Title

Chromosome-level genome of the three-spot damselfish, Dascyllus trimaculatus

Journal

G3: Genes, Genomes, Genetics, 13(4)

ISSN

2160-1836

Authors

Roberts, May B
Schultz, Darrin T
Gatins, Remy
et al.

Publication Date

2023-04-11

DOI

10.1093/g3journal/jkac339

Peer reviewed

G3, 2023, 13(4), jkac339
https://doi.org/10.1093/g3journal/jkac339
Advance Access Publication Date: 11 March 2023
Genome Report

Chromosome-level genome of the three-spot damselfish, *Dascyllus trimaculatus*

May B. Roberts [1], Darrin T. Schultz,[2,3,4,*] Remy Gatins [5], Merly Escalona [4], Giacomo Bernardi [1,*]

[1]Department of Ecology and Evolutionary Biology, University of California, Santa Cruz, Santa Cruz, CA 95060, USA
[2]Department of Molecular Evolution and Development, University of Vienna, Vienna 1010, Austria
[3]Monterey Bay Aquarium Research Institute, Moss Landing, CA 95039, USA
[4]Department of Biomolecular Engineering and Bioinformatics, University of California, Santa Cruz, Santa Cruz, CA 95060, USA
[5]Department of Marine Sciences, Northeastern University, Boston, MA 02115, USA

*Corresponding authors: Department of Neurosciences and Developmental Biology, University of Vienna, Universitätsring 1, 1010 Vienna, Austria. Email: darrin.schultz91@gmail.com; 115 McAllister Way, Santa Cruz CA 93117, USA. Email: bernardi@ucsc.edu

Abstract

Damselfishes (Family: Pomacentridae) are a group of ecologically important, primarily coral reef fishes that include over 400 species. Damselfishes have been used as model organisms to study recruitment (anemonefishes), the effects of ocean acidification (spiny damselfish), population structure, and speciation (*Dascyllus*). The genus *Dascyllus* includes a group of small-bodied species, and a complex of relatively larger bodied species, the *Dascyllus trimaculatus* species complex that is comprised of several species including *D. trimaculatus* itself. The three-spot damselfish, *D. trimaculatus*, is a widespread and common coral reef fish species found across the tropical Indo-Pacific. Here, we present the first-genome assembly of this species. This assembly contains 910 Mb, 90% of the bases are in 24 chromosome-scale scaffolds, and the Benchmarking Universal Single-Copy Orthologs score of the assembly is 97.9%. Our findings confirm previous reports of a karyotype of $2n = 47$ in *D. trimaculatus* in which one parent contributes 24 chromosomes and the other 23. We find evidence that this karyotype is the result of a heterozygous Robertsonian fusion. We also find that the *D. trimaculatus* chromosomes are each homologous with single chromosomes of the closely related clownfish species, *Amphiprion percula*. This assembly will be a valuable resource in the population genomics and conservation of Damselfishes, and continued studies of the karyotypic diversity in this clade.

Keywords: hybrid genome assembly, Robertsonian polymorphism, chromosome fusion, domino damselfish, ONT, Hi-C Chicago, illumina shotgun, coral reef fish, Pomacentridae

Introduction

Damselfishes (Pomacentridae) are a group of small-bodied species found across all coral reef regions and most temperate marine systems where they are often the most visibly abundant fishes on the reef (Hiatt and Strasburg 1960; Allen 1991; Allen and Werner 2002; Bellwood and Wainwright 2002). This family includes more than 400 species that, despite their small size (max 30 cm), play important ecological roles (Allen 1991; Tang *et al.* 2021). Within this large family, the genus *Dascyllus* comprises 11 species, four of which, make up the *Dascyllus trimaculatus* species complex. This species complex includes three described species with restricted geographic ranges, *D. albisella* in the Hawaiian Islands, *D. strasburgi* in the Marquesas Islands, and *D. auripinnis* in the Line Islands. In contrast, *D. trimaculatus* has the broadest range, extending from the Red Sea, where it was first described (Rüppell,1828), across the tropical and subtropical Indo-Pacific (Fig. 1).

Three-spot damselfish is an abundant and common species, which exhibits a typical bipartite life history, with a site-attached adult phase, where mate pairs lay, fertilize, and care for demersal eggs, followed by a pelagic larval phase. Larvae hatch after ~6 days and feed in the water column on zooplankton where their pelagic larval duration lasts 23–30 days until they recruit back to the reef (Wellington and Victor 1989; Robitzch *et al.* 2016). Larvae settle primarily into anemones for protection often sharing this shelter with different species of the popular anemonefish (in Hawai'i, where anemonefishes and anemones are absent, *D. albisella* recruits to branching coral). As subadults, they leave the anemone and live nearby in small to large groups.

There has also been considerable effort in understanding the chromosomes architecture and variation of *Dascyllus* and other damselfishes. Chromosome number varies between species of *Dascyllus* as well as within species (Ojima and Kashiwagi 1981; Kashiwagi *et al.* 2005; Getlekha *et al.* 2017) giving insight into chromosomal drivers of evolution (Galetti *et al.* 2000; Hardie and Hebert 2004; Molina and Galetti 2004) and how this variation is manifested ecologically (Molina and Galetti 2004; Martinez *et al.* 2015). As we shift into the age of genomic natural history where genomic tools offer vastly more detail and statistical power, a reference genome will aid in further refining our understanding of wildlife biology (Hotaling *et al.* 2021). There are currently 14 Pomacentrid reference genomes, five of which are publicly available through the National

Received: June 24, 2022. **Accepted:** September 14, 2022

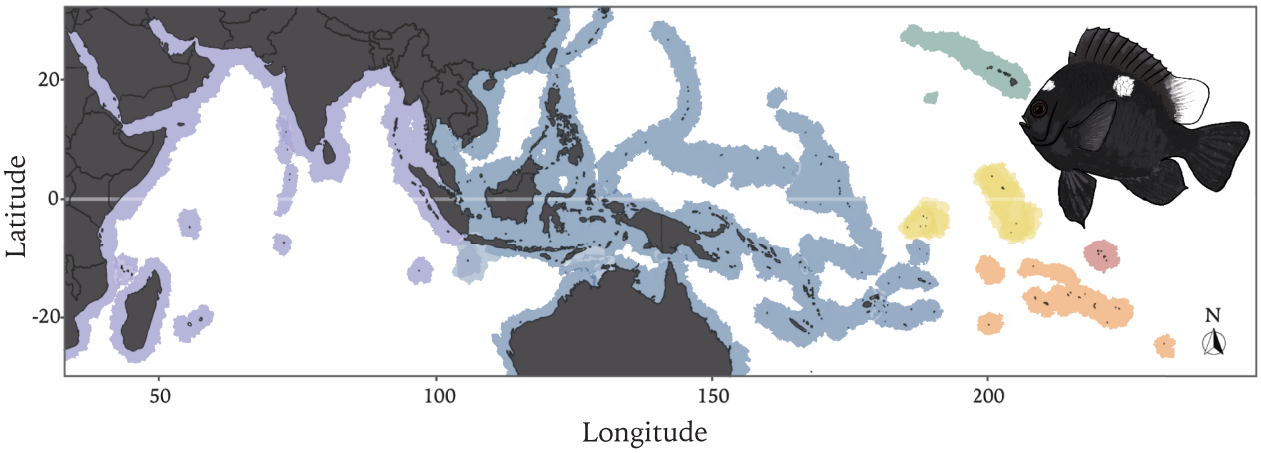

Fig. 1. *Dascyllus trimaculatus* and species complex global distribution. The three-spot damselfish (Pacific morph) is shown on the right of the map adapted from Leray *et al.* (2009) and Salas *et al.* (2020). The map shows in blue the broad distribution of *D. trimaculatus*. Differences in blue reflect results of Salas *et al.* (2020) which showed Indian Ocean differentiation from Pacific populations as well as a sub population in Cocos Keeling and a hybridization zone in Christmas Island. Similarly, the darker blue patch in the Central Pacific shows another divergent population of *D. trimaculatus* identified in Leray *et al.* (2009). The other colors show the distributions of the other species within the *D. trimaculatus* complex: green for *D. albisella*, yellow for *D. auripinnis*, and red for *D. strasburgi*.

Center for Biotechnology Information (NCBI; https://www.ncbi.nlm.nih.gov/) (*A. ocellaris*, (Tan *et al.* 2018); *Acanthochromis polyacanthus*, (Schunter *et al.* 2016); *Amphiprion percula*, (Lehmann *et al.* 2019); *Amphiprion ocellaris* (Ryu *et al.* 2022); *Acanthochromis polyacanthus*, Lehmann in review), and another nine from a single study (Marcionetti *et al.* 2019). Of these, only one is of species other than genus *Amphiprion* and only three of those listed above (*A. ocellaris*, *A. percula*, and *A. polyacanthus*) are chromosome-scale genomes. Of the Pomacentrid chromosome-scale genomes, all had $2n = 48$, with genome sizes ranging between 863 and 956 Mb. The two published genomes, *A. ocellaris* (Ryu *et al.* 2022) and *A. percula*, were highly complete with published Benchmarking Universal Single-Copy Orthologs (BUSCO) values of 97.01 and 97.2%, respectively. Chromosome-scale genomes provide a more complete sequence and locations of genes and allow for research into how chromosome architecture influences ecology, population dynamics, and adaptive evolution. Here, we present the first-genome assembly within the genus *Dascyllus* and add to the short, but growing list of Pomacentrid chromosome-scale genomes.

Materials and methods

Biological materials

The *D. trimaculatus* individual used for this genome assembly was ordered from an online pet fish supplier (liveaquaria.com), sourced from the West Pacific Rim population (Limon *et al.* 2023). It was euthanized following an approved IACUC protocol animal use. Liver, muscle, gill, and brain tissue were harvested from the right side of the individual and each placed in separate, preweighed Covaris cryogenic vials, flash frozen in liquid nitrogen, and stored at −80°C until further processing. The remaining intact left side of the specimen is stored in −80°C at University of California Santa Cruz. *Dascyllus trimaculatus* exhibit nonfunctional protogyny (Asoh and Kasuya 2002), and this individual was determined to be male based on presence of testis.

Nucleic acid library preparation and sequencing

Whole-genome shotgun library preparation

DNA was extracted from 13 mg of muscle tissue using a DNeasy Blood and Tissue kit (Qiagen), quantified using Qubit dsDNA HS Assay kit (Thermo Fisher Scientific) and Qubit 4.0 Fluorometer, then assayed with 1.0% agarose gel electrophoresis to determine molecular weight. DNA was sheared for 26 cycles of shearing (15 seconds on, 30 seconds chilling) using a Bioruptor sonicator (Diagenode), then size selected using SPRI beads (Beckman) to select for fragments between 200 and 500 bp.

The NEBNext UltraII DNA Library Prep Kit for Illumina (New England Bio Labs) was used according to manufacturer's protocol except that KAPA Hot Mix Ready Start Master Mix (Roche Diagnostics) was used for library amplification instead of NEB Q5 Master Mix. Paired-end sequencing was done at the University of California Davis Genome Center on a HiSeq4000 sequencer on a 2×150PE cycle.

Chicago library preparation

High molecular weight (HMW) DNA was isolated from the *Dascyllus trimaculatus* individual by lysing gill tissue in low-EDTA TE buffer (Dawson *et al.* 1998), then purifying with a chloroform, phenol:chloroform, chloroform and ethanol precipitation protocol (Sambrook and Russell 2006). The quality of the HMW DNA was assayed with 1.0% agarose gel electrophoresis. This DNA was used in the preparation of the Chicago, Hi-C, and for Oxford Nanopore Technologies sequencing libraries.

From this DNA, three Chicago libraries were prepared using a published method (Putnam *et al.* 2016), each using a different restriction enzyme: one with DpnII cutting at GATC sites, one with MluCI cutting at AATT sites, and one with FatI cutting at CATG sites. These libraries were sequenced on a 2×150PE cycle at Fulgent Genetics on a HiSeq400 sequencer.

Hi-C library preparation

Two Hi-C libraries were generated from approximately 100 ng of LN_2-flash-frozen muscle. The libraries were constructed using a published protocol (Adams *et al.* 2020). One library was constructed using the enzyme DpnII, and the other library was constructed with the enzyme MluCI.

Oxford nanopore library

Next, 1500 ng of the HMW DNA prepared for Chicago libraries was also used to prepare two Oxford nanopore library (ONT) WGS

libraries with the SQK-LSK109 modified protocol "versionGDE_9063_v109_revT_14Aug2019". The DNA repair steps at 20°C and 65°C were carried out for 20 minutes each, instead of 5 minutes each. We ran each of the resulting libraries on two separate MinION flow cells (FLO-MIN106), each for 72 hours. Raw fast5 files from the two MinION runs were basecalled using Guppy ("Guppy Basecalling Software" 2019) v3.3.

A summary of sequencing information for the various libraries can be found in Supplementary Table 1.

Genome assembly

All programs and versions used for the assembly are listed in Table 1.

Sequencing adapters were removed from the Illumina whole-genome shotgun (WGS) reads with Trimmomatic (Bolger *et al.* 2014) v0.39 with parameters: "ILLUMINACLIP: all_seqs.-fa:2:30:10:8:TRUE SLIDINGWINDOW:4:20 MINLEN:50'. We used jellyfish (Bolger *et al.* 2014) v2.2.10 to make a k-21 k-mer count vs abundance histogram and used the histogram with Genome Scope (Vurture *et al.* 2017) v2.0 to estimate *D. trimaculatus* genome size, heterozygosity, and repeat content. MaSuRCA (Zimin *et al.* 2013; Jiang *et al.* 2019; Wang *et al.* 2020) was used to assemble the first version of the genome using both the ONT and WGS reads.

Table 1. List of programs and program versions in order of use for the genome assembly of the three-spot damselfish, *Dascyllus trimaculatus*.

Purpose		Software and version
Estimate genome size		Jellyfish v2.2.10
Reference-free characterization		GenomeScope v2
Basecalling		Guppy v3.3
De novo assembly		MaSuRCA downloaded Sept 2020
Map HiC/Chicago reads		Arima HiC pipeline Release date: 05. 2019 BWA v0.7.17-r1188; samtools v1.13; Picard v2.26.2
File processing		Samtools v1.13, bedtools v2.30.0
Proximity ligation scaffolding		SALSA2 v2.3
Trim adapters and filter		Trimmomatic v0.39
Polish assembly (2x)		Pilon v1.23
Check for contamination		Blobtools2 v3.1.0
Manual curation	Map Hi-C to genome	Chromap v0.2.2
	File conversion	Juicebox Assembly Tools v2.14.0 Artisinal tools 9a79889 https://github.com/conchoecia/genome_assembly_pipelinesb0cda60
	Chromosome assignment	D-Genies Accessed 2022 April
Assembly statistics	Completeness	BUSCO v5.2.2, assembly-stats v0.0.1, merqury v1.3
	General stats	Assembly-stats v0.0.1
	Quality and error rate	Merqury v1.3

We followed the Arima-HiC mapping pipeline (https://github.com/ArimaGenomics/mapping_pipeline/blob/master/Arima_Mapping_UserGuide_A160156_v02.pdf) to prepare the data for scaffolding. The pipeline aligns the sequencing data from each the Hi-C and Chicago dataset against the assembly from MaSuRCA, it then filters ligation adapters and removes PCR duplicates from the resulting alignments. These alignments were then processed with samtools (Li *et al.* 2009; Danecek *et al.* 2021) v1.13 and converted into BED files with bedtools (Quinlan *et al.* 2010) v2.30.

The MaSuRCA assembly was scaffolded with SALSA (Ghurye *et al.* 2019) v2.3 with ligation junction parameter -e AATT, GATC, CATG. Iteration number was set to 10 (-i 10) and we allowed for Hi-C/Chicago data to also correct assembly errors (-m yes).

We aligned the trimmed Illumina WGS reads to the scaffolded output of SALSA with bwa mem (Li and Durbin 2009) v0.7.17-r1188 and used that alignment to polish the assembly with Pilon (Walker *et al.* 2014) v1.23. We repeated the alignment and polishing steps once. The error-corrected assembly was then screened for possible contaminants, using Blobtools2 (Laetsch and Blaxter 2017) v3.1.0. Any contigs assigned to phyla other than Chordata were removed. However, any sequences categorized as "No hits' were kept. The assembly was then manually curated by mapping the DpnII and MluCI Hi-C reads to the genome assembly with chromap (Zhang *et al.* 2021) v0.2.2 with a quality filter of 0 and converted to a .hic file with Juicebox Assembly Tools (JBAT) (Durand *et al.* 2016) v2.14.00. Artisanal tools commit 9a79889 (https://bitbucket.org/bredeson/artisanal) was used to generate a JBAT assembly file. We used the Juicebox GUI (Dudchenko *et al.* 2018) v1.11.08 to manually curate the assembly with the .hic and .assembly files. Modifications made to the assembly included ordering and orienting scaffolds into chromosome-scale scaffolds, removing duplicated regions, and making manual assembly breaks to place misassembled contig pieces onto the correct scaffold. Artisanal was used to generate an updated genome assembly FASTA file. Scaffolds not placed on chromosomes were sorted by the strongest Hi-C connection to chromosome-scale scaffolds with genome assembly tools commit b0cda60 (https://github.com/conchoecia/genome_assembly_pipelines. D-Genies (Cabanettes and Klopp 2018), accessed 2022 April 30, was used to align the manually curated assembly to the chromosome-scale assembly of the closely related *Amphiprion percula* genome assembly (Lehmann *et al.* 2019). The evidence from this analysis was used to assign chromosome numbers to the *D. trimaculatus* scaffolds based on homology with *A. percula* chromosomes.

Genome quality assessment

BUSCO (Simão *et al.* 2015; Waterhouse *et al.* 2018) v5.2.2 was used to evaluate genome completeness by comparing number of orthologous genes found in the assembly to the 3,640 genes in the actinopterygii_odb10 database. Assembly statistics (assembly-stats; https://github.com/sanger-pathogens/assembly-stats) were generated to track N50, L50, contigs, gaps, and lengths at each step. We used merqury (Rhie *et al.* 2020) v1.3, to calculate the genome completeness and error rates.

Results

Sequencing

We sequenced four library types: a WGS library which resulted in 314.6 Mb paired-end 150 bp reads, representing 103x coverage, and 3.52 M (4.84 Gb) and 8.57 M (19.77 Gb) ONT reads from the

Table 2. A comparison of genome metrics between *D. trimaculatus* assembly stages.

		Assembly version				
Assembly step		**MaSuRCA** De novo hybrid assembly	**SALSA2** Scaffolding-proximity ligation	**Pilon 1st iteration** Error correcting	**Pilon 2nd iteration** Error correcting	**Final assembly** Manual assembly curation
Input data		**ONT + Illumina shotgun**	**Hi-C + Chicago**	**Illumina shotgun**	**Illumina shotgun**	**Hi-C dpnII and mlucl data**
Genome length (bp)		919,275,268	919,814,768	919,296,307	919,364,398	910,763,285
Number of contigs		3,501	2,467	2,467	2,467	2,156
Average contig length (bp)		262,575.06	372,847.49	372,637.34	372,664.94	422,431.95
Largest contig length (bp)		8,494,454	34,904,224	34,891,597	34,891,597	41,400,476
N50 length: *n*		1,108,072 : 200	16,013,467 : 20	16,006,006 : 20	16,006,752 : 20	34,909,338 : 13
N90 length: *n*		124,315 : 1,137	218,694 : 274	218,662 : 274	218,658 : 274	22,085,708 : 24
N100 length: *n*		1,790 : 3,501	1,790 : 2,467	1,790 : 2,467	1,790 : 2,467	1,790 : 2,156
Gaps		18	1,097	1,090	1,088	1,411
N count		1,800	541,300	471,149	448,527	141,100
BUSCO scores	C	97.3	97.9	97.9	97.9	97.9
(% of 3,640 orthologs)	S	94.5	95.7	95.8	95.7	96.2
	D	2.8	2.2	2.1	2.2	1.7
	F	0.5	0.6	0.6	0.6	0.7
	M	2.2	1.5	1.5	1.5	1.4
Completeness (%)		86.18	86.18	86.16	86.19	85.96
Quality value		44.7342	44.7342	44.2101	44.6053	44.61
Error rate		0.000033619	0.000033619	3.79307E-05	0.000046314	0.000034597

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.

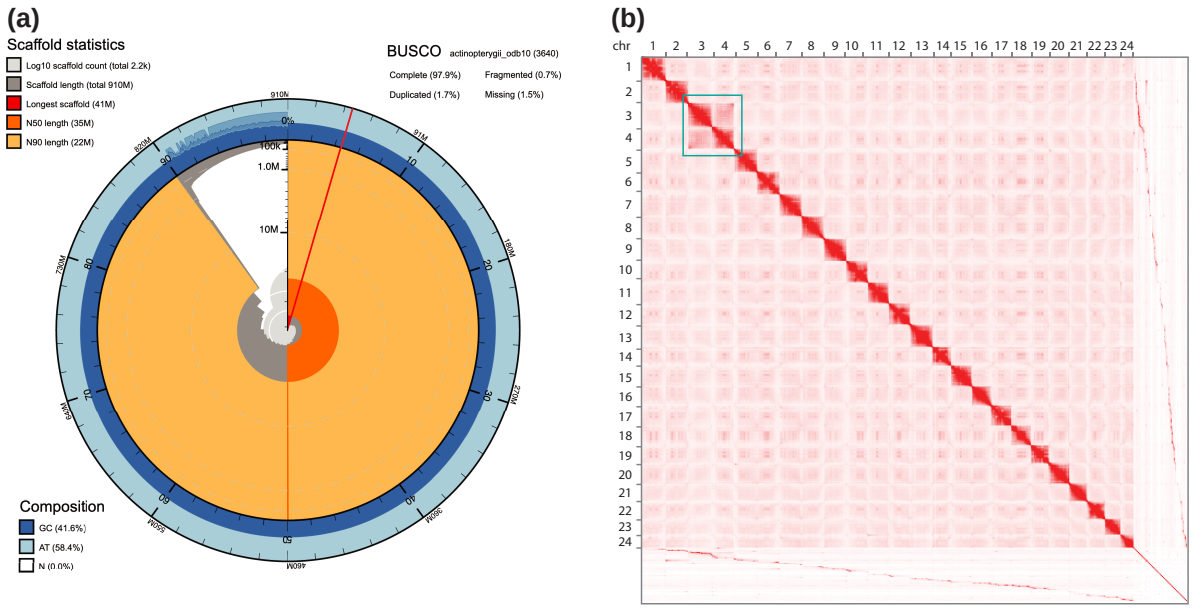

Fig. 2. Genome statistics and chromosome map. Panel A: The outer circumference of the main plot represents the full length of the 910,763,285 bp chromosome-scale assembly of *Dascyllus trimaculatus*. The outer ring of blues depicts GC (dark blue) and AT (light blue) content along the assembly which is summarized in the lower left. The second ring is demarcated by percentage of the total contigs of the genome. Orange and pale-orange arcs show the N50 and N90 record lengths (34,909,338 and 22,085,708 bp), respectively, overlying the dark gray, which arranges scaffolds in order by size starting from the largest scaffold (41,400,476 bp and ~4% genome, shown in red). A summary of BUSCO statistics for complete (97.9%), fragmented (0.7%), duplicated (1.7%), and missing (1.5%), orthologous genes in the actinopterygii_odb10 set is shown in the top right. Panel B: A Hi-C contact map made with the MluCI and the DpnII libraries showing 24 chromosome clusters and the unscaffolded contigs. In the green square, chromosomes 3 and 4 show strong interchromosomal connections at roughly half coverage indicating Robertsonian fusion in one set of chromosomes contributed a parent with 2n = 47 while the other parent contributed 2n = 48.

two runs on the minION flowcells, representing 22x ONT coverage for the initial hybrid assembly. The five proximity ligation libraries used for scaffolding, two Hi-C (restriction enzymes DpnII and MluCl), and three Chicago libraries (restriction enzymes DpnII, MluCI, and FatI) yielded ~108 M, ~152 M, ~65 M, ~74 M, and ~67 M, reads, respectively, for a total proximity ligation coverage

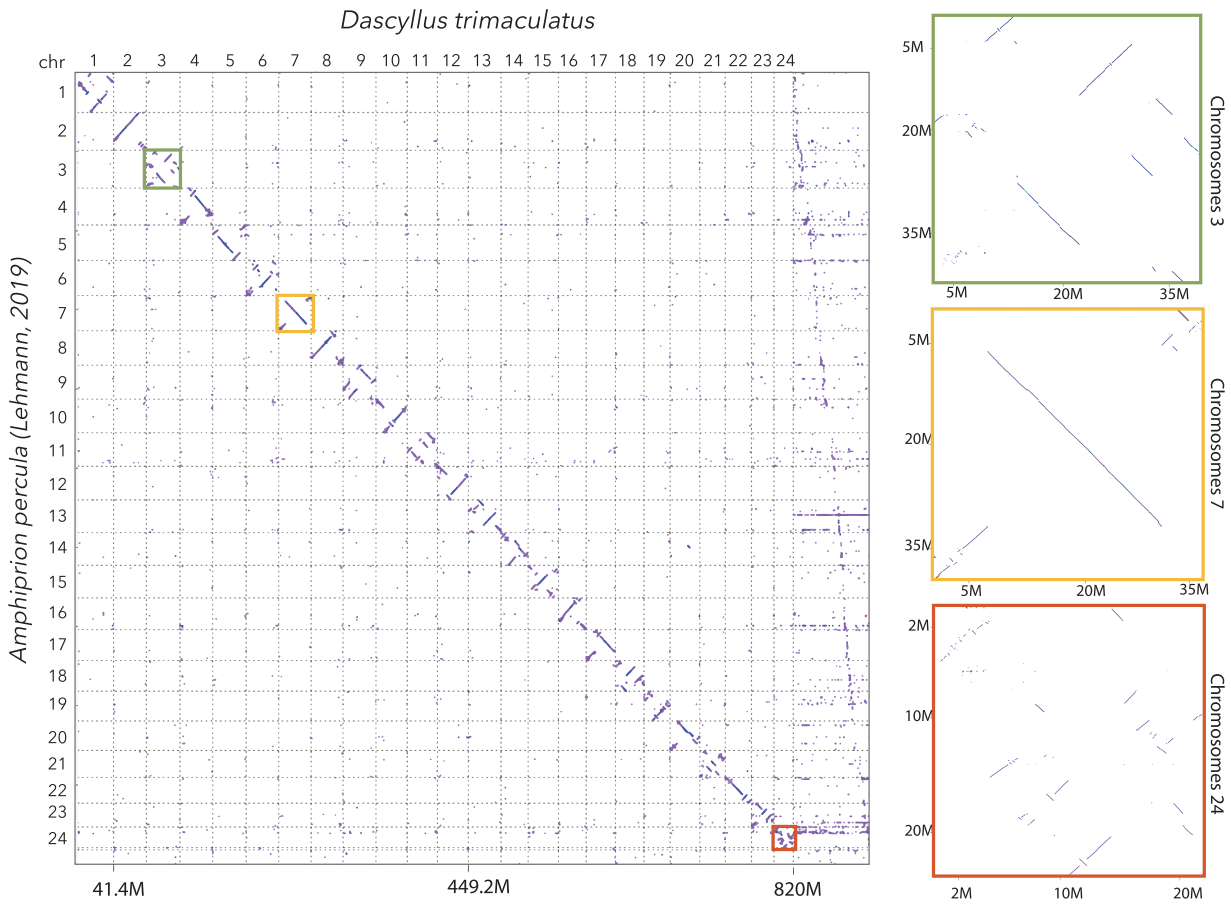

Fig. 3. *Dascyllus trimaculatus* mapped against chromosome-level genome of *Amphiprion percula* (Pomacentridae). The main panel shows a dot-plot of the assemblies of *Dascyllus trimaculatus* (presented in this manuscript) and the genome of the anemone fish *Amphiprion percula* (Lehmann *et al.* 2019). The remaining unscaffolded contigs are shown in the last column. The right three panels show a close-up dot-plot of the color-coded boxes in the main panel of chromosome 3, chromosome 7, and chromosome 24. Chromosome 3, one of the chromosomes involved in Robertsonian fusions within the species, shows many rearrangements as well as regions of repeat sequences near one end. Chromosome 7 seems to be one of the most architecturally conserved chromosomes between *Dascyllus* and *Amphiprion,* whereas Chromosome 24 shows an example of a highly rearranged chromosome.

of 154x. In total, across all data types, we had a final coverage of 280x (See Supplementary Table 1 for sequencing details).

Heterozygosity and repetitive sequence estimation

GenomeScope estimated the genome size to be 809 Mb, with 84% unique and 16% repetitive sequences, and 1.02% heterozygosity (Supplementary Fig. 1).

Genome assembly

Genome quality metrics for each step of the assembly are listed in Table 2. The initial de novo assembly by MaSuRCA with ONT and Illumina shotgun data had a total length of 919,275,268 bp in 3,501 contigs with an N50 of 1,108 Kb. Scaffolding with the HiC and Chicago libraries dropped the number of contigs to 2,467 and increased N50 to 16,013 Kb. After two rounds of polishing with trimmed Illumina shotgun reads gaps decreased from 1,097 to 1,088. Blobtools2 showed that of the 2,467 contigs, none matched other taxa in NCBI databases of bacteria, invertebrates, mammals, phages, plants, and fungi, or environmental samples. Four hundred seventy-eight contigs did not match any databases (no-hits) and were left in the genome.

The manual curation of the genome assembly yielded 24 scaffolds consistent with chromosome-scale scaffolds (Fig. 2). A dot-plot comparison (Fig. 3) with the *Amphiprion percula* (Lehmann *et al.* 2019) genome revealed that each of the *D. trimaculatus* chromosome-scale scaffolds had a one-to-one corresponding homologous, albeit rearranged, chromosome in the *Amphiprion percula* genome.

The final assembly (GenBank accession: JAMOIN000000000) has a length of 910.7 Mb, 90% of which was on chromosome-scale scaffolds and BUSCO score of 97.9%. Merqury calculated 86.19% completeness, QV of 44.6, and an estimated error rate 0.0000346, or a single nucleotide error every 28.9 Kb.

Discussion

The biology, evolution, and biogeography of the three-spot damselfish is relatively well studied using genetic (Bernardi and Crane 1999; Bernardi *et al.* 2001; McCafferty *et al.* 2002; Leray *et al.* 2009, 2010; Liggins *et al.* 2016; Getlekha *et al.* 2017; Crandall *et al.* 2019) and genomic tools (Salas *et al.* 2019, 2020) and, as we shift further into the age of WGS data and tools, a reference genome is an invaluable resource. Here, we present the chromosome-scale genome assembly of a three-spot damselfish, *Dascyllus trimaculatus,* collected from the Indonesian/Philippine population (Limon *et al.* in review). It is the first within the genus *Dascyllus* of the widely studied, and large Pomacentridae family. This high-quality de novo assembly of a nonmodel coral reef fish is a valuable reference for furthering studies of evolutionary, ecological,

and conservation studies for the species and for coral reef fish in general.

We report sequences for 24 chromosomes of the *D. trimaculatus* genome with total length and repetitive content (Fig. 2a, Supplementary Fig. 1) that is expected for this species (Arai 1976; Getlekha *et al.* 2017; Yuan *et al.* 2018). Interestingly, our Hi-C data also show that chromosomes three and four have strong connections at half the depth of other intra-chromosomal connections (Fig. 2b). This pattern can be explained by a hemizygous state wherein one parental gamete contributed a Robertsonian fusion of chromosomes three and four, and the other parental gamete contributed chromosomes three and four as separate chromosomes making the individual sequenced here, a $2n = 47$ individual. This finding is consistent with previous studies that report both $2n = 47$ and $2n = 48$ for *Dascyllus trimaculatus* (Arai 1976; Ojima and Kashiwagi 1981; Kashiwagi *et al.* 2005). Chromosome numbers vary both within and among species of *Dascyllus*. One report on several *Dascyllus* species collected in the Philippines and the Ryukyu Archipelago of southern Japan demonstrated polymorphic karyotypes in all but one of the species (Ojima and Kashiwagi, 1981). *Dascyllus aruanus* had the most karyotypic variation—between $2n = 27$–33 chromosomes, *D. reticulatus* $2n = 34$–37, *D. trimaculatus* $2n = 47$–48, and *D. melanurus* with $2n = 48$.

In addition to confirming variation in chromosome number, the dot-plot comparison between this genome and of the closest relative with an available chromosome-scale assembly, *Amphiprion percula* (Lehmann *et al.* 2019), revealed several rearrangements in every chromosome between corresponding chromosomes (Fig. 3). The Pomacentrid subfamilies Chrominae and Amphiprionini are estimated to have diverged over 50 million years ago (mya) (McCord *et al.* 2021). The estimated number of rearrangements within chromosomes ranged from 2+ in chromosome 7 of *D. trimaculatus* which was the most like its counterpart in *A. percula* to over 35 in chromosome 24 (Fig. 3). This pattern of rearrangements has not been characterized between chromosome-scale genome assemblies of Pomacentridae. The role of variation in chromosome number has been the subject of several cytogenic studies which have found that chromosome diversity inversely related to mobility of the fish and that chromosome rearrangements can serve to either promote or prevent recombination events (Galetti *et al.* 2000; Molina and Galetti 2004; Kirkpatrick and Barton 2006; Martinez *et al.* 2015). Interestingly, chromosome 3 in the genome presented in this paper is one of the most rearranged while also being one of the chromosomes involved in the Robertsonian fusion mentioned above. This assembly will be a useful starting point to study how this type of genome structure varies at a meta-population scale, and how this influences recombination and adaptation.

This assembly represents the first chromosome-level genome of the genus *Dascyllus* as well as the first non*Amphiprion* chromosome-scale genome published in the Pomacentridae family. Damselfishes are excellent model species due to their relatively small size, ease to manage in the wild and lab, and those interested in this group will benefit from this addition to the available genomic resources. *Dascyllus trimaculatus* itself, is has had a dynamic evolutionary trajectory across the Indo-Pacific, evident in species complex that is continuing to reveal its complexity and provide insight into evolutionary mechanisms. In addition to providing a high-quality reference genome to further our understanding of genomic architecture, this assembly will serve to leverage information stored across the genome to better understand the population dynamics, phylogeny, biogeography, demographics, of *Dascyllus trimaculatus*, as well as gain insight into historical, current, and future response to changes in climate.

Data availability

The assembly and genomic sequencing reads generated for this study have all been deposited in the NCBI GenBank database under BioProject ID PRJNA828170. The accession for the genome is JAMOIN000000000, WGS data (SRX17663068), proximity ligation data (SRX17663069 - SRX17663073), and ONT data (SRX17742644, SRX177426445).

Supplemental material available at G3 online.

Acknowledgements

The authors are grateful for access to the Hummingbird computational cluster and the team behind it at University of California Santa Cruz. The whole-genome shotgun sequencing was carried out at the DNA Technologies and Expression Analysis Core at the University of California Davis Genome Center, supported by NIH Shared Instrumentation Grant 1S10OD010786-01. We also appreciate Jonas Oppenheimer and Robert Lehmann for helpful discussions and advice.

Funding

This work was supported in part by Marilyn C. Davis Scholarship and American Association of University Women (through STARS Reentry program at UCSC), Burnand-Partridge Foundation, tuition support in part by a Dissertation Quarter Fellowship (UCSC) to M.B.R. D.T.S was supported by the US National Science Foundation GRFP DGE 1339067, the US National Science Foundation DEB-1542679 to Steven Haddock, and the European Research Council's Horizon 2020: European Union Research and Innovation Programme, grant no. 945026 to Oleg Simakov.

Conflicts of interest

None declared.

Literature cited

Adams M, McBroome J, Maurer N, Pepper-Tunick E, Saremi NF, Green RE, Vollmers C, Corbett-Detig RB. One fly–one genome: chromosome-scale genome assembly of a single outbred *Drosophila melanogaster*. Nucleic Acids Res. 2020;48(13):e75. doi:10.1093/nar/gkaa450.

Allen GR. Damselfishes of the World. Melle, Germany; Mentor, Ohio: Mergus ; Aquarium Systems [distributor]./z-wcorg/; 1991. http://catalog.hathitrust.org/api/volumes/oclc/24436027.html.

Allen GR, Werner TB. Coral reef fish assessment in the 'coral triangle' of Southeastern Asia. Environ Biol Fishes. 2002;65(2):209–214. doi:10.1023/A:1020093012502.

Arai R. Chromosomes of four Species of coral fishes from Japan. Bull Natn Sci Mus Tokyo Ser A. 1976;2(2):137–141.

Asoh K, Kasuya M. Gonadal development and mode of sexuality in a coral-reef damselfish, *Dascyllus trimaculatus*. J Zool. 2002;256(3):301–309. doi:10.1017/S0952836902000341.

Bellwood DR, Wainwright PC. Chapter 1: the history and biogeography of fishes on coral reels. In: Coral Reef Fishes. Elsevier Science; 2002. p. 28.

Bernardi G, Crane NL. Molecular phylogeny of the humbug damselfishes inferred from MtDNA sequences. J Fish Biol. 1999;54(6):1210–1217. doi:10.1111/j.1095-8649.1999.tb02049.x.

Bernardi G, Holbrook SJ, Schmitt RJ. Gene flow at three spatial scales in a coral reef fish, the three-spot dascyllus, *Dascyllus trimaculatus*. Mar Biol. 2001;138(3):457–465. doi:10.1007/s002270000484.

Bolger AM, Lohse M, Usadel B. Trimmomatic: a flexible trimmer for illumina sequence data. Bioinformatics. 2014;30(15):2114–2120. doi:10.1093/bioinformatics/btu170.

Cabanettes F, Klopp C. D-GENIES: dot plot large genomes in an interactive, efficient and simple way. PeerJ. 2018;6(June):e4958. doi:10.7717/peerj.4958.

Crandall ED, Riginos C, Bird CE, Liggins L, Treml E, Beger M, Barber PH, Connolly SR, Cowman PF, DiBattista JD, *et al.* The molecular biogeography of the Indo-Pacific: testing hypotheses with multispecies genetic patterns. Global Ecol Biogeogr. 2019;28(7):943–960. doi:10.1111/geb.12905.

Danecek P, Bonfield JK, Liddle J, Marshall J, Ohan V, Pollard MO, Whitwham A, Keane T, McCarthy SA, Davies RM, *et al.* Twelve years of SAMtools and BCFtools. GigaScience. 2021;10(2):giab008. doi:10.1093/gigascience/giab008.

Dawson MN, Raskoff KL, Jacobs DK. Field preservation of marine invertebrate tissue for DNA analyses. Mol Mar Biol Biotechnol. 1998;7(2):145–152.

Dudchenko O, Shamim MS, Batra SS, Durand NC, Musial NT, Mostofa R, Pham M, Glenn St Hilaire B, Yao W, Stamenova E, *et al.* 2018. The juicebox assembly tools module facilitates de novo assembly of Mammalian genomes with chromosome-length scaffolds for under $1000. preprint at bioRxiv. doi:10.1101/254797.

Durand NC, Robinson JT, Shamim MS, Machol I, Mesirov JP, Lander ES, Aiden EL. Juicebox provides a visualization system for hi-C contact maps with unlimited zoom | Elsevier Enhanced Reader. Cell Syst. 2016;3(July):99–101. doi:10.1016/j.cels.2015.07.012.

Galetti PM, Aguilar CT, Molina WF. An overview of marine fish cytogenetics" In: Solé-Cava AM, Russo CAM, Thorpe JP, editors. Marine Genetics. Dordrecht: Springer Netherlands; 2000. p. 55–62. doi:10.1007/978-94-017-2184-4_6.

Getlekha N, Supiwong W, Yeesin P, Pengseng P, Kasiroek W, Tanomtong A. Chromosomal characteristics of the three-spot damselfish, *Dascyllus trimaculatus* (Perciformes, Pomacentridae) in Thailand. Cytologia (Tokyo). 2017;82(1):51–57. doi:10.1508/cytologia.82.51.

Ghurye J, Rhie A, Walenz BP, Schmitt A, Selvaraj S, Pop M, Phillippy AM, Koren S. Integrating Hi-C links with assembly graphs for chromosome-scale assembly. PLOS Comput Biol. 2019;15(8):e1007273. doi:10.1371/journal.pcbi.1007273.

Hardie DC, Hebert PD. Genome-size evolution in fishes. Can J Fish Aquatic Sci. 2004;61(9):1636–1646. doi:10.1139/f04-106.

Hiatt RW, Strasburg DW. Ecological relationships of the fish fauna on coral reefs of the Marshall Islands. Ecol Monogr. 1960;30(1):65–127. doi:10.2307/1942181.

Hotaling S, Kelley JL, Frandsen PB. Toward a genome sequence for every animal: where are we now? Proc Natl Acad Sci U S A. 2021;118(52):e2109019118. doi:10.1073/pnas.2109019118.

Jiang JB, Quattrini AM, Francis WR, Ryan JF, Rodríguez E, McFadden CS. A hybrid *de novo* assembly of the sea pansy (*Renilla Muelleri*) genome. GigaScience. 2019;8(4):4. doi:10.1093/gigascience/giz026.

Kashiwagi E, Takai A, Ojima Y. Chromosomal distribution of constitutive heterochromatin and nucleolus organizer regions in four dascyllus fishes (Pomacentridae, Perciformes). Cytologia (Tokyo). 2005;70(3):345–349. doi:10.1508/cytologia.70.345.

Kirkpatrick M, Barton N. Chromosome inversions, local adaptation and speciation. Genetics. 2006;173(1):419–434. doi:10.1534/genetics.105.047985.

Laetsch DR, Blaxter ML. Blobtools: interrogation of genome assemblies. F1000Res. 2017;6(July):1287. doi:10.12688/f1000research.12232.1.

Lehmann R, Lightfoot DJ, Schunter C, Michell CT, Ohyanagi H, Mineta K, Foret S, Berumen ML, Miller DJ, Aranda M, *et al.* Finding Nemo's genes: a chromosome-scale reference assembly of the genome of the orange Clownfish *Amphiprion percula*. Mol Ecol Resour. 2019;19(3):570–585. doi:10.1111/1755-0998.12939.

Leray M, Beldade R, Holbrook SJ, Schmitt RJ, Planes S, Bernardi G. Isolation and characterization of 13 polymorphic nuclear microsatellite primers for the widespread Indo-Pacific three-spot damselfish, *Dascyllus Trimaculatus*, and closely related *D. Auripinnis*. Mol Ecol Resour. 2009;9(1):213–215. doi:10.1111/j.1755-0998.2008.02380.x.

Leray M, Beldade R, Holbrook SJ, Schmitt RJ, Planes S, Bernardi G. Allopatric divergence and speciation in coral reef fish: the three-spot dascyllus, *Dascyllus trimaculatus*, Species Complex. Evolution. 2010;64(5):1218–1230. doi:10.1111/j.1558-5646.2009.00917.x.

Li H, Durbin R. Fast and accurate short read alignment with Burrows-Wheeler transform. Bioinformatics. 2009;25(14):1754–1760. doi:10.1093/bioinformatics/btp324.

Li H, Handsaker A, Wysoker T, Fennell J, Ruan N., Homer G., Marth G., Abecasis R., Durbin, 1000 Genome Project Data Processing Subgroup. The sequence alignment/map format and SAMtools. Bioinformatics 2009;25(16):2078–2079. doi:10.1093/bioinformatics/btp352.

Liggins L, Treml EA, Possingham HP, Riginos C. Seascape features, rather than dispersal traits, predict spatial genetic patterns in co-distributed reef fishes. J Biogeogr. 2016;43(2):256–267. doi:10.1111/jbi.12647.

Limon J, Roberts MB, Schultz DT, Bernardi G. The complete mitochondrial genome of *Dascyllus trimaculatus*. *Mitochondrial DNA Part B* 2023;8(1):105–106. doi:10.1080/23802359.2022.2161838.

Marcionetti A, Rossier V, Roux N, Salis P, Laudet V, Salamin N. Insights into the genomics of clownfish adaptive radiation: genetic basis of the mutualism with sea anemones. Genome Biol Evol. 2019;11(3):869–882.

Martinez PA, Zurano JP, Amado TF, Penone C, Betancur-R R, Bidau CJ, Jacobina UP. Chromosomal diversity in tropical reef fishes is related to body size and depth range. Mol Phylogenet Evol. 2015;93-(December):1–4. doi:10.1016/j.ympev.2015.07.002.

McCafferty S, Bermingham E, Quenouille B, Planes S, Hoelzer G, Asoh K. Historical biogeography and molecular systematics of the indo-pacific genus *Dascyllus* (Teleostei: Pomacentridae). Mol Ecol. 2002;11(8):1377–1392. doi:10.1046/j.1365-294X.2002.01533.x.

McCord CL, Nash CM, James Cooper W, Westneat MW. Phylogeny of the damselfishes (Pomacentridae) and patterns of asymmetrical diversification in body size and feeding ecology. PLoS One. 2021;16(10):e0258889. doi:10.1371/journal.pone.0258889.

Molina WF, Galetti PM. Karyotypic changes associated to the dispersive potential on Pomacentridae (Pisces, Perciformes). J Exp Mar Biol Ecol. 2004;309(1):109–119. doi:10.1016/j.jembe.2004.03.011.

Ojima Y, Kashiwagi E. Chromosomal evolution associated with Robertsonian fusion in the genus *Dascyllus* (Chrominae, Pisces). Chrominae, Pisces. Proc Japan Acad Ser B. 1981;57(10):368–370. doi:10.2183/pjab.57.368.

"Guppy Basecalling Software." Oxford Nanopore Technologies; 2019. https://nanoporetech.com/.

Putnam NH, O'Connell BL, Stites JC, Rice BJ, Blanchette M, Calef R, Troll CJ, Fields A, Hartley PD, Sugnet CW, *et al.* Chromosome-scale shotgun assembly using an in vitro method for long-range linkage. Genome Res. 2016;26(3):342–350. doi:10.1101/gr.193474.115.

Quinlan AR, Clark RA, Sokolova S, Leibowitz ML, Zhang Y, Hurles ME, Mell JC, Hall IM. Genome-wide mapping and assembly of

structural variant breakpoints in the mouse genome. Genome Res. 2010;20(5):623–635. doi:10.1101/gr.102970.109.

Rhie A, Walenz BP, Koren S, Phillippy AM. Merqury: reference-free quality, completeness, and phasing assessment for genome assemblies. Genome Biol. 2020;21(1):245. doi:10.1186/s13059-020-02134-9.

Robitzch VSN, Lozano-Cortés D, Kandler NM, Salas E, Berumen ML. Productivity and sea surface temperature are correlated with the pelagic larval duration of damselfishes in the Red Sea. Mar Pollut Bull. 2016;105(2):566–574. doi:10.1016/j.marpolbul.2015.11.045.

Rüppell WPES. Atlas zu der Reise im nördlichen Africa. Fischedes Rothen Meeres. 1828-30;1:141–143.

Ryu T, Herrera M, Moore B, Izumiyama M, Kawai E, Laudet V, Ravasi T. A chromosome-scale genome assembly of the false Clownfish, *Amphiprion ocellaris*. G3 (Bethesda). 2022:12(5):jkac074. doi:10.1093/g3journal/jkac074.

Salas E, Giacomo M, Bernardi M, Berumen L, Gaither MR, Rocha LA. RADseq analyses reveal concordant Indian Ocean biogeographic and phylogeographic boundaries in the reef fish *Dascyllus Trimaculatus*. R Soc Open Sci. 2019;6(5):172413. doi:10.1098/rsos.172413.

Salas EM, Hobbs JA, Bernal MA, Simison WB, Berumen ML, Bernardi G, Rocha LA. Distinct patterns of hybridization across a suture zone in a coral reef fish (*Dascyllus Trimaculatus*). Ecol Evol. 2020;10(6):2813–2837. doi:10.1002/ece3.6068.

Sambrook J, Russell DW. Purification of nucleic acids by extraction with phenol:chloroform. CSH Protoc. 2006;2006(1):pdb.prot4455. doi:10.1101/pdb.prot4455.

Schunter C, Welch MJ, Ryu T, Zhang H, Berumen ML, Nilsson GE, Munday PL, Ravasi T. Molecular signatures of transgenerational response to ocean acidification in a species of reef fish. Nat Climate Change. 2016;6(11):1014. doi:10.1038/nclimate3087.

Simão FA, Waterhouse RM, Ioannidis P, Kriventseva EV, Zdobnov EM. BUSCO: assessing genome assembly and annotation completeness with single-copy orthologs. Bioinformatics. 2015;31(19):3210–3212. doi:10.1093/bioinformatics/btv351.

Tan MH, Austin CM, Hammer MP, Lee YP, Croft LJ, Gan HM. Finding nemo: hybrid assembly with Oxford nanopore and illumina reads greatly improves the Clownfish (*Amphiprion ocellaris*) genome assembly. GigaScience. 2018;7(3):gix137. doi:10.1093/gigascience/gix137.

Tang KL, Stiassny MLJ, Mayden RL, DeSalle R. Systematics of damselfishes. Ichthyol Herpetol. 2021;109(1):258–318. doi:10.1643/i2020105.

Vurture GW, Sedlazeck FJ, Nattestad M, Underwood CJ, Fang H, Gurtowski J, Schatz MC. Genomescope: fast reference-free genome profiling from short reads. Bioinformatics. 2017;33(14):2202–2204. doi:10.1093/bioinformatics/btx153.

Walker BJ, Abeel T, Shea T, Priest M, Abouelliel A, Sakthikumar S, Cuomo CA, Zeng Q, Wortman J, Young SK, *et al.* Pilon: an integrated tool for comprehensive microbial variant detection and genome assembly improvement. PLoS One. 2014;9(11):e112963. doi:10.1371/journal.pone.0112963.

Wang W, Das A, Kainer D, Schalamun M, Morales-Suarez A, Schwessinger B, Lanfear R. The draft nuclear genome assembly of Eucalyptus Pauciflora: a pipeline for comparing de Novo assemblies. GigaScience. 2020;9(1):giz160. doi:10.1093/gigascience/giz160.

Waterhouse RM, Seppey M, Simão FA, Manni M, Ioannidis P, Klioutchnikov G, Kriventseva EV, Zdobnov EM. BUSCO applications from quality assessments to gene prediction and phylogenomics. Mol Biol Evol. 2018;35(3):543–548. doi:10.1093/molbev/msx319.

Wellington GM, Victor BC. Planktonic larval duration of one hundred Species of Pacific and Atlantic damselfishes (Pomacentridae). Mar Biol. 1989;101(4):557–567. doi:10.1007/BF00541659.

Yuan Z, Liu S, Zhou T, Tian C, Bao L, Dunham R, Liu Z. Comparative genome analysis of 52 fish Species suggests differential associations of repetitive elements with their living aquatic environments. BMC Genomics. 2018;19(1):141. doi:10.1186/s12864-018-4516-1.

Zhang H, Song L, Wang X, Cheng H, Wang C, Meyer CA, Liu T, Tang M, Aluru S, Yue F, *et al.* Fast alignment and preprocessing of chromatin profiles with chromap. Nat Commun. 2021;12(1):6566. doi:10.1038/s41467-021-26865-w.

Zimin AV, Marçais G, Puiu D, Roberts M, Salzberg SL, Yorke JA. The MaSuRCA genome assembler. Bioinformatics. 2013;29(21):2669–2677. doi:10.1093/bioinformatics/btt476.

Editor: D. J. de Koning