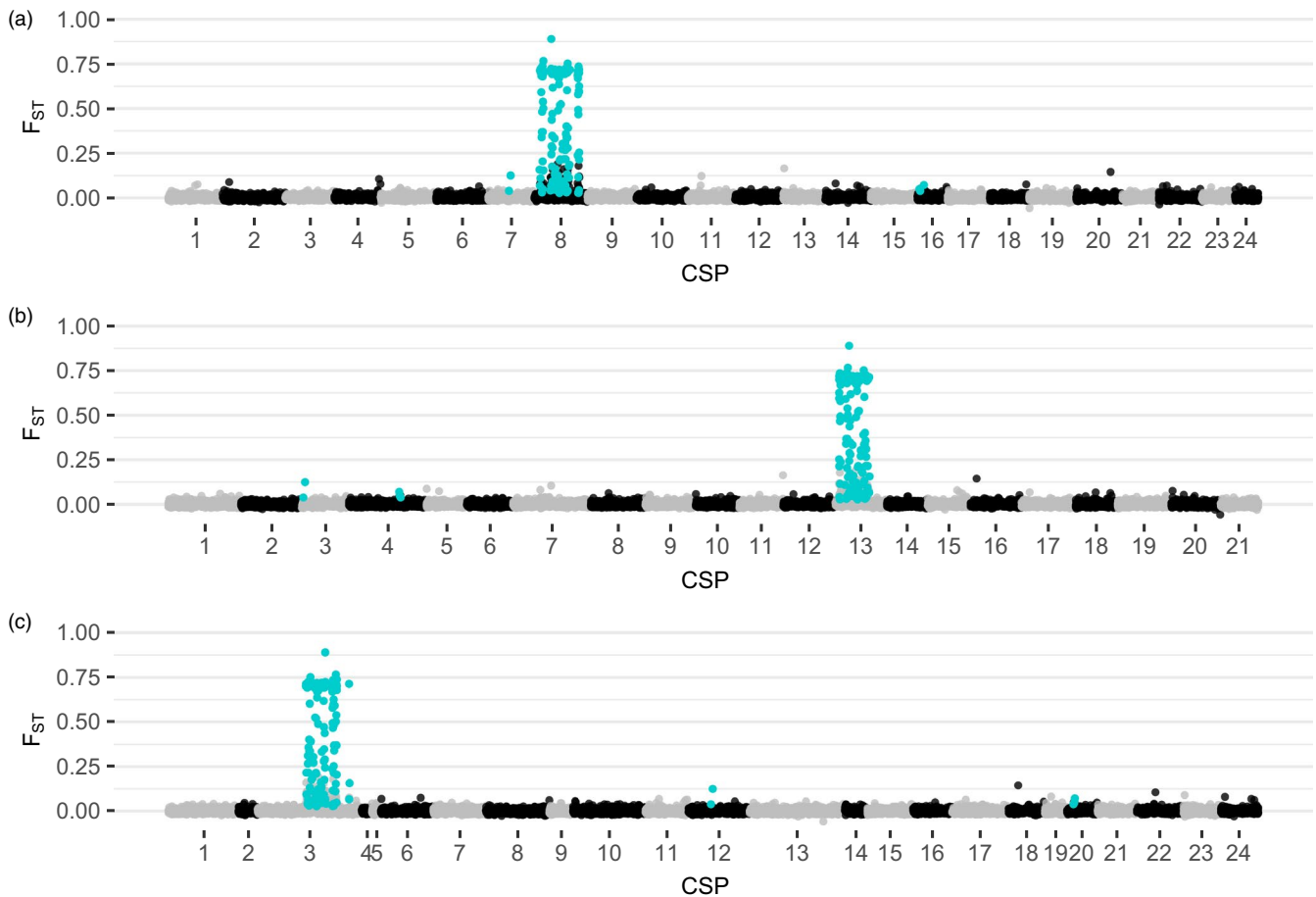
Sex-biased marker analysis revealed a localized region on lingcod CSP23, in the *Sebastes*-guided assembly, that exhibits a statistically significant association with sex (Figure S14). Notably, CSP8 in the *Sebastes*-guided assembly harbors the outlier and fixed loci driving the observed differentiation pattern.

## 4 | DISCUSSION

In this study, we identify two distinct genetic clusters of lingcod showing a latitudinal cline in the Eastern Pacific, with a genetic break centered off Northern California between Stewarts Point and Point Reyes. This novel delineation of population boundaries differs from previous lingcod population genetic studies, which have generally suggested coastwide panmixia (Jagiello et al., 1996; LeClair



**FIGURE 4** mtDNA D-loop haplotype networks using 45 individuals (15 northern cluster, 15 admixed, and 15 southern cluster) from different regions color coded by (a) region and (b) cluster assignment. Haplotype frequency is proportional to circle size with the smallest circle representing a single representation. Number of nucleotide differences separating haplotypes is shown by perpendicular strikes through connecting line



**FIGURE 5** Manhattan plots aligning RADseq contigs with locus-specific  $F_{ST}$  values based on 42 sampling sites to the draft lingcod genome assembly scaffolded into chromosomal-scale pseudomolecules (CSPs) using the genomes of (a) *Sebastes schlegelii*, Korean rockfish; (b) *Gasterosteus aculeatus*, three-spined stickleback; and (c) *Larimichthys crocea*, large yellow croaker, as reference guides. Outlier loci are highlighted in turquoise

et al., 2006; Marko et al., 2007). Compared to the single mitochondrial sequence or 19–32 nuclear markers used in these prior studies, the sampling of more sites across the genome with >16,000 SNPs empowered the ability to detect this signal of population structure. Genetic clustering analyses strongly assigned individuals to either a northern or southern cluster or as admixed, with admixed individuals roughly assigned evenly to both clusters with no evidence of backcrossed individuals. Pairwise  $F_{ST}$  comparisons corroborate strong differentiation between sampling sites located on opposite sides of the latitudinal break, while low and often nonsignificant differentiation was observed when comparing sites on the same side of the latitudinal break, with the exception of the Salish Sea. There, some sites show weak, yet significant differentiation from other northern sampling sites. The strong genetic structure between northern and southern clusters is driven by outlier loci, as the signal of differentiation largely disappears in analyses excluding these loci. Alignments of RADseq loci to the three reference-guided lingcod CSP assemblies revealed a clear pattern of outlier loci overwhelmingly aligning to a single CSP. These results point to evidence that these loci are likely colocalized on a single lingcod chromosome. A reasonable explanation for these observations given our data is a chromosomal

rearrangement such as an inversion, which have recently been detected in a number of other taxa (discussed below). However, additional genome sequencing data from a southern haplotype individual is needed to confirm that this structural variation exists.

#### 4.1 | Strong population differentiation is driven by outlier loci

Prior genetic studies in lingcod did not detect the strong population differentiation observed in the present study. Importantly, previous studies included large sample sizes from Central California, which would almost certainly include individuals identified in this study as southern cluster or admixed; however, no striking signals of differentiation were observed along the outer coast (Jagiello et al., 1996; LeClair et al., 2006; Marko et al., 2007). These discordant results are most likely due to the differences in the molecular markers used and the extent to which these markers sampled the genome in this species. For example, Jagiello et al. (1996) and LeClair et al. (2006) used allozyme or allozyme and microsatellite markers, respectively, which sample less than a few dozen sites in the nuclear genome, while Marko et al. (2007)

specifically evaluated population structure using cytochrome oxidase I in the mitochondrial genome. Sampling thousands of markers distributed widely throughout the genome allowed us to detect a very strong signal of divergence driven by a small number of outlier loci likely colocalized to a single region in the genome. In fact, when we excluded outlier loci from analyses our results corroborate previous findings of weak genetic structure, as does our mtDNA analysis using D-loop, which showed no obvious haplotype lineage sorting based on cluster assignment or sampling region. Although males exhibit territorial behavior as nest guards and are generally thought to move less than females (Stahl, Green, & Vaughn, 2014), we saw no evidence for bias in the distribution of sexes among genetic clusters (Figure S7); moreover, the putative sex CSP is different than the CSP harboring all of the outlier loci (Figure 5; Figure S14). Similarly, we did not detect bias in these results among age classes (Figure S8). These findings suggest cluster assignment is not sex dependent nor associated with a recruitment anomaly.

With the exception of Salish Sea sites, genetic clustering analyses within northern and southern clusters lack patterns of differentiation. Although Jagiello et al. (1996) and Marko et al. (2007) did not detect strong structure, both suggested weak genetic differentiation between the inside waters of the Salish Sea and the outer coast. However, LeClair et al. (2006) failed to find the same genetic signal with additional individuals from the Salish Sea and the increased power of microsatellite markers in conjunction with allozyme data from Jagiello et al. (1996). Although we observed major structure associated with outlier loci, we also detect weak signals of differentiation and substructure between lingcod in the Salish Sea and the Pacific Ocean both with and without outlier loci. Marko et al. (2007) estimated asymmetrical migration rates in this region with net emigration from the inner waters of the Salish Sea to the outer coast, which coincides with surface-water circulation patterns (Ebbesmeyer et al., 1984) and larval life-history behavior—lingcod larvae remain in surface waters for the first two weeks of their 90-day PLD (Phillips & Barraclough, 1977). With reduced connectivity due to the net outflow of surface water and migrants, the inland waters of the Salish Sea may rely on self-replenishment as suggested by Marko et al. (2007). That hypothesis is supported by our results, which show weak but significant population differentiation between the Salish Sea and outer coast.

## 4.2 | Possible mechanisms for differentiation

There are a number of evolutionary mechanisms that can lead to the strong differentiation observed between northern and southern cluster lingcod; here, we evaluate several possibilities before discussing evidence for a chromosomal inversion. It is conceivable that the northern and southern clusters represented cryptic species interbreeding and producing sterile F1 hybrids (admixed individuals); however, the lack of divergence in mtDNA and the heterogeneous genomic patterns do not support this alternate hypothesis. If pre- or postzygotic barriers prevented backcrossing of admixed individuals with pure cluster individuals, we would expect no gene

flow between clusters and therefore significant differentiation of mitochondrial haplotypes as well as strong patterns of differentiation throughout the entirety of the nuclear genome. Yet, we saw no evidence of D-loop haplotype segregation in northern or southern clusters (Figure 4b) corroborating the high levels of gene flow in the mitochondrial genome reported by Marko et al. (2007). Although patterns of mitochondrial and nuclear discordance have been observed between sister rockfish (*Sebastes*) species, a high proportion of haplotypes were found to segregate in these cases (Burford & Bernardi, 2008; Kai, Park, & Nakabo, 2012; Muto, Kai, & Nakabo, 2011; Narum, Buonaccorsi, Kimbrell, & Vetter, 2004). Additionally, these studies detected strong nuclear divergence using amplified fragment length polymorphism (AFLP) and microsatellite markers, which sample orders of magnitude fewer regions of the genome, suggesting the signal of nuclear divergence was widespread.

Another alternative hypothesis for the differentiation pattern we observed in lingcod is divergent selection in the absence of a chromosomal rearrangement, which is expected to produce a distinctive region of elevated divergence or “genomic islands of divergence.” However, islands derived from divergent selection are expected to be dispersed throughout the genome and on the order of kilobases not megabases (Kim & Stephan, 2002; Lamichhaney & Andersson, 2019; Nosil, Funk, & Ortiz-Barrientos, 2009). This scenario appears unlikely in lingcod where we consistently observed the colocalization of outlier and fixed loci to a region around 15 megabases in size on a single lingcod CSP across reference-guided assemblies (Figure 5; Figure S10).

Yet another possibility is that the observed clusters represent recently isolated populations characterized by strong sexual selection on the sex-determining chromosome (Martin et al., 2013; Nosil et al., 2009; Qvarnström & Bailey, 2009; Ravinet et al., 2017; Sæther et al., 2007), but this hypothesis is inconsistent with our findings as the outlier loci driving the strong differentiation patterns localize on CSP8 in the *Sebastes*-guided assembly (Figure 5a; Figure S10a) while the sex-determining markers localize on CSP23 (Figure S14). Interestingly, this result corroborates previous work on sex differentiation in lingcod that suggested males are the heterogametic sex (Rondeau, Laurie, Johnson, & Koop, 2016).

Chromosomal inversions, which have been detected at an increasing rate due to the proliferation of genomewide sequencing techniques, have been implicated in a number of recent studies evaluating population differentiation and the genetic basis of life histories (Berg et al., 2016; Knief et al., 2016; Lundberg et al., 2017; Sodeland et al., 2016; Wellenreuther & Bernatchez, 2018). Here in lingcod, we observed similar patterns detected across inversion studies: Highly localized genomic regions explain most observed variation, where individuals have strong affinities to discrete genetic groups. For example, the PCA pattern we observed in lingcod (Figure 3) has been seen among migratory and nonmigratory Atlantic cod, *Gadus morhua*, (Barth et al., 2019; Berg et al., 2016), Australian zebra finch, *Taeniopygia guttata*, (Knief et al., 2016), and the willow warbler, *Phylloscopus trochilus*, (Lundberg et al., 2017), where homokaryotypes for inverted and noninverted haplotypes,

respectively, grouped separately along a PC axis with heterokaryotypes (individuals heterozygous for the inverted and noninverted region) clustering between the respective homokaryotype groups. Lundberg et al. (2017) failed to differentiate among warblers with different inversion haplotypes when loci located on inversion blocks were removed. In lingcod, we observed partial overlap among individuals from separate clusters (i.e., northern, southern, and admixed) when we removed outliers, although some differentiation was still observed (Figure 3b). The distinct *STRUCTURE* pattern we observed for  $K = 2$  for all loci and outlier loci only analyses is also consistent with results seen in known inversion systems such as the Atlantic cod, where heterokaryotypic individuals assigned roughly equally to clusters homozygous for inverted and noninverted haplotypes, respectively (Kess et al., 2019). These patterns can result from systems where the colinear regions of the genome are generally characterized by high levels of gene flow, and consequentially low divergence and inversion haplotypes may be the genomic region driving differentiation patterns among individuals.

Genomic island of divergence can also be produced through localized recombination suppression between inverted and noninverted regions in the presence of otherwise unimpeded gene flow throughout the colinear regions (Nosil et al., 2009). We consistently observed a pattern of outlier loci colocalizing to a single genome region when lingcod RADseq markers were aligned to three reference-guided lingcod CSP assemblies. The pattern was also seen when aligning loci to the chromosome-level reference genome assemblies, which includes species that span three teleost orders. Additionally, the lingcod outlier markers show strong linkage disequilibrium (LD), a pattern also seen in diverse taxa where inversions have been found (Joron et al., 2011; Knief et al., 2016; Sinclair-Waters et al., 2018). Inter-chromosomal synteny appears to be well conserved in divergent teleosts lacking recent whole-genome duplications (Pettersson et al., 2019; Rondeau et al., 2014). Indeed, the fact that lingcod outlier loci overwhelmingly map to a single lingcod CSP and to a single chromosome in all three reference teleost genomes suggests a high level of chromosomal synteny and some conservation of intra-chromosomal syntenic blocks. We interpret this consistent pattern as evidence that genetic differentiation in lingcod is attributed to loci colocalized to a region, likely characterized by an inversion, on a single chromosome. Future work to construct a southern cluster derived genome assembly will be important to properly characterize the size, nature, and location of the putative inversion harboring these loci promoting genetic differentiation between the northern and southern genetic groups.

### 4.3 | Potential factors driving divergence and latitudinal cline

Lingcod exhibit a very clear latitudinal cline with respect to genetic cluster assignment (Figure 2). Endler (1977) postulates that clines can arise from one or more of four scenarios; random genetic drift, secondary contact between previously isolated populations, steep

or discontinuous differences in environmental conditions, and continuous change in the environment. Although these scenarios are not mutually exclusive, the evidence suggests that the distinct genetic cline in lingcod is most likely a product of selection driven by change (continuous and/or discontinuous) in environmental conditions. It is possible that genetic drift may have contributed to the establishment of the cline in the presence of a historical physical barrier; however, the cline appears to be maintained by selection. For one, outlier loci are almost entirely found on a single CSP. If drift were responsible for the clinal distribution of haplotype clusters, one would also expect other genomic regions to show differentiation patterns due to drift. However, most of the genome outside the region of differentiation shows signs of high gene flow, including in the mitochondrial genome. Although the realized larval dispersal may be less extensive than the potential dispersal potential (Lester & Ruttenberg, 2005; Lester, Ruttenberg, Gaines, & Kinlan, 2007), even low levels of gene flow, in the absence of selection, would rapidly homogenize cluster haplotype distributions and quickly erode the latitudinal cline (Slatkin, 1987). Secondary contact between previously isolated populations also is not supported; our lingcod D-loop haplotype networks based on cluster assignment showed no evidence of fixation or segregation nor do previous lingcod mitochondrial haplotype analyses (Marko et al., 2007). However, the pattern we observe in lingcod does suggest that cluster haplotypes, which are defined by the colocalized outlier loci, may carry a fitness component with respect to latitude. That is, northern cluster haplotypes appear to be advantageous in the northern portion of the range while southern cluster haplotypes appear to carry an advantage in the southern portion of the range. The California coastline, where we see the steepest change in frequency of cluster assignment, has also produced selection-driven population genetic breaks in marine invertebrates (Kelly & Palumbi, 2010; Sotka, Wares, Barth, Grosberg, & Palumbi, 2004) and fish (Sivasundar & Palumbi, 2010). In these examples, the selective pressures driving the patterns are unknown but the authors point out well-known differences in abiotic factors, such as sea surface temperature and upwelling (Checkley & Barth, 2009; Huyer, 1983; Magnell et al., 1990), and note that Central California represents a broad transition zone between northern and southern taxa with many species range limits found in the area (Allen, Horn, & Pondella, 2006). However, the effects of lingcod cluster haplotypes on fitness with respect to environment need to be formally tested and the presence of a chromosomal inversion needs to be demonstrated in future work.

### 4.4 | Fishery management implications

Our study sampled lingcod throughout their geographic range (with a limited sample from the waters of Mexico) and found evidence for two distinct genetic clusters with the break centered between Stewarts Point and Point Reyes, California. These results emphasize the need to consider transboundary stock assessments and to reconsider the location of the current stock assessment and

management boundary located at the U.S. Oregon–California border (Haltuch et al., 2018). Although most of the genome is characterized by low differentiation, suggesting significant levels of connectivity, the putative chromosomal inversion haplotypes are on different demographic trajectories and meet the criteria for distinct and independent stocks (Waples, 1998). Specifically, the putative inversion haplotypes are genetically isolated from each other and the strong latitudinal cline suggests they may be associated with adaptive differentiation. In managed species characterized by inversions, such as Atlantic cod (Kirubakaran et al., 2016), optimal strategies will aim to maintain haplotype polymorphisms that confer adaptive advantages (Bernatchez et al., 2017; Funk, McKay, Hohenlohe, & Allendorf, 2012). Although the adaptive nature of the putative lingcod inversion is unknown, there is more risk in managing discrete stocks as a single population as opposed to managing a panmictic population as distinct populations (Hauser, Adcock, Smith, Ramírez, & Carvalho, 2002; Kerr et al., 2017; Laikre, Palm, & Ryman, 2005; Okamoto et al., 2019; Spies & Punt, 2015; Sterner, 2007).

The southern lingcod stock (California), as it is currently assessed, has experienced slower recovery and greater fishing pressure than the northern stock (Haltuch et al., 2018). The southern stock consists of individuals from both the northern and southern genetic cluster, but is dominated by the southern cluster south of the Farallon Islands (Figure 2; Table S2). Estimates of effective population size ( $N_e$ ) are large for both clusters, indicating population sizes are large or were in the recent past. Although replenishment via dispersal of northern cluster larvae to Southern California seems highly plausible based on a 90-day PLD, the latitudinal cline suggests that the realized dispersal may be less extensive than potential dispersal (Lester & Ruttenberg, 2005; Lester et al., 2007), and/or the northern cluster individuals have lower fitness in Southern California. Signals of high gene flow in regions of the genome outside the putative inversion, excluding Salish Sea sites, and the steepness of the latitudinal cline suggest the latter plays a larger role. Regardless, it appears that Southern California lingcod predominantly rely on recruitment of the southern cluster haplotype. A management strategy that maintains southern cluster recruitment will likely be important for continued southern stock recovery. Within the northern cluster, the Salish Sea shows signs of weak but significant differentiation from outer coast sites suggesting relatively restricted connectivity. These results suggest that prudence should be taken in changing the current management boundaries separating the Salish Sea from the outer coast. More work is needed to elucidate how management of Salish Sea lingcod as a separate population could influence long-term sustainability. Taken together, international, national, and state management agencies may want to consider stock productivity and assessment modeling in the context of these genetic and other ecological factors that may contribute to differences in their productivity and survival.

Finally, a key quandary in drawing management boundaries for lingcod is the consideration of individuals that are admixed between the northern and southern genetic groups, and the historical and future location of these boundaries with changing ocean

conditions. Our working hypothesis is that these individuals are heterokaryotypic for an inversion on a single chromosome, and the fitness consequences for these individuals are uncertain. In lingcod, positive identification of the inversion and understanding the role putative heterokaryotypes play is an important future line of investigation. Specifically, understanding their influence on productivity and connectivity of the northern and southern clusters will reduce uncertainty in future assessment outputs. How cluster haplotypes may affect observed regional variation in growth rates, size at maturity, longevity, and mortality rates (Lam, 2019) is currently being investigated.

## 5 | SUMMARY

Across geopolitical boundaries of the eastern Pacific Ocean, we identified two notably distinctive genetic clusters of lingcod, with a strong genetic break centered off Northern California (~38.3°N latitude). This genetic break is driven entirely by a small number of loci in linkage disequilibrium that are likely colocalized on a single chromosome that may be characterized by an inversion. Future work is necessary to confirm the presence and nature of the putative inversion; if confirmed, additional work will need to evaluate the ecological life-history significance and fitness implications of such an inversion. We also found evidence for substructure within the northern cluster based on weak but significant differentiation between the inland waters of the Salish Sea and the outer coast that is likely driven by reduced connectivity between these areas. Current stock boundaries for management of the species differ somewhat from the patterns we observed. Our results imply that for lingcod, as with many species, an evaluation of how genetic population structure, jurisdictional boundaries, and life-history considerations come together to influence assessment accuracy is likely to provide important insights for management of the species.

## ACKNOWLEDGMENTS

Funding for this work was provided by a number of mechanisms within NOAA Fisheries, including the NMFS Habitat Assessment Improvement Plan (HAIP) (#NA15-NMF4550353) and through Cooperative Research Proposals with partners that included numerous fishing boat captains for the collection of samples. The authors thank Robin Waples for insightful discussions and assistance on  $N_e$  analyses; Romain Feron for assistance with RADsex analyses; and Giacomo Bernardi, Mike Ford, Melissa Haltuch, and Garrett McKinney for excellent discussions on analyses and interpreting results. We thank Kathryn Temple and the DFO Special Collections for providing Canadian tissue samples; Jay Orr and RACE sampling from the Gulf of Alaska, Mike Vaughn, and ADFG for allowing us to subsample; Armand Barilotti for donating tissue samples from a recreational fishing trip in Baja, Mexico; and Tim Carpenter at the Seattle Aquarium for allowing us to test sampling procedures. We thank the external reviewers for comments and suggestions that improved the quality of the manuscript. We also thank staff at the University of

Washington Northwest Genomics Center for facilitation HiSeqX10 genome sequencing, and Porecamp Vancouver for initial nanopore genome sequencing data. The work was approved on SJSU IACUC protocol #964. Collections were permitted by NOAA SRP-31-2016 and SRP-21-2017, CDFW SCP-6477, ODFW ORSTP #20237 and 21074, and WDFW permit SCP-16-138 and SCP-17-024.

## CONFLICT OF INTEREST

None declared.

## DATA AVAILABILITY STATEMENT

All sequences are openly available and deposited at NCBI. Demultiplexed, quality-filtered forward and reverse DNA sequence reads for lingcod individuals sequenced in this study, and the draft lingcod assembly are available at the sequence read archive (SRA) under the BioProject PRJNA595583 <http://www.ncbi.nlm.nih.gov/bioproject/PRJNA595583>. The mtDNA sequences can be found on GenBank with accession numbers MT498360-MT498404.

## ORCID

Gary C. Longo  <https://orcid.org/0000-0002-4282-9111>

Krista M. Nichols  <https://orcid.org/0000-0003-3453-7239>

## REFERENCES

- Ali, O. A., O'Rourke, S. M., Amish, S. J., Meek, M. H., Luikart, G., Jeffres, C., & Miller, M. R. (2016). RAD Capture (Rapture): Flexible and efficient sequence-based genotyping. *Genetics*, 202(2), 389–400. <https://doi.org/10.1101/028837>
- Allen, L., Horn, M., & Pondella, D. (2006). *The ecology of marine fishes: California and adjacent waters*. Berkeley, CA: University of California Press.
- Alonge, M., Soyk, S., Ramakrishnan, S., Wang, X., Goodwin, S., Sedlazeck, F. J., ... Schatz, M. C. (2019). RaGOO: Fast and accurate reference-guided scaffolding of draft genomes. *Genome Biology*, 20(1), 1–17. <https://doi.org/10.1186/s13059-019-1829-6>
- Andrews, K. S., Nichols, K. M., Elz, A., Tolimieri, N., Harvey, C. J., Pacunski, R., ... Tonnes, D. M. (2018). Cooperative research sheds light on population structure and listing status of threatened and endangered rockfish species. *Conservation Genetics*, 19(4), 865–878. <https://doi.org/10.1007/s10592-018-1060-0>
- Ao, J., Mu, Y., Xiang, L.-X., Fan, D. D., Feng, M. J., Zhang, S., ... Chen, X. (2015). Genome sequencing of the Perciform Fish *Larimichthys crocea* provides insights into molecular and genetic mechanisms of stress adaptation. *PLoS Genetics*, 11(4), 1–25. <https://doi.org/10.1371/journal.pgen.1005118>
- Barth, J. M. I., Villegas-Ríos, D., Freitas, C., Moland, E., Star, B., André, C., ... Jentoft, S. (2019). Disentangling structural genomic and behavioural barriers in a sea of connectivity. *Molecular Ecology*, 28, 1394–1411. <https://doi.org/10.1111/mec.15010>
- Bassett, M., Lindholm, J., Garza, C., Kvittek, R., & Wilson-Vandenberg, D. (2018). Lingcod (*Ophiodon elongatus*) habitat associations in California: Implications for conservation and management. *Environmental Biology of Fishes*, 101(1), 203–213. <https://doi.org/10.1007/s10641-017-0692-0>
- Beamish, R. J., & Chilton, D. (1977). Age determination of Lingcod (*Ophiodon elongatus*) using dorsal fin rays and scales. *Journal of the Fisheries Board of Canada*, 34(9), 1305–1313.
- Benestan, L., Quinn, B. K., Maaroufi, H., Laporte, M., Clark, F. K., Greenwood, S. J., ... Bernatchez, L. (2016). Seascape genomics provides evidence for thermal adaptation and current-mediated population structure in American lobster (*Homarus americanus*). *Molecular Ecology*, 25(20), 5073–5092. <https://doi.org/10.1111/mec.13811>
- Benjamini, Y., & Hochberg, Y. (1995). Controlling the false discovery rate: A practical and powerful approach to multiple testing. *Journal of the Royal Statistical Society. Series B (Methodological)*, 57(1), 289–300.
- Berg, P. R., Star, B., Pampoulie, C., Sodeland, M., Barth, J. M. I., Knutsen, H., ... Jentoft, S. (2016). Three chromosomal rearrangements promote genomic divergence between migratory and stationary ecotypes of Atlantic cod. *Scientific Reports*, 6(3), 1–12. <https://doi.org/10.1038/srep23246>
- Bernatchez, L., Wellenreuther, M., Araneda, C., Ashton, D. T., Barth, J. M. I., Beacham, T. D., ... Withler, R. E. (2017). Harnessing the power of genomics to secure the future of seafood. *Trends in Ecology and Evolution*, 32(9), 665–680. <https://doi.org/10.1016/j.tree.2017.06.010>
- Burford, M. O., & Bernardi, G. (2008). Incipient speciation within a sub-genus of rockfish (*Sebastosomus*) provides evidence of recent radiations within an ancient species flock. *Marine Biology*, 154(4), 701–717. <https://doi.org/10.1007/s00227-008-0963-6>
- Cass, A. J., & Beamish, R. J. (1983). First evidence of validity of the Fin-Ray method of age determination for marine fishes. *North American Journal of Fisheries Management*, 3(2), 182–188. [https://doi.org/10.1577/1548-8659\(1983\)3<182:feovot>2.0.co;2](https://doi.org/10.1577/1548-8659(1983)3<182:feovot>2.0.co;2)
- Cass, A. J., Beamish, R. J., & McFarlane, G. A. (1990). Lingcod (*Ophiodon elongatus*). *Canadian Journal of Fisheries and Aquatic Sciences*, 109.
- Catchen, J., Hohenlohe, P. A., Bassham, S., Amores, A., & Cresko, W. A. (2013). Stacks: An analysis tool set for population genomics. *Molecular Ecology*, 22(11), <https://doi.org/10.1111/mec.12354>
- Chang, C. C., Chow, C. C., Tellier, L. C. A. M., Vattikuti, S., Purcell, S. M., & Lee, J. J. (2015). Second-generation PLINK: Rising to the challenge of larger and richer datasets. *GigaScience*, 4(1), 1–16. <https://doi.org/10.1186/s13742-015-0047-8>
- Checkley, D. M., & Barth, J. A. (2009). Patterns and processes in the California Current System. *Progress in Oceanography*, 83(1–4), 49–64. <https://doi.org/10.1016/j.pocean.2009.07.028>
- Chilton, D. E., & Beamish, R. J. (1982). Age determination methods for fishes studied by the groundfish program at the Pacific Biological Station. *Canadian Special Publication of Fisheries and Aquatic Sciences*, 60, 34–38.
- Claiborne, A., Rosenfield, S., Topping, J., Downs, D., & Tsou, T. (2016). *Comparison of ages determined from vertebrae, dorsal fin rays, and otoliths in lingcod (Ophiodon elongatus)*. Oregon: Newport.
- Crow, K. D., Powers, D. A., & Bernardi, G. (1997). Evidence for multiple maternal contributors in nests of kelp greenling (*Hexagrammos decagrammus*, Hexagrammidae). *Copeia*, 1997(1), 9–15. <https://doi.org/10.2307/1447834>
- Danecek, P., Auton, A., Abecasis, G., Albers, C. A., Banks, E., DePristo, M. A., ... Durbin, R. (2011). The variant call format and VCFtools. *Bioinformatics*, 27(15), 2156–2158. <https://doi.org/10.1093/bioinformatics/btr330>
- Do, C., Waples, R. S., Peel, D., Macbeth, G. M., Tillett, B. J., & Ovenden, J. R. (2014). NeEstimator v2: Re-implementation of software for the estimation of contemporary effective population size (Ne) from genetic data. *Molecular Ecology Resources*, 14, 209–214. <https://doi.org/10.1111/1755-0998.12157>
- Earl, D. A., & vonHoldt, B. M. (2012). STRUCTURE HARVESTER: A website and program for visualizing STRUCTURE output and implementing the Evanno method. *Conservation Genetics Resources*, 4(2), 359–361. <https://doi.org/10.1007/s12686-011-9548-7>
- Ebbesmeyer, C. C., Coomes, C. A., Cox, J. M., Helseth, J. M., Hinchey, L. R., Cannon, G. A., & Barnes, C. A. (1984). Synthesis of current measurements in Puget Sound WA - Vol 3: circulation in puget sound



- an interpretation based on historical records of currents. NOAA Technical Memorandum NOS OMS 5.
- Endler, J. A. (1977). *Geographic variation, speciation, and clines, Monographs in Population Biology*. Princeton, NJ: Princeton University Press.
- Evanno, G., Regnaut, S., & Goudet, J. (2005). Detecting the number of clusters of individuals using the software STRUCTURE: A simulation study. *Molecular Ecology*, 14(8), 2611–2620. <https://doi.org/10.1111/j.1365-294X.2005.02553.x>
- Foll, M., & Gaggiotti, O. (2008). A genome-scan method to identify selected loci appropriate for both dominant and codominant markers: A Bayesian perspective. *Genetics*, 180(2), 977–993. <https://doi.org/10.1534/genetics.108.092221>
- Funk, W. C., McKay, J. K., Hohenlohe, P. A., & Allendorf, F. W. (2012). Harnessing genomics for delineating conservation units. *Trends in Ecology and Evolution*, 27(9), 489–496. <https://doi.org/10.1016/j.tree.2012.05.012>
- Goudet, J. (2005). HIERFSTAT, a package for R to compute and test hierarchical F-statistics. *Molecular Ecology Notes*, 5, 184–186. <https://doi.org/10.1111/j.1471-8278>
- Haltuch, M. A., Wallace, J., Akselrud, C. A., Nowlis, J., Lewis, A. K., Valero, J. L., ... Lam, L. (2018). 2017 Lingcod Stock Assessment. Portland, OR: Pacific Fishery Management Council.
- Hart, J. L. (1973). *Pacific fishes of Canada (Bulletin N)*. Ottawa, ON: Fisheries Research Board of Canada.
- Hauser, L., Adcock, G. J., Smith, P. J., Ramirez, J. H. B., & Carvalho, G. R. (2002). Loss of microsatellite diversity and low effective population size in an overexploited population of New Zealand snapper (*Pagrus auratus*). *Proceedings of the National Academy of Sciences*, 99(18), 11742–11747. <https://doi.org/10.1073/pnas.172242899>
- Hauser, L., & Seeb, J. E. (2008). Advances in molecular technology and their impact on fisheries genetics. *Fish and Fisheries*, 9(4), 473–486. <https://doi.org/10.1111/j.1467-2979.2008.00306.x>
- He, Y., Chang, Y., Bao, L., Yu, M., Li, R., Niu, J., ... Qi, J. (2019). A chromosome-level genome of black rockfish, *Sebastes schlegelii*, provides insights into the evolution of live birth. *Molecular Ecology Resources*, 19, 1309–1321. <https://doi.org/10.1111/1755-0998.13034>
- Huyer, A. (1983). Coastal upwelling in the California current system. *Progress in Oceanography*, 12(3), 259–284. [https://doi.org/10.1016/0079-6611\(83\)90010-1](https://doi.org/10.1016/0079-6611(83)90010-1)
- Jagiello, T. (1990). Movement of tagged lingcod *Ophiodon elongatus* at Neah Bay. *Washington Fishery Bulletin*, 88(4), 815–820.
- Jagiello, T. H., Leclair, L. L., & Vorderstrasse, B. A. (1996). Genetic variation and population structure of lingcod. *Transactions of the American Fisheries Society*, 125(3), 372–386. [https://doi.org/10.1577/1548-8659\(1996\)125<0372](https://doi.org/10.1577/1548-8659(1996)125<0372)
- Jakobsson, M., & Rosenberg, N. A. (2007). CLUMPP: A cluster matching and permutation program for dealing with label switching and multimodality in analysis of population structure. *Bioinformatics*, 23(14), 1801–1806. <https://doi.org/10.1093/bioinformatics/btm233>
- Jombart, T. (2008). ADEGENET: A R package for the multivariate analysis of genetic markers. *Bioinformatics*, 24(11), 1403–1405. <https://doi.org/10.1093/bioinformatics/btn129>
- Jombart, T., Devillard, S., Balloux, F., Falush, D., Stephens, M., Pritchard, J., ... Nei, M. (2010). Discriminant analysis of principal components: A new method for the analysis of genetically structured populations. *BMC Genetics*, 11(1), 94. <https://doi.org/10.1186/1471-2156-11-94>
- Joron, M., Frezal, L., Jones, R. T., Chamberlain, N. L., Lee, S. F., Haag, C. R., ... French-Constant, R. H. (2011). Chromosomal rearrangements maintain a polymorphic supergene controlling butterfly mimicry. *Nature*, 477(7363), 203–206. <https://doi.org/10.1038/nature10341>
- Kai, Y., Park, K. D., & Nakabo, T. (2012). The incomplete history of mitochondrial lineages between two rockfishes, *Sebastes longispinis* and *Sebastes hubbsi* (Scorpaeniformes: Scorpaenidae). *Journal of Fish Biology*, 81(3), 954–965. <https://doi.org/10.1111/j.1095-8649.2012.03326.x>
- Kelly, R. P., & Palumbi, S. R. (2010). Genetic structure among 50 species of the northeastern Pacific rocky intertidal community. *PLoS One*, 5(1), <https://doi.org/10.1371/journal.pone.0008594>
- Kerr, L. A., Hintzen, N. T., Cadrin, S. X., Clausen, L. W., Dickey-Collas, M., Goethel, D. R., ... Nash, R. D. M. (2017). Lessons learned from practical approaches to reconcile mismatches between biological population structure and stock units of marine fish. *ICES Journal of Marine Science*, 74(6), 1708–1722. <https://doi.org/10.1093/icesjms/fsw188>
- Kess, T., Bentzen, P., Lehnert, S. J., Sylvester, E. V. A., Lien, S., Kent, M. P., ... Bradbury, I. R. (2019). A migration-associated supergene reveals loss of biocomplexity in Atlantic cod. *Science Advances*, 5, eaav2461. <https://doi.org/10.1101/361691>
- Kim, Y., & Stephan, W. (2002). Detecting a local signature of genetic hitchhiking along a recombining chromosome. *Genetics*, 160(2), 765–777.
- King, J. R., & Withler, R. E. (2005). Male nest site fidelity and female serial polyandry in lingcod (*Ophiodon elongatus*, Hexagrammidae). *Molecular Ecology*, 14(2), 653–660. <https://doi.org/10.1111/j.1365-294X.2005.02438.x>
- Kirubakaran, T. G., Grove, H., Kent, M. P., Sandve, S. R., Baranski, M., Nome, T., ... Andersen, Ø. (2016). Two adjacent inversions maintain genomic differentiation between migratory and stationary ecotypes of Atlantic cod. *Molecular Ecology*, 25(10), 2130–2143. <https://doi.org/10.1111/mec.13592>
- Knief, U., Hemmrich-Stanisak, G., Wittig, M., Franke, A., Griffith, S. C., Kempnaers, B., & Forstmeier, W. (2016). Fitness consequences of polymorphic inversions in the zebra finch genome. *Genome Biology*, 17(1), 1–22. <https://doi.org/10.1186/s13059-016-1056-3>
- Laikre, L., Palm, S., & Ryman, N. (2005). Genetic population structure of fishes: Implications for coastal zone management. *Ambio*, 34(2), 111–119. <https://doi.org/10.1579/0044-7447-34.2.111>
- Lam, L. S. (2019). Geographic and habitat-based variation in lingcod (*Ophiodon elongatus*) demography and life-history along the U.S. west coast. Capstone Projects and Master's Theses, 657. Retrieved from [https://digitalcommons.csumb.edu/caps\\_thes\\_all/657](https://digitalcommons.csumb.edu/caps_thes_all/657)
- Lamichhaney, S., & Andersson, L. (2019). A comparison of the association between large haplotype blocks under selection and the presence/absence of inversions. *Ecology and Evolution*, 9(8), 4888–4896. <https://doi.org/10.1002/ece3.5094>
- Langmead, B., & Salzberg, S. L. (2012). Fast gapped-read alignment with Bowtie 2. *Nature Methods*, 9(4), 357–359. <https://doi.org/10.1038/nmeth.1923>
- LeClair, L. L., Young, S. F., & Shaklee, J. B. (2006). Allozyme and microsatellite DNA analyses of lingcod from Puget sound, Washington, and adjoining waters. *Transactions of the American Fisheries Society*, 135(6), 1631–1643. <https://doi.org/10.1577/T05-116.1>
- Lester, S. E., & Ruttenberg, B. I. (2005). The relationship between pelagic larval duration and range size in tropical reef fishes: A synthetic analysis. *Proceedings of the Royal Society B: Biological Sciences*, 272(1563), 585–591. <https://doi.org/10.1098/rspb.2004.2985>
- Lester, S. E., Ruttenberg, B. I., Gaines, S. D., & Kinlan, B. P. (2007). The relationship between dispersal ability and geographic range size. *Ecology Letters*, 10(8), 745–758. <https://doi.org/10.1111/j.1461-0248.2007.01070.x>
- Low, C. J., & Beamish, R. J. (1978). *A study of the nesting behavior of lingcod (Ophiodon elongatus) in the Strait of Georgia, British Columbia* (p. 843). Can Fish Mar Serv Tech Rep.
- Lundberg, M., Liedvogel, M., Larson, K., Sigeman, H., Grahn, M., Wright, A., ... Bensch, S. (2017). Genetic differences between willow warbler migratory phenotypes are few and cluster in large haplotype blocks. *Evolution Letters*, 1(3), 155–168. <https://doi.org/10.1002/evl3.15>
- Luu, K., Bazin, E., & Blum, M. G. B. (2017). pcadapt: An R package to perform genome scans for selection based on principal component

- analysis. *Molecular Ecology Resources*, 17(1), 67–77. <https://doi.org/10.1111/1755-0998.12592>
- Lynch, P. D., Methot, R. D., & Link, J. S. (2018). Implementing a Next Generation Stock Assessment Enterprise. An Update to the NOAA Fisheries Stock Assessment Improvement Plan. U.S. Dep. Commer., NOAA Tech. Memo. NMFS-F/SPO-183. <https://doi.org/10.7755/TMSPO.183>
- Magnell, B. A., Bray, N. A., Winant, C. D., Greengrove, C. L., Largier, J., Borchardt, J. F., Dorman, C. E. (1990). Convergent shelf flow at Cape Mendocino. *Oceanography*, 4, 4–11. <https://doi.org/10.5670/oceanog.1990.14>
- Marko, P. B., Rogers-Bennett, L., & Dennis, A. B. (2007). MtDNA population structure and gene flow in lingcod (*Ophiodon elongatus*): Limited connectivity despite long-lived pelagic larvae. *Marine Biology*, 150(6), 1301–1311. <https://doi.org/10.1007/s00227-006-0395-0>
- Martin, S. H., Dasmahapatra, K. K., Nadeau, N. J., Salazar, C., Walters, J. R., Simpson, F., ... Jiggins, C. D. (2013). Genome-wide evidence for speciation with gene flow in *Heliconius* butterflies. *Genome Research*, 23(11), 1817–1828. <https://doi.org/10.1101/gr.159426.113>
- Meyer, A., Kocher, T. D., Basasibwaki, P., & Wilson, A. C. (1990). Monophyletic origin of Lake Victoria cichlid fishes suggested by mitochondrial DNA sequences. *Nature*, 347, 550–553. <https://doi.org/10.1038/347550a0>
- Miller, D. J., & Geibel, J. J. (1973). Summary of blue rockfish and lingcod life histories, a reef ecology study, and giant kelp, *Macrocystis pyrifera*, experiments in Monterey Bay, California. *Fish Bulletin*, 158.
- Muto, N., Kai, Y., & Nakabo, T. (2011). Genetic and morphological differences between *Sebastes vulpes* and *S. zonatus* (Teleostei: Scorpaeniformes: Scorpaenidae). *Fishery Bulletin*, 109(4), 429–439.
- Narum, S. R., Buonaccorsi, V. P., Kimbrell, C. A., & Vetter, R. D. (2004). Genetic divergence between Gopher Rockfish (*Sebastes carnatus*) and Black and Yellow Rockfish (*Sebastes chrysomelas*). *Copeia*, 2004(4), 926–931.
- Nei, M., & Chesser, R. K. (1983). Estimation of fixation indices and gene diversities. *Annals of Human Genetics*, 47(3), 253–259. <https://doi.org/10.1111/j.1469-1809.1983.tb00993.x>
- Nosil, P., Funk, D. J., & Ortiz-Barrientos, D. (2009). Divergent selection and heterogeneous genomic divergence. *Molecular Ecology*, 18(3), 375–402. <https://doi.org/10.1111/j.1365-294X.2008.03946.x>
- Okamoto, D. K., Hessing-Lewis, M., Samhuri, J. F., Shelton, A. O., Stier, A., Levin, P. S., & Salomon, A. K. (2019). Spatial variation in exploited metapopulations obscures risk of collapse. *Ecological Applications*, 30(3), 1–16. <https://doi.org/10.1002/eap.2051>
- Paradis, E. (2010). Pegas: An R package for population genetics with an integrated-modular approach. *Bioinformatics*, 26(3), 419–420. <https://doi.org/10.1093/bioinformatics/btp696>
- Paradis, E., & Schliep, K. (2019). Ape 5.0: An environment for modern phylogenetics and evolutionary analyses in R. *Bioinformatics*, 35(3), 526–528. <https://doi.org/10.1093/bioinformatics/bty633>
- Paris, J. R., Stevens, J. R., & Catchen, J. M. (2017). Lost in parameter space: A road map for STACKS. *Methods in Ecology and Evolution*, 8(10), 1360–1373. <https://doi.org/10.1111/2041-210X.12775>
- Peichel, C. L., Sullivan, S. T., Liachko, I., & White, M. A. (2017). Improvement of the Threespine Stickleback genome using a Hi-C-based proximity-guided assembly. *Journal of Heredity*, 108(6), 693–700. <https://doi.org/10.1093/jhered/esx058>
- Petterson, M. E., Rochus, C. M., Han, F., Chen, J., Hill, J., Wallerman, O., ... Andersson, L. (2019). A chromosome-level assembly of the Atlantic herring genome-detection of a supergene and other signals of selection. *Genome Research*, 29(11), 1919–1928. <https://doi.org/10.1101/gr.253435.119>
- Phillips, A. C., & Barraclough, W. E. (1977). On the early life history of lingcod (*Ophiodon elongatus*). *Can Fish Mar Serv Tech Rep*, 756, 35.
- Pritchard, J. K., Stephens, M., & Donnelly, P. (2000). Inference of population structure using multilocus genotype data. *Genetics*, 155(2), 945–959.
- Qvarnström, A., & Bailey, R. I. (2009). Speciation through evolution of sex-linked genes. *Heredity*, 102(1), 4–15. <https://doi.org/10.1038/hdy.2008.93>
- Ravinet, M., Faria, R., Butlin, R. K., Galindo, J., Bierne, N., Rafajlović, M., ... Westram, A. M. (2017). Interpreting the genomic landscape of speciation: A road map for finding barriers to gene flow. *Journal of Evolutionary Biology*, 30(8), 1450–1477. <https://doi.org/10.1111/jeb.13047>
- Reiss, H., Hoarau, G., Dickey-Collas, M., & Wolff, W. J. (2009). Genetic population structure of marine fish: Mismatch between biological and fisheries management units. *Fish and Fisheries*, 10(4), 361–395. <https://doi.org/10.1111/j.1467-2979.2008.00324.x>
- Rochette, N. C., & Catchen, J. M. (2017). Deriving genotypes from RAD-seq short-read data using STACKS. *Nature Protocols*, 12(12), 2640–2659. <https://doi.org/10.1038/nprot.2017.123>
- Rondeau, E. B., Laurie, C. V., Johnson, S. C., & Koop, B. F. (2016). A PCR assay detects a male-specific duplicated copy of Anti-Müllerian hormone (*amh*) in the lingcod (*Ophiodon elongatus*). *BMC Research Notes*, 9(1), 230. <https://doi.org/10.1186/s13104-016-2030-6>
- Rondeau, E. B., Minkley, D. R., Leong, J. S., Messmer, A. M., Jantzen, J. R., von Schalburg, K. R., ... Koop, B. F. (2014). The genome and linkage map of the northern pike (*Esox lucius*): Conserved synteny revealed between the salmonid sister group and the neoteleostei. *PLoS One*, 9(7), e102089. <https://doi.org/10.1371/journal.pone.0102089>
- Rosenberg, N. A. (2004). DISTRUCT: A program for the graphical display of population structure. *Molecular Ecology Notes*, 4(1), 137–138. <https://doi.org/10.1046/j.1471-8286.2003.00566.x>
- Saether, S. A., Saetre, G.-P., Borge, T., Wiley, C., Svedin, N., Andersson, G., ... Qvarnstrom, A. (2007). Sex chromosome-linked species recognition and evolution of reproductive isolation in flycatchers. *Science*, 318(5847), 95–97. <https://doi.org/10.1126/science.1141506>
- Sandoval-Castillo, J., Robinson, N. A., Hart, A. M., Strain, L. W. S., & Beheregaray, L. B. (2018). Seascape genomics reveals adaptive divergence in a connected and commercially important mollusc, the greenlip abalone (*Haliotis laevis*), along a longitudinal environmental gradient. *Molecular Ecology*, 27(7), 1603–1620. <https://doi.org/10.1111/mec.14526>
- Seppely, M., Manni, M., & Zdobnov, E. M. (2019). BUSCO: Assessing genome assembly and annotation completeness. In: M. Kollmar (Ed.), *Gene prediction. Methods in molecular biology*, vol. 1962. New York, NY: Humana.
- Sinclair-Waters, M., Bradbury, I. R., Morris, C. J., Lien, S., Kent, M. P., & Bentzen, P. (2018). Ancient chromosomal rearrangement associated with local adaptation of a postglacially colonized population of Atlantic Cod in the northwest Atlantic. *Molecular Ecology*, 27(2), 339–351. <https://doi.org/10.1111/mec.14442>
- Sivasundar, A., & Palumbi, S. R. (2010). Life history, ecology and the biogeography of strong genetic breaks among 15 species of Pacific rockfish. *Marine Biology*, 157(7), 1433–1452. <https://doi.org/10.1007/s00227-010-1419-3>
- Slatkin, M. (1987). Gene flow and the geographic structure of natural populations. *Science*, 236(4803), 787–792.
- Sodeland, M., Jorde, P. E., Lien, S., Jentoft, S., Berg, P. R., Grove, H., ... Knutsen, H. (2016). “Islands of Divergence” in the Atlantic Cod Genome represent polymorphic chromosomal rearrangements. *Genome Biology and Evolution*, 8(4), 1012–1022. <https://doi.org/10.1093/gbe/evw057>
- Sotka, E. E., Wares, J. P., Barth, J. A., Grosberg, R. K., & Palumbi, S. R. (2004). Strong genetic clines and geographical variation in gene flow in the rocky intertidal barnacle *Balanus glandula*. *Molecular Ecology*, 13(8), 2143–2156. <https://doi.org/10.1111/j.1365-294X.2004.02225.x>
- Spies, I., & Punt, A. E. (2015). The utility of genetics in marine fisheries management: A simulation study based on Pacific cod off Alaska. *Canadian Journal of Fisheries and Aquatic Sciences*, 72(9), 1415–1432. <https://doi.org/10.1139/cjfas-2014-0050>

- Stahl, J., Green, K., & Vaughn, M. (2014). Examination of Lingcod, *Ophiodon elongatus*, Movements in Southeast Alaska Using Traditional Tagging Methods. Game, Fishery Data Series, No. 14–28.
- Starr, R. M., O'Connell, V., & Ralston, S. (2004). Movements of lingcod (*Ophiodon elongatus*) in southeast Alaska: Potential for increased conservation and yield from marine reserves. *Canadian Journal of Fisheries and Aquatic Sciences*, 61(7), 1083–1094. <https://doi.org/10.1139/f04-054>
- Sterner, T. (2007). Unobserved diversity, depletion and irreversibility: The importance of subpopulations for management of cod stocks. *Ecological Economics*, 61(2–3), 566–574. <https://doi.org/10.1016/j.ecolecon.2006.05.015>
- Storey, J. D., Bass, A. J., Dabney, A., & Robinson, D. (2018). qvalue: Q-value estimation for false discovery rate control.
- Tolimieri, N., Andrews, K., Williams, G., Katz, S., & Levin, P. S. (2009). Home range size and patterns of space use by lingcod, copper rockfish and quillback rockfish in relation to diel and tidal cycles. *Marine Ecology Progress Series*, 380, 229–243. <https://doi.org/10.3354/meps07930>
- Turner, S. D. (2018). qqman: An R package for visualizing GWAS results using Q-Q and manhattan plots. *Journal of Open Source Software*, 3(25), 731. <https://doi.org/10.21105/joss.00731>
- Waples, R. S. (1998). Separating the wheat from the chaff: Patterns of genetic differentiation in high gene flow species. *Journal of Heredity*, 89(5), 438–450. <https://doi.org/10.1093/jhered/89.5.438>
- Waples, R. S., & Do, C. (2008). LDNE: A program for estimating effective population size from data on linkage disequilibrium. *Molecular Ecology Resources*, 8(4), 753–756. <https://doi.org/10.1111/j.1755-0998.2007.02061.x>
- Waples, R. S., & Gaggiotti, O. (2006). What is a population? An empirical evaluation of some genetic methods for identifying the number of gene pools and their degree of connectivity. *Molecular Ecology*, 15(6), 1419–1439. <https://doi.org/10.1111/j.1365-294X.2006.02890.x>
- Waples, R. K., Larson, W. A., & Waples, R. S. (2016). Estimating contemporary effective population size in non-model species using linkage disequilibrium across thousands of loci. *Heredity*, 117(4), 233–240. <https://doi.org/10.1038/hdy.2016.60>
- Waples, R. S., Punt, A. E., & Cope, J. M. (2008). Integrating genetic data into management of marine resources: How can we do it better? *Fish and Fisheries*, 9(4), 423–449. <https://doi.org/10.1111/j.1467-2979.2008.00303.x>
- Weir, B. S., & Cockerham, C. C. (1984). Estimating F-statistics for the analysis of population structure. *Evolution*, 38(6), 1358–1370. <https://doi.org/10.2307/2409936>
- Wellenreuther, M., & Bernatchez, L. (2018). Eco-evolutionary genomics of chromosomal inversions. *Trends in Ecology & Evolution*, 33(6), 427–440. <https://doi.org/10.1016/j.tree.2018.04.002>
- Winter, D. J. (2012). MMod: An R library for the calculation of population differentiation statistics. *Molecular Ecology Resources*, 12(6), 1158–1160. <https://doi.org/10.1111/j.1755-0998.2012.03174.x>
- Withler, R. E., King, J. R., Marliave, J. B., Beath, B., Li, S., Supernault, K. J., & Miller, K. M. (2004). Polygamous mating and high levels of genetic variation in lingcod, *Ophiodon elongatus*, of the Strait of Georgia, British Columbia. *Environmental Biology of Fishes*, 69(1–4), 345–357. <https://doi.org/10.1023/B:EBFI.0000022898.49007.bc>
- Zimin, A. V., Marçais, G., Puiu, D., Roberts, M., Salzberg, S. L., & Yorke, J. A. (2013). The MaSuRCA genome assembler. *Bioinformatics*, 29(21), 2669–2677. <https://doi.org/10.1093/bioinformatics/btt476>

## SUPPORTING INFORMATION

Additional supporting information may be found online in the Supporting Information section.

**How to cite this article:** Longo GC, Lam L, Basnett B, et al. Strong population differentiation in lingcod (*Ophiodon elongatus*) is driven by a small portion of the genome. *Evol Appl.* 2020;13:2536–2554. <https://doi.org/10.1111/eva.13037>