Genome Resources

Reference Genome of the California Sheephead, *Semicossyphus pulcher* (Labridae, Perciformes), A Keystone Fish Predator in Kelp Forest Ecosystems

Giacomo Bernardi®, Melissa DeBiasse, Merly Escalona®, Mohan P. A. Marimuthu®, Oanh Nguyen, Samuel Sacco, Eric Beraut, Courtney Miller, Erin Toffelmier, and H. Bradley Shaffer®

From the Department of Ecology and Evolutionary Biology, University of California Santa Cruz, CA 95060, USA (Bernardi, Sacco and Beraut); School of Natural Sciences, University of California Merced, CA 95343, USA (DeBiasse); Department of Biomolecular Engineering, University of California Santa Cruz, CA 95064, USA (Escalona); DNA Technologies and Expression Analysis Core Laboratory, Genome Center, University of California-Davis, CA 95616, USA (Marimuthu and Nguyen); Department of Ecology & Evolutionary Biology, University of California, Los Angeles, CA 90095-7239, USA (Miller, Toffelmier and Bradley Shaffer); and La Kretz Center for Conservation Science, Institute of the Environment and Sustainability, University of California, Los Angeles, CA 90095-7239, USA (Bradley Shaffer).

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

Corresponding Editor: William Murphy

Abstract

Keystone species are known to play a critical role in kelp forest health, including the well-known killer whales, sea otter, sea urchin, kelp trophic cascade in the Aleutian Islands, Alaska, USA. In California, a major player in the regulation of sea urchin abundance, and in turn, the health of kelp forests ecosystems, is a large wrasse, the California Sheephead, *Semicossyphus pulcher*. We present a reference genome for this ecologically important species that will serve as a key resource for future conservation research of California’s inshore marine environment utilizing genomic tools to address changes in life-history traits, dispersal, range shifts, and ecological interactions among members of the kelp forest ecological assemblages. Our genome assembly of *S. pulcher* has a total length of 0.794 Gb, which is similar to many other marine fishes. The assembly is largely contiguous (N50 = 31.9 Mb) and nearly complete (BUSCO single-copy core gene content = 98.1%). Within the context of the California Conservation Genomics Project (CCGP), the genome of *S. pulcher* will be used as an important reference resource for ongoing whole genome resequencing efforts of the species.

Key words: California Conservation Genomics Project, CCGP

Marine ecosystems are experiencing unprecedented environmental change (Oliver et al. 2018), which has already resulted in widespread shifts in species distributions and patterns of connectivity (Sanford et al. 2019). Many marine algae, invertebrates, and vertebrates exist within well-characterized networks of ecological interactions (Burt et al. 2018), and the health of both marine ecosystems and commercial fisheries depends on those interactions remaining intact. In California, where the coastline is primarily oriented along a north–south axis, climate change is likely to result in a northward shift in the distribution range of marine species.

The California Conservation Genomics Project (CCGP, Shaffer et al. 2022) is a large, multi-investigator initiative that uses the inferences derived from landscape, including seascape, genomics to document current patterns of genomic variation across 235 species of plants and animals, including 29 marine species. Here, we present the reference genome of the kelp forest keystone fish species, the California Sheephead as part of the CCGP initiative.

Ray-finned fishes comprise more than 20,000 species of fishes, and include the majority of coral reef fishes. The family Labridae (wrasses and parrotfishes) include more than 600 species, primarily found on coral reefs, but also in semi-tropical and temperate reefs. The California Sheephead, *Semicossyphus pulcher*, is a large, protogynous hermaphroditic wrasse (Poortvliet et al. 2013). California Sheephead feed on invertebrates, with a preference for purple urchin, *Strongylocentrotus purpuratus* (Hamilton et al. 2007), which, in turn, play an essential role in regulating the abundance of giant kelp, *Macrocystis pyrifera* (all three species being investigated within the CCGP framework). The recent decline of California kelp forests due to the explosion of urchin populations underscores the essential role of urchin regulators such as California Sheephead (Smith et al. 2020).
In California, younger females and larger males are targeted by live fish fisheries and anglers/spearfishers, respectively (Hamilton et al. 2007). These biased harvesting practices, when applied to a sequential hermaphrodite like the California Sheephead, have the potential to fundamentally alter life-history characteristics and population stability. In this case, preferentially targeting younger females (for live-fish fisheries) and larger males (for recreational spear and pole-fishing) tends to accelerate the shift of females into males at smaller sizes, making them less able to protect and mate with large harem, and negatively affecting kelp forest dynamics.

Historically, California Sheephead have predominantly been found in southern California, south of Point Conception (Poortvliet et al. 2013; Love and Passarelli 2020). However, in recent years, records of California Sheephead north of Point Conception, with established populations in Monterey Bay, have been increasing. Here, we present a chromosome-scale reference genome for *S. pulcher*. This genome assembly is a critical resource for ongoing and future analyses of the genomic underpinnings of the ecology, life history, dispersal capability, and distribution dynamics of this keystone species.

**Methods**

**Biological Materials**

One adult female California Sheephead, *S. pulcher*, was collected by spear at Leo Carrillo State Park, Los Angeles County, California (N 34.0436 W −118.9338) in August 2020 by GB under California Department of Fish and Wildlife permit GM-20184006-20191-001 (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 until DNA extraction.

**PacBio HiFi Library Preparation and Sequencing**

High molecular weight (HMW) DNA was extracted from 62 mg of fin tissue (SPULCO1G) using the Nanobind Tissue Big DNA kit (Pacific BioSciences), following the manufacturer’s instructions. We assessed DNA purity using absorbance ratios (260/280 = 1.80 and 260/230 = 2.29) on a NanoDrop ND-1000 spectrophotometer. We quantified DNA yield (310 ng/μl; 58 μg total) using the Quantus Fluorometer (QuantiFluor ONE dsDNA Dye assay, Promega). We estimated the size distribution of the HMW DNA using the Femto Pulse system (Agilent) and found that >45% of the DNA fragments were >50Kb.

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 genomic DNA (gDNA) was sheared to a target DNA size distribution between 15 kb and 20 kb. The sheared gDNA was concentrated using 0.45X of AMPure PB beads (PacBio Cat. #100-265-900) to remove 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-hour movie on a PacBio Sequel II sequencer.

**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. Specimen tissue was thoroughly ground with a mortar and pestle in liquid nitrogen, followed by in situ chromatin fixation. The suspended chromatin solution was then passed through 100 μm 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, cross-links were reversed, and the DNA purified from proteins and treated to remove biotin that was not internal to ligated fragments. A 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.

---

**Figure 1.** Distribution (dark area) of California Sheephead, *Semicossyphus pulcher*. California Sheephead are found on rocky reefs of California, USA, and Baja California, Mexico, including the isolated offshore Guadalupe island (represented as a dot), and the Sea of Cortez. The collection site of the sequenced individual, Leo Carrillo State Beach, is indicated by the black star on the map. Drawings represent male (top) and female (bottom) California Sheephead (artwork by Amadeo Bachar, www.abachar.com).
Nuclear Genome Assembly
We assembled the genome of the California Sheephead following the CCGP assembly protocol Version 3.0 (Lin et al. unpublished data; Shaffer et al. 2022). The final output corresponds to a diploid assembly that consists of two pseudo haplotypes (primary and alternate). The primary assembly is more complete and consists of longer phased blocks. The alternate consists of haplotigs (contigs that come from 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/)

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). 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 if major misassemblies 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; Table 1. Assembly pipeline and software usage. Software citations are listed in the text

| 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 |
| Omni-C mapping for SALSA | Arima Genomics mapping pipeline | Commit 2c74ea4 |
| Omni-C Scaffolding | SALSA | 2 |
| Gap closing | YAGCloser | Commit 20e2769 |
| Omic-C Contact map generation | 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 | MitroHiFi | 2 Commit c06ed3e |
| Genome quality assessment | QUAST | 5.0.2 |
| Assembly completeness | BUSCO | 5.0.0 |
| | Merqury | 1 |
| Contamination screening | BLAST+ | 2.10 |
| Local alignment tool | BlobToolKit | 2.3.3 |
Mitochondrial Genome Assembly

We assembled the mitochondrial genome of the California Sheeprhead from the PacBio HiFi reads using the reference-guided pipeline MitoHiFi (https://github.com/marcelauliano/MitoHiFi). The mitochondrial sequence of Thalassoma lunare (NC_048980), another member of the Labridae, 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.

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 mercury (Rhi et al. 2020). We further estimated genome assembly accuracy via BUSCO gene set frameshift analysis using a pipeline previously described (Korlach et al. 2017).

Results

Mitochondrial Assembly

Final mitochondrial genome size was 16 549 bp. The base composition of the final assembly version is A = 27.32%, C = 29.64%, G = 17.75%, T = 25.28%, and consists of 22 unique transfer RNAs and 13 protein coding genes. This is similar in organization to the mitochondrial genome of the wrasse Thalassoma lunare (Kukai et al. 2019). The genome of Thalassoma lunare is 524 bp larger than S. pulcher, with much of the difference in size in the highly variable control region. The remainder of the mt genome was 15 606 and 15 781 bp for S. pulcher and T. lunare, respectively, indicating that the two congeneric species differed by only 175 bp in length. Not including the control region, the two genomes differed by 3745 bp substitutions, which included 2045 transitions (1289 Y, 756 R), and 1700 transversions (755 M, 551 W, 208 S, 186 K). The observed sequence divergence of 23.7% between S. pulcher (tribe Hypsigenyines) and T. lunare (tribe Julidines) is consistent with the tribe-level divergence of 22% reported by Westneat and Alfaro (2005).

Nuclear Assembly

We generated a de novo nuclear genome assembly of the California Sheeprhead (fSemPu1) using 67.3 million read pairs of Omni-C data and 1.5 million PacBio HiFi reads. The latter yielded ~54.4 fold coverage (N50 read length 1459 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 794.1 Mb. We only closed 1 gap, and no further sequences were introduced. This final genome size is very similar to that estimated 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). We did not find any major misassemblies in the primary assembly as it was generated from the scaffold. Assembly statistics are reported in tabular and graphical form in Table 2 and Figure 2B, respectively.

The primary assembly consists of 179 scaffolds spanning 794.1 Mb with contig N50 of 31 948 211 bp, a nearly identical scaffold N50 of 32 091 781 bp, longest contig of 38.5 Mb, and largest scaffold of 38.5 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 98.7% using the Actinopterygii gene set, a per base quality (QV) of 67.8, a kmer completeness of 94.33%, and a frameshift indel QV of 48.9.

Discussion

Early genetic work on S. pulcher dealt with its population genetics (Bernardi et al. 2003; Poortvliet et al. 2013), and taxonomy, where Semicossyphus was shown to belong to the Hypsigenyine tribe, an early branch in the wrasse (Labridae) phylogenetic tree, in a group that includes hogfishes (genus Bodians) and creole wrasses (genus Clepticus) (Westneat and Alfaro 2005; Beldade et al. 2009). Genome size has only been reported for other wrasses using cytological methods (0.91–0.98 pg, equivalent to ~0.890-0.958 Gb, Hinegardner and Rosen 1972), and karyotypes are also only known for other wrasses (the majority being 2n = 48, including the closely related hogfishes) (Molina et al. 2012; Almeida et al. 2017). An ultracentrifugation analysis showed that the genome of S. pulcher had an average GC content of 40.8% (Bucciarelli et al. 2002).

In this study, we have found that the genome of S. pulcher is 0.794 Gb, a result that is consistent with these earlier c-value estimates. The sizes of the 24 largest scaffolds in the current assembly change from one contig to the next in smaller increments than the remaining scaffolds (Figure 3), consistent with the karyotype of the closely related genus Bodians.
with \(2n = 48\) chromosomes. The largest 24 scaffolds comprise 0.721 Gb, which corresponds to approximately 91% of the genome. This suggests that the genome presented here is very close to a chromosome-level assembly. The GC content was 41.6%, a value similar to that reported by Bucciarelli et al. (2002). Finally, we note that the contig and scaffold N50 values reported here are nearly identical, consistent with the overall excellent quality of the HiFi reads for this species.

The high quality of the genome we are presenting here (contig N50 = 32.1 Mb, BUSCO = 98.7%) will be an important reference for the medium-coverage, whole genome resequencing projects underway to evaluate population

**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 frequency correspond to differences between haplotypes, whereas the higher coverage and higher frequency k-mers correspond to the similarities between haplotypes. (B) BlobToolKit Snail plot showing a graphical representation of the quality metrics presented in Table 2 for the *Semicossyphus pulcher* primary assembly (fSemPul1). The plot circle represents the full size of the assembly. From the inside-out, the central plot covers length-related metrics. The line represents the size of the longest scaffold; all other scaffolds are arranged in size-order moving clockwise around the plot and drawn in grey starting from the outside of the central plot. Dark and light arcs show the scaffold N50 and scaffold N90 values. The central light grey 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 versus light areas around it shows mean, maximum, and minimum GC versus 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.
genomic variation of *S. pulcher* across California in the next phase of the CCGP (Shaffer et al. 2022). Our long-term goal is to draw a clear picture of the genetic boundaries between potential management units in California, and design relevant protected areas supported by strong genetic data. Assembling high-contiguity reference genomes is a critical step in this important endeavor that will ultimately result in sound protection plans for California’s marine resources.

**Funding**

This work was supported by the California Conservation Genomics Project, with funding provided to the University
of California by the State of California, State Budget Act of 2019 [UC Award ID RSI-19-690224].

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

Acknowledgments
This study is a contribution of the Marine Networks Consortium (PIs Michael N Dawson, Rachael A. Bay) funded by the California Conservation Genomics Project (PI: H. Bradley Shaffer). We would like to thank Daniel Wright (UCSC) for help in the field during collection of the sample and for discussions and help in the lab. PacBio Sequel II library prep and sequencing was carried out at the DNA Technologies and Expression Analysis Cores at the UC Davis Genome Center, supported by National Institutes of Health 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 National Institutes of Health (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.

Data Availability
Data generated for this study are available under NCBI BioProject PRJNA763860. Raw sequencing data for sample SPU_LCO1_2020 (NCBI BioSamples SAMN25656429, SAMN25656430) are deposited in the NCBI Short Read Archive (SRA) under SRR18540358 for PacBio HiFi sequencing data and SRR18540356-7 for Omni-C Illumina Short read sequencing data. GenBank accessions for both sequencing data and SRR18540356-7 for Omni-C Illumina Short read sequencing data. GenBank accessions for both primary and alternate assemblies are GCA_022749685.1 and GCA_022749735.1; and for genome sequences JAKSZQ0000000000 and JAKSZR0000000000. The GenBank organelle genome assembly for the mitochondrial genome is JAKSZQ010000179.1. Assembly scripts and other data for the analyses presented can be found at the following GitHub repository: www.github.com/ccgproject/ccgp_assembly, 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.

Almeida LAH, Nunes LA, Bitencourt JA, Molina WF, Affonso PRAM. 2017. Chromosomal evolution and cytotaxonomy in wrasses (perciformes; Labridae). J Hered. 108:239–253.

Beldade R, Heiser JB, Robertson DR, Gasparini JL, Floeter SR, Bernardi G. 2009. Historical biogeography and speciation in the creole wrasses (labridae, clepticus). Mar Biol. 156:679–687.

Bernardi G, Findley L, Rocha-Olivares A. 2003. Vicariance and dispersal across Baja California in disjunct marine fish populations. Evolution (N Y). 57:1599–1609.

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

Burt JM, Tim Tinker M, Okamoto DK, Demes KW, Holmes K, Salomon AK. 2018. Sudden collapse of a mesopredator reveals its complementary role in mediating rocky reef regime shifts. Proc R Soc B Biol Sci. doi:10.1098/rspb.2018.0533.

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.

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

Ghurie J, Riche A, Walenz BP, Schmitt A, Selvaraj S, Pop M, Philippy AM, Koren S. 2019. Integrating Hi-C links with assembly graphs for chromosome-scale assembly. PLoS Comput Biol. 15:e10072731–e10072719. doi:10.1371/journal.pcbi.1007273.

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

Guán D, Guán D, McCarthy SA, Wood J, Howe K, Wang Y, Durbin R, Durbin R. 2020. Identifying and removing haplotype 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.

Hamilion SL, Caselle JE, Standish JD, Schroeder DM, Love MS, Rosaless-Casian JA, Sosa-Nishizaki O. 2007. Size-selective harvesting alters life histories of a temperate sex-changing fish. Ecol Appl. 17:2268–2280.

Hinegardner R, Rosen DE. 1972. Cellular DNA content and the evolution of teleostean fishes. Am Nat. 106:621–644.

Kerpedjiev P, Abdennur N, Lekshas F, McCullam 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, Kirvan 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.

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

Love MS, Passarelli JK. 2020. Miller and Leci’s Guide to the coastal marine fishes of California. University of California Press. Berkeley, California, USA.

Molina WF, Motta Neto CC, Senna DC, Cioffi MB, Bertollo LAC. 2012. Karyoevolutionary aspects of Atlantic hogfishes (Labridae-Bodianinae), with evidence of an atypical decondensed argentophilic heterochromatin. Mar Genomics. 6:25–31.

Oliver ECJ, Donat MG, Burrows MT, Moore PJ, Smale DA, Alexander AM, Koren S. 2019. Integrating Hi-C links with assembly graphs for chromosome-scale assembly. GigaScience. 10:1–9.

Porpviit M, Longo GC, Selkoe K, Barber PH, White C, Caselle JE, Perez-Matus A, Gaines SD, Bernardi G. 2013. Phylogeography of the California sheephead, Semicossyphus pulcher: The role of deep reefs as stepping stones and pathways to antitropicality. Ecol Evol. 3:4538–4571.

Ramirez F, Bhaward V, Arrigoni L, 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. 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. doi:10.1038/s41467-020-14998-3.
Rhie A, Walenz BP, Koren S, Phillippy AM. 2020. Merqury: reference-free quality, completeness, and phasing assessment for genome assemblies. *Genome Biol.* 21:1–27. doi:10.1186/s13059-020-02134-9.

Sanford E, Sones JL, Garcia-Reyes M, Goddard JHR, Largier J. 2019. Widespread shifts in the coastal biota of northern California during the 2014–2016 marine heatwaves. *Sci Rep.* 9:1–14. doi:10.1038/s41598-019-40784-3.

Shaffer HB, Toffelmier E, Corbett-Detig RB, Escalona M, Erickson B, Fiedler P, Gold M, Harrigan RJ, Hodges S, Luckau TK, et al. 2022. Landscape genomics to enable conservation actions: The California Conservation Genomics Project. *J Hered.* 113:577–588.

Sim S. 2021. sheinasim/HiFiAdapterFilt: first release. doi: 10.5281/zenodo.4716418.

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

Smith JG, Tomoleoni J, Staedler M, Lyon S, Fujii J, Tinker MT. 2021. Behavioral responses across a mosaic of ecosystem states restructure a sea otter-urchin trophic cascade. *Proc Natl Acad Sci USA.* doi:10.1073/pnas.2012493118.

Westneat MW, Alfaro ME. 2005. Phylogenetic relationships and evolutionary history of the reef fish family Labridae. *Mol Phylogenet Evol.* 36:370–390.

Yukai Y, Xiaolin H, Heizhao L, Tao L, Wei Y, Zhong H. 2019. The complete mitochondrial genome of Thalassoma lunare (Labriformes, Labridae). *Mitochondrial DNA Part B Resour.* 4:3147–3148.