*Molecular Ecology*

WILEY

MOLECULAR ECOLOGY

ORIGINAL ARTICLE OPEN ACCESS

# Crossing the Pacific: Genomics Reveals the Presence of Japanese Sardine (*Sardinops melanosticta*) in the California Current Large Marine Ecosystem

Gary C. Longo[1] | Jeremiah J. Minich[2] | Nicholas Allsing[2] | Kelsey James[3] | Ella S. Adams-Herrmann[1,4,5] | Wes Larson[6] | Nolan Hartwick[2] | Tiffany Duong[2] | Barbara Muhling[3,7] | Todd P. Michael[2] | Matthew T. Craig[3]

[1]National Marine Fisheries Service, Southwest Fisheries Science Center, Ocean Associates, Inc., Under Contract to the National Oceanic and Atmospheric Administration, La Jolla, California, USA | [2]The Plant Molecular and Cellular Biology Laboratory, The Salk Institute for Biological Studies, La Jolla, California, USA | [3]National Marine Fisheries Service, Southwest Fisheries Science Center, National Oceanographic and Atmospheric Administration, La Jolla, California, USA | [4]University of San Diego, San Diego, California, USA | [5]University of Central Florida, Department of Biology, Orlando, FL, USA | [6]National Oceanographic and Atmospheric Administration, National Marine Fisheries Service, Alaska Fisheries Science Center, Auke Bay Laboratories, Juneau, Alaska, USA | [7]Institute of Marine Sciences Fisheries Collaborative Program, University of California, Santa Cruz, Santa Cruz, California, USA

**Correspondence:** Gary C. Longo (garycharleslongo@gmail.com)

**Received:** 25 May 2024 | **Accepted:** 12 September 2024

**Handling Editor:** Michael M. Hansen

**Funding:** The authors received no specific funding for this work.

**Keywords:** dispersal | fisheries | genome | lcWGS | Pacific Sardine | range shift

## ABSTRACT

Recent increases in frequency and intensity of warm water anomalies and marine heatwaves have led to shifts in species ranges and assemblages. Genomic tools can be instrumental in detecting such shifts. In the early stages of a project assessing population genetic structure in Pacific Sardine (*Sardinops sagax*), we detected the presence of Japanese Sardine (*Sardinops melanosticta*) along the west coast of North America for the first time. We assembled a high quality, chromosome-scale reference genome of the Pacific Sardine and generated low coverage, whole genome sequence (lcWGS) data for 345 sardine collected in the California Current Large Marine Ecosystem (CCLME) in 2021 and 2022. Fifty individuals sampled in 2022 were identified as Japanese Sardine based on strong differentiation observed in lcWGS SNP and full mitogenome data. Although we detected a single case of mitochondrial introgression, we did not observe evidence for recent hybridization events. These findings change our understanding of *Sardinops* spp. distribution and dispersal in the Pacific and highlight the importance of long-term monitoring programs.

## 1 | Introduction

Spanning the waters from Vancouver Island, Canada, to Baja California Sur, Mexico, the California Current Large Marine Ecosystem (CCLME) is one of the four major eastern boundary upwelling systems across the globe. These systems are extraordinarily productive due to high nutrient fluxes towards surface waters that are driven by wind induced upwelling. This productivity has a profound effect on fisheries. Fueled by the abundance of species at the base of the food web, including many planktivorous coastal pelagic species such as sardines and anchovies, and despite only comprising ~1% of ocean waters, eastern boundary upwelling systems may produce up to 20% of the global catch (Pauly and Christensen 1995; Pikitch et al. 2014). Systemic perturbations causing a reduction in the productivity of the CCLME and other upwelling systems are of

particular concern as climate change continues on its current trajectory (Bakun et al. 2015).

Increasing evidence shows that climate change induced range shifts are pervasive across systems (Parmesan and Yohe 2003; Pinsky et al. 2013; Lenoir et al. 2020). However, marine species are experiencing faster range shifts and tracking range edges better than terrestrial species, which may be due to reduced barriers to dispersal and stricter physiological constraints (Fredston et al. 2021; Lenoir et al. 2020). In some instances, marine range shifts are ephemeral in nature, such as the rapid but temporary poleward expansion of California market squid (*Doryteuthis opalescens*), which was attributed to increased populations in conjunction with climate-associated temperature increase and oxygen loss that constrained aerobic activity (Burford et al. 2022). Other shifts appear more stable, such as the appearance of Atlantic Bluefin Tuna (*Thunnus thynnus*) in 2012 as bycatch in the seasonal Atlantic Mackerel (*Scomber scombrus*) fishery in East Greenland waters, hundreds of kilometres north of their previously documented summer feeding ground following increased abundance and warmer oceanographic conditions (Jansen et al. 2021). Most marine species range shifts track isotherms to higher latitudes (i.e., poleward) or to greater depths, however complex hydrographic conditions can result in species moving in other directions as well (Pinsky et al. 2013; Pinsky, Selden, and Kitchel 2020; Dulvy et al. 2008). Certain life-history traits, such as ecological generalism, short generation time, high adult mobility, and long pelagic larval duration (PLD), predispose taxa for fast responses to shifting isotherms (Pinsky, Selden, and Kitchel 2020), making many small pelagic fishes excellent indicators for climate driven changes (Peck et al. 2013).

As a key member of forage fish assemblages, sardine (*Sardinops* spp.) are distributed globally with two species occurring in the North Pacific. The Pacific Sardine (*Sardinops sagax*) occurs in the eastern Pacific from the coast of Chile to the Gulf of Alaska (Love et al. 2021; Robertson and Allen 2024) and the Japanese Sardine (*Sardinops melanosticta*) occurs in the Northwest Pacific and the Bering Sea (Dyldin et al. 2022). Cold and warm temperature boundaries in the higher and lower latitudes of the North Pacific, respectively, have been thought to act as effective dispersal barriers and may have contributed to the divergence of Japanese Sardine and Pacific Sardine (Bowen and Grant 1997; Grant, Clark, and Bowen 1998). Anthropogenic climate change, however, has led to increased and persistent warm water temperature anomalies or marine heatwaves (MHWs; Werb and Rudnick 2023), providing the potential to alter species distributions (Pinsky, Selden, and Kitchel 2020; Wilson et al. 2016). Indeed, a recent MHW (2014–2016) in the North Pacific resulted in unprecedented shifts in fish species abundances and assemblages in the CCLME (Gold et al. 2023; Thompson et al. 2022). Increases in the frequency and duration of MHWs, along with other shorter periods of anomalously warm sea surface temperature (SST) over the past two decades, may be setting the stage for distributional shifts of marine organisms including coastal pelagic fish species such as sardine, a group with high ecological and historically economic importance.

In the early stages of a project designed to assess population genetic structure in Pacific Sardine (*S. sagax*) using low coverage whole genome sequence (lcWGS) data, we observed two, highly distinct genetic groups off the west coast of North America that differed at levels normally observed between species. Upon further examination, it was determined that the dataset indeed represented the two phenotypically similar sardine species that occur in the North Pacific (*S. sagax* and *S. melanosticta*). Herein, we present the first chromosome level reference genome for Pacific Sardine, *S. sagax*, report the presence of Japanese Sardine, *S. melanostica*, in the CCLME for the first time, and provide hypotheses for their dispersal into the eastern Pacific Ocean. The analysis and results of population genetic structure in Pacific Sardine will be reported in a forthcoming publication.

## 2 | Materials and Methods

### 2.1 | Sample Collection

To generate a high-quality reference genome, post mortem blood samples were taken from eight Pacific Sardine, which were sourced from the Everingham Brothers bait barge in Mission Bay, California. Whole blood was placed on dry ice immediately after collection and stored at −80°C. For lcWGS, samples were primarily collected over the course of the 2021 and 2022 California Current Ecosystem Surveys (CCES) conducted by the SWFSC. The CCES runs from late June through late September or early October and is typically conducted aboard a NOAA fishery survey vessel using a Nordic 264 trawl (see Dorval et al. 2022; Renfree et al. 2022, 2023, for a summary of methods). Tissue samples (caudal muscle) were taken at sea immediately after capture and stored in 100% ethanol. Samples were collected from Tillamook, OR, USA, to Ensenada, Baja California, MX, in 2021 and from Cape Mendocino, CA, USA, to Punta Colonet, Baja California, MX, in 2022. Additional samples were obtained in 2022 from Long Beach, CA, USA for a forthcoming temporal study and from Bahía Magdalena, Baja California, MX, through collaboration with Instituto Mexicano de Investigación en Pesca y Acuacultura Sustentables. Following detection of Japanese Sardine in the lcWGS data, CCES samples from 2013 to 2023 were sequenced in an ad hoc GTseq (Campbell, Harmon, and Narum 2015) panel targeting fixed differences between Japanese and Pacific Sardine mitochondrial genome haplotypes as a means of understanding the potential impact to past biomass estimates of Pacific Sardine (see Supporting Information for details).

### 2.2 | Reference Genome

#### 2.2.1 | High Molecular Weight DNA Extraction, Library Preparation, and Sequencing

Approximately 20 μL of whole blood was thawed on ice and processed through a high molecular weight DNA extraction protocol. Specifically, genomic DNA was extracted using the Nanobind Nucleated blood protocol (Circulomics: EXT-NBU-001), which uses a variety of Circulomics kits (Circulomics CBB Big DNA kit: NB-900-001-01; Circulomics UHMW DNA Aux kit: NB-900-101-01). DNA was allowed to rest at 4°C for 1 week to allow DNA to go into solution and then quantified in triplicate (top, middle, and bottom of the tube) using the Qubit Broad Range kit (ThermoFisher:

Cat# Q32850). DNA was checked for impurities using the 260/230 nm and 260/280 nm ratios on a Nanodrop spectrophotometer. The sample (hereafter referred to as SsagS4) with the best DNA quality was used to generate the final reference genome. We followed the manufacturer's instructions for DNA extraction, library preparation, and DNA sequencing.

For the reference genome of sample SsagS4, long reads were used to generate an initial genome assembly that was then scaffolded using a short-read Hi-C approach. Approximately 3 μg genomic DNA was sheared using the Megaruptor 3 (Cat# B06010003) at speed 40. A library was prepared using the Pacific Biosciences (PacBio) SMRTbell 3.0 kit (Cat# 102-141-700, 102-158-300, 102-178-400). As per the protocol, library DNA fragments of less than 5 kb were removed. The library was processed on two separate Sequel IIe runs that generated a total of 11.9 and 14.2 Gb of HiFi data, respectively. PacBio data were concatenated prior to assembly. For short-read sequencing, standard Illumina libraries were prepared from gDNA using the KAPA Hyper Plus prep (Roche Cat #07962380001). The Hi-C library was prepared using the original frozen whole blood aliquot with the Proximo Hi-C animal kits (Phase Genomics Proximo Hi-C Kit Animal).

#### 2.2.2 | Genome Assembly, Scaffolding, Quality Evaluation, and Gene Model Prediction

Genome size was estimated using the Illumina reads in GenomeScope2 v2.0 (Ranallo-Benavidez, Jaron, and Schatz 2020). The PacBio HiFi data was assembled using HiFiasm v0.19.8 (Cheng et al. 2021, 2022), which was also given the HiC reads to aid in phasing. Hi-C reads were then aligned to each contig/draft assembly using Juicer v1.6.2 (Durand, Shamim, et al. 2016), a draft scaffolding was performed with 3ddna v180419 (Durand, Shamim, et al. 2016; Dudchenko et al. 2017), with manual QC using Juicebox Assembly Tools v2.20.00 (Durand, Robinson, et al. 2016; Dudchenko et al. 2018). Genome completeness was evaluated by comparing the predicted gene orthologs to the Actinopterygii odb10 database using BUSCO v 5.2.2. (Manni, Berkeley, Seppey, Simão, et al. 2021; Manni, Berkeley, Seppey, and Zdobnov 2021). The base level accuracy for each haplotype was estimated using Merqury v1.3 (Rhie et al. 2020).

Gene models were predicted using a pipeline based on TSEBRA v1.1.2.3 (Gabriel et al. 2021). Repeats were called by RepeatModeler v2.0.5 (Flynn et al. 2020). The resultant repeat library was used to softmask the genome using repeatMasker v4.1.5 (Tarailo-Graovac and Chen 2009). Sixteen RNAseq libraries from NCBI bioproject PRJNA701779 were aligned to the genome using hisat2 v2.2.1 (Emami-khoyi et al. 2021). Related proteins from seven previously annotated Clupeiformes were collected from NCBI RefSeq. Braker v2.1.6 was run twice, once with related protein evidence and once with aligned RNAseq (Brůna et al. 2021). The results were then processed by TSEBRA. Any gene model whose CDS was not found to be completely softmasked was treated as a high confidence gene model. Primary transcripts were selected based on maximising the CDS length. The resultant proteomes were then evaluated by BUSCO v5.4.3 in protein mode (Seppey, Manni, and Zdobnov 2019).

### 2.3 | Library Preparation and Low Coverage Whole Genome Sequencing

#### 2.3.1 | DNA Extraction, Library Preparation, and Sequencing

Genomic DNA was extracted from muscle tissue stored in 100% ethanol using Qiagen DNAeasy Blood & Tissue 96 extraction kits (Qiagen, Inc., Valencia, CA) following manufacturer protocol. Extractions were run on a standard 2% agarose gel to screen for high molecular weight DNA and were then quantified using a PicoGreen fluorescence on a BioTek Synergy HTX microplate reader; only samples with > 5 ng/μL were selected. After 10 ng of DNA from each high-quality extraction was plated, the 96-well plate was sealed with a microporous sealing film and stored at room temperature until liquid evaporated from all wells. DNA was then fragmented and tagged with a universal Nextera overhang following the Nextera DNA Library Prep Kit protocol (Illumina, Inc., San Diego, CA) with some modifications (i.e., using 1/20th of recommended reagents). Tagmented libraries were then amplified with low-cycle PCR and barcoded using Illumina Nextera dual-indices at concentrations of 5 μM. Additional amplification and the attachment of Illumina P5 and P7 sequencing primers was carried out using another round of low-cycle PCR. Tagmented and indexed samples were then normalised (≤ 25 ng) using 96-well SequelPrep Normalisation Plates following manufacturer protocol and then pooled for each plate. Pooled libraries were cleaned using AMPure XP beads (Beckman Coulter, Inc., Brea, CA) and eluted in 20 μL of TLE buffer. Final lcWGS sequencing libraries were then visualised on an E-Gel to determine whether the ideal size range (200–1000 bp) was achieved and quantified using a Qubit 2.0 dsDNA HS Assay (ThermoFisher, Inc., Waltham, MA). Four lcWGS libraries, each containing 96 individuals, were sequenced on two lanes of 2 × 150 bp paired-end Illumina NovaSeq 6000 at the University of Oregon Genomics and Cell Characterisation Core Facility.

#### 2.3.2 | lcWGS Data Filtering and Analyses

We generally followed Laura Timm's lcWGS analysis pipeline (see https://github.com/letimm/WGSfqs-to-genolikelihoods for scripts). In preparation for lcWGS analyses, haplotype 1 (hap 1) of the Pacific Sardine reference genome (BioProject PRJNA1094947) was indexed using BWA v0.7.17 (Li and Durbin 2009) and Samtools v1.11 *faidx* (Li et al. 2009) after dropping contigs that were not incorporated into putative chromosomes. Raw lcWGS data were de-multiplexed into forward and reverse fastq files for each individual. We used FastQC v0.11.9 (Andrews 2010) and MultiQC v1.14 (Ewels et al. 2016) to check sequence quality on individual raw reads. We trimmed adapters and polyG tails from raw fastq files using Trimmomatic v0.39 (Bolger, Lohse, and Usadel 2014) and fastp v0.23.2 (S. Chen et al. 2018), respectively, and again assessed sequence quality on trimmed reads using FastQC and MultiQC. Next, we aligned trimmed reads to the reference genome using BWA. Samtools was then used to clean up read pairings and flags from BWA with *fixmate*, convert sam to bam files, filter non-unique and poor-quality mappings before sorting read pairs by mapping coordinate. After bam

files were built, duplicate reads were detected and removed with Picard *MarkDuplicates* v2.23.9 (http://broadinstitute.github.io/picard/) and overlapping paired-end reads were clipped with bamtools *clipOverlap* v2.5.1 (Barnett et al. 2011) to generate final bam files. We then used Samtools *depth* to tally alignment depth in all individuals. Individuals with $<1\times$ mean depth of coverage were filtered from downstream analyses. To reduce potential sequencing depth bias, we performed targeted down-sampling. Target down-sampling depths were drawn from the distribution of mean individual depths calculated from the data.

#### 2.3.3 | Mitogenome Analyses

During preliminary exploration of lcWGS data, we observed strong differentiation in principle component analyses (PCA), which led us to investigate differentiation in mitochondrial DNA. Due to the high copy number of mitochondrial genomes relative to nuclear genomes, we were able to generate full mitochondrial genome sequences with high coverage for all samples passing lcWGS filtering parameters. To achieve this, we repeated the nuclear genome alignment methods described above but mapped trimmed and clipped reads from each individual to a publicly available mitochondrial reference genome accessioned under the species name "*Sardinops sagax*" (MW338734; Tang and Chen 2021). However, this individual was collected in the Northwest (NW) Pacific (36.425 N, 158.6026 E) and is thus likely a Japanese Sardine (*S. melanostica*). We then used Samtools v1.11 (Danecek et al. 2021) to generate consensus fasta sequences with a minimum depth of 10×. Muscle v3.42 (Edgar 2004) was then used to generate multiple sequence alignments that also included the following mitochondrial reference genomes: the Japanese Sardine individual used for the alignment, MW338734; a second Japanese Sardine (NC_002616) collected in the Western Pacific (location not described; Inoue et al. 2000); a Pacific Sardine from the Eastern Pacific off the coast of Baja Sur, Mexico (OR482441.1; 23.41 N, 110.23 W); and a European Pilchard, *Sardina pilchardus* (NC_009592; Lavoué et al. 2007), which is the sister taxon to *Sardinops* (Egan et al. 2024; Jérôme et al. 2003). A pairwise distance matrix was computed using Kimura's 2-parameter distance (K80; Kimura 1980) and neighbour-joining tree estimation (Saitou and Nei 1987) was performed as implemented in ape v5.7 (Paradis and Schliep 2019). Finally, we visualised trees using ggtree v3.8.2 (Yu et al. 2017).

#### 2.3.4 | lcWGS Genotype Likelihood Calls and Analyses

Genotype likelihoods for all sites were calculated using ANGSD v0.933 (Korneliussen, Albrechtsen, and Nielsen 2014). Low-quality base calls and mapped reads were excluded with minimum quality and mapping quality set to 15 (-minQ 15 and -minMapQ 15). We set the minimum depth to the total number of individuals (-setminDepth 345) and the maximum depth to the total number of individuals multiplied by 20 (-setmaxDepth 6900), which should exclude mtDNA but still retain regions sequenced at high coverage. We set the threshold for minor allele frequency to 5% (-minMaf 0.05) and the *p*-value filter for polymorphic sites to $10^{-8}$ (-SNP_pval 1e-10).

To explore potential genetic structure in our data, we conducted principal component analysis (PCA) using PCAngsd (Meisner and Albrechtsen 2018) based on SNPs from the full genome as well as for each chromosome independently. The covariance matrices were then imported into R (R Core Team 2023) to perform eigen decomposition and visualisation. We also estimated individual admixture proportions with NgsAdmix (Skotte, Korneliussen, and Albrechtsen 2013) testing k values 1-10 with 3 iterations. The Evanno method (Evanno, Regnaut, and Goudet 2005) was used to identify the most likely k value (number of genetic clusters).

To understand the level of genetic divergence observed in PCAs, we placed individuals into two groups based on PC1 separation and estimated population-level $F_{ST}$ using genotype likelihood data. In order to determine weighted pairwise $F_{ST}$ for the two groups, site allele frequency likelihoods were calculated in ANGSD using the same filtering criteria as above. Global and genome-wide $F_{ST}$ were calculated between the two groups using the folded site frequency spectrum (-realSFS). To assess significance of global $F_{ST}$, we tested if the observed $F_{ST}$ value fell significantly outside a distribution from permuting individuals, assuming $F_{ST}$ values follow an exponential distribution (Elhaik 2012). We then generated Manhattan plots in R to visualise genetic differentiation between the two groups across the genome.

### 2.4 | GTseq Species Identification

Following the detection of Japanese Sardine in the lcWGS data, we designed an ad hoc GTseq (Campbell, Harmon, and Narum 2015) species identification panel to target fixed interspecific differences observed in the full mitochondrial genomes. This panel was then used to retroactively sequence 4008 sardine samples collected in the CCES from 2013 to 2023 to better understand timing of Japanese Sardine dispersal and potential impact to past biomass estimates of Pacific Sardine (see Supporting Information for GTseq panel design and analysis details).

### 2.5 | Aging

To aid in determining the timing and possible dispersal mechanisms for Japanese Sardine into the CCLME, 33 fish that were sequenced and identified as Japanese Sardine were also aged using sagittal otoliths. Otoliths were extracted, cleaned with water, placed in 0.6 mL microcentrifuge tubes, allowed to dry overnight, and assigned individual barcodes. Whole otoliths were submerged in water in a small dish with a black background and viewed under reflected white light using a Leica MZ10 F stereomicroscope. Otoliths were aged using white light within 3 min of submersion without knowledge of species, month of capture, sex, length, or weight following Yaremko (1996).

### 2.6 | Habitat Model

Species distribution modelling studies of Japanese Sardine are generally specific to ocean conditions in the western North Pacific (Shi et al. 2023), and are likely not transferrable to the central and eastern North Pacific. We therefore used a simple envelope model to estimate the limits of suitable habitat at a

basin scale. We assumed that the lower tolerable temperature limit for sardine was approximately 4°C, based on catch locations (Muko et al. 2018; Yang, Zhang et al. 2023), and laboratory studies (Pribyl et al. 2016). Waters where SST was warmer than 8°C were defined as "favourable", while waters 4°C–8°C were defined as "marginal". We used previous species distribution modelling studies and collection locations (Shi et al. 2023; Yang, Zhang et al. 2023) to approximate tolerable ranges of surface chlorophyll, as a proxy for planktonic prey availability. We assumed that surface chlorophyll concentrations $>0.2\,mg/m^3$ surface chlorophyll were favourable for Japanese sardine, while values 0.1–0.2 $mg/m^3$ were marginal. Regions with SST < 4°C or surface chlorophyll $<0.1\,mg/m^3$ were defined as unsuitable.

SST was extracted from the NOAA 0.25° Daily Optimum Interpolation Sea Surface Temperature (OISST) product, version 2.1 (Huang et al. 2021). Surface chlorophyll was extracted from the Copernicus-GlobColour level multi-sensor interpolated gap-free product (https://doi.org/10.48670/moi-00281). The native resolution of 4 km was coarsened to 0.25° to match OISST.

The precise route and mechanism by which Japanese Sardine arrived in the eastern North Pacific is not known. However, sardine are occasionally recorded in pelagic trawl gear in the Bering Sea and northwest Gulf of Alaska by the NOAA North Pacific Observer Program. Observers have been collecting data in the region since 1973 (Ganz et al. 2020). We used these occurrences (50 positive catch locations from 1985 to 2022) to identify regions in the Bering Sea and Gulf of Alaska that may have served as overwintering habitat for Japanese Sardine *en route* from the western to eastern North Pacific. Daily SST time-series were used to identify times and areas where temperatures were above assumed lower thermal limits.

## 3 | Results

### 3.1 | Reference Genome

#### 3.1.1 | Genome Sequencing

Long-read and short-read sequencing was generated for a single *S. sagax* sample referred to as SsagS4. The SsagS4 sample was run on two PacBio HiFi SMRT cells, generating a total of 45.7 Gbp of data. There were 46.1 Gbp of Illumina paired-end short-read data and 107.5 Gbp of Hi-C data. The scaffolded, phased genome assembly resulted in haplotype lengths of 917.0 Mb (hap 1) and 908.6 Mb (hap 2), scaffold N50 of 35.2 Mb (hap 1) and 35.1 Mb (hap 2), and BUSCO protein scores above 80% (Table 1). Ninety percent of the genome assembly was contained within 23 (hap 1) and 22 (hap 2) scaffolds. The genome has 24 chromosomes (Figure S1). The overall HiFi sequencing coverage was approximately 50× or 25× for each haplotype. The total number of primary high confidence genes was 36,223 (hap 1) and 35,629 (hap 2) (Figure S1b). The repeat content was similar across haplotypes (Table S1). We compared the kmer profiles from the whole genome assemblies of both haplotypes of SsagS4 with 11 other high quality Clupeiform genomes including (*Denticeps clupeoides*, *Coilia nasus*, *Clupea harengus*, *S. pilchardus*, *Alosa sapidissima*, *Alosa fallax*, *Alosa alosa*, *Sardinella longiceps*, *Tenualosa ilisha*, *Tenualosa thibaudeaui*, and *Limnothrissa miodon*). Specifically, we compare the values derived from the Pankmer output which is the Jaccard similarity matrix where 0 = different, 1 = similar of the samples, derived from the kmers. Additional details as to how this is calculated can be found in the Pankmer manuscript (Aylward et al. 2023). The two haplotypes from the same individual had a similarity of 0.5097. Based on the whole genome kmer comparison to other Clupeiformes, *S. sagax* was more similar (although very weak) to the *S. pilchardus* (Jaccard similarity 0.0678) which corresponds to an estimated divergence time of 29.2 my based on TimeTree (Kumar et al. 2022). The *S. sagax* genomes were next most similar to the three *Alosa* spp. (*A. sapidissima* 0.046, *A. fallax* 0.0447, and *A. alosa* 0.0428) which corresponds to an estimated divergence time of 41 my based on TimeTree. Interestingly, the *A. alosa* and *A. fallax* genomes had the highest similarity (0.6309). Fishes from within the same genus had more similar genomes. This was apparent for the three *Alosa* spp. and two *Tenualosa* spp. included in the analysis (Figure S1c). The *S. sagax* genome is the first, haplotype-resolved genome assembly of any Clupeiformes. Both of the haplotypes scaffold N50s are the fourth highest (Hap1 35.22, Hap2 = 35.06) behind *A. sapidissima* (38.44 Mb), *Coilia nasus* (35.42 Mb), and *A. alosa* (35.35 Mb). The mean scaffold length of the haplotypes for *S. sagax* was (54.09 Mb and 53.30 Mb, respectively), which is second only to the American Shad *A. sapidissima* (56.50 Mb) (GCA_018492685.1; Figure S2).

**TABLE 1** | Phased haplotype assembly results of *Sardinops sagax* (SsagS4.phased_hifiasm_hic).

| Haplotype | Total (Mb) | Scaffolds (#) | Scaffold N50 (Mb) | Mean scaffold length (Mb) | Longest scaffold (Mb) |
|---|---|---|---|---|---|
| 1 | 917.0 | 703 | 35.2 | 1.3 | 54.1 |
| 2 | 908.6 | 453 | 35.1 | 2 | 53.3 |
| **Haplotype** | | **Gaps** | **QV** | **BUSCO** | |
| 1 | | 1164 | 47 | C: 84.0% [S: 81.8%, D: 2.2%], F: 2.3%, M: 13.7% | |
| 2 | | 954 | 47.4 | C: 83.7% [S: 81.6%, D: 2.1%], F: 2.4%, M: 13.9% | |

*Note:* QV: Merqury QV base level accuracy estimation. BUSCO: Actinopterygii_odb10 (3640 BUSCOs) protein mode (braker)—CS (complete single-copy), CD (complete duplicated), F (fragmented), M (missing).

### 3.2 | lcWGS

Depth of coverage for the 384 lcWGS samples ranged from 0.01 to 7.85 with a mean of 2.77. After QF, 345 samples remained with a mean coverage of 3.01. After targeted downsampling, mean coverage for the 345 QF samples was 1.53 (range 1.01 to 4.19). SNP filtering parameters resulted in 4,821,933 polymorphic sites. Genome-wide PCA revealed two distinct genetic clusters separated on PC1, which explained 6.57% of the variation, with 50 individuals collected in 2022 from Oregon to Central California, USA (PC1 > 0.1), being clearly distinct from 295 individuals collected in 2021 and 2022 from throughout the sampling range (PC1 < 0; Figure 1). As previously mentioned, in preliminary analyses these clearly differentiated groups were determined to represent Japanese and Pacific Sardine, respectively (see mitochondrial genome analyses for details below), and we refer to them as such for clarity moving forward. Separation of Japanese and Pacific Sardine along PC1 was also observed in every chromosome-specific PCA (Figure S3). PC2 in the genome-wide PCA explained 0.33% of the variation and separated Pacific Sardine (PC1 < 0) into three groups. Notably, there was no apparent correlation with latitude as all three groups contained individuals from throughout the sampling range. Pacific Sardine also clearly clustered into three groups in PCAs for chromosomes 2, 11, and 15, again with no apparent geographic correlation (Figure S3). Admixture results for k = 2, which was the best supported k value, corroborated PCA results and definitively separated Pacific and Japanese Sardine (Figures 1A and 2, Figure S3). Signals of introgression were unidirectional and relatively weak with only Pacific Sardine showing evidence of admixture with the highest reported value being ~4% (mean = 0.3%).

The neighbour joining analysis for mitochondrial genomes yielded a tree with two well-separated clusters within *Sardinops* samples (Figure S4). The 50 individuals observed together in PCAs and admixture results (Figures 1 and 2, Figure S3) formed a cluster that includes the two Japanese Sardine (*S. melanostica*) reference sequences (MW338734 and NC_002616) and a single individual from the larger PCA and Admixture group (Figures 1 and 2, Figure S3). The remaining 294 samples formed the other cluster that included the Pacific Sardine (*S. sagax*) reference mitochondrial genome (OR482441.1). These results indicate that the 50 samples collected in 2022 from north of Central California, USA, represent Japanese Sardine (*S. melanostica*) and the other cluster collected in 2021–2022 from Oregon, CA, USA, to Baja California Sur, Mexico, represent Pacific Sardine (*S. sagax*) with a single apparent case of mitochondrial introgression (*S. sagax* nuclear DNA and *S. melanostica* mtDNA). Notably, there was no sign of elevated nuclear introgression in the mitochondrial introgressed individual (Figures 1A and 2).

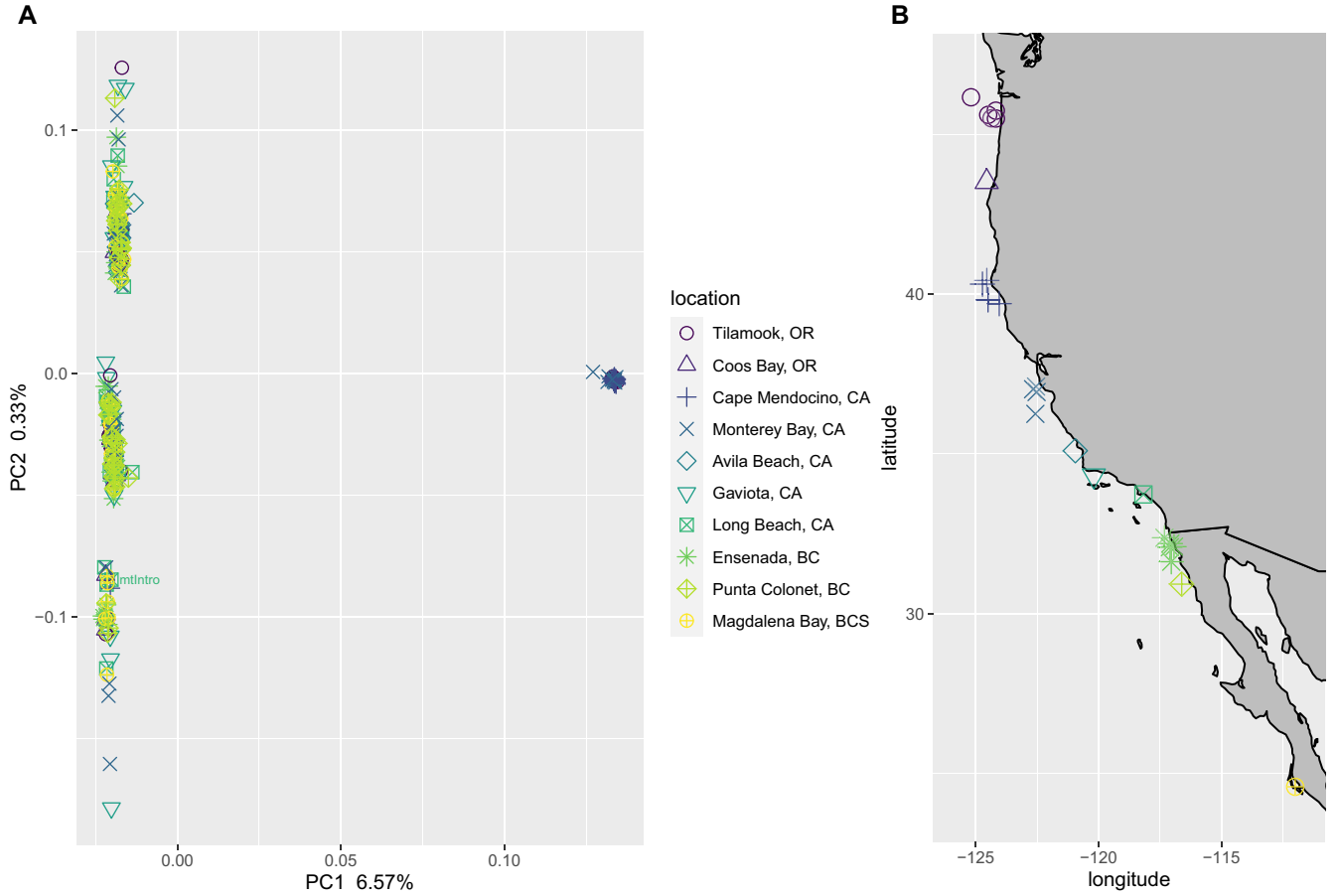


**FIGURE 1** | (A). Principal component analysis on 4,821,933 polymorphic sites from 345 *Sardinops* samples collected off the west coast of North America. Mitochondrial genome analysis suggests the right grouping (PC1 > 0.1; 50 individuals) represent Japanese Sardine (*S. melanostica*) and the left grouping (PC1 < 0; 295 individuals) represent Pacific Sardine (*S. sagax*). Mitochondrial introgressed individual is labelled "mtIntro". (B) *Sardinops* sampling sites.

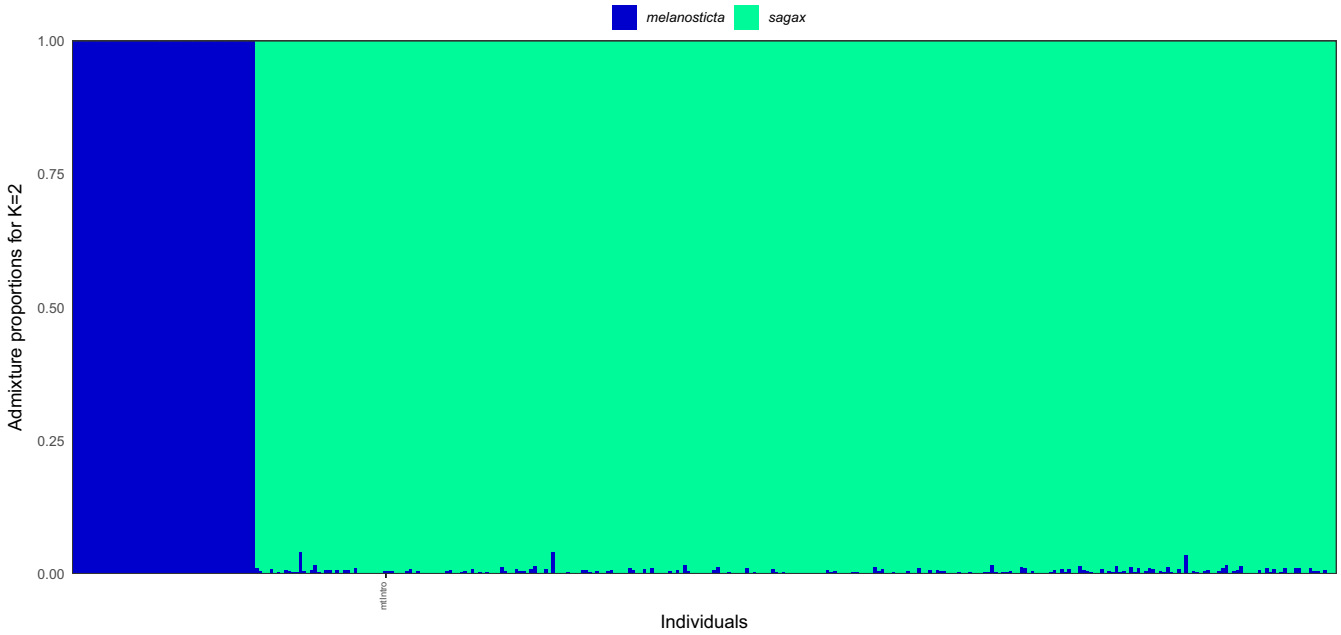


**FIGURE 2** | Admixture results for $k=2$ on 4,821,933 polymorphic sites from 345 *Sardinops* samples collected off the west coast of North America. Individuals are arranged based on identification as Japanese Sardine (*S. melanostica*) and Pacific Sardine (*S. sagax*). Mitochondrial introgressed individual is labelled "mtIntro".

The global weighted $F_{ST}$ between Japanese and Pacific Sardine (the distinct PC1 groups observed in the PCA) was 0.0827 ($p<0.0001$). Manhattan plots with locus specific $F_{ST}$ revealed prolific differentiation throughout the genome with all chromosomes containing widespread regions characterised by high $F_{ST}$ SNPs (Figure S5).

### 3.3 | GTseq Species Identification

Overall, 3484 of the 4008 CCES sardine samples from 2013 to 2023 sequenced with the mitochondrial GTseq panel (Table S2) passed quality filters (86.9%), ranging from 78.5% to 98.5% per year (Table S3). Japanese Sardine haplotypes were only detected in 2022 and 2023, aside from a single individual in 2014, which was subsequently identified as an introgressed Pacific Sardine (*S. sagax* nuclear DNA and *S. melanosticta* mtDNA) in a lcWGS library for another project (Longo et al., in prep). A subset of samples from 2021 and 2022 were sequenced both with lcWGS and GTseq (20 and 42, respectively), and yielded consistent species calls. A portion of CCES samples from 2021 and 2022 were genotyped only with lcWGS data (93 and 92, respectively). Collectively, all 174 genotyped samples from 2021 (81 GTseq, 93 lcWGS, 20 both) were identified as Pacific Sardine. Of the 172 genotyped samples from 2022 (80 GTseq, 92 lcWGS, 42 both), 72 (41.9%) were genotyped as Japanese Sardine, and were collected from Santa Cruz, CA, to Astoria, OR (Figure S6). For the 825 samples passing quality filters from 2023, 334 (40.5%) samples were genotyped as Japanese Sardine and were collected from the Southern California Bight to near Cape Flattery, WA (Figure S7). Notably, a portion of the samples with Japanese Sardine mitochondrial genotypes may represent mitochondrial introgressed individuals; however, these cases appear rare based on observed cases in the 2022 lcWGS data.

### 3.4 | Aging

Thirty-three fish identified as Japanese Sardine were aged by reading otolith annuli. Samples comprised three age classes (1–3). The dominant age class was age two ($N=25$), followed by age one ($N=7$), and age three ($N=1$).

### 3.5 | Habitat Model

The sardine habitat model showed a band of favourable habitat across the North Pacific Transition Zone region, which was the most continuous longitudinally during warmer months. Favourable habitat also extended northwards into the Bering Sea and Gulf of Alaska during summer (Figure 3). The Aleutian Islands were marginal during winter, while the rest of the Bering Sea was too cold to be suitable for sardine. Positive catch locations for sardine in the Bering Sea and Gulf of Alaska were clustered around the Pribilof Islands, the central Aleutian Islands, the eastern Aleutian Islands near Unimak Island, and Kodiak Island (Figure S8). SST time-series from within these areas of interest showed that winter SSTs were typically well below the approximate lower thermal limit for sardine near the Pribilof Islands and eastern Aleutian Islands. However, winter SSTs near the central Aleutian Islands and Kodiak Island have stayed near or above 4°C since 2015 (Figure 4). In recent years, marginal and favourable habitat in these areas has become available earlier in the spring, and persisted later in the fall, when compared to the 1980s—a previous period of high Japanese Sardine biomass (Figure S9).

## 4 | Discussion

This study added to the growing literature that demonstrates that lcWGS data can be a powerful resource in biogeography

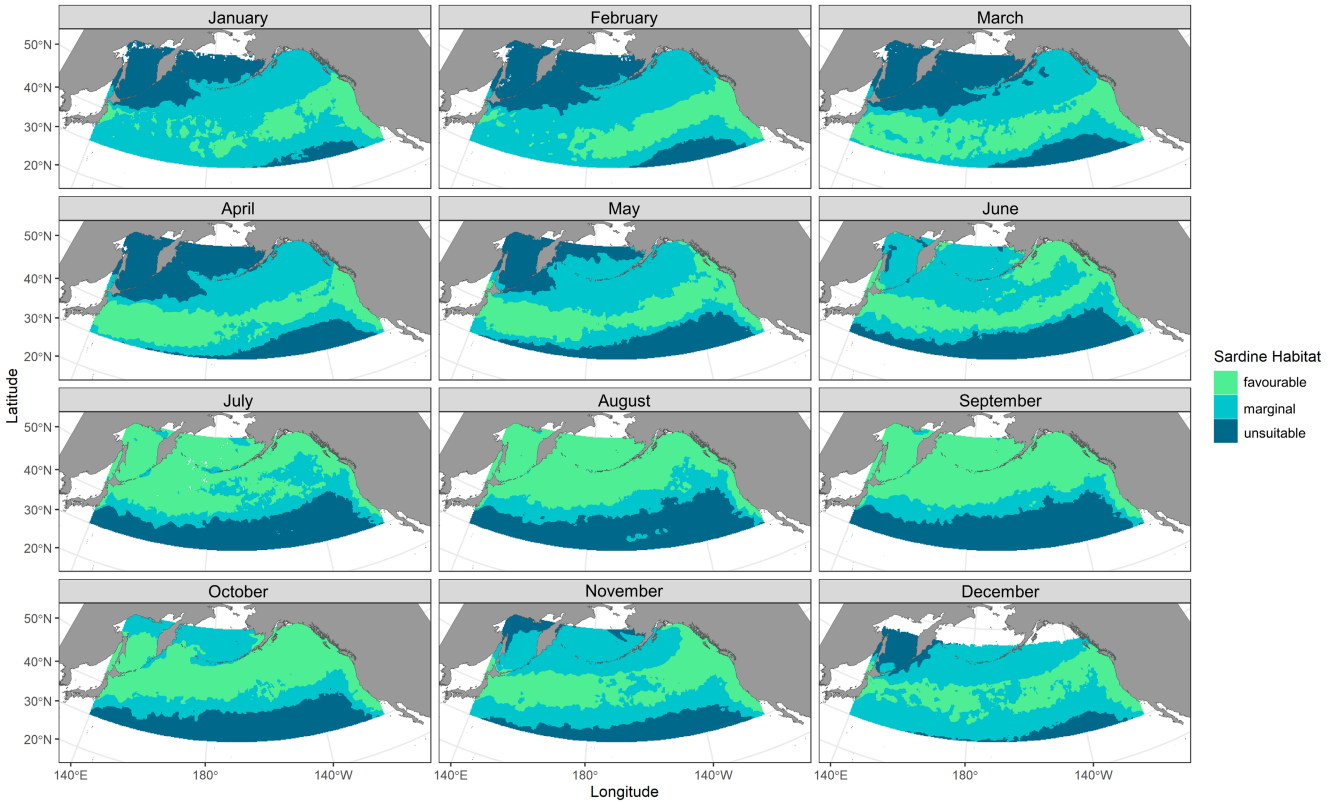


**FIGURE 3** | Mean sardine habitats for years 2020–2022 by month. Favourable habitats were areas with SST ≥ 8°C and surface chlorophyll ≥ 0.2 mg/m$^3$, marginal habitats had SSTs of > 4°C and surface chlorophyll > 0.1 mg/m$^3$, while unsuitable habitats had SST < 4°C or chlorophyll < 0.1 mg/m$^3$.

and fisheries management. Here, we used lcWGS to generate millions of genome-wide SNPs and to assemble full mitochondrial genomes in sardine from the northeastern (NE) Pacific. The unexpectedly strong levels of nuclear differentiation detected across all chromosomes in conjunction with the mitogenome assemblies allowed for the first known detection of Japanese Sardine in the Eastern Pacific. This unexpected finding may have implications for sardine population monitoring off the coast of North America as it changes our understanding of *Sardinops* spp. distribution and dispersal in the Pacific.

Several conspicuous patterns in the lcWGS data will be briefly noted here but will require a more thorough investigation in future work. Although our sampling is relatively limited, admixture results and the single case of mitochondrial introgression in these data (another case observed in GTseq and unpublished lcWGS data shows the same pattern) suggest introgression has likely occurred historically and was unidirectional from Japanese Sardine into Pacific Sardine. Indeed, we detect no evidence of admixture in any Japanese Sardine and the level of admixture in Pacific Sardine is quite low. The mitochondrial introgressed individual (Pacific Sardine with a Japanese Sardine mitochondrial haplotype) clearly clusters with Pacific Sardine in all PCAs and does not exhibit elevated levels of admixture compared to other admixed Pacific Sardine (0.5% compared to highest observed admixture of 4%), suggesting introgression occurred in the evolutionary past corroborating previous genetic studies (Bowen and Grant 1997; Grant, Clark, and Bowen 1998). Furthermore, the relatively widespread distribution of high $F_{ST}$ loci across all chromosomes observed in the Manhattan plot with limited evidence of discrete islands of divergence suggests that interspecific geneflow has likely not occurred recently (c.f. Shi et al. 2021). Notably, Pacific Sardine in PCAs for chromosomes 2, 11, and 15 and Japanese Sardine for chromosome 9 (Figure S3) show a clear pattern often associated with inversions where homokaryotypes for inverted and uninverted karyotypes group separately with heterokaryotypes (i.e., individuals heterozygous for inverted and uninverted regions) falling out between (Wellenreuther and Bernatchez 2018). Chromosomal inversions can act as barriers to geneflow in recently diverged taxa (Faria et al. 2019; Noor et al. 2001) and these may have played a role in the divergence of Pacific and Japanese Sardine. We also note that each putative karyotype group of the candidate inversions in Pacific Sardine contain individuals from throughout the broad latitudinal sampling range. This suggests that the respective inversion haplotypes likely do not harbour adaptive alleles related to environmental variables correlated with latitude, such as sea surface temperature (Anderson et al. 2005; Wellenreuther and Bernatchez 2018).

The presence of Japanese Sardine in the CCLME is a surprising finding and may have implications for both the community ecology and management of coastal pelagic species along the west coast of North America. Their detection follows a period of noticeably increased MHWs in the northeast Pacific (Carvalho, Smith, and Wang 2021; Chen et al. 2021; Werb and Rudnick 2023). These warm conditions were implicated in several biological phenomena including a shift in distributions of marine organisms (Cavole et al. 2016; Gold et al. 2023; O'Leary et al. 2022; Thompson et al. 2022) and may have provided

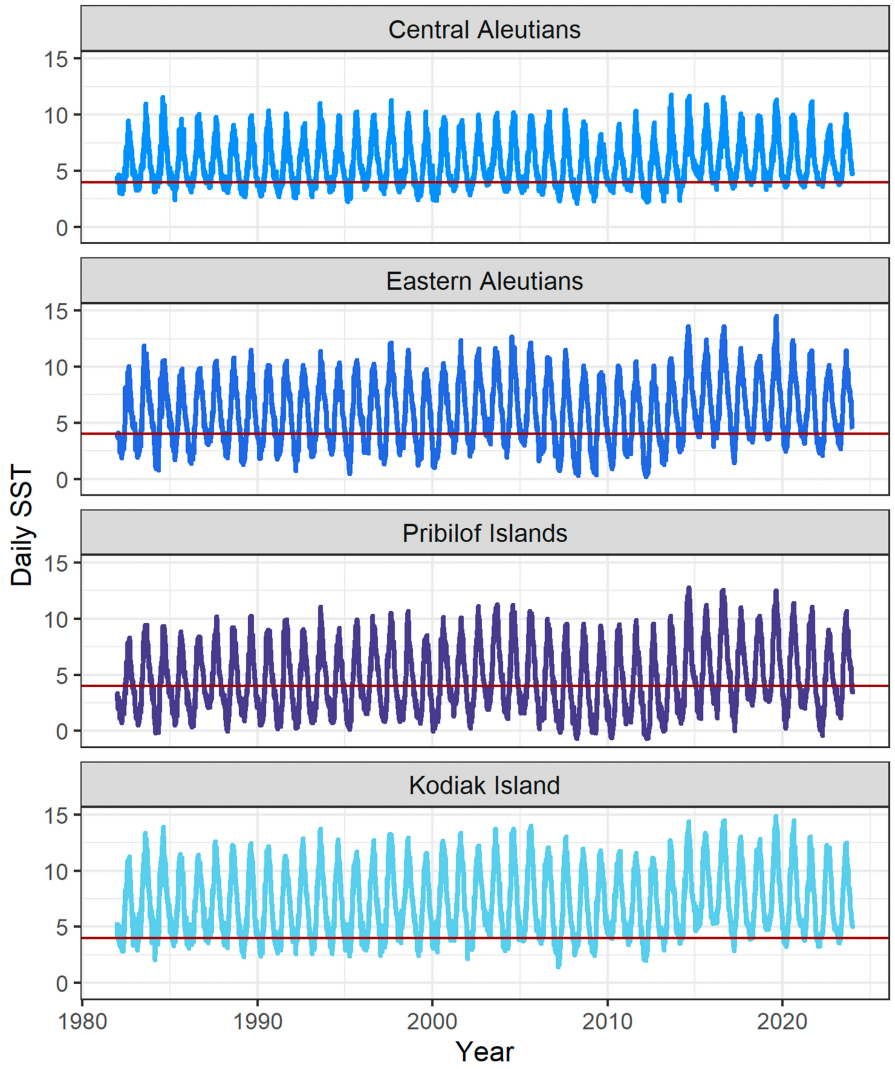


**FIGURE 4** | Daily time-series of SST 1982–2023 within areas where sardine have been recorded by fisheries observers (see Figure S8). The horizontal red line shows 4°C.

favourable conditions for the dispersal of Japanese Sardine to the CCLME. Indeed, sardine are characterised by life-history traits associated with rapid leading-edge range shifts in marine taxa—ecological generalists, short generation times, high adult mobility, and a long PLD (Pinsky, Selden, and Kitchel 2020). Notably, the ages of Japanese Sardine (1–3), which were collected in 2022, coincide with widespread heatwave conditions in the Bering Sea and eastern North Pacific (Amaya et al. 2020; Carvalho, Smith, and Wang 2021; Chen et al. 2021).

The mechanism by which Japanese sardine moved to the CCLME is not yet clear. As with many Clupeiform fishes, the geographic range of the Japanese Sardine expands and contracts during periods of high and low abundance, respectively (Lluch-Belda et al. 1989; Sarr, Kindong, and Tian 2021). During periods of high abundance, Japanese Sardine may range into the Bering Sea (Sarr, Kindong, and Tian 2021) and the Pacific Sardine is known to range as far north as Alaska (Parrish, Serra, and Grant 1989). During synchronous periods of high abundance, it is therefore reasonable to assume that the ranges of these species could overlap if ecological conditions are favourable. While this may explain historical patterns of mixing between these two species, it may not explain the current situation. Pacific Sardine are at relatively low biomass compared to historical levels and have contracted to its core distributional range off of southern California, USA, and northern Baja California, Mexico (Tran 2023). However, after crashing in the 1990s, Japanese Sardine abundance has been steadily increasing since 2010 and they have been detected as far east as 170°W since 2017 (Fisheries Stock Assessment Center 2020; Yang, Han et al. 2023).

Larval Japanese Sardine are dispersed eastwards into the western Pacific via the Kuroshio Current during spring spawning. Shi et al. (2023) modelled potential habitat on the main fishing grounds and showed that it roughly followed the warm Kuroshio Extension, expanding eastward in late summer and early fall. Plausibly, favourable conditions in the Kuroshio Extension and western North Pacific could have provided the recovering Japanese Sardine population an intermittent habitat corridor following the North Pacific Current east to cross the Pacific basin. However, studies of eastwards debris dispersion after the 2011 Japan tsunami suggest that without active swimming, it would likely take more than 1 year for sardine to reach the CCLME via passive advection (Maximenko

et al. 2018). While the simple sardine habitat model suggested that favourable habitat can stretch across much of the North Pacific during most months, it is uncertain whether satellite-derived surface chlorophyll can capture abundance of the planktonic organisms that sardine require for growth and survival. Although sampling is sparse, several historical studies have also noted that while Japanese Sardine can be collected in the western North Pacific as far east as ~180°, they have not been collected further east in the offshore eastern North Pacific (Pearcy et al. 1996).

It is also possible that Japanese Sardine arrived at the CCLME using a more northward route. Unusually warm ocean conditions may have facilitated the entry of Japanese Sardine into the Bering Sea from the south, as an extension of their usual northwards movements in summer and fall (Sakamoto et al. 2023). If warm temperatures allowed them to survive the winter in the Aleutian Islands or western Gulf of Alaska, they would then be able to move into the CCLME the following spring. A west to east dispersal pattern in the north Pacific is supported by genetic studies in another Clupeiformes, Pacific Herring (*Clupea pallasii*), where NW Pacific mitochondrial lineages appear in the NE Pacific but no NE lineages appear in the NW Pacific (Liu et al. 2011, 2012).

How frequent are such Trans-Pacific *Sardinops* dispersal events? Although the lcWGS data only include 2021 and 2022 samples, the ad hoc GTseq species panel ran on thousands of samples from 2013 to 2023 only detected Japanese Sardine in 2022 and 2023 (a single Japanese Sardine haplotype was detected in 2014 but was subsequently identified as a mitochondrial introgressed Pacific Sardine; see Supporting Information for details). Additionally, a recent mtDNA phylogeographic analysis of 434 Pacific Sardine samples collected between 2002 and 2006 from British Columbia, Canada, to the Gulf of California, Mexico, did not detect Japanese Sardine (Adams and Craig 2024). Taken together, we do not see evidence for Japanese Sardine outside of 2022 and 2023. However, it is plausible that dispersal occurs somewhat regularly when conditions are favourable across the North Pacific. The proposed divergence time of ~200,000–300,000 years between Japanese and Pacific Sardine (Bowen and Grant 1997; Grant, Clark, and Bowen 1998) encompassed prolonged glacial periods (Kawamura et al. 2007) when the North Pacific would have acted as a formidable barrier to dispersal. During these extended periods of low geneflow, species barrier loci (e.g., chromosomal inversions) could have evolved, which may be why we detected no evidence of recent introgression in our data although further investigation is needed to assess this possibility.

The detection of Japanese Sardine in the CCLME shifts the range edge of the species thousands of kilometres east. Empirical evidence suggests that most marine species range edges, particularly in the North Pacific, follow temperature boundaries and that winter temperatures can be integral components of range edge dynamics (Fredston et al. 2021; Pinsky, Selden, and Kitchel 2020). Our habitat model suggests that recent warming trends in the North Pacific resulted in winter temperatures reaching marginal conditions (i.e., above critical thermal minima) for sardine, possibly opening a habitat corridor across higher latitudes. This potentially emergent corridor for sardine may also enable dispersal in other temperate species of the eastern and/or western North Pacific that were previously restricted. Warming of poleward waters globally may be increasing dispersal likelihood for temperate marine taxa across (i.e., longitudinally) previously unfavourable habitat in other ocean basins as well (e.g., North Atlantic). Continued monitoring through surveys such as NOAA's CCES will be instrumental in detecting such dispersal events and determining whether Japanese Sardine in the CCLME represent a temporary or more permanent range shift.

### Author Contributions

G.C.L., M.T.C. and W.L. designed the study. Laboratory work was performed by G.C.L., J.J.M., N.A., K.J., E.S.A.-H., N.H., T.D., and M.T.C. Bioinformatic data generation and analyses were performed by G.C.L. for lcWGS and J.J.M., N.A., N.H., and T.D. for the reference genome. M.T.C., W.L. and T.P.M. provided laboratory support and analytical tools for lcWGS and reference genome work. K.J. designed and performed the otolith aging study. B.M. designed and performed the habitat model study. G.C.L., J.J.M., and B.M. generated the figures of the manuscript. G.C.L. and M.T.C. wrote the paper with analysis specific contributions from J.J.M., K.J., and B.M. All authors edited and provided feedback to manuscript drafts.

### Acknowledgements

We thank the numerous NMFS staff, NOAA Corps officers and crew of the F/V *Reuben Lasker*, and several volunteers who facilitated collection of genetic samples used in this study. We are grateful for assistance in lcWGS library preparation from Katie D'Amelio and to Laura Timm and Sara Schaal for guidance on lcWGS analyses. We thank the Everingham Brothers Bait Co. for assistance in supplying specimens for reference genome creation, and Glenn Campbell and Pearl Rojas from the Alaska Fisheries Science Center for observer program data. We thank Concepción Enciso Enciso (IMIPAS) for providing samples from southern Baja California.

### Conflicts of Interest

The authors declare no conflicts of interest.

### Data Availability Statement

All genomic data have been deposited at NCBI and are accessible under BioProject PRJNA1094947.

### References

Adams, E. S., and M. T. Craig. 2024. "Phylogeography of the Pacific Sardine, *Sardinops sagax*, across its Northeastern Pacific range." *Bulletin of the Southern California Academy of Sciences*.

Amaya, D. J., A. J. Miller, S. P. Xie, and Y. Kosaka. 2020. "Physical Drivers of the Summer 2019 North Pacific Marine Heatwave." *Nature Communications* 11, no. 1: 1903. https://doi.org/10.1038/s41467-020-15820-w.

Anderson, A. R., A. A. Hoffmann, S. W. Mckechnie, P. A. Umina, and A. R. Weeks. 2005. "The Latitudinal Cline in the in(3R)Payne Inversion Polymorphism Has Shifted in the Last 20 Years in Australian *Drosophila melanogaster* Populations." *Molecular Ecology* 14, no. 3: 851–858. https://doi.org/10.1111/j.1365-294X.2005.02445.x.

Andrews, S. 2010. *FastQC: A Quality Control Tool for High Throughput Sequence Data*. Cambridge: Babraham Bioinformatics, Babraham Institute.

Aylward, A. J., S. Petrus, A. Mamerto, N. T. Hartwick, and T. P. Michael. 2023. "PanKmer: K-Mer-Based and Reference-Free Pangenome Analysis." *Bioinformatics* 39, no. 10: btad621.

Bakun, A., B. A. Black, S. J. Bograd, et al. 2015. "Anticipated Effects of Climate Change on Coastal Upwelling Ecosystems." *Current Climate Change Reports* 1, no. 2: 85–93. https://doi.org/10.1007/s40641-015-0008-4.

Barnett, D. W., E. K. Garrison, A. R. Quinlan, M. P. Strömberg, and G. T. Marth. 2011. "BamTools: A C++ API and Toolkit for Analyzing and Managing BAM Files." *Bioinformatics* 27, no. 12: 1691–1692. https://doi.org/10.1093/bioinformatics/btr174.

Bolger, A. M., M. Lohse, and B. Usadel. 2014. "Trimmomatic: A Flexible Trimmer for Illumina Sequence Data." *Bioinformatics* 30, no. 15: 2114–2120. https://doi.org/10.1093/bioinformatics/btu170.

Bowen, B. W., and W. S. Grant. 1997. "Phylogeography of the Sardines (*Sardinops* spp.): Assessing Biogeographic Models and Population Histories in Temperate Upwelling Zones." *Evolution* 51, no. 5: 1601–1610.

Brůna, T., K. J. Hoff, A. Lomsadze, M. Stanke, and M. Borodovsky. 2021. "BRAKER2: Automatic Eukaryotic Genome Annotation With GeneMark-EP+ and AUGUSTUS Supported by a Protein Database." *NAR Genomics and Bioinformatics* 3, no. 1: 1–11. https://doi.org/10.1093/nargab/lqaa108.

Burford, B. P., L. A. Wild, R. Schwarz, et al. 2022. "Rapid Range Expansion of a Marine Ectotherm Reveals the Demographic and Ecological Consequences of Short-Term Variability in Seawater Temperature and Dissolved Oxygen." *American Naturalist* 199, no. 4: 523–550.

Campbell, N. R., S. A. Harmon, and S. R. Narum. 2015. "Genotyping-In-Thousands by Sequencing (GT-Seq): A Cost Effective SNP Genotyping Method Based on Custom Amplicon Sequencing." *Molecular Ecology Resources* 15, no. 4: 855–867. https://doi.org/10.1111/1755-0998.12357.

Carvalho, K. S., T. E. Smith, and S. Wang. 2021. "Bering Sea Marine Heatwaves: Patterns, Trends and Connections With the Arctic." *Journal of Hydrology* 600: 126462. https://doi.org/10.1016/j.jhydrol.2021.126462.

Cavole, L., A. Demko, R. Diner, et al. 2016. "Biological Impacts of the 2013–2015 Warm-Water Anomaly in the Northeast Pacific: Winners, Losers, and the Future." *Oceanography* 29, no. 2: 273–285. https://doi.org/10.5670/oceanog.2016.32.

Chen, S., Y. Zhou, Y. Chen, and J. Gu. 2018. "Fastp: An Ultra-Fast All-In-One FASTQ Preprocessor." *Bioinformatics* 34, no. 17: i884–i890. https://doi.org/10.1093/bioinformatics/bty560.

Chen, Z., J. Shi, Q. Liu, H. Chen, and C. Li. 2021. "A Persistent and Intense Marine Heatwave in the Northeast Pacific During 2019–2020." *Geophysical Research Letters* 48, no. 13: e2021GL093239. https://doi.org/10.1029/2021GL093239.

Cheng, H., G. T. Concepcion, X. Feng, H. Zhang, and H. Li. 2021. "Haplotype-Resolved de Novo Assembly Using Phased Assembly Graphs With Hifiasm." *Nature Methods* 18, no. 2: 170–175. https://doi.org/10.1038/s41592-020-01056-5.

Cheng, H., E. D. Jarvis, O. Fedrigo, et al. 2022. "Haplotype-Resolved Assembly of Diploid Genomes Without Parental Data." *Nature Biotechnology* 40, no. 9: 1332–1335. https://doi.org/10.1038/s41587-022-01261-x.

Danecek, P., J. K. Bonfield, J. Liddle, et al. 2021. "Twelve Years of SAMtools and BCFtools." *GigaScience* 10, no. 2: giab008. https://doi.org/10.1093/gigascience/giab008.

Dorval, E., D. Porzio, B. D. Schwartzkopf, K. C. James, L. Vasquez, and B. Erisman. 2022. "Sampling Methodology for Estimating Life History Parameters of Coastal Pelagic Species Along the U.S. Pacific Coast." https://doi.org/10.25923/e8wq-xf39.

Dudchenko, O., S. S. Batra, A. D. Omer, et al. 2017. "De Novo Assembly of the Aedes Aegypti Genome Using Hi-C Yields Chromosome-Length Scaffolds." *Science* 356, no. 6333: 92–95.

Dudchenko, O., M. S. Shamim, S. S. Batra, et al. 2018. "The Juicebox Assembly Tools Module Facilitates De Novo Assembly of Mammalian Genomes with Chromosome-Length Scaffolds for Under $1000." *BioRxiv*, 254797. https://doi.org/10.1101/254797.

Dulvy, N. K., S. I. Rogers, S. Jennings, V. Stelzenmüller, S. R. Dye, and H. R. Skjoldal. 2008. "Climate Change and Deepening of the North Sea Fish Assemblage: A Biotic Indicator of Warming Seas." *Journal of Applied Ecology* 45, no. 4: 1029–1039.

Durand, N. C., J. T. Robinson, M. S. Shamim, et al. 2016. "Juicebox Provides a Visualization System for Hi-C Contact Maps With Unlimited Zoom." *Cell Systems* 3, no. 1: 99–101.

Durand, N. C., M. S. Shamim, I. Machol, et al. 2016. "Juicer Provides a One-Click System for Analyzing Loop-Resolution Hi-C Experiments." *Cell Systems* 3, no. 1: 95–98.

Dyldin, Y. V., A. M. Orlov, L. Hanel, V. I. Romanov, R. Fricke, and E. D. Vasil'eva. 2022. "Ichthyofauna of the Fresh and Brackish Waters of Russia and Adjacent Areas: Annotated List With Taxonomic Comments. 1. Families Petromyzontidae–Pristigasteridae." *Journal of Ichthyology* 62, no. 3: 385–414.

Edgar, R. C. 2004. "MUSCLE: Multiple Sequence Alignment With High Accuracy and High Throughput." *Nucleic Acids Research* 32, no. 5: 1792–1797. https://doi.org/10.1093/nar/gkh340.

Egan, J. P., A. M. Simons, M. S. Alavi-Yeganeh, et al. 2024. "Phylogenomics, Lineage Diversification Rates, and the Evolution of Diadromy in Clupeiformes (Anchovies, Herrings, Sardines, and Relatives)." *Systematic Biology*: syae022. https://doi.org/10.1093/sysbio/syae022.

Elhaik, E. 2012. "Empirical Distributions of FST From Large-Scale Human Polymorphism Data." *PLoS One* 7, no. 11: e49837. https://doi.org/10.1371/journal.pone.0049837.

Emami-khoyi, A., R. Le Roux, M. G. Adair, et al. 2021. "Transcriptomic Diversity in the Livers of South African Sardines Participating in the Annual Sardine Run." *Genes* 12, no. 3: 1–15. https://doi.org/10.3390/genes12030368.

Evanno, G., S. Regnaut, and J. Goudet. 2005. "Detecting the Number of Clusters of Individuals Using the Software STRUCTURE: A Simulation Study." *Molecular Ecology* 14, no. 8: 2611–2620. https://doi.org/10.1111/j.1365-294X.2005.02553.x.

Ewels, P., M. Magnusson, S. Lundin, and M. Käller. 2016. "MultiQC: Summarize Analysis Results for Multiple Tools and Samples in a Single Report." *Bioinformatics* 32, no. 19: 3047–3048. https://doi.org/10.1093/bioinformatics/btw354.

Faria, R., K. Johannesson, R. K. Butlin, and A. M. Westram. 2019. "Evolving Inversions." In *Trends in Ecology and Evolution*, vol. 34, no. 3, 239–248. Elsevier Ltd. https://doi.org/10.1016/j.tree.2018.12.005.

Fisheries Stock Assessment Center. 2020. "*Stock Assessment of Japanese Sardine Pacific Stock in 2020. FRA-SA2020-SC01-1*."

Flynn, J. M., R. Hubley, C. Goubert, et al. 2020. "RepeatModeler2 for Automated Genomic Discovery of Transposable Element Families." *Proceedings of the National Academy of Sciences of the United States of America* 117, no. 17: 9451–9457. https://doi.org/10.1073/pnas.1921046117.

Fredston, A., M. Pinsky, R. L. Selden, et al. 2021. "Range Edges of North American Marine Species Are Tracking Temperature Over Decades." *Global Change Biology* 27, no. 13: 3145–3156.

Gabriel, L., K. J. Hoff, T. Brůna, M. Borodovsky, and M. Stanke. 2021. "TSEBRA: Transcript Selector for BRAKER." *BMC Bioinformatics* 22, no. 1: 566. https://doi.org/10.1186/s12859-021-04482-0.

Ganz, P., C. Faunce, G. Mayhew, et al. 2020. "Deployment Performance Review of the 2019 North Pacific Observer Program." *U.S. Dep. Commer., NOAA Tech. Memo., NMFS-AFSC-411*.

Gold, Z., R. P. Kelly, A. O. Shelton, et al. 2023. "Message in a Bottle: Archived DNA Reveals Marine Heatwave-Associated Shifts in Fish Assemblages." *Environmental DNA* 6: e400. https://doi.org/10.1002/edn3.400.

Grant, W. S., A.-M. Clark, and B. W. Bowen. 1998. "Why Restriction Fragment Length Polymorphism Analysis of Mitochondrial DNA Failed to Resolve Sardine (*Sardinops*) Biogeography: Insights From Mitochondrial DNA Cytochrome *b* Sequences." *Canadian Journal of Fisheries and Aquatic Sciences* 55, no. 12: 2539–2547. https://doi.org/10.1139/f98-127.

Huang, B., C. Liu, V. Banzon, et al. 2021. "Improvements of the Daily Optimum Interpolation Sea Surface Temperature (DOISST) Version 2.1." *Journal of Climate* 34, no. 8: 2923–2939. https://doi.org/10.1175/JCLI-D-20-0166.1.

Inoue, J. G., M. Miya, K. Tsukamoto, and M. Nishida. 2000. "Complete Mitochondrial DNA Sequence of the Japanese Sardine *Sardinops melanostictus* +." *Fisheries Science* 66, no. 5: 924–932. https://doi.org/10.1046/j.1444-2906.2000.00148.x.

Jansen, T., E. E. Nielsen, N. Rodriguez-Ezpeleta, H. Arrizabalaga, S. Post, and B. R. MacKenzie. 2021. "Atlantic Bluefin Tuna (*Thunnus thynnus*) in Greenland—Mixed-Stock Origin, Diet, Hydrographic Conditions, and Repeated Catches in This New Fringe Area." *Canadian Journal of Fisheries and Aquatic Sciences* 78, no. 4: 400–408.

Jérôme, M., C. Lemaire, J. M. Bautista, J. Fleurence, and M. Etienne. 2003. "Molecular Phylogeny and Species Identification of Sardines." *Journal of Agricultural and Food Chemistry* 51, no. 1: 43–50. https://doi.org/10.1021/jf020713w.

Kawamura, K., F. Parrenin, L. Lisiecki, et al. 2007. "Northern Hemisphere Forcing of Climatic Cycles in Antarctica Over the Past 360,000 Years." *Nature* 448, no. 7156: 912–916. https://doi.org/10.1038/nature06015.

Kimura, M. 1980. "A Simple Method for Estimating Evolutionary Rates of Base Substitutions Through Comparative Studies of Nucleotide Sequences." *Journal of Molecular Evolution* 16, no. 2: 111–120. https://doi.org/10.1007/BF01731581.

Korneliussen, T. S., A. Albrechtsen, and R. Nielsen. 2014. "ANGSD: Analysis of Next Generation Sequencing Data." *BMC Bioinformatics* 15, no. 1: 1–13. https://doi.org/10.1186/s12859-014-0356-4.

Kumar, S., M. Suleski, J. M. Craig, et al. 2022. "TimeTree 5: An Expanded Resource for Species Divergence Times." *Molecular Biology and Evolution* 39, no. 8. https://doi.org/10.1093/molbev/msac174.

Lavoué, S., M. Miya, K. Saitoh, N. B. Ishiguro, and M. Nishida. 2007. "Phylogenetic Relationships Among Anchovies, Sardines, Herrings and Their Relatives (Clupeiformes), inferred From Whole Mitogenome Sequences." *Molecular Phylogenetics and Evolution* 43, no. 3: 1096–1105. https://doi.org/10.1016/j.ympev.2006.09.018.

Lenoir, J., R. Bertrand, L. Comte, et al. 2020. "Species Better Track Climate Warming in the Oceans Than on Land." *Nature Ecology & Evolution* 4, no. 8: 1044–1059.

Li, H., and R. Durbin. 2009. "Fast and Accurate Short Read Alignment With Burrows-Wheeler Transform." *Bioinformatics* 25, no. 14: 1754–1760. https://doi.org/10.1093/bioinformatics/btp324.

Li, H., B. Handsaker, A. Wysoker, et al. 2009. "The Sequence Alignment/Map Format and SAMtools." *Bioinformatics* 25, no. 16: 2078–2079. https://doi.org/10.1093/bioinformatics/btp352.

Liu, J. X., A. Tatarenkov, T. D. Beacham, V. Gorbachev, S. Wildes, and J. C. Avise. 2011. "Effects of Pleistocene Climatic Fluctuations on the Phylogeographic and Demographic Histories of Pacific Herring (*Clupea pallasii*)." *Molecular Ecology* 20, no. 18: 3879–3893. https://doi.org/10.1111/j.1365-294X.2011.05213.x.

Liu, M., L. Lin, T. Gao, T. Yanagimoto, Y. Sakurai, and W. S. Grant. 2012. "What Maintains the Central North Pacific Genetic Discontinuity in Pacific Herring?" *PLoS One* 7, no. 12: e50340. https://doi.org/10.1371/journal.pone.0050340.

Lluch-Belda, D., R. J. M. Crawford, T. Kawasaki, et al. 1989. "World-Wide Fluctuations of Sardine and Anchovy Stocks: The Regime Problem." *South African Journal of Marine Science* 8, no. 1: 195–205. https://doi.org/10.2989/02577618909504561.

Love, M. S., J. J. Bizzarro, A. M. Cornthwaite, B. W. Frable, and K. P. Maslenikov. 2021. "Checklist of Marine and Estuarine Fishes From the Alaska–Yukon Border, Beaufort Sea, to Cabo San Lucas, Mexico." *Zootaxa* 5053, no. 1: 1–285.

Manni, M., M. R. Berkeley, M. Seppey, F. A. Simão, and E. M. Zdobnov. 2021. "BUSCO Update: Novel and Streamlined Workflows Along With Broader and Deeper Phylogenetic Coverage for Scoring of Eukaryotic, Prokaryotic, and Viral Genomes." *Molecular Biology and Evolution* 38, no. 10: 4647–4654. https://doi.org/10.1093/molbev/msab199.

Manni, M., M. R. Berkeley, M. Seppey, and E. M. Zdobnov. 2021. "BUSCO: Assessing Genomic Data Quality and Beyond." *Current Protocols* 1, no. 12: e323. https://doi.org/10.1002/cpz1.323.

Maximenko, N., J. Hafner, M. Kamachi, and A. MacFadyen. 2018. "Numerical Simulations of Debris Drift From the Great Japan Tsunami of 2011 and Their Verification With Observational Reports." *Marine Pollution Bulletin* 132: 5–25. https://doi.org/10.1016/j.marpolbul.2018.03.056.

Meisner, J., and A. Albrechtsen. 2018. "Inferring Population Structure and Admixture Proportions in Low-Depth NGS Data." *Genetics* 210, no. 2: 719–731. https://doi.org/10.1534/genetics.118.301336.

Muko, S., S. Ohshimo, H. Kurota, T. Yasuda, and M. A. Fukuwaka. 2018. "Long-Term Change in the Distribution of Japanese Sardine in the Sea of Japan During Population Fluctuations." *Marine Ecology Progress Series* 593: 141–154. https://doi.org/10.3354/meps12491.

Noor, M. A. F., K. L. Gratos, L. A. Bertucci, and J. Reiland. 2001. "Chromosomal Inversions and the Reproductive Isolation of Species." *Proceedings of the National Academy of Sciences of the United States of America* 98, no. 21: 12084–12088. https://doi.org/10.1073/pnas.221274498.

O'Leary, C. A., L. B. Defilippo, J. T. Thorson, et al. 2022. "Understanding Transboundary stocks' Availability by Combining Multiple Fisheries-Independent Surveys and Oceanographic Conditions in Spatiotemporal Models." *ICES Journal of Marine Science* 79, no. 4: 1063–1074. https://doi.org/10.1093/icesjms/fsac046.

Paradis, E., and K. Schliep. 2019. "Ape 5.0: An Environment for Modern Phylogenetics and Evolutionary Analyses in R." *Bioinformatics* 35, no. 3: 526–528. https://doi.org/10.1093/bioinformatics/bty633.

Parmesan, C., and G. Yohe. 2003. "A Globally Coherent Fingerprint of Climate Change Impacts Across Natural Systems." *Nature Cell Biology* 421: 37–42.

Parrish, R. H., R. Serra, and W. S. Grant. 1989. "The Monotypic Sardines, Sardina and Sardinops: Their Taxonomy, Distribution, Stock Structure, and Zoogeography." *Journal of Fisheries and Aquatic Science* 46: 2019–2036.

Pauly, D., and V. Christensen. 1995. "Primary Production Required to Sustain Global Fisheries." *Nature* 374, no. 6519: 255–257. https://doi.org/10.1038/374255a0.

Pearcy, W. G., J. P. Fisher, G. Anma, and T. Meguro. 1996. "Species Associations of Epipelagic Nekton of the North Pacific Ocean, 1978–1993." In *Fisheries Oceanography*, vol. 5, no. 1, 1–20. Blackwell Publishing Ltd. https://doi.org/10.1111/j.1365-2419.1996.tb00013.x.

Peck, M. A., P. Reglero, M. Takahashi, and I. A. Catalán. 2013. "Life Cycle Ecophysiology of Small Pelagic Fish and Climate-Driven Changes in Populations." *Progress in Oceanography* 116: 220–245.

Pikitch, E. K., K. J. Rountos, T. E. Essington, et al. 2014. "The Global Contribution of Forage Fish to Marine Fisheries and Ecosystems." *Fish and Fisheries* 15: 43–64. https://doi.org/10.1111/faf.12004.

Pinsky, M. L., R. L. Selden, and Z. J. Kitchel. 2020. "Climate-Driven Shifts in Marine Species Ranges: Scaling From Organisms to Communities." *Annual Review of Marine Science* 12, no. 1: 153–179.

Pinsky, M. L., B. Worm, M. J. Fogarty, J. L. Sarmiento, and S. A. Levin. 2013. "Marine Taxa Track Local Climate Velocities." *Science* 341, no. 6151: 1239–1242.

Pribyl, A. L., J. R. Hyde, L. Robertson, and R. Vetter. 2016. "Defining an Ideal Temperature Range for the Northern Subpopulation of Pacific Sardine, *Sardinops sagax* Caeruleus." *Environmental Biology of Fishes* 99, no. 2–3: 275–291. https://doi.org/10.1007/s10641-016-0473-1.

R Core Team. 2023. "*R: A Language and Environment for Statistical Computing*." https://www.R-project.org/.

Ranallo-Benavidez, T. R., K. S. Jaron, and M. C. Schatz. 2020. "GenomeScope 2.0 and Smudgeplot for Reference-Free Profiling of Polyploid Genomes." *Nature Communications* 11, no. 1: 1432. https://doi.org/10.1038/s41467-020-14998-3.

Renfree, J. S., A. Beittel, N. M. Bowlin, et al. 2023. "*Report on the Summer 2022 California Current Ecosystem Survey (CCES) (2207RL), 27 June to 30 September 2022, Conducted Aboard NOAA Ship Reuben Lasker, Fishing Vessels Lisa Marie and Long Beach Carnage, and Uncrewed Surface Vehicles*." https://doi.org/10.25923/66p9-hc28.

Renfree, J. S., N. M. Bowlin, B. E. Erisman, et al. 2022. "*Report on the Summer 2021 California Current Ecosystem Survey (CCES) (2107RL), 6 July to 15 October 2021, Conducted Aboard NOAA Ship Reuben Lasker, Mexican Research Vessel Dr. Jorge Carranza Fraser, Fishing Vessels Lisa Marie and Long Beach Carnage, and Uncrewed Surface Vehicles*." https://doi.org/10.25923/cadh-x974.

Rhie, A., B. P. Walenz, S. Koren, and A. M. Phillippy. 2020. "Merqury: Reference-Free Quality, Completeness, and Phasing Assessment for Genome Assemblies." *Genome Biology* 21, no. 1: 245. https://doi.org/10.1186/s13059-020-02134-9.

Robertson, D. R., and G. R. Allen. 2024. "*Shorefishes of the Tropical Eastern Pacific: Online Information System* (Vol. 3). Version."

Saitou, N., and M. Nei. 1987. "The Neighbor-Joining Method: A New Method for Reconstructing Phylogenetic Trees." *Molecular Biology and Evolution* 4: 406–425. https://doi.org/10.1093/oxfordjournals.molbev.a040454.

Sakamoto, T., T. Kodama, S. Horii, et al. 2023. "Geographic, Seasonal and Ontogenetic Variations of δ15N and δ13C of Japanese Sardine Explained by Baseline Variations and Diverse Fish Movements." *Progress in Oceanography* 219: 103163. https://doi.org/10.1016/j.pocean.2023.103163.

Sarr, O., R. Kindong, and S. Tian. 2021. "Knowledge on the Biological and Fisheries Aspects of the Japanese Sardine, *Sardinops melanostictus* (Schlegel, 1846)." In *Journal of Marine Science and Engineering*, vol. 9, no. 12. MDPI. https://doi.org/10.3390/jmse9121403.

Seppey, M., M. Manni, and E. M. Zdobnov. 2019. "BUSCO: Assessing Genome Assembly and Annotation Completeness." *Methods in Molecular Biology (Clifton, N.J.)*: 227–245. https://doi.org/10.1007/978-1-4939-9173-0_14.

Shi, Y., K. L. Bouska, G. J. McKinney, et al. 2021. "Gene Flow Influences the Genomic Architecture of Local Adaptation in Six Riverine Fish Species." *Molecular Ecology* 32, no. 7: 1549–1566.

Shi, Y., B. Kang, W. Fan, et al. 2023. "Spatio-Temporal Variations in the Potential Habitat Distribution of Pacific Sardine (*Sardinops sagax*) in the Northwest Pacific Ocean." *Fishes* 8, no. 2: 86. https://doi.org/10.3390/fishes8020086.

Skotte, L., T. S. Korneliussen, and A. Albrechtsen. 2013. "Estimating Individual Admixture Proportions From Next Generation Sequencing Data." *Genetics* 195, no. 3: 693–702. https://doi.org/10.1534/genetics.113.154138.

Tang, F., and W. Chen. 2021. "The Complete Mitochondrial Genome of *Sardinops sagax* (Jenyns, 1842) (Clupeiformes: Clupeidae) and Phylogenetic Analyses of Sardines." *Mitochondrial DNA Part B Resources* 6, no. 3: 1080–1081. https://doi.org/10.1080/23802359.2021.1899867.

Tarailo-Graovac, M., and N. Chen. 2009. "Using RepeatMasker to Identify Repetitive Elements in Genomic Sequences." *Current Protocols in Bioinformatics* 4: 4.10.1–4.10.14. https://doi.org/10.1002/0471250953.bi0410s25.

Thompson, A. R., N. J. Ben-Aderet, N. M. Bowlin, D. Kacev, R. Swalethorp, and W. Watson. 2022. "Putting the Pacific Marine Heatwave Into Perspective: The Response of Larval Fish Off Southern California to Unprecedented Warming in 2014–2016 Relative to the Previous 65 Years." *Global Change Biology* 28, no. 5: 1766–1785. https://doi.org/10.1111/gcb.16010.

Tran, B. 2023. "*Sardine and Anchovy Larvae Biogeography in the Southern California Current Ecosystem*." [Master Thesis]. UC San Diego.

Wellenreuther, M., and L. Bernatchez. 2018. "Eco-Evolutionary Genomics of Chromosomal Inversions." *Trends in Ecology & Evolution* 33, no. 6: 427–440. https://doi.org/10.1016/j.tree.2018.04.002.

Werb, B. E., and D. L. Rudnick. 2023. "Remarkable Changes in the Dominant Modes of North Pacific Sea Surface Temperature." *Geophysical Research Letters* 50, no. 4. https://doi.org/10.1029/2022GL101078.

Wilson, L. J., C. J. Fulton, A. M. Hogg, K. E. Joyce, B. T. M. Radford, and C. I. Fraser. 2016. "Climate-Driven Changes to Ocean Circulation and Their Inferred Impacts on Marine Dispersal Patterns." In *Global Ecology and Biogeography*, vol. 25, no. 8, 923–939. Blackwell Publishing Ltd. https://doi.org/10.1111/geb.12456.

Yang, C., H. Han, H. Zhang, et al. 2023. "Assessment and Management Recommendations for the Status of Japanese Sardine *Sardinops melanostictus* Population in the Northwest Pacific." *Ecological Indicators* 148: 110111. https://doi.org/10.1016/j.ecolind.2023.110111.

Yang, C., H. Zhang, H. Han, et al. 2023. "Spatio-Temporal Distribution of *Sardinops sagax* in the North Pacific: Optimal Environmental Characteristics." *Progress in Fishery Sciences* 44, no. 4: 99–110.

Yaremko, M. L. 1996. "Age Determination in Pacific Sardine, *Sardinops sagax*."

Yu, G., D. K. Smith, H. Zhu, Y. Guan, and T. T. Lam. 2017. "GGTREE: An R Package for Visualization and Annotation of Phylogenetic Trees With Their Covariates and Other Associated Data." *Methods in Ecology and Evolution* 8, no. 1: 28–36. https://doi.org/10.1111/2041-210X.12628.

## Supporting Information

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