Genome Resources

Reference Genome of the Black Surfperch, *Embiotoca jacksoni* (Embiotiocidae, Perciformes), a California Kelp Forest Fish That Lacks a Pelagic Larval Stage

Giacomo Bernardi*†, Jason A. Toy*†, Merly Escalona‡, Mohan P.A. Marimuthu‡, Ruta Sahasrabudhe, Oanh Nguyen, Samuel Sacco, Eric Beraut, Erin Toffelmier, Courtney Miller, and H. Bradley Shaffer‡

From the Department of Ecology and Evolutionary Biology, University of California–Santa Cruz, Santa Cruz, CA 95060, USA (Bernardi, Toy, Sacco, and Beraut); Department of Biomolecular Engineering, University of California–Santa Cruz, Santa Cruz, CA 95064, USA (Escalona); DNA Technologies and Expression Analysis Core Laboratory, Genome Center, University of California–Davis, Davis, CA 95616, USA (Marimuthu, Sahasrabudhe, and Nguyen); Department of Ecology and Evolutionary Biology, University of California–Los Angeles, Los Angeles, CA 90095-7239, USA (Toffelmier, Miller, and Shaffer); and La Kretz Center for California Conservation Science, Institute of the Environment and Sustainability, University of California–Los Angeles, Los Angeles, CA 90095-7239, USA (Shaffer).

*These authors contributed equally to the work.

Address correspondence to G. Bernardi at the address above, or e-mail: bernardi@ucsc.edu

Corresponding Editor: Arun Sethuraman

Abstract

Surfperches (Family Embiotocidae) are viviparous temperate reef fishes that brood their young. This life history trait translates into limited dispersal, strong population structure, and an unusually strong potential for local adaptation in a marine fish. As part of the California Conservation Genomics Project (CCGP), we sequenced the genome of the Black Surfperch, *Embiotoca jacksoni*, to establish a genomic model for understanding phylogeographic patterns of marine organisms in California. These patterns, in turn, may inform the design of marine protected areas using dispersal models based on genomic data. The genome of *E. jacksoni* is typical of marine fishes at less than 1Gb (genome size = 635 Mb), and our assembly is near-chromosome level (contig N50 = 6.5Mb, scaffold N50 = 15.5Mb, BUSCO = 98.1%). Within the context of the CCGP, the genome will be used as a reference for future whole genome resequencing projects aimed at enhancing our knowledge of the population structure of the species, and efficacy of Marine Protected Areas across the state.

Key Words: California Conservation Genomics Project, CCGP, Marine Protected Areas

Most marine fish exhibit a bipartite life history, with a pelagic larval stage, which typically lasts from a few days to a few weeks, and a more sedentary adult stage (Leis 1991). Because of the broad dispersal that often results from the larval stage, many marine fish have (or are presumed to have) relatively modest population structure across coastal seaescapes. In California, one family of nearshore fish, the surfperches (Embiotiocidae), lack a pelagic larval phase. Females brood approximately 10–50 embryos in a uterine pouch where the mother feeds the offspring via vascularized tissues before giving birth to fully developed juveniles that are ready to recruit to the parental habitat (Tarp 1952; Longo and Bernardi 2015). Surfperches comprise approximately 23 species, with 3 species found around Japan, 19 species in eastern Pacific coastal waters from Alaska, USA to Baja California, Mexico, and 1 freshwater species in California (Longo et al. 2018).

As a result of their unusual life history traits, surfperches exhibit low levels of dispersal, resulting in a high potential for local adaptation and strong within-species phylogeographic structure (Bernardi 2000, 2005; Johnson et al. 2016). Given this, surfperches can be used as predictors of phylogeographic breaks along the California coast, and may be important umbrella species that help optimize the design and boundaries of potential areas of marine conservation priorities. To further this important goal, which is one of the missions of the California Conservation Genomics Project (CCGP, Shaffer et al. 2022), we sequenced and assembled a near chromosome-level reference genome for the black surfperch, *Embiotoca jacksoni*, within the CCGP framework.

Black Surfperch are medium sized (usually about 25–30 cm TL) kelp forest fish that are found from Fort Bragg, Mendocino County, CA, to Punta Abreojos, Baja California, Mexico (Figure 1). Black surfperch are not IUCN listed, as they are currently neither endangered nor threatened. Although they are caught by recreational fishers, they are not specifically targeted and have not been the subject of a formal fisheries
assessment. A major genetic break in the Los Angeles region has previously been identified in this species based on a relatively limited set of mitochondrial genetic markers, indicating that northern and southern populations are genetically distinct management units (Bernardi 2000, 2005).

The assembled genome of *E. jacksoni* described here will serve as a valuable resource for studying the ecology, life history, adaptation, dispersal capability, and distribution dynamics of this ecologically and recreationally important species, as well as establish a useful model species for the study of evolutionary dynamics along the California Current Large Marine Ecosystem (CCLME).

Methods

Biological Materials

One adult male Black Surfperch, *E. jacksoni*, was collected by spear at Leo Carrillo State Beach (lat 34.0436° N, long 118.9338° W) in August 2020 by GB under California Department of Fish and Wildlife permit GM-201840006-20191-001 (Figure 1). The fish was dissected in the field, and liver, muscle, fin, and gill tissues were immediately placed in liquid nitrogen. Samples were later transferred to a –80 °C freezer at the lab until DNA extraction.

Omni-C Library Preparation

The Omni-C library was prepared using the Dovetail™ Omni-C™ Kit (Dovetail Genomics, CA) according to the manufacturer's protocol with slight modifications. First, specimen tissue was thoroughly ground with a mortar and pestle while cooled with liquid nitrogen. Subsequently, chromatin was fixed in place in the nucleus. The suspended chromatin solution was then passed through 100 and 40 μm cell strainers to remove large debris. Fixed chromatin was digested under various conditions of DNase I until a suitable fragment length distribution of DNA molecules was obtained. Chromatin ends were repaired and ligated to a biotinylated bridge adapter followed by proximity ligation of adapter-containing ends. After proximity ligation, crosslinks were reversed and the DNA purified from proteins. Purified DNA was treated to remove biotin that was not internal to ligated fragments. An NGS library was generated using an NEB Ultra II DNA Library Prep kit (NEB, Ipswich, MA) with an Illumina compatible y-adaptor. Biotin-containing fragments were then captured using streptavidin beads. The post-capture product was split into two replicates prior to PCR enrichment to preserve library complexity with each replicate receiving unique dual indices. The library was sequenced at Vincent J. Coates Genomics Sequencing Lab (Berkeley, CA) on an Illumina (San Diego, CA) NovaSeq platform to generate approximately 67 million reads.

PacBio HiFi Library Preparation and Sequencing

High molecular weight (HMW) DNA was extracted from 27 mg of liver tissue (#EJA LCO1 G) using a Nanobind Tissue Big DNA kit (Circulomics, Baltimore, MD) following the manufacturer's instructions. The purity of the DNA (260/280 = 1.87 and 260/230 = 2.24) was assessed on a NanoDrop spectrophotometer. DNA yield (130 ng/µl; 12.3 µg total) was quantified using Quans Fluorometer (QuantiFluor ONE dsDNA Dye assay, Promega, Madison, WI; cat. E6150). The integrity of the HMW DNA (>49% of the DNA fragments were > 50 kb) was estimated using the Femto Pulse system (Agilent Technologies, Santa Clara, CA).

The HiFi SMRTbell library was constructed using the SMRTbell Express Template Prep Kit v2.0 (Pacific Biosciences—PacBio, Menlo Park, CA; Cat. #100-938-900) according to the manufacturer's instructions. HMW gDNA was sheared to a target DNA size distribution between 15 and 20 kb. The sheared gDNA was concentrated using 0.45x of AMPure PB beads (PacBio Cat. #100-265-900) for the removal of single-strand overhangs at 37 °C for 15 min, followed by further enzymatic steps of DNA damage repair at 37 °C for 30 min, end repair and A-tailing at 20 °C for 10 min and 65 °C for 30 min, ligation of overhang adapter v3 at 20 °C for 60 min and 65 °C for 10 min to inactivate the ligase, then nuclease treated at 37 °C for 1 h. The SMRTbell library was purified and concentrated with 0.45x Ampure PB beads (PacBio Cat. #100-265-900) for size selection using the BluePippin system (Sage Science, Beverly, MA; Cat #BLF7510) to collect fragments greater than 9 kb. The 15–20 kb average HiFi SMRTbell library was sequenced at UC Davis DNA Technologies Core (Davis, CA) using one 8M SMRT cell, Sequel II sequencing chemistry 2.0, and 30-h movie on a PacBio Sequel II sequencer.

Nuclear Genome Assembly

We assembled the genome of the Black Surfperch following the CCGP assembly protocol Version 3.0 (Lin et al. 2022; Shaffer et al. 2022). We removed remnant adapter sequences from the PacBio HiFi dataset using HiFiAdapterFilt [Version 1.0] (Sim 2021, see Table 1 for assembly pipeline and relevant software) and generated the initial diploid assembly with the filtered PacBio reads using HiFiasm [Version 0.16.1-r375] (Cheng et al. 2021). The diploid assembly consists of two pseudo haplotypes (primary and alternate), where the primary assembly is more complete and consists of longer phased blocks, and...
the alternate consists of haplotigs (contigs with the same haplotype) in heterozygous regions, and is not as complete and more fragmented. Given the characteristics of the latter, it cannot be considered on its own but as a complement of the primary assembly (https://lh3.github.io/2021/04/17/concepts-in-phased-assemblies, https://www.ncbi.nlm.nih.gov/grc/help/definitions/).

Next, we identified sequences corresponding to haplotypic duplications and contig overlaps on the primary assembly with purge_dups [Version 1.2.6] (Guan et al. 2020) and transferred them to the alternate assembly. We scaffolded both assemblies using the Omni-C data with SALSA [Version 2.2] (Ghurye et al. 2019).

The primary assembly was manually curated by generating and analyzing Omni-C contact maps and breaking the assembly where major mis-assemblies were found. No further joins were made after this step. To generate the contact maps, we aligned the Omni-C data against the corresponding reference with bwa mem [Version 0.7.17-r1188, options-5SP] (Li 2013), identified ligation junctions, and generated Omni-C pairs using pairtools [Version 0.3.0] (Goloborodko et al. 2018). We generated a multi-resolution Omni-C matrix with Cooler [Version 0.8.10] (Abdennur and Mirny 2020) and balanced it with hicExplorer [Version 3.6] (Ramírez et al. 2018). We used HiGlass [Version 2.1.11] (Kerpedjiev et al. 2018) and the PretextSuite (https://github.com/wtsi-hpag/PretextView; https://github.com/wtsi-hpag/PretextMap; https://github.com/wtsi-hpag/PretextSnapshot) to visualize the contact maps.

We closed the remaining gaps generated during scaffolding with the PacBio HiFi reads and YAGCloser [commit 20e2769] (https://github.com/merlyescalona/yagcloser) to visualize the contact maps.

We then checked for contamination using the BlobToolKit Framework [Version 2.3.3] (Challis et al. 2020). Finally, we trimmed remnants of sequence adaptors and mitochondrial contamination based on NCBI contamination screening.

### Mitochondrial Genome Assembly

We assembled the mitochondrial genome of the black surfperch from the PacBio HiFi reads using the reference-guided

| Assembly                                             | Software                                              | Version         |
|------------------------------------------------------|-------------------------------------------------------|-----------------|
| Filtering PacBio HiFi adapters                       | HiFiAdapterFilt                                       | Commit 64d1c7b  |
| K-mer counting                                       | Meryl                                                 | 1               |
| Estimation of genome size and heterozygosity         | GenomeScope                                           | 2               |
| De novo assembly (contiging)                         | HiFiasm                                               | 0.16.1-r375     |
| Long read, genome-genome alignment                   | minimap2                                               | 2.16            |
| Remove low-coverage, duplicated contigs               | purge_dups                                            | 1.2.6           |
| Scaffolding                                          | Arima Genomics mapping pipeline                       | Commit 2e74ea4  |
| Omni-C mapping for SALSA                              | Arima Genomics mapping pipeline                       | Commit 2e74ea4  |
| Omni-C Scaffolding                                   | SALSA                                                 | 2               |
| Gap closing                                          | YAGCloser                                             | 20e2769         |
| Omni-C Contact map generation                        | bwa                                                   | 0.7.17-r1188    |
| Short-read alignment                                 | bwa                                                   | 0.7.17-r1188    |
| SAM/BAM processing                                   | samtools                                              | 1.11            |
| SAM/BAM filtering                                    | pairtools                                             | 0.3.0           |
| Pairs indexing                                       | pairix                                                | 0.3.7           |
| Matrix generation                                    | Cooler                                                | 0.8.10          |
| Matrix balancing                                     | hicExplorer                                           | 3.6             |
| Contact map visualization                            | HiGlass                                               | 2.1.11          |
|                                          | PretextMap                                             | 0.1.4           |
|                                          | PretextView                                            | 0.1.5           |
|                                          | PretextSnapshot                                        | 0.0.3           |
| Organelle assembly                                   | MitoHiFi                                              | 2 Commit c06ed3e|
| Mitogenome assembly                                  | MitoHiFi                                              | 2 Commit c06ed3e|
| Genome quality assessment                            | QUAST                                                 | 5.0.2           |
| Basic assembly metrics                               | BUSCO                                                 | 5.0.0           |
| Assembly completeness                                | Merqury                                               | 1               |
| Contamination screening                              | BLAST+                                                | 2.10            |
| Local alignment tool                                 | BlobToolKit                                           | 2.3.3           |
| General contamination screening                      | BlobToolKit                                           | 2.3.3           |
pipeline MitoHiFi (https://github.com/marcelauliano/MitoHiFi). The mitochondrial sequence of *E. jacksoni* (NC_029362.1) was used as the starting reference sequence. After completion of the nuclear genome, we searched for matches of the resulting mitochondrial assembly sequence in the nuclear genome assembly using BLAST + [Version 2.10] (Camacho et al. 2009) and filtered out contigs and scaffolds from the nuclear genome with a percentage of sequence identity >99% and size smaller than the mitochondrial assembly sequence.

Figure 2. Visual overview of genome assembly metrics. (A) K-mer spectra output generated from PacBio HiFi data without adapters using GenomeScope2.0. The bimodal pattern observed corresponds to a diploid genome. K-mers covered at lower coverage and lower frequency correspond to differences between haplotypes, whereas the higher coverage and higher frequency k-mers correspond to the similarities between haplotypes. (B) BlobTooKit Snail plot showing a graphical representation of the quality metrics presented in Table 2 for the *Embiotoca jacksoni* primary assembly (fEmbJac1). The plot circle represents the full size of the assembly. From the inside-out, the central plot covers length-related metrics. The red line represents the size of the longest scaffold; all other scaffolds are arranged in size-order moving clockwise around the plot and drawn in gray starting from the outside of the central plot. Dark and light orange arcs show the scaffold N50 and scaffold N90 values. The central light gray spiral shows the cumulative scaffold count with a white line at each order of magnitude. White regions in this area reflect the proportion of Ns in the assembly. The dark vs. light blue area around it shows mean, maximum and minimum GC vs. AT content at 0.1% intervals (Challis et al. 2020). (C–D) Omni-C Contact maps for the primary (C) and alternate (D) genome assembly generated with PretextSnapshot. Hi-C contact maps translate proximity of genomic regions in 3-D space to contiguous linear organization. Each cell in the contact map corresponds to sequencing data supporting the linkage (or join) between two of such regions.
Genome Size Estimation and Quality Assessment

We generated k-mer counts (k = 21) from the PacBio HiFi reads using meryl [Version 1] (https://github.com/marbl/meryl). The generated k-mer database was then used in GenomeScope2.0 [Version 2.0] (Ranallo-Benavidez et al. 2020) to estimate genome features including genome size, heterozygosity, and repeat content. To obtain general contiguity metrics, we ran QUAST [Version 5.0.2] (Gurevich et al. 2013). To evaluate genome quality and completeness we used BUSCO [Version 5.0.0] (Simão et al. 2015) with the actinopterygii ortholog database (actinopterygii_odb10) which contains 3640 genes. Assessment of base level accuracy (QV) and kmer completeness was performed using the previously generated meryl database and merqury (Rhie et al. 2020). We further estimated genome assembly accuracy via BUSCO gene set frameshift analysis using a pipeline previously described (Korlach et al. 2017).

Finally, using Repeat Masker (v4.1.2-p1) (Smit, Hubley, and Green) we identified the repeat content of the assembled sequence by running a slow search (-s parameter) with the species parameter set as “actinopterygii.”

Results

Mitochondrial Assembly

Final mitochondrial genome size was 16,515 bp. The base composition of the final assembly version is A = 34.62%, C = 26.37%, G = 12.65%, T = 26.34%, and consists of 22

| Bio Projects and Vouchers | NCBI Genome accessions | Assembly accession | Genome sequences |
|---------------------------|------------------------|--------------------|------------------|
| CCGP NCBI BioProject      | PRJNA720569            | GCA_022577435.1    | JAKOON000000000 |
| Genera NCBI BioProject    | PRJNA765818            |                   |                  |
| Species NCBI BioProject   | PRJNA777170            |                   |                  |
| NCBI BioSample            | SAMN24959158           |                   |                  |
|                            |                        |                   |                  |

| Genome Sequence           | Assembly identifier (Quality code) | HiFi Read coverage |
|---------------------------|------------------------------------|-------------------|
| PacBio HiFi reads         | fEmbJac1(6.6.Q54)                  | 42.57X            |
| OmnI-C Illumina reads     |                                    |                   |
|                           |                                    |                   |

| Genome Assembly Quality Metrics | Primary | Alternate |
|---------------------------------|---------|-----------|
| Number of contigs               | 463     | 63,422    |
| Contig N50 (bp)                 | 6 500  778 | 24 901   |
| Longest Contigs                 | 22 960 984 | 456 295  |
| Number of scaffolds             | 229     | 62 783    |
| Scaffold N50 (bp)               | 15 510 558 | 24 915   |
| Largest scaffold                | 26 794 396 | 1 021 151 |
| Size of final assembly (bp)     | 634 761 826 | 1 687 120 827 |
| Gaps per Gp                     | 369     | 1007      |
| Indel QV (Frame shift)          | 45.23   | 45.23     |
| Base pair QV                    | 54.7375 | 53.9269   |
|                                 | Full assembly = 54.134            |         |
| k-mer completeness              | 97.3237 | 97.8577   |
| BUSCO completeness (actinopterygii) |          |          |
|                                 | C       | S         | D         | F         | M         |
|                                 | P       | 98.10%   | 97.30%    | 0.80%     | 0.80%     | 1.10%     |
|                                 | A       | 96.10%   | 46.30%    | 49.80%    | 1.90%     | 2.00%     |
| Organelles                      | 1 complete mitochondrial sequence | JAKOON0100000230.1 |

aAssembly quality code x.y.Q derived notation, from (Rhie et al. 2021). x = log10[contig NG50]; y = log10[scaffold NG50]; Q = Phred base accuracy QV (Quality value). BUSCO Scores. (C)omplete and (S)ingle; (C)omplete and (D)uplicated; (F)ragmented and (M)issing BUSCO genes. n, number of BUSCO genes in the set/database. Bp: base pairs.

bRead coverage has been calculated based on a genome size of 634.7 Mb.
unique transfer RNAs and 13 protein coding genes. This was identical in size and organization to a previously published mitochondrial genome (Longo et al. 2016). The two genomes differed by 15 base pair substitutions, which included 12 transitions and 3 transversions.

Nuclear Assembly

We generated a de novo nuclear genome assembly of the black surfperch (fEmbJac1) using 67.3 million read pairs of Omni-C data and 1.5-million PacBio HiFi reads. The latter yielded ~42.57-fold coverage (N50 read length 15 459 bp; minimum read length 43 bp; mean read length 15 332 bp; maximum read length of 49 720 bp) based on the Genomescope2.0 genome size estimation of 634.7 Mb. The final genome size (634.7 Mb) is essentially identical to the estimated values from the Genomescope2.0 k-mer spectra. The k-mer spectrum output shows a bimodal distribution with two major peaks, at ~18 and ~39-fold coverage, where peaks correspond to homozygous and heterozygous states respectively of a diploid species (Figure 2A). Assembly statistics are reported in tabular and graphical form in Table 2 and Figure 2B, respectively.

The primary assembly consists of 229 scaffolds spanning 634.7 Mb with contig N50 of 6.5 Mb, scaffold N50 of 15.5 Mb, largest contig of 22.9 Mb, and largest scaffold of 26.7 Mb. The Omni-C contact map suggests that the primary assembly is highly contiguous (Figure 2C). As expected, the alternate assembly, which consists of sequence from heterozygous regions, is less contiguous (Figure 2D). Because the primary assembly is not fully phased, we have deposited scaffolds corresponding to the alternate haplotype in addition to the primary assembly.

Based on PacBio HiFi reads, we estimated 0.00217% sequencing error rate and 0.71% nucleotide heterozygosity rate. The assembly has a BUSCO completeness score of 98.1% using the actinopterygii gene set, a per base quality (QV) of 54.73, a k-mer completeness of 97.32% and a frameshift indel QV of 45.23.

In total, RepeatMasker identified 53 134 428 bp of repeat sequence (8.37% of the genome). Retroelements were estimated to make up 1.51% of the genome and DNA transposons were estimated to make up 2.12%. Simple repeats were the largest repeat group, making up 4.05% of the genome, while low complexity regions, satellites, and small RNA (rRNA, snRNA, tRNA) accounted for 0.45%, 0.04%, and 0.03%, respectively (Supplementary Table).

Discussion

Early genetic work on *E. jacksoni* has mostly dealt with its population genetics and taxonomy, where populations were shown to display very limited gene flow (Bernardi 2000, 2005), and where *E. jacksoni* was shown to belong to a genus that comprises 3 closely related species: *E. jacksoni*, *E. caryi*, and *E. lateralis*, which themselves belong to a larger group of rock dwelling, kelp forest associated species (Bernardi and Bucciarelli 1999; Bernardi 2009; Longo and Bernardi 2015; Longo et al. 2018). Early work suggested that the genome size was 1pg (based on c value where 1 pg–0.978 Gb) with karyotype of 2n = 48 (Hinegardner and Rosen 1972). The karyotype cited by Hinegardner and Rosen, however, was from a PhD thesis from Chen (1967), with this karyotype only presented as a table entry with no supporting information. A karyotype has been published for the Japanese surfperch species *Neoditrema ransonneti*, and is given as 2n = 48, but the authors claim that they “could not obtain clear chromosome figures sufficient for determining the karyotype of this species” (Arai and Yamamoto 1981). Finally, an ultracentrifugation analysis showed that the genome of *E. jacksoni* had an average GC content of 40.0% (Bucciarelli et al. 2002).

In this study, we have found that the genome size of *E. jacksoni* is 635 Mb, which is considerably lower than the c-value based estimates of Hinegardner and Rosen (1972). We also found that the size of the first 23 largest scaffolds slowly decreased in size in a regular fashion, with a difference between two adjacent scaffolds always being less than 9% of the scaffold length (average = 3.8%, Standard deviation = 3.0%). In contrast, the 24th scaffold is smaller than the 23rd scaffold by 27.1%, suggesting that it may not be a complete chromosome, but rather an incompletely assembled one, or even a contig to be included in one of the previous scaffolds (Figure 3, highlighted scaffolds). This would mean...
that the actual karyotype of *E. jacksoni* might be 2\(n = 46\). Taken together, the largest 23 scaffolds comprise 0.425 Gb, which corresponds to approximately 66.9% of the genome, so it is difficult to exclude the possibility that the actual karyotype is 2\(n = 48\). Further work establishing the karyotype for this species is clearly warranted. Finally, the genomic GC content was 41.6%, a value similar, but not identical to the 40.0% estimate based on ultracentrifugation analysis (Bucciarelli et al. 2002).

The high quality of the genome that we are presenting here (contig N50 = 6.5Mb, BUSCO completeness = 98.1%) will allow us to use it as a reference for the medium-coverage whole genome resequencing project for *E. jacksoni* that comprises the next phase of the CCGP data collection pipeline (Shafer et al. 2022). Our long-term goal is to draw a clear picture of the genetic boundaries between adjacent regions in California, as well as determine the degree of local adaptation among regions, and to use these data to delineate relevant protected areas that are grounded in strong genetic data. This genome is the first step in an important endeavor that will ultimately result in a sound protection plan for California’s natural marine resources.

**Supplementary Material**

Supplementary data are available at *Journal of Heredity* online.

**Funding**

Supported by the California Conservation Genomics Project; the State of California, State Budget Act of 2019 (UC Award ID RSI-19-690224) to the University of California.

**Acknowledgments**

We would like to thank Kristy Kroeker for help during the elaboration of the project, and Daniel Wright (UCSC) for help in the field during the collection of the sample and for discussions and help in the lab. PacBio Sequel II library preparation and sequencing was carried out at the DNA Technologies and Expression Analysis Cores at the UC Davis Genome Center, supported by NIH Shared Instrumentation Grant 1S10OD010786-01. Deep sequencing of Omni-C libraries used the NovaSeq S4 sequencing platforms at the Vincent J. Coates Genomics Sequencing Laboratory at UC Berkeley, supported by NIH S10 OD018174 Instrumentation Grant. We thank the staff at the UC Davis DNA Technologies and Expression Analysis Cores and the UC Santa Cruz Paleogenomics Laboratory for their diligence and dedication to generating high quality sequence data.

**Conflict of Interest**

The authors declare that by publishing this manuscript they have no conflicts of interest.

**Data Availability**

Data generated for this study are available under NCBI BioProject PRJNA806479. Raw sequencing data for sample EJA_LCO1_2020 (NCBI Bio Sample—SAMN24959158) are deposited in the NCBI Short Read Archive (SRA) under SRR18365597 for PacBio HiFi sequencing data and SRR18365595-6 for Omni-C Illumina Short read sequencing data. GenBank accessions for both primary and alternate assemblies are GCA_022577435.1 and GCA_022578405.1; and for genome sequences JAKOON000000000.1 and JAKOOO0000000000. The GenBank organelle genome assembly for the mitochondrial genome is JAKOON010000230.1. Assembly scripts and other data for the analyses presented can be found at the following GitHub repository: www.github.com/ccgproject/ccg_p, including estimated genome size, N50 (and/or k-mer) statistics for contigs and scaffolds, longest contigs, number of gaps, and BUSCO scores. This is also summarized in Table 2.

**References**

Abdennur N, Mirny LA. 2020. Cooler: scalable storage for Hi-C data and other genomically labeled arrays. Bioinformatics. 36: 311–316.

Arai R, Yamamoto T. 1981. Chromosomes of six species of percoid fishes from Japan. Bull Natl Sci Museum Tokyo, Ser A. 7:87–100.

Bernardi G. 2000. Barriers to gene flow in Embiotoca jacksoni, a marine fish lacking a pelagic larval stage. Evolution. 54:226–237.

Bernardi G. 2005. Phylogeography and demography of sympatric sister surfperch species, *Embiotoca jacksoni* and *E. lateralis* along the California coast: historical versus ecological factors. Evolution. 59:386–394.

Bernardi G. 2009. The name of the father: conflict between Louis and Alexander Agassiz and the Embiotoca surfperch radiation. J Fish Biol. 74:1049–1055. doi:10.1111/j.1095-8649.2008.02127.x

Bernardi G, Bucciarelli G. 1999. Molecular phylogeny and speciation of the surfperches (Embiotocidae, Perciformes). Mol Phylogenet Evol. 13:77–81.

Bucciarelli G, Bernardi G, Bernardi G. 2002. An ultracentrifugation analysis of two hundred fish genomes. Gene. 295:153–162.

Camacho C, Coulouris G, Avagyan V, Ma N, Papadopoulos J, Bealer K, Madden TL. 2009. BLAST+: architecture and applications. BMC Bioinf. 10:1–9.

Challis R, Richards E, Rajan J, Cochrane G, Blaxter M. 2020. BlobToolKit—interactive quality assessment of genome assemblies. G3 Genomes Genomes Genet. 10:1361–1374.

Chen TR. 2016. Comparative karyology of selected deep-sea and shallow water Teleost fishes. Ph.D. thesis, Yale Univ.

Cheng H, Jarvis ED, Fedrigo O, Koepfli KP, Urban L, Gemmell NJ, Li H. 2021. Robust haplotype-resolved assembly of diploid individuals without parental data. arXiv. http://arxiv.org/abs/2109.04785

Ghurye J, Rieh A, Walenz BP, Schmitt A, Selvaraj S, Pop M, Phillippy AM, Koren S. 2019. Integrating Hi-C links with assembly graphs for chromosome-scale assembly. PLoS Comput Biol. 15:e10072731–e10072719.

Goloborodko A, Abdennur N, Venev V, Brandao H, Fudenberg G. 2018. mirnylab/pairtools: v0.2.0. doi:10.5281/zenodo.1490831

Guan D, Guan D, McCarthy SA, Wood J, Howe K, Wang Y, Durbin R, Durbin R. 2020. Identifying and removing haplotypic duplication in primary genome assemblies. Bioinformatics. 36:2896–2898.

Gurevich A, Saveliev V, Vyahhi N, Tesler G. 2013. QUAST: quality assessment tool for genome assemblies. Bioinformatics. 29:1072–1075.

Hinagardner R, Rosen DE. 1972. The cytology, morphology and evolution of teleostean fishes. Am Nat. 106:621–644.

Johnson DW, Freiwald J, Bernardi G. 2016. Genetic diversity affects the strength of population regulation in a marine fish. Ecology. 97:627–639.

Kerpedjiev P, Abdennur N, Lekschas F, McCullum C, Dinkla K, Strobel H, Luber JM, Ouellette SB, Azhir A, Kumar N, et al. 2018. HiGlass: web-based visual exploration and analysis of genome interaction maps. Genome Biol. 19:1–12.
Korlach J, Gedman G, Kingan SB, Chin CS, Howard JT, Audet JN, Cantin L, Jarvis ED. 2017. De novo PacBio long-read and phased avian genome assemblies correct and add to reference genes generated with intermediate and short reads. *GigaScience*. 6:1–16.

Leis J. 1991. The pelagic stage of reef fishes. In: Sales P, editor. *The ecology of fishes on coral reefs*. San Diego (CA): Academic Press Inc. p. 182–229.

Li H. 2013. Aligning sequence reads, clone sequences and assembly contigs with BWA-MEM. arXiv. [http://arxiv.org/abs/1303.3997](http://arxiv.org/abs/1303.3997)

Lin M, Escalona M, Sahasrabudhe R, Nguyen O, Beraut E, Buchalski MR, Wayne RK. 2022. A reference genome assembly of the bobcat, *Lynx rufus*. *J Hered*. 113:615–623.

Longo G, Bernardi G. 2015. The evolutionary history of the embiotocid surfperch radiation based on genome-wide RAD sequence data. *Mol Phylogenet Evol*. 88:55–63.

Longo GC, Bernardi G, Lea RN. 2018. Taxonomic revisions within Embiotocidae (Teleostei, Perciformes) based on molecular phylogenetics. *Zootaxa*. 4482:591–596.

Longo GC, O’Connell B, Green RE, Bernardi G. 2016. The complete mitochondrial genome of the black surfperch, *Embiotoca jacksoni*: selection and substitution rates among surfperches (Embiotocidae). *Mar Genomics*. 28:107–112. doi:10.1016/j.margen.2016.03.006

Ramírez F, Bhardwaj V, Arrigoni I, Lam KC, Grüning BA, Villaveces J, Habermann B, Akhtar A, Manke T. 2018. High-resolution TADs reveal DNA sequences underlying genome organization in flies. *Nat Commun*. 9:189 doi:10.1038/s41467-017-02525-w

Ranallo-Benavidez TR, Jaron KS, Schatz MC. 2020. GenomeScope 2.0 and Smudgeplot for reference-free profiling of polyploid genomes. *Nat Commun*. 11:1432. doi:10.1038/s41467-020-14998-3

R.Misc. 2021. sheinasim/HiFiAdapterFilt: first release. doi:10.5281/zenodo.4716418

Simão FA, Waterhouse RM, Ioannidis P, Kriventseva EV, Zdobnov EM. 2015. BUSCO: assessing genome assembly and annotation completeness with single-copy orthologs. *Bioinformatics*. 31:3210–3212.

Tarp FH. 1952. A revision of the family Embiotocidae (the surfperches). *Fish Bulletin No. 88*. State of California Fish and Game, Sacramento, California, USA.