GBE

Comparative Genomics of Transisthmian Damselfishes (*Abudefduf saxatilis* and *A. troschelii*)

Claire B. Tracy [1,2,*], Carlos F. Arias [2,3], Eirlys Tysall [2,4], Marc P. Hoeppner [5], W. Owen McMillan [2], Oscar Puebla [2,6,7], Moisés A. Bernal [1,2,*]

[1]Auburn University, Auburn, AL 36830, USA
[2]Smithsonian Tropical Research Institute, Panama City, Panamá
[3]Data Science Lab, Office of the Chief Information Officer, Smithsonian Institution, Washington, DC, USA
[4]University of Cambridge, Cambridge, UK
[5]Institute of Clinical Molecular Biology, Christian-Albrechts-University of Kiel, Kiel 24105, Germany
[6]Coastal Resources, Leibniz Centre for Tropical Marine Research (ZMT), Bremen 28359, Germany
[7]Institute for Chemistry and Biology of the Marine Environment, Carl von Ossietzky Universität Oldenburg, Oldenburg 26111, Germany

*Corresponding author: E-mails: cbt0022@auburn.edu; mab0205@auburn.edu.

Accepted: April 13, 2026

Abstract

The uplift of the Central American Isthmus (CAI) represents a natural laboratory for the study of allopatric speciation in marine organisms. Several geminate species pairs of both vertebrates and invertebrates formed following the uplift of the CAI, including damselfishes of the Pomacentridae family. However, to date few studies have explored the genomic differences among trainsisthmian species in the Pacific and Atlantic Oceans. In this study, we present genome assemblies for the transisthmian species pair consisting of the Sergeant Major *Abudefduf saxatilis* (Tropical Atlantic) and the Panamic Sergeant Major *Abudefduf troschelii* (Tropical Eastern Pacific) derived from PacBio long-read sequencing. The new genomes are near-chromosome level, and among the highest-quality genomes currently available for coral reef fishes. We show that large structural variants distinguish the two species, including a nine-megabase inversion in linkage group (putative chromosome) six. Additionally, we show through an analysis of demographic history that alleles within the two genomes have different coalescence time distributions, which may be due to different effective population sizes, population structures, and/or selection regimes in the two oceans. Finally, we highlight gene families that were significantly expanded or contracted between *A. troschelii* and *A. saxatilis*. Some of these are related to the environmental differences between the two oceans, including gamma crystallin M (*crygm*), which is linked to vision, and vitellogenin (*vtg*), which is associated with egg provisioning. These genomes set the stage for comparative analyses of genetic structure and selection on the marine organisms that originated with the formation of the CAI.

Key words: allopatric speciation, geminate species, gene family expansion, genome evolution, marine fishes, Isthmus of Panama.

Significance

Understanding speciation in the ocean has represented a historical challenge, given the high rates of dispersal mediated by planktonic larval stages of most marine groups. There are few geologic events that have propelled marine biodiversity while completely blocking genetic exchange during the speciation process. The final closure of the Central American Isthmus (CAI), 2.8 million years ago, is one of these rare cases where marine populations from the tropical Pacific and Atlantic oceans have been completely isolated, providing the opportunity to study how genomes diverge in the absence of gene flow. Here, we present the first comparative genomics study of marine organisms separated by the CAI, showing that genomes on the two sides of the CAI differ structurally. We also identify contractions and expansions of various gene families between the two species, which are potentially associated with adaptations to the different environmental conditions in the two oceans. This represents a relevant first step for comparative genomics of geminate species isolated by a major geologic event.

Introduction

The uplift of the Central American Isthmus (CAI) is a major geological event that led to speciation in many marine taxa. This occurred over millions of years as landmasses gradually uplifted via tectonic movements, leading initially to the formation of islands (the Panama Arc) throughout the Central American Seaway, and making oceanic connections between present day Pacific and Atlantic oceans shallower (O'Dea et al. 2016). This process caused gradual changes in the biotic and abiotic environments that built up over time and culminated in the final closure of the CAI approximately 2.8 million years ago (MYA; O'Dea et al. 2016) causing the complete isolation of the Tropical Pacific Ocean from the Tropical Atlantic Ocean.

This physical closure caused the cessation of gene flow between the two ocean basins, while simultaneously promoting major shifts in ocean currents, resulting in starkly different environments in the Tropical Eastern Pacific Ocean (TEP) and Tropical Western Atlantic Ocean (TWA). The TEP experienced an increase in productivity due to the uplift of cooler waters driven by trade winds, compared to the TWA which experienced a reduction in upwellings changing nutrient availability and causing the Caribbean to become warmer, more saline, and less productive (Jackson and O'Dea 2023). These changes have been shown to also have biotic effects, decreasing presence of suspension feeders and predators in the Caribbean (Jackson and O'Dea 2023). Today several areas of the TEP are still characterized by seasonal upwellings (i.e. January through April; D'Croz and O'Dea 2007; O'Dea et al. 2012). In contrast, most of the TWA is characterized by warmer waters and oligotrophic conditions year-round, with only a few small areas of upwelling as the exceptions (D'Croz and O'Dea 2007; Rueda-Roa and Muller-Karger 2013; Jackson and O'Dea 2023). The formation of the CAI is recognized as one of the most consequential events in marine biogeography, representing a natural laboratory for studying the processes of genetic divergence, speciation, and adaptation in marine species.

This vicariant event led to the formation of many geminate species pairs (sister species separated by the formation of a barrier; Jordan 1908). While the timing of the uplift and final closure of the CAI has been debated, with some estimates of final closure closer to 10 to 30 MYA (Bacon et al. 2015; Montes et al. 2015), an estimate of final closure at 2.8 MYA is now largely accepted (O'Dea et al. 2016). This closure, and the formation of geminate species pairs, has been the topic of extensive research. A large body of literature has sought to elucidate the exact timing of species divergence to better understand whether speciation occurred upon final closure, or whether it occurred earlier throughout the millions of years of gradual uplift leading up to the final closure (Lessios 2008). As a result, it is now generally understood that the timing of divergence can vary substantially among taxa. Some species have diverged prior to the final closure of the CAI, including molluscs (*Arca mutabilis & A. imbricata*; >30 MYA; Marko 2002), fishes (Eleotridae, Apogonidae, 5 to 15 MYA; Thacker 2017, and Haemulidae; Tavera et al. 2012), and snapping shrimps (genus *Alpheus*; 3 to 18 MYA; Knowlton and Weigt 1998; Hurt et al. 2009). These instances were likely a result of restriction of gene flow due to the gradual uplift of islands and a presumed change in depth across the Central American Seaway. Fewer molecular studies have evaluated the potential for reproductive isolation among species on either side of the CAI. These few examples include shrimps (Knowlton et al. 1993) and sea urchins (Lessios 1984; Lessios and Cunningham 1990), and in many cases hybridization is still possible despite millions of years of separation (Lessios 2008). Despite these advances, questions remain on the role geographic isolation and adaptation to local conditions have played in the differentiation of species separated by the CAI at the genome level.

One pair of geminate taxa that was formed by the uplift of the CAI are Sergeant Majors (damselfishes, Pomacentridae), with the Panamic Sergeant Major (*Abudefduf troschelii*) in the TEP and the Sergeant Major (*Abudefduf saxatilis*) in the Tropical Atlantic (Fig. 1). We note that these two species are not geminate species strictly speaking since *A. saxatilis* is closely related to the African Sergeant *A. hoefleri* in the Eastern Atlantic, and this process appears to have taken place after the divergence from *A. troschelii* (Campbell et al. 2018; Tang et al. 2021). To reflect this, we do not refer to *A. saxatilis* and *A. troschelii* as geminate species but transisthmian species instead. Estimates date the divergence between *A. troschelii* and *A. saxatilis* around 2.4 MYA to 2.5 MYA, coinciding with the final closure of the CAI at 2.8 MYA (Quenouille et al. 2004; Campbell et al. 2018; Rabosky et al. 2018; McCord et al. 2021; Tang et al. 2021). Today, *A. saxatilis* is distributed throughout the tropical and subtropical Atlantic Ocean, in warm, tropical waters (Fishelson 1970), and *A. troschelii* is more narrowly distributed throughout the TEP and experiences colder and more variable temperatures throughout its range (Robertson and Allen 2024; Fig. 1). Prior to their divergence, the common ancestor was likely distributed throughout the Central American Seaway, and the uplift of the CAI formed two isolated populations that underwent allopatric speciation. The different environments experienced by this transisthmian species pair provide an opportunity to evaluate the role of geographic isolation and adaptation in the divergence of closely related species.

Despite decades of molecular analyses of geminate species separated by the CAI (Knowlton and Weigt 1998; Lessios 2008; Thacker 2017), few studies have leveraged whole-genome sequencing to study their evolutionary history. Long-read technologies increase the reliability of de novo assemblies in non-model taxa, leading to higher quality reference genomes for subsequent analysis of population divergence, gene expression, and epigenetics. Long-read sequences are particularly well suited for the assembly of long repetitive regions, and to identify large structural variants (SVs) by spanning these regions—including SV breakpoints—entirely (Luan et al. 2020; Mérot et al. 2020). This is relevant as structural variants can play a role in speciation and have been shown to be associated with environmental adaptation in both plants and animals (Kirkpatrick 2010; Fuller et al. 2018; Tigano et al. 2018; Mahmoud et al. 2019; Cayuela et al. 2021; Zhang et al. 2021; Akopyan et al. 2022).

This study presents the first long-read genomes for a transisthmian species pair, and the first genomes for the genus *Abudefduf*. These fishes were chosen based on their role in coastal food webs, abundance in shallow marine ecosystems, as well as their territorial behavior, representing a group of great ecological relevance in tropical coastal regions of the Americas (Cooper et al. 2009; Villegas-Hernández et al. 2022). Within this group, the *A. troschelii*/*A. saxatilis* pair was selected because it most likely diverged close to the final closure of the CAI. The main aims of this study were to: (i) compare the structure of the genomes of a transisthmian species pair; (ii) compare the history of effective population sizes of the two species; and (iii) identify gene families that have expanded or contracted during the differentiation of the two species. This study represents a first step forward in understanding the complex impact of the rise of the CAI on the divergence of marine organisms, while also developing new genomic resources for the study of damselfishes.

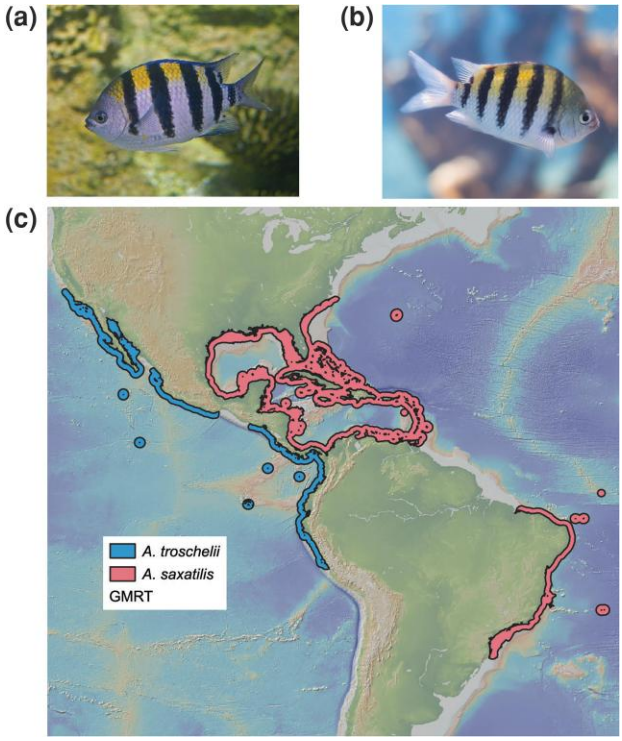

Fig. 1. Photographs and new-world distributions of the two species considered in this study. a) Image of *Abudefduf troschelii*, image credit: Hectonichus, License CC-BY-SA 3.0, b) Image of *Abudefduf saxatilis*, image credit: Matthew T Rader, MatthewTRader.com, License CC-BY-SA 4.0. c) Map of the distribution of *A. troschelii* (blue) and A. saxatilis (pink) limited to the Americas, obtained from the IUCN RedList.

Results and Discussion

Genome Assembly and Annotation

The genome of one individual of *A. saxatilis* from Florida was sequenced using two PacBio cells. The resulting de novo genome assembly had a total length of 811 Mb

across 287 contigs with an N50 of 25.8 Mb, a maximum contig length of 36 Mb, a minimum contig length of 13 kb, a GC content of 40.96%, a heterozygous read coverage of 44X, and a homozygous read coverage of 88X (Table 1). The heterozygosity was determined to be 0.563% (Fig. S1) which is similar to previously reported measures of heterozygosity in other teleost species (Tigano et al. 2021). The Benchmarking Universal Single-Copy Orthologs (BUSCO) analysis of genome completeness identified 3,598 (98.85%) complete genes, 30 missing genes (0.82%) and 12 (0.33%) partial genes. The mitochondrial genome was detected in one of the contigs resulting from the assembly, resulting in 16,695 bp, with 13 genes and 45%GC content (Fig. S2). Decontamination using Blobtools (Laetsch and Blaxter 2017; Challis et al. 2020) identified that 86.57% of scaffolds mapped to Chordata (124 reads, spanning 799,260,409 bp), 17% had no hit (146 reads, spanning 10,361,765 bp), and <0.05% mapped to Bacteria (5 reads), Proteobacteria (6 reads), or algal material. Scaffolds matching the last three categories were removed before preparation of the final assembly.

The genome of one individual of *A. troschelii* from Panama was sequenced using a single PacBio cell. The de novo genome assembly of *A. troschelii* had a total length of 802 Mb across 126 contigs with an N50 of 23 Mb, a heterozygous read coverage of 20X, and a homozygous read coverage of 39X (Table 1). The heterozygosity was determined to be 0.593% (Fig. S1). The BUSCO analysis identified 3,595 (98.76%) complete genes, 32 missing genes (0.88%) and 13 (0.36%) partial genes. Both genomes had higher completeness than many other published teleost genomes to date (Table 2). The partial mitochondrial genome of *A. troschelii* was obtained, and resulted in 15,716 bp, 13 genes and 47% GC content (Fig. S2). For *A. troshelii*, BlobTools identified 90% of scaffolds as Chordata (81 scaffolds, spanning 795,569,153 bp), while the remaining 10% (9 scaffolds, spanning 591,150 bp) were classified as no-hit. No scaffolds were assigned to Bacteria, Proteobacteria, or algal lineages in the *A. troshelii* assembly and no scaffolds were removed during final assembly preparation.

Raw reads mapped back to the assembled genomes were examined using Merqury (Rhie et al. 2020). We found a consensus quality value (QV) of Q63.88 and a k-mer completeness of 90.68% for *A. troschelii*, and a QV of Q61.08 and a k-Mer completeness of 91.83% for *A. saxatilis* (Figs. S3, S4). These slight differences in consensus quality, mitochondrial and nuclear genome length, contig number and N50 are most likely due to differences in sequencing effort (two PacBio cells for *A. saxatilis* vs. one for *A. troschelii*), rather than processes such as drift, selection, or mutations.

Table 1 Assembly statistics of the de novo assemblies of *A. saxatilis* and *A. troschelii* compared to previously published and annotated fish genomes

Scientific name	Common name	Seq counts	Total bases	GC%	avgLen	medianLen	maxLen	minLen	N50	L50	Source
Abudefduf saxatilis	Sergeant Major	270	809,622,174	40.96	2,998,600	69,846	36,082,483	13,662	25,792,322	13	This study
Abudefduf troschelii	Panamic Sergeant Major	126	802,853,132	40.97	6,371,850	768,645	36,674,582	20,499	23,322,547	14	This study
Amphiprion percula	Orange Clownfish	365	908,900,000	39.53	2,483,486	39,850	46,100,000	16,638	38,400,000	12	Lehmann et al., (2019)
Lates calcarifer	Asian Seabass	3808	668,481,366	40.75	175,547	25,397	18,910,200	3232	1,191,366	119	Vij et al., (2016)
Fugu rubripes	Fugu	128	384,126,662	45.66	3,000,990	71,604	29,232,231	16,447	16,705,553	10	Aparicio et al., (2002)
Gasterosteus aculeatus	Threespine Stickleback	2937	471,894,361	44.66	160,672	3573	34,181,212	226	20,445,003	10	Nath et al., (2021)
Danio rerio	Zebrafish	1923	1,679,203,469	36.60	873,221	146,921	78,093,715	650	52,186,027	14	Howe et al., (2013)
Plagiognathops microlepis	Small-scale Yellowfin	98	976,347,075	37.5	14,793,138	47,676	57,807,229	1,000	38,400,000	11	Liang et al., (2024)
Lagodon rhomboides	Pinfish	2704	785.3 Mb	42.5	290,433	4,045	39,332,093	482	31,400,000	12	Eaton et al., (2024)

Table 2 BUSCO statistics detailing the completeness of the genome of *A. saxatilis* and *A. troschelii* compared to previously published and annotated fish genomes

Scientific name	Core genes queried	Complete	Complete + Partial	Partial	Missing	Average # orthologs per core genes	% detected core genes with more than 1 ortholog
Abudefduf saxatilis	3,640	3,598 (98.85%)	3,610 (99.18%)	12	30 (0.82%)	1.02	1.81
Abudefduf troschelii	3,640	3,595 (98.76%)	3,608 (99.12%)	13	32 (0.88%)	1.01	0.58
Amphiprion percula	3,640	3,575 (98.21%)	3,593 (98.71%)	18	47 (1.29%)	1.01	0.87
Lates calcarifer	3,640	3,585 (98.5%)	3,606 (99.07%)	21	32 (0.93%)	1.04	2.85
Gasterosteus aculeatus	3,640	3,530 (96.98%)	3,566 (97.97%)	36	74 (2.03%)	1.03	2.49
Danio rerio	3,640	3,492 (95.93%)	3,554 (97.64%)	62	86 (2.36%)	1.10	8.82
Fugu rubripes	3,640	3,525 (96.84%)	3,549 (97.50%)	24	91 (2.50%)	1.02	1.90
Plagiognathops microlepis	3,640	3,570 (98.09%)	3,590 (98.63%)	20	50 (1.37%)	NA	NA
Lagodon rhomboides	3,640	3,598 (98.85%)	3,607 (99.09%)	9	33 (0.9%)	NA	NA

All statistics based on BUSCO v.5 to enable comparison among genomes, using the Actinopterygii dataset (actinopterygii_odb10; Simão et al. 2015).

Synteny

The de novo assemblies of both species produced large contigs that were near chromosome length. To confirm the contig lengths and compare them to the known chromosomes of another damselfish with a chromosome-level genome available, the genome of *A. saxatilis* was mapped to the genome of the orange clownfish *Amphiprion percula* (ENSMBL GCA_003047355.2). This mapping revealed the genome of *A. saxatilis* was close to chromosomal scale: one to four contigs from the Sergeant major assembly mapped to 16 chromosomes of *A. percula*. Meanwhile, the number of contigs of *A. saxitilis* was higher (8 to 31 contigs) for eight chromosomes of *A. percula* (Table S1). The mapping of *Abudefduf* reads to *A. percula* was not used for further downstream analyses for *A. saxatilis* or *A. troschelii*.

As expected, there is high synteny between *A. saxatilis* and *A. troschelii*, with large blocks of synteny and some inversions or chromosomal rearrangements between the two (Fig. 2). Three large potential inversions were identified (Fig. 2): a 28 Mb inversion hereafter called Inversion 1 (*A. saxatilis*: ptg000023; *A. troschelii*: ptg000042); an 8 Mb inversion called Inversion 2 (*A. saxatilis*: ptg000026; *A. troschelii*: ptg000001); and another 8 Mb inversion termed Inversion 3 (*A. saxatilis*: ptg000019; *A. troschelii*: ptg000019).

We manually examined the genes around the inversion breakpoints (5 genes before, first and last 5 within inversion, and 5 after) for both species. Around the Inversion 1 breakpoints, genes that stood out include gonadotropin releasing hormone receptor (*GNRHR*) and three high choriolytic enzyme 1-like genes in *A. saxatilis*, however these genes were not annotated in our *A. troschelii* genome. No notable genes seeming to relate to the environment stood out around the Inversion 2 breakpoints. Meanwhile, around the Inversion 3 breakpoints, genes that stood out include empty spiracles 3 homeobox (*emx3*), and fibroblast growth factor receptor-like 1 (*fgfrl1*) in *A. saxatilis*, while fork-head box (*FOXl1*) was found in *A. troschelii*.

A search for specific gene ontology (GO) categories that may be over-represented in each inversion compared to the rest of the genome found no significant genes above the 10% false discovery rate (FDR) for any of the three possible GO categories (Molecular Function, Biological Process, or Cellular Component) for *A. saxatilis* in any of the three inversions, or for *A. troschelii* on Inversion 1. For Inversion 2 in *A. troschelii* one significant GO category above the 10% FDR was detected for MF (4/8 amine transmembrane transporter activity, P-value = 0.006), BP (4/8 amine transmembrane transporter activity, P-value = 0.006), and CC (7/65 brush border membrane, P-value = 0.095). The amine transmembrane transporters allow amines to be transported across a membrane which could include neurotransmitters such as serotonin, dopamine, and norepinephrine impacting synapses (Purves et al. 2001). The analysis on *A. troschelii* Inversion 3 found one significant GO category above the 10% FDR for BP (activation of transmembrane receptor protein tyrosine kinase activity, $P = 0.009$) and MF (transmembrane receptor protein tyrosine kinase activator activity, $P = 0.032$), but none were found in CC. These transmembrane receptor tyrosine kinases play a role in cell signaling and growth (Li and Hristova 2010).

Inversions and chromosomal rearrangements may contribute to local adaptation (Kirkpatrick 2010; Akopyan et al. 2022). However, these genomes were

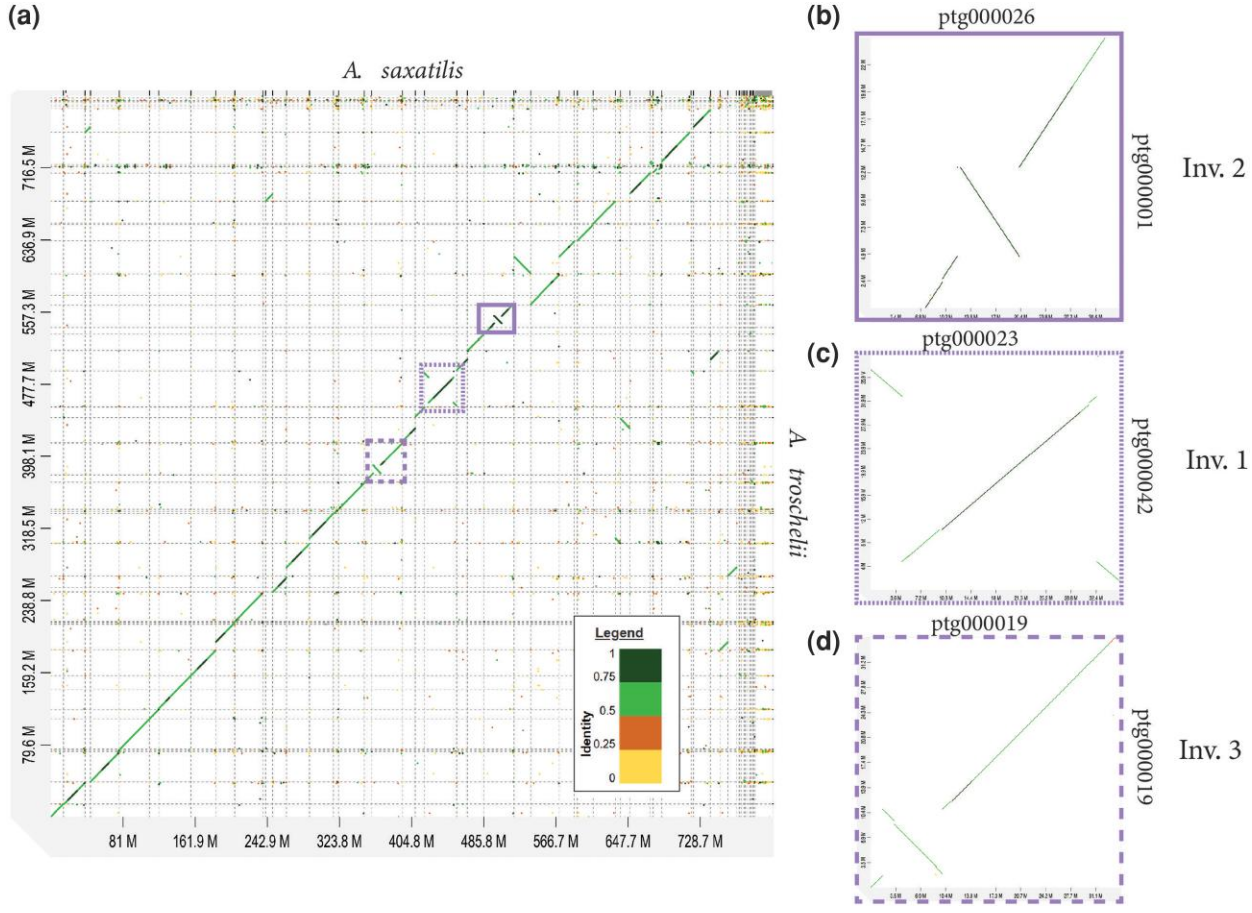

Fig. 2. Analysis of synteny mapping the *A. saxatilis* assembly (*x*-axis) to the *A. troschelii* assembly (*y*-axis) to visualize large rearrangements or structural variants between the two genomes. Genes falling along the diagonal are matches to one another, and potential structural rearrangements are seen when genes fall off the diagonal or when the diagonal is in the opposite direction. a) Synteny analysis between the entire genome of *A. saxatilis* (*x*-axis) and *A. troschelii* (*y*-axis) from D-GENIES. The analysis shows three potential inversions across the genome shown in panels B–D.

based on a single individual for each species, and whole genome sequencing within and across populations of each group will be required to determine the prevalence and age of the inversions, and whether they may contribute to local adaptation.

Demographic History

The results of the Pairwise Sequentially Markovian Coalescence (PSMC; Li and Durbin 2011) analysis are presented in Fig. 3. The first thing to note is that the analysis fails to go back until the divergence point between the two species, potentially due to a scarcity of alleles with deep coalescence times in *A. troschelii*. This illustrates the limitation of this approach when it comes to addressing the initial divergence of transisthmian species, even when it occurred upon final closure of the CAI. Nevertheless, our objective was not to date the divergence time between the two species with the PSMC analysis since such an estimate would rely heavily on mutation rate and generation time estimates, which are uncertain. In this regard divergence estimates provided by phylogenetic methods, in particular the ones that are independent of the timing of the final closure of the CAI, appear more reliable (Quenouille et al. 2004; Campbell et al. 2018; Rabosky et al. 2018; McCord et al. 2021; Tang et al. 2021).

The PSMC analysis identified a peak in the effective population size (Ne) of both *A. saxatilis* and *A. troschelii* around 2.8 MYA. This was followed by a decline in effective population size for *A. troschelii,* which remained very low until the limit of the resolution of our analysis approximately 3 thousand years ago (KYA, Fig. 3). Meanwhile, *A. saxatilis* showed a decline, followed by a subsequent increase that occurred around 700 KYA, which was followed by a decline until the limit of the resolution of our analyses. Overall, the estimates showed a trend of larger population size of *A. troschelii*

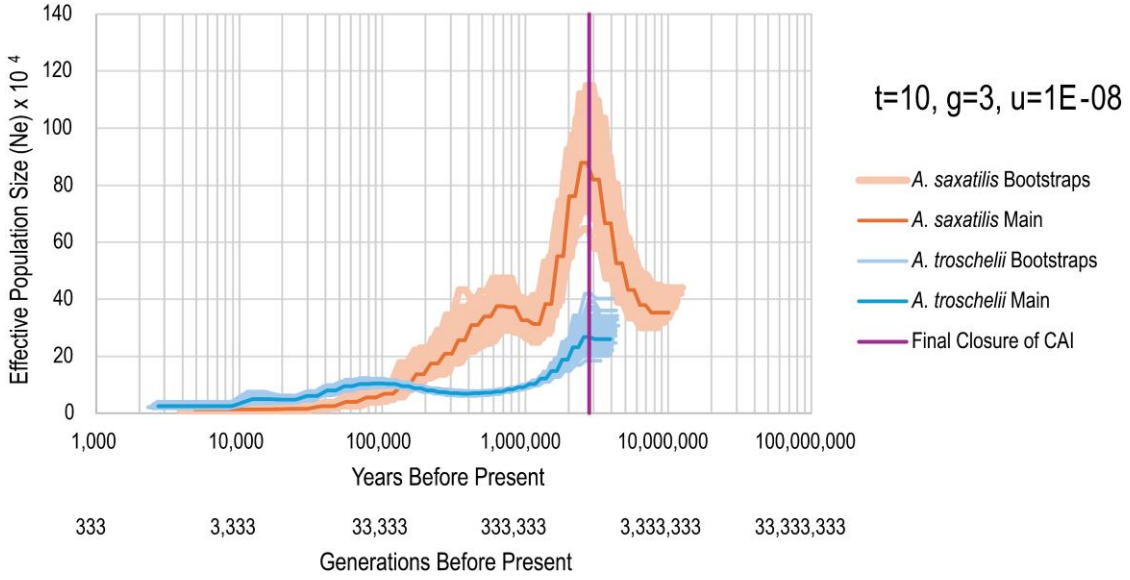

Fig. 3. PSMC estimates of effective population size of the genomes of *A. saxatilis* (orange) and *A. troschelii* (blue). Dark lines indicate the average posterior estimate, while lighter colored lines indicate a bootstrap estimate. The purple vertical line represents the time of final closure of the CAI at 2.8 MYA.

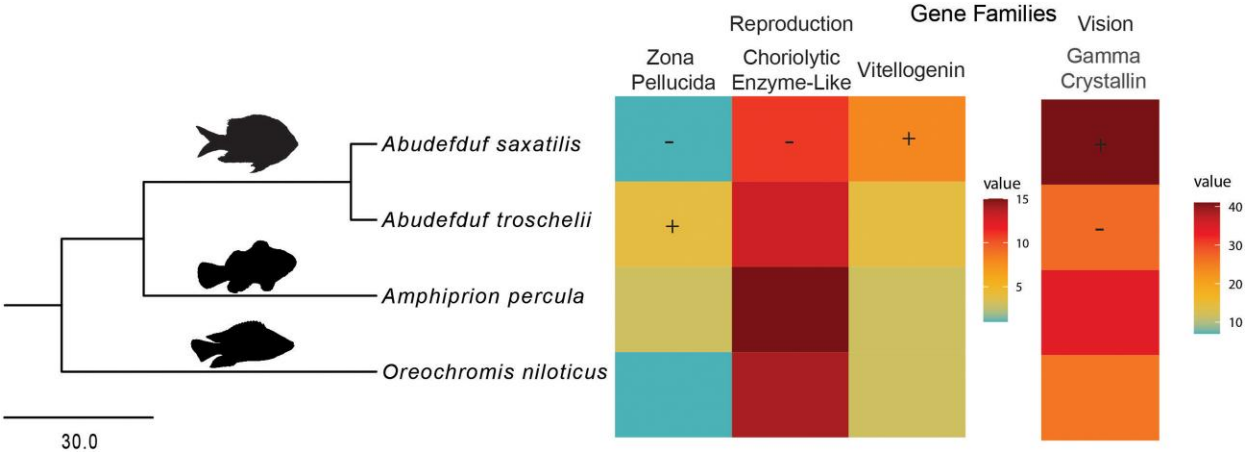

Fig. 4. Gene family expansion of genes related to reproduction (zona pellucida, high and low choriolytic enzyme-like, and vitellogenin) and vision (gamma crystalin M family) across the phylogeny. Significant expansions and contractions are indicated with a + and −, respectively.

than *A. saxatilis* beginning at 100 KYA. Similar patterns were found in the shape of estimates when PSMC was run with different parameters, however the timing and values for effective population size varied significantly when changing parameters such as mutation rate (*u*) and generation time (*g*) for the species, or PSMC parameters such as *–t* or *–p* (Figs S5 to S7).

The increase and subsequent decline in effective population size for both species could be associated with the oceanographic changes experienced in the TWA and the TEP following the closure of the CAI. Both species also had an overall plateau or slight decline in effective population size the past 100 KY. This is similar to the patterns observed in intertidal limpets (Giles et al. 2025) but it contrasts with recent studies in fishes that have documented rapid expansions around the last glacial maximum (Eaton et al. 2024; Gatins et al. 2024; Hirase et al. 2025). Furthermore, the low estimates of effective population sizes from present back to 100 KYA are unexpected, as these species are currently abundant in reefs throughout the Caribbean and Pacific (Froese and Pauly 2024). It is possible that these methods lack the resolution to show more recent changes in effective population size after the last glacial maximum

(Nadachowska-Brzyska et al. 2016). Further, PSMC results are likely to reflect variation in not just effective population size but also population genetic structure and selection regimes in the two sides of the CAI since these two processes are also expected to affect coalescence times (Mather et al. 2020).

Analysis of Gene Family Expansion

An analysis of gene family size was conducted using CAFE (Mendes et al. 2020) to identify gene families that have expanded or contracted in the two species. This analysis found 118 gene families significantly expanded or contracted in *A. saxatilis*, 104 in *A. troschelii*, and 63 in their most recent common ancestor. Among the gene families significantly expanded/contracted in *A. saxatilis* and *A. troschelii*, 41 were significant in both species.

Many of the gene families that were significantly expanded or contracted (Table S2) were expected due to their fast evolutionary rates and previously reported variability among taxa. These included genes associated with immunity, odorant and olfactory receptors, and metabolic related genes. Nevertheless, a few functional groups stood out as candidates for adaptation to the respective environments of the TWA and TEP, such as vision, reproduction, olfaction, stress response, and biosynthetic pathways. One gene family that plays a role in vision is gamma crystallin M, which is involved in eye lens development (Bloemendal et al. 2004). The expansion of gamma crystallin M in *A. saxatilis* (41 genes) and contraction in *A. troschelii* (27 genes; Figs 4 and S8) may be an indicator of their evolution in different light environments. The TWA has clearer water due to lower productivity and fewer nutrients than the TEP, and gene families influencing vision may be under stronger selection in clear-water environments compared to the turbid waters of the Pacific.

Three findings could be associated with reproductive differences between the transisthmian species (Figs 4 and S9). Vitellogenins (*vtg*) are the dominant contributors to vertebrate egg yolk and were significantly expanded in *A. saxatilis* (8 genes) compared to *A. troschelii* (4 genes). This expansion of proteins may indicate increased resources to egg development, which is consistent with previous findings that *A. saxatilis* has higher nutritional provisioning to hatchlings and a faster growth rate than *A. troschelii* (Wellington and Robertson 2001). This is consistent with the regional-productivity hypothesis demonstrating that Caribbean species produce larger eggs than their close relatives in the Pacific due to lower larval food supplies in the Caribbean (Robertson and Collin 2015; Jackson and O'Dea 2023). In a similar vein, a zona pellucida protein gene family (*zpax4*), which plays a role in forming the egg casing, was significantly expanded in *A. troschelii* (4 genes) and contracted in *A. saxatilis* (1 gene). Zona pellucida genes have been shown to vary across the fish phylogeny and multiple gene duplication events have been previously identified (Sano et al. 2022). High and low choriolytic enzyme-like (*hce*), which are the two types of hatching enzymes which cooperatively digest the egg envelope (Yasumasu et al. 1989; Kawaguchi et al. 2014), were also significantly contracted in *A. saxatilis* (11 genes) compared to *A. troschelii* (13 genes). Thus, two gene families that relate to egg casing formation and digestion, respectively, are contracted in *A. saxatilis* compared to *A. troschelii*.

While these gene family expansions and contractions show an interesting tie to the environment in which the species evolved, there are important caveats. First, while the gene families discussed were significantly expanded or contracted between this transisthmian species pair, this is not an exhaustive search of these gene types within the genomes. Secondly, this analysis relies on methods of gene family clustering that can lead to false identification of orthologs, and this is susceptible to different results based on the input parameters and clustering method employed. Finally, while we were able to identify significant expansions and contractions of gene families and speculate on how that may be significant given the species' environment, additional analyses are needed to link these expansions and contractions to functional changes between these species. With this in mind, we recommend future studies to directly link the hypothesized expanded gene families with the environmental differences in the TWA and TEP.

Conclusion

Despite the rich history of applying molecular tools to the study of species separated by the CAI, this manuscript represents the first effort to characterize the genomic divergence of a transisthmian species pair in the TWA and TEP. The analyses presented here showed high conservation of the genome structure between transisthmian *Abudefduf*, with three large inversions between the two species. These inversions might have played a role in adaptation and speciation in *Abudefduf*, but further analyses with broader sampling are necessary to test this hypothesis. Many gene families that were significantly expanded or contracted are seemingly a result of adaptation to specific environmental conditions and the development of reproductive isolation. These findings, particularly the expansion of vitellogenin in *A. saxatilis*, are consistent with previous findings and hypotheses of differential resource allocation to egg development among species in the TEP

and TWA. These genomes serve as resources for future studies on transisthmian species, to improve our understanding of the evolutionary processes promoted by the closure of the CAI.

Materials and Methods

Sampling

One individual of *A. saxatilis* was obtained from an aquarium distributor (www.saltwaterfish.com) in July 2020, through whom the fish had been collected from Florida. The fish was shipped to the laboratory and was euthanized via cervical transection and tissue samples dissected (Auburn University IACUC Protocol Number PRN 2020-3708).

One adult individual of *A. troschelii* was collected on Isla Cocos (Coiba, Panama) in November 2020 using a pole spear. Muscle tissue was immediately dissected and fast frozen on liquid nitrogen. Upon arrival to Panama City tissue sample was transferred to −80 °C for storage at the Smithsonian Tropical Research Institute Collection (Mi Ambiente collection permit: SE/AO-4-19).

Genome Sequencing

Gill and muscle tissue of one individual of *A. saxatilis* was dissected and immediately used for extraction of genomic DNA using the Qiagen Genomic-tip 500/G Kit. The genomic DNA extracts were sent to Novogene for PacBio Sequel II DNA HiFi library preparation. Briefly, to generate the SMRTbell libraries, the DNA fragments were ligated to hairpin adapters, to generated double stranded fragments. The hairpin dimers and the products that were not ligated were removed through AMPure beads and exonuclease treatment, respectively. After purification, the final primer annealing and binding of annealed SMRTbell templates was completed. There libraries were sequenced on two flow cells of PacBio Sequel II system (HiFi/CCS mode).

For *A. troschelii,* high molecular weight DNA was extracted from frozen muscle tissue using a phenol:chloroform method (Sambrook and Russell 2006). The DNA was cleaned with 3X KAPA pure Beads (Roche Sequencing) and concentration/purity of the extracted DNA was assessed using a Qubit fluorometer (Thermo Fisher Scientific). The integrity of the DNA was evaluated with a field inversion gel (Pippin Pulse, Sage Science). On average, DNA fragments were ~50 kb in length. The PacBio HiFi library was prepared with the SMRTbell Express Template Prep Kit 2.0 (PacBio 100-938-900) using the PacBio low input protocol (DNA sheared to 15 kb) for HiFi sequencing followed by AMPure bead cleanup. The genomic library was sequenced on a single 30-h movie 8 M SMRT cells in CCS mode on the PacBio Sequel II system at the Brigham Young University (BYU) DNA Sequencing Center.

Genome de Novo Assembly and Annotation

For both species, the raw genomic reads were assembled de novo using PacBio's assembler HifiASM v. 0.16.1 (Cheng et al. 2021). Contamination in the genome was identified using blobtools v. 3.1.0 (Laetsch and Blaxter 2017; Challis et al. 2020) and removed prior to annotation and comparative analyses. The completeness of the assembled genome was evaluated using BUSCO v.5 Actinopterygii dataset (actinopterygii_odb10; Simão et al. 2015). Heterozygosity of the raw reads was determined by counting k-mers in Jellyfish (Marçais and Kingsford 2011) and calculating heterozygosity in GenomeScope (Vurture et al. 2017). Reads were mapped back to the assembled genome to examine completeness using Merqury (Rhie et al. 2020). Genome annotation was conducted with ESGA (Torres and Höppner 2021). This pipeline identifies and masks repeats using RepeatModeler v. 2.0.2 (Flynn et al. 2020) and RepeatMasker v. 4.1.2 (Smit et al. 2013) and then proceeds with Augustus v. 3.4.0 (Stanke et al. 2008) for gene model identification and training. An assembled transcriptome of *A. saxatilis* was also provided to Augustus to build gene models (Swank et al. 2025). The third dataset provided to this pipeline were known protein data from three species of the Pomacentridae family: *Acanthochromis polyacanthus* (ENSMBL GCA_002109545.1), *Amphiprion ocellaris* (ENSMBL GCA_022539595.1), and *Stegastes partitus* (ENSMBL GCA_000690725.1). Finally, this pipeline uses EVidenceModeler v. 1.1.1 (Haas et al. 2008) and PASA v. 2.5.1 (Haas et al. 2003) to build consensus gene models. The consensus gene models were functionally annotated using eggNOG-mapper v2.1.9 (Cantalapiedra et al. 2021), based on eggNOG orthology data (Huerta-Cepas et al. 2019), with the flag –tax_scope 7898 to restrict the taxonomic scope used for annotation to Actinopterygii.

Mitochondrial Genomes

To evaluate if the assembled contigs contained the mitochondrial reads, the mitochondrial genome was used as reference. For this, the program BLASTn was used, with an *e*-value of e^{-10}, using the of *Amphiprion percula* (Lehmann et al. 2019) reads as reference. Before the decontamination step, the genomes of *A. saxatilis* and *A. troschelii* assembled with HifiASM v. 0.16.1 (Cheng et al. 2021) were used as queries in the BLASTn search. The contigs matching the longest mitochondrial fragments were selected, they trimmed

manually, and annotated using MitoFish v2025.06 (Sato et al. 2018). A figure of the genome and the corresponding annotation was generated with the program Chloroplot (Zheng et al. 2020).

Comparative Genomic Analyses

A pairwise comparison of synteny was conducted between the genomes of *A. saxatilis* and *A. troschelii* to visualize structural variation in the genomes of the geminate sister species. Synteny analyses were conducted using D-GENIES online portal (Cabanettes and Klopp 2018) using the aligner Minimap2 v2.24 (Li, 2018) in D-GENIES.

Following the synteny analysis, the large inversions identified between *A. saxatilis* and *A. troschelii* (see results) was examined for the genes within and immediately surrounding the inversion. To see whether there was a disproportionate number of genes with similar GOs within the inversion, we extracted GO terms from the annotation and used these to run a rank-based gene ontology analysis using the program GO-MWU (https://github.com/z0on/GO_MWU) using a Fisher's test (Wright et al. 2015). We tested whether specific GO terms were overrepresented in the inversion compared to the rest of the genome using all three GO categories (Molecular Function, Biological Process, and Cellular Component).

Analyses of demographic history were conducted using PSMC, which scans a genome sequentially for recombination breakpoints and, using the number of coalescence events along with estimates of mutation rate and generation time, estimates the effective population size throughout time in each non-recombining fragment (Nadachowska-Brzyska et al. 2016). Multiple PSMC analyses were conducted, varying the values for input parameters to examine the robustness of Ne estimates to input parameters. Specifically, they were conducted using a rate of mutation (*-u*) spanning from $1\times\ 10^{-8}$ to 5.97×10^{-9}, *–t* values of 5 versus 10, and *–p* flags of "4 + 5*3 + 4", "4 + 20*3 + 4", and "4 + 25*2 + 4 + 6", and generation times (*–g*) between 2 and 7 years. The *–p* and *–t* flags can be used together to estimate overfitting by obtaining the expected number of segments in the interval and ensuring they were above 20, following Li and Durbin (2011). In all tests, there was no overfitting found, so we used the more complex model with more free parameters as it provided a more detailed overview of changes in Ne over time. Mutation rates used were based on estimates from previously published mutation rates for a variety of fish species. The mutation rate of Atlantic herring has been estimated to be 2.0×10^{-9} per site per generation (Feng et al. 2017), whereas in sticklebacks the mutation rate has been estimated to be 3.7×10^{-8} (Liu et al. 2016). A more recent study estimated the mutation rate averaged across all fishes in their study to be 5.7×10^{-09} using pedigree analyses (Bergeron et al. 2023), however this pushed estimates of divergence back significantly later than well-supported phylogenetic estimates. In this same study, the published estimates for *Amphiprion ocellaris* ranged from.5×10^{-08} to 1.2×10^{-08}, and given the close relationship to *Abudefduf*, we used a mutation rate within this range of 1×10^{-08} as the mutation rate used in our study. Generation times ranging from 2 to 7 years were used in analyses and presented in the supplement, however we present results for a generation time of 3 years. This estimate was based on the age *A. saxatilis* reaches sexual maturity, which is ~2.5 years (Villegas-Hernández et al. 2022). In the PSMC context, generation time refers to the mean age of reproducing individuals, which may be different from age at sexual maturity, and would be closer to 3 years.

A gene family expansion analysis was conducted using CAFE v5.0 (Mendes et al. 2020). A total of 17 taxa were included in the analysis to represent various clades across the fish tree of life (Table S3). Protein data for each species included was downloaded from Ensembl biomart. Proteins for each species were all compared to one another using BLAST to obtain similarity scores, with an *e*-value cutoff of 1e−6, and they were clustered by similarity using mcl (Van Dongen 2008), with an inflation parameter of 2 ($I = 2$).

A time-calibrated phylogenetic tree was created by downloading sequences used in the fish tree of life (Rabosky et al. 2018; Chang et al. 2019) and subsampling it to only include the taxa used in the CAFE analysis. This alignment was then edited manually to remove loci for which data was missing for >58% of the taxa (removed genes with 7 or fewer samples). The tree was created using BEAST v2.5 (Bouckaert et al. 2019) with bmodeltest (Bouckaert and Drummond 2017) to infer the model of evolution for each loci and was run for 50,000,000 generations sampling the posterior every 5,000 generation. Two fossil calibrations used in the Fish Tree of Life that applied to the nodes in the subset tree were used. These are on the most recent common ancestor (MRCA) of all Pomacentridae with a gamma distribution offset 44 and an alpha and beta parameter of 5 and 3, respectively (including *Abudefduf saxatilis, Abudefduf troschelii,* and *Amphiprion percula*), and the second on the MRCA of Otomorpha with a gamma distribution offset 145, with an alpha parameter of 5 and a beta parameter of 3 (including *Astyanax mexicanus, Clupea harengus, Danio rerio, and Ictalurus punctatus*).

Supplementary Material

Supplementary material is available at *Genome Biology and Evolution* online.

Acknowledgments

We would like to thank Adam Hallaj for his assistance with DNA extractions, Katie Eaton for her helpful guidance on analyses, and members of the Bernal Lab and the Phyletica Lab. Special thanks to Jamie Oaks and Matthew Buehler for discussions and feedback on the project. Computational work was possible through resources provided by the Easley Computing Cluster at Auburn University and the Alabama Super Computer Authority (ASC).

Funding

This work was supported by Auburn University through startup funds to MAB, by the Smithsonian Tropical Research Institute through funds provided to WOM, and a German Research Foundation individual grant to OP and MH (PU 571/14-1).

Data Availability

Raw reads and assembled genomes are available through NCBI (BioProject: PRJNA1244908, *A. saxatilis* BioSample: SAMN47735137, SRA read ID: SRR32938816, Genome Accession: JBUAPX000000000, and *A. troschelii* BioSample: SAMN47735138, SRA read ID: SRR32938815, Genome Accession: JBUAPW000000000). The mitochondrial genomes are available through NCBI accession numbers PX923159 for *A. saxatilis*, and PX923160 for *A. troschelii*. Scripts used for analyses and selected results files are available on github: https://github.com/C-Tracy/GenomeAlignmentAnnotation.

Literature Cited

Akopyan M, et al. Comparative linkage mapping uncovers recombination suppression across massive chromosomal inversions associated with local adaptation in Atlantic silversides. Mol Ecol. 2022:31:3323–3341. https://doi.org/10.1111/MEC.16472.

Aparicio S, et al. Whole-genome shotgun assembly and analysis of the genome of *Fugu rubripes*. Science. 2002:297:1301–1310. https://doi.org/10.1126/SCIENCE.1072104.

Bacon CD, et al. Biological evidence supports an early and complex emergence of the Isthmus of Panama. Proc Natl Acad Sci U S A. 2015:112:6110–6115. https://doi.org/10.1073/pnas.1423853112.

Bergeron LA, et al. Evolution of the germline mutation rate across vertebrates. Nature. 2023:615:285–291. https://doi.org/10.1038/s41586-023-05752-y.

Bloemendal H, et al. Ageing and vision: structure, stability and function of lens crystallins. Prog Biophys Mol Biol. 2004:86:407–485. https://doi.org/10.1016/J.PBIOMOLBIO.2003.11.012.

Bouckaert R, et al. BEAST 2.5: an advanced software platform for Bayesian evolutionary analysis. PLOS Comput Biol. 2019:15:e1006650. https://doi.org/10.1371/JOURNAL.PCBI.1006650.

Bouckaert RR, Drummond AJ. bModelTest: bayesian phylogenetic site model averaging and model comparison. BMC Evol Biol. 2017:17:42. https://doi.org/10.1186/S12862-017-0890-6/FIGURES/6.

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

Campbell MA, Ross Robertson D, Vargas MI, Allen GR, McMillan WO. Multilocus molecular systematics of the circumtropical reef-fish genus *Abudefduf* (Pomacentridae): history, geography and ecology of speciation. PeerJ. 2018:2018:e5357. https://doi.org/10.7717/peerj.5357.

Cantalapiedra CP, Herņandez-Plaza A, Letunic I, Bork P, Huerta-Cepas J. eggNOG-mapper v2: functional annotation, orthology assignments, and domain prediction at the metagenomic scale. Mol Biol Evol. 2021:38:5825–5829. https://doi.org/10.1093/MOLBEV/MSAB293.

Cayuela H, et al. Thermal adaptation rather than demographic history drives genetic structure inferred by copy number variants in a marine fish. Mol Ecol. 2021:30:1624–1641. https://doi.org/10.1111/MEC.15835.

Challis R, Richards E, Rajan J, Cochrane G, Blaxter M. BlobToolKit—interactive quality assessment of genome assemblies. G3-Genes Genom Genet. 2020:10:1361–1374. https://doi.org/10.1534/g3.119.400908.

Chang J, Rabosky DL, Smith SA, Alfaro ME. An r package and online resource for macroevolutionary studies using the ray-finned fish tree of life. Methods Ecol Evol. 2019:10:1118–1124. https://doi.org/10.1111/2041-210X.13182.

Cheng H, Concepcion GT, Feng X, Zhang H, Li H. Haplotype-resolved de novo assembly using phased assembly graphs with hifiasm. Nat Methods. 2021:18:170–175. https://doi.org/10.1038/s41592-020-01056-5.

Cooper WJ, Smith LL, Westneat MW. Exploring the radiation of a diverse reef fish family: phylogenetics of the damselfishes (Pomacentridae), with new classifications based on molecular analyses of all genera. Mol Phylogenet Evol. 2009:52:1–16. https://doi.org/10.1016/J.YMPEV.2008.12.010.

D'Croz L, O'Dea A. Variability in upwelling along the Pacific shelf of Panama and implications for the distribution of nutrients and chlorophyll. Estuar Coast Shelf S. 2007:73:325–340. https://doi.org/10.1016/J.ECSS.2007.01.013.

Eaton KM, Krabbenhoft TJ, Backenstose NJ, Bernal MA. The chromosome-scale reference genome for the pinfish (*Lagodon rhomboides*) provides insights into their evolutionary and demographic history. G3 (Bethesda). 2024:14:jkae096. https://doi.org/10.1093/g3journal/jkae096.

Feng C, et al. Moderate nucleotide diversity in the Atlantic herring is associated with a low mutation rate. Elife. 2017:6:e23907. https://doi.org/10.7554/ELIFE.23907.

Fishelson L. Behaviour and ecology of a population of *Abudefduf saxatilis* (Pomacentridae, Teleostei) at Eilat (Red Sea). Anim Behav. 1970:18:225–237. https://doi.org/10.1016/S0003-3472(70)80032-X.

Flynn, et al. RepeatModeler2 for automated genomic discovery of transposable element families. Proc Natl Acad Sci U S A. 2020:117:9451–9457. https://doi.org/10.1073/pnas.1921046117.

Froese R, Pauly D. *FishBase*. Version (10/2024). 2024. https://www.fishbase.org/

Fuller ZL, Leonard CJ, Young RE, Schaeffer SW, Phadnis N. Ancestral polymorphisms explain the role of chromosomal inversions in speciation. PLoS Genet. 2018:14:e1007526. https://doi.org/10.1371/journal.pgen.1007526.

Gatins R, Arias CF, Sánchez C, Bernardi G, De León LF. Whole genome assembly and annotation of the King Angelfish (*Holacanthus passer*) gives insight into the evolution of marine fishes of the Tropical Eastern Pacific. GigaByte. 2024:2024: 1–18. https://doi.org/10.46471/gigabyte.115.

Giles EC, et al. Comparative genomics points to ecological drivers of genomic divergence among intertidal limpets. Mol Ecol Resour. 2025:25:e14075. https://doi.org/10.1111/1755-0998.14075.

Haas BJ, et al. Improving the Arabidopsis genome annotation using maximal transcript alignment assemblies. Nucleic Acids Res. 2003:31:5654–5666. https://doi.org/10.1093/nar/gkg770.

Haas BJ, et al. Automated eukaryotic gene structure annotation using EVidenceModeler and the program to assemble spliced alignments. Genome Biol. 2008:9:R7. https://doi.org/10.1186/gb-2008-9-1-r7.

Hirase S, Nagano AJ, Nohara K, Kikuchi K, Kokita T. Phenotypic and genomic signatures of latitudinal local adaptation along with prevailing ocean current in a coastal goby. Mol Ecol. 2025:34: e17599. https://doi.org/10.1111/mec.17599.

Howe K, et al. The zebrafish reference genome sequence and its relationship to the human genome. Nature. 2013:496:498–503. https://doi.org/10.1038/nature12111.

Huerta-Cepas J, et al. eggNOG 5.0: a hierarchical, functionally and phylogenetically annotated orthology resource based on 5090 organisms and 2502 viruses. Nucleic Acids Res. 2019:47: D309–D314. https://doi.org/10.1093/NAR/GKY1085.

Hurt C, Anker A, Knowlton N. A multilocus test of simultaneous divergence across the Isthmus of Panama using snapping shrimp in the genus Alpheus. Evolution. 2009:63:514–530. https://doi.org/10.1111/J.1558-5646.2008.00566.X.

Jackson JBC, O'Dea A. Evolution and environment of Caribbean coastal ecosystems. Proc Natl Acad Sci U S A. 2023:120: e2307520120. https://doi.org/10.1073/PNAS.2307520120/SUPPL_FILE/PNAS.2307520120.SD01.XLSX.

Jordan DS. The law of geminate species. Am Nat. 1908:42:73–80. https://doi.org/10.1086/278905.

Kawaguchi M, et al. Comparison of hatching mode in pelagic and demersal eggs of two closely related species in the order Pleuronectiformes. Zool Sci. 2014:31:709–715.

Kirkpatrick M. How and why chromosome inversions evolve. PLOS Biol. 2010:8:e1000501. https://doi.org/10.1371/JOURNAL.PBIO.1000501.

Knowlton N, Weigt LA. New dates and new rates for divergence across the Isthmus of Panama. Proc R Soc Lond B Biol Sci. 1998:265:2257–2263. https://doi.org/10.1098/RSPB.1998.0568.

Knowlton N, Weigt LA, Solórzano LA, Mills DK, Bermingham E. Divergence in proteins, mitochondrial DNA, and reproductive compatibility across the Isthmus of Panama. Science. 1993:260: 1629–1632. https://doi.org/10.1126/SCIENCE.8503007.

Laetsch DR, Blaxter ML. BlobTools: interrogation of genome assemblies. F1000Res. 2017:6:1287. https://doi.org/10.12688/f1000research.12232.1.

Lehmann R, et al. Finding Nemo's genes: a chromosome-scale reference assembly of the genome of the orange clownfish Amphiprion percula. Mol Ecol Resour. 2019:19:570–585. https://doi.org/10.1111/1755-0998.12939.

Lessios HA. Possible prezygotic reproductive isolation in sea urchins separated by the Isthmus of Panama. Evolution. 1984:38: 1144. https://doi.org/10.2307/2408446.

Lessios HA. The great American schism: divergence of marine organisms after the rise of the central American isthmus. Annu Rev Ecol Evol S. 2008:39:63–91. https://doi.org/10.1146/ANNUREV.ECOLSYS.38.091206.095815.

Lessios HA, Cunningham CW. Gametic incompatibility between species of the sea urchin *Echinometra* on the two sides of the Isthmus of Panama. Evolution. 1990:44:933–941. https://doi.org/10.1111/J.1558-5646.1990.TB03815.X.

Li E, Hristova K. Receptor tyrosine kinase transmembrane domains: function, dimer structure and dimerization energetics. Cell Adhes Migr. 2010:4:249–254. https://doi.org/10.4161/CAM.4.2.10725.

Li H. Minimap2: pairwise alignment for nucleotide sequences. Bioinformatics. 2018:34:3094–3100. https://doi.org/10.1093/bioinformatics/bty191.

Li H, Durbin R. Inference of human population history from individual whole-genome sequences. Nature. 2011:475:493–496. https://doi.org/10.1038/nature10231.

Liang Y, et al. Chromosome-level genome assembly of the smallscale yellowfin (*Plagiognathops microlepis*). Sci Data. 2024:11: 1234. https://doi.org/10.1038/s41597-024-04105-2.

Liu S, Hansen MM, Jacobsen MW. Region-wide and ecotype-specific differences in demographic histories of threespine stickleback populations, estimated from whole genome sequences. Mol Ecol. 2016:25:5187–5202. https://doi.org/10.1111/MEC.13827.

Luan MW, Zhang XM, Zhu ZB, Chen Y, Xie SQ. Evaluating structural variation detection tools for long-read sequencing datasets in *Saccharomyces cerevisiae*. Front Genet. 2020:11:159. https://doi.org/10.3389/fgene.2020.00159.

Mahmoud M, et al. Structural variant calling: the long and the short of it. Genome Biol. 2019:20:246. https://doi.org/10.1186/s13059-019-1828-7.

Marçais G, Kingsford C. A fast, lock-free approach for efficient parallel counting of occurrences of k-mers. Bioinformatics. 2011:27:764–770. https://doi.org/10.1093/BIOINFORMATICS/BTR011.

Marko PB. Fossil calibration of molecular clocks and the divergence times of geminate species pairs separated by the Isthmus of Panama. Mol Biol Evol. 2002:19:2005–2021. https://doi.org/10.1093/oxfordjournals.molbev.a004024.

Mather N, Traves SM, Ho SYW. A practical introduction to sequentially Markovian coalescent methods for estimating demographic history from genomic data. Ecol Evol. 2020:10: 579–589. https://doi.org/10.1002/ece3.5888.

McCord CL, Cooper JW, Nash CM, Westneat MW. Phylogeny of the damselfishes (Pomacentridae) and patterns of asymmetrical diversification in body size and feeding ecology. PLoS One. 2021: 16:e0258889. https://doi.org/10.1101/2021.02.07.430149.

Mendes FK, Vanderpool D, Fulton B, Hahn MW. CAFE 5 models variation in evolutionary rates among gene families. Bioinformatics. 2020:36:5516–5518. https://doi.org/10.1093/bioinformatics/btaa1022.

Mérot C, Oomen RA, Tigano A, Wellenreuther M. A roadmap for understanding the evolutionary significance of structural genomic variation. Trends Ecol Evol. 2020:35:561–572. https://doi.org/10.1016/j.tree.2020.03.002.

Montes C, et al. Middle miocene closure of the central American seaway. Science. 2015:348:226–229. https://doi.org/10.1126/science.aaa2815.

Nadachowska-Brzyska K, Burri R, Smeds L, Ellegren H. PSMC analysis of effective population sizes in molecular ecology and its application to black-and-white Ficedula flycatchers. Mol Ecol. 2016:25:1058–1072. https://doi.org/10.1111/mec.13540.

Nath S, Shaw DE, White MA. Improved contiguity of the threespine stickleback genome using long-read sequencing. G3 (Bethesda). 2021:11:jkab007. https://doi.org/10.1093/G3JOURNAL/JKAB007.

O'Dea A, et al. Formation of the Isthmus of Panama. Sci Adv. 2016:2:e1600883. https://doi.org/10.1126/SCIADV.1600883.
O'Dea A, Hoyos N, Rodríguez F, Degracia B, De Gracia C. History of upwelling in the tropical eastern Pacific and the paleogeography of the Isthmus of Panama. Palaeogeogr Palaeoclimatol Palaeoecol. 2012:348:59–66. https://doi.org/10.1016/J.PALAEO.2012.06.007.
Purves D, et al. The biogenic amines. Sinauer Associates; 2001. https://www.ncbi.nlm.nih.gov/books/NBK11035/
Quenouille B, Bermingham E, Planes S. Molecular systematics of the damselfishes (Teleostei: Pomacentridae): bayesian phylogenetic analyses of mitochondrial and nuclear DNA sequences. Mol Phylogenet Evol. 2004:31:66–88. https://doi.org/10.1016/S1055-7903(03)00278-1.
Rabosky DL, et al. An inverse latitudinal gradient in speciation rate for marine fishes. Nature. 2018:559:392–395. https://doi.org/10.1038/S41586-018-0273-1.
Rhie A, Walenz BP, Koren S, Phillippy AM. Merqury: reference-free quality, completeness, and phasing assessment for genome assemblies. Genome Biol. 2020:21:1–27. https://doi.org/10.1186/S13059-020-02134-9/FIGURES/6.
Robertson DR, Allen GR. Shorefishes of the Tropical Eastern Pacific: online information system. Version 3.0 Smithsonian Tropical Research Institute, Balboa, Panamá; 2024.
Robertson DR, Collin R. Inter- and intra-specific variation in egg size among reef fishes across the isthmus of Panama. Front Ecol Evol. 2015:2:84. https://doi.org/10.3389/fevo.2014.00084.
Rueda-Roa DT, Muller-Karger FE. The southern Caribbean upwelling system: sea surface temperature, wind forcing and chlorophyll concentration patterns. Deep-Sea Res I. 2013:78:102–114. https://doi.org/10.1016/J.DSR.2013.04.008.
Sambrook J, Russell DW. Purification of nucleic acids by extraction with phenol:chloroform. CSH Protoc. 2006:2006:pdb.prot4455. https://doi.org/10.1101/PDB.PROT4455.
Sano K, et al. Lineage-specific evolution of zona pelluci da genes in fish. J Exp Zool B Mol Dev Evol. 2022:338:181–191.
Sato Y, Miya M, Fukunaga T, Sado T, Iwasaki W. MitoFish and MiFish pipeline: a mitochondrial genome database of fish with an analysis pipeline for environmental DNA metabarcoding. Mol Biol Evol. 2018:35:1553–1555.
Simão FA, Waterhouse RM, Ioannidis P, Kriventseva EV, Zdobnov EM. BUSCO: assessing genome assembly and annotation completeness with single-copy orthologs. Bioinformatics. 2015:31:3210–3212. https://doi.org/10.1093/BIOINFORMATICS/BTV351.
Smit A, Hubley R, Green P. *RepeatMasker Open-4.0*. 2013. http://www.repeatmasker.org
Stanke M, Diekhans M, Baertsch R, Haussler D. Using native and syntenically mapped cDNA alignments to improve de novo gene finding. Bioinformatics. 2008:24:637–644. https://doi.org/10.1093/bioinformatics/btn013.
Swank AR, Tracy CB, Mendonça MT, Bernal MA. Molecular plasticity to ocean warming and habitat loss in a coral reef fish. J Hered. 2025:116:126–138. https://doi.org/10.1093/JHERED/ESAE024.
Tang KL, Stiassny MLJ, Mayden RL, DeSalle R. Systematics of damselfishes. Ichthyology & Herpetology. 2021:109:258–318. https://doi.org/10.1643/I2020105.
Tavera JJ, Acero PA, Balart EF, Bernardi G. Molecular phylogeny of grunts (Teleostei, Haemulidae), with an emphasis on the ecology, evolution, and speciation history of new world species. BMC Evol Biol. 2012:12:57. https://doi.org/10.1186/1471-2148-12-57.
Thacker CE. Patterns of divergence in fish species separated by the Isthmus of Panama. BMC Evol Biol. 2017:17:1–14. https://doi.org/10.1186/S12862-017-0957-4/TABLES/4.
Tigano A, et al. Chromosome-level assembly of the Atlantic Silverside genome reveals extreme levels of sequence diversity and structural genetic variation. Genome Biol Evol. 2021:13:evab098. https://doi.org/10.1093/gbe/evab098.
Tigano A, Reiertsen TK, Walters JR, Friesen VL. 2018. A complex copy number variant underlies differences in both colour plumage and cold adaptation in a dimorphic seabird [preprint]. BioRxiv: 507384. https://doi.org/10.1101/507384.
Torres M, Höppner M. *ESGA*. 2021. https://github.com/ikmb/esga
Van Dongen S. Graph clustering via a discrete uncoupling process. SIAM J Matrix Anal A. 2008:30:121–141. https://doi.org/10.1137/040608635.
Vij S, et al. Chromosomal-level assembly of the Asian seabass genome using long sequence reads and multi-layered scaffolding. PLoS Genet. 2016:12:e1005954. https://doi.org/10.1371/JOURNAL.PGEN.1005954.
Villegas-Hernández H, et al. Life-history traits facilitate the population success of the sergeant major *Abudefduf saxatilis* (Linnaeus, 1758, Perciformes: Pomacentridae) in the Caribbean sea. Mar Biol Res. 2022:18:230–251. https://doi.org/10.1080/17451000.2022.2117827.
Vurture GW, et al. GenomeScope: fast reference-free genome profiling from short reads. Bioinformatics. 2017:33:2202–2204. https://doi.org/10.1093/BIOINFORMATICS/BTX153.
Wellington GM, Robertson DR. Variation in larval life-history traits among reef fishes across the Isthmus of Panama. Mar Biol. 2001:138:11–22. https://doi.org/10.1007/s002270000449.
Wright RM, Aglyamova GV, Meyer E, Matz MV. Gene expression associated with white syndromes in a reef building coral, Acropora hyacinthus. BMC Genomics. 2015:16:1–12. https://doi.org/10.1186/S12864-015-1540-2/FIGURES/5.
Yasumasu S, Katow S, Umino Y, Iuchi I, Yamagami K. A unique proteolytic action of HCE, a constituent protease of a fish hatching enzyme: tight binding to its natural substrate, egg envelope. Biochem Biophys Res Commun. 1989:162:58–63.
Zhang L, Reifová R, Halenková Z, Gompert Z. How important are structural variants for speciation? Genes (Basel). 2021:12:1084. https://doi.org/10.3390/GENES12071084.
Zheng S, Poczai P, Hyvönen J, Tang J, Amiryousefi A. Chloroplot: an online program for the versatile plotting of organelle genomes. Front Genet. 2020:11:576124.

Associate editor: Bonnie Fraser