scientific data

Check for updates

OPEN

DATA DESCRIPTOR

A chromosome-level genome assembly for the brown sea cucumber from the Galápagos Islands, *Isostichopus fuscus*

Jaime D. Ortiz-Pachar[1,2]✉, Jorge Ramírez-González[2], Nicolas Moity[2], Solange Andrade-Vera[2], Wilson D. Rivadeneira[2] & Nina Overgaard Therkildsen[1]

The Galápagos Islands face threats from human activities and climate change, with significant consequences for native fauna like the brown sea cucumber, *Isotichopus fuscus* (Ludwig, 1875). Once abundant in the region's shallow coastal waters, this species has suffered dramatic population declines due to overexploitation and is now listed as endangered by the IUCN and protected under CITES. To support genomics-informed conservation efforts, we employed Pacific Biosciences HiFi long sequencing reads and Dovetail Omni-C proximity reads to create a high-quality chromosome-level genome assembly, which was subsequently annotated using RNA-seq-informed gene prediction. The resulting assembly spans 898.4 Mb, with a contig N50 of 21.5 Mb, a scaffold N50 of 36.9 Mb, and a BUSCO completeness score of 98.4%—quality statistics comparable to those of other chromosome-scale assemblies for sea cucumbers in the Stichopodidae family. Consistent with those species, our assembly features 23 chromosomes. The newly available genomic resources for *I. fuscus* will be important for studies on population genomics, comparative genomics, and the impact of environmental stressors on local adaptation.

Background & Summary

Sea cucumbers are marine invertebrates of the phylum Echinodermata and class Holothuroidea, with the species *Isostichopus fuscus* (Ludwig, 1875) being both economically and ecologically important in the Galápagos Islands, a globally recognized biodiversity hotspot[1]. Although seemingly uncharismatic, sea cucumbers play a crucial ecological role in marine benthic communities[2]. Like other holothurians, *I. fuscus* populations in the Galapagos[3] structure and regulate nutrient fluxes through the consumption of organic material, promoting the growth of microalgae[4], and also provide critical protective benefits to coral reefs by reducing disease prevalence through their sediment-feeding activities that suppress pathogenic microbes[5,6]. The protection and recovery of this species is therefore of utmost importance in the Galápagos, as it supports crucial ecosystem services and provides a livelihood for local fishers in this unique region[7]. This keystone species mainly inhabits rocky reef areas from 2 to 40 m depth[8]. The recruitment of *I. fuscus* is closely related to local water temperatures, which makes it susceptible to climate change[9,10]. Sea cucumbers are also highly vulnerable to overexploitation due to their ease of capture, slow population growth, and dependence on dense populations for reproduction[11]. An intense fishery for *I. fuscus*, which began in the early 1990s and rapidly expanded, has reduced populations to critically low levels in the Galápagos Islands[7,12]. The population declines are so severe that the International Union for Conservation of Nature (IUCN) has classified *I. fuscus* as endangered[13], and it has also become listed in Appendix III of The Convention on International Trade in Endangered Species[14].

Stakeholders of the Galápagos Marine Reserve (GMR) have collectively prioritized the recovery of this important benthic species as a key conservation goal[15]. To most effectively leverage conservation efforts, however, the management authorities in Galápagos require a thorough understanding of the genetic structure and potential local adaptation among populations of *I. fuscus* across the vast archipelago to manage connectivity and identify potential source populations for re-stocking programs. Therefore, as part of the Galápagos Marine

[1]Department of Natural Resources and the Environment, Cornell University, Ithaca, NY, USA. [2]Charles Darwin Research Station, Charles Darwin Foundation, Santa Cruz, Galapagos Islands, Ecuador. ✉e-mail: jdo53@cornell.edu

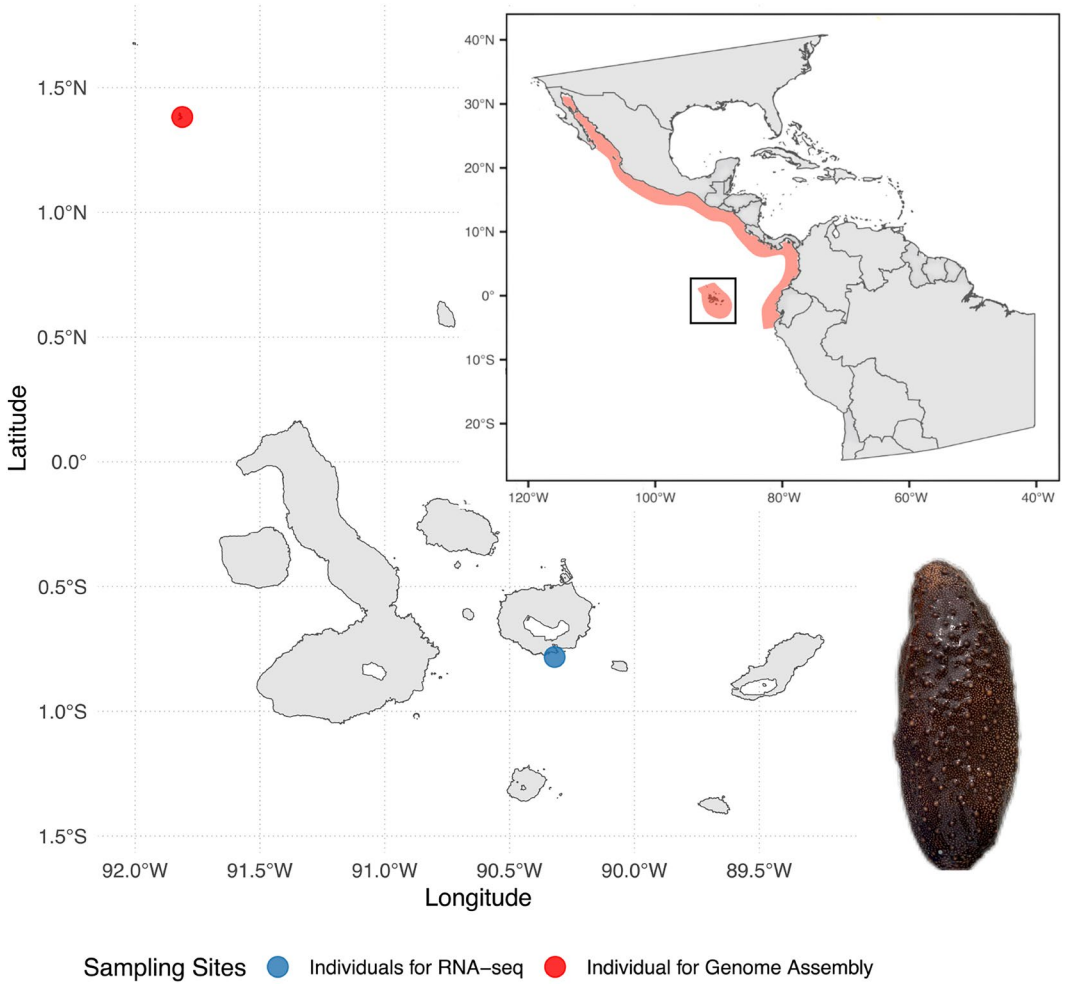

Fig. 1 Main: Sampling sites for individuals of *Isostichopus fuscus* used for PacBio long-read sequencing (N = 1; red point) and RNA-seq (N = 2; blue point). Inset: Map of Central and South America with the distribution range of the brown sea cucumber, *I. fuscus*, indicated in coral color[13]. Photograph of *I. fuscus* from the Galapagos, courtesy of Nicolas Moity, Charles Darwin Foundation.

Genomics Project, we generated a high-quality genome assembly that will be a critical resource for characterizing geographic patterns of genetic variation and evidence for local adaptation among the sea cucumber populations in the Galápagos Islands to inform and prioritize conservation actions.

The genome assembly for the brown sea cucumber, *I. fuscus*, presented here significantly enriches the genomic resources available for marine invertebrates, particularly holothurians. It is only the eighth chromosome-level genome assembly[16–22] among more than 1,800 described sea cucumber species and represents the fourth such resource for the commercially important Stichopodidae[20–22] family. Notably, it is the only high-quality chromosome-level genome available for an endangered sea cucumber species. This reference genome provides a critical foundation for exploring genomic variation in *I. fuscus*, a species facing considerable conservation pressures due to commercial exploitation. It will facilitate the development of genetic tools for population management, such as stock assessment, parentage analysis for aquaculture, and conservation genomics initiatives. Furthermore, the high contiguity and completeness of this assembly enhance its utility for comparative genomic studies with other echinoderms and for investigating the unique physiological adaptations of sea cucumbers. Overall, this resource will advance research on holothurian biology, evolution, and conservation.

Methods

Sample collection. In February 2022, two adult *I. fuscus* individuals were collected on Santa Cruz Island (0.78 S 90.32 W; Fig. 1) at 15 m depth, on a rocky area mixed with sandy patches (length: 21.6 and 23.5 cm; sea temp: 22 °C). The individuals were kept alive in a water tank and subsequently transported back to the Charles Darwin Research Station on Santa Cruz Island, where they were dissected. Tissue samples from gonads, upper tentacles, internal muscle, and respiratory system were flash-frozen with dry ice and stored at −80 °C. Frozen tissue samples were shipped on dry ice to Dovetail Genomics (Scotts Valley, California) for DNA extraction and preparation of libraries for PacBio long-read sequencing, proximity ligation libraries using Omni-C, and RNA-seq libraries for annotation. However, due to transportation constraints, the DNA quality requirements for HiFi

long-read sequencing was not met, and these samples were used only for RNA-seq libraries for annotation (NCBI BioSamples: SAMN49786374-SAMN49786381).

During a second sample collection trip in December 2022, an adult *I. fuscus* individual (30 cm long; NCBI BioSample SAMN49722880) was collected at Wolf Island, one of the northernmost islands of the Galapagos Archipelago (1.3826 N 91.8112 W; Fig. 1) at 23 m depth, on a rocky area mixed with hard corals (25.6 °C, 33.3 ± 0.7 PSU). The individual was kept alive in a water tank filled with seawater (renewed every other day) and an aerator and transported back to the Charles Darwin Research Station on Santa Cruz Island, where it was dissected. Tissue samples from gonads, upper tentacles, internal muscle, and respiratory system were flash-frozen with dry ice and stored at −80 °C. Frozen tissue samples were shipped on dry ice to Dovetail Genomics for DNA extraction and preparation of libraries for PacBio long-read sequencing and proximity ligation libraries using Omni-C. Due to logistical constraints during field sample processing, weights were not measured, and we were unable to determine sex from visual inspection of the gonads (either because the individuals were sexually immature or sampled outside their reproductive season).

High molecular weight genomic DNA preparation. High-molecular-weight genomic DNA was successfully extracted from gonad tissue from the individual *I. fuscus* collected at Wolf Island (NCBI BioSample SAMN49722880) using a modified cetyltrimethylammonium bromide (CTAB) protocol. Fresh tissue samples were homogenized in 10 mL of CTAB extraction buffer supplemented with 0.5% β-mercaptoethanol and incubated at 68 °C for 20 minutes. Subsequent purification involved two rounds of phenol-chloroform (25:24:1 v/v) extraction followed by a single chloroform (24:1 v/v) extraction, with phase separation achieved through centrifugation at 12,000 × g for 15 minutes at 4 °C. Nucleic acids were precipitated by adding 0.7 volumes of isopropanol and pelleted by centrifugation at 15,000 × g for 20 minutes. The DNA pellet was then resuspended in 9.5 mL of G2 buffer (QIAamp DNA Mini Kit; Qiagen, Aarhus, Denmark) containing 200 μL of proteinase K and 19 μL of RNase A (100 μg/mL), followed by incubation at 50 °C for 30 minutes. QIAGEN column-based purification was performed according to the manufacturer's specifications (QIAGEN Genomic-tips; Qiagen), with final DNA elution achieved through overnight incubation in 100 μL of TE buffer (10 mM Tris-HCl, 1 mM EDTA, pH 8.0) at 50 °C. Genomic DNA integrity and concentration were assessed using the Agilent 4200 TapeStation system (Agilent Technologies, Santa Clara, California) with the Genomic DNA ScreenTape assay. One microliter of extracted DNA was combined with 10 μL of Genomic DNA Sample Buffer and analyzed following the manufacturer's instructions.

PacBio sequencing. The DNA concentration was quantified using a Qubit 2.0 Fluorometer (Life Technologies, Carlsbad, CA, USA). The PacBio SMRTbell library (~20 kb) for PacBio Sequel was constructed using a SMRTbell Express Template Prep Kit 2.0 (PacBio, Menlo Park, CA, USA) according to the manufacturer's recommended protocol. The library was bound to polymerase using the Sequel II Binding Kit 2.0 (PacBio) and loaded onto the PacBio Sequel II. Sequencing was performed on 2 PacBio Sequel II 8 M SMRT cells.

The PacBio CCS sequencing library generated 3.09 million reads (50.2 Gb), corresponding to 56-fold sequencing coverage, and the reads had an N50 read length of 15,962 bp (minimum read length 107 bp; mean read length 15,850 bp; maximum read length 37,088 bp).

Omni-C library preparation. The same gonad tissue from the individual collected from Wolf Island (NCBI BioSample SAMN49722880) was also used to generate data for scaffolding with an Omni-C proximity ligation library prepared using the Dovetail Omni-C Kit (Dovetail Genomics, Scotts Valley, CA) following the manufacturer's protocol. First, chromatin was fixed in place with formaldehyde in the nucleus and then extracted. Fixed chromatin was digested with DNase I, chromatin ends were repaired and ligated to a biotinylated bridge adapter, followed by proximity ligation of adapter-containing ends. After proximity ligation, crosslinks were reversed, and the DNA was purified. Purified DNA was treated to remove biotin that was not internal to the ligated fragments. Sequencing libraries were generated using NEBNext Ultra enzymes and Illumina-compatible adapters. Biotin-containing fragments were isolated using streptavidin beads before PCR enrichment of each library. The library was sequenced on an Illumina HiSeqX platform to produce approximately 30x coverage, generating 47.87 million read pairs (after filtering) and aligning to coverage levels used in recent genome assemblies for similar genomes[23]. Omni-C's endonuclease-based approach provides more uniform coverage than restriction enzyme-based Hi-C, particularly beneficial for regions with repetitive sequences. Contact maps showed clear chromosome-scale interactions, confirming scaffolding accuracy.

RNA-seq. Tissue samples from the gonads, upper tentacles, internal muscle, and respiratory system from the two individuals collected in Santa Cruz were used for RNA-seq (NCBI BioSamples SAMN49786374 to SAMN49786381). Total RNA was extracted separately from each tissue type using the QIAGEN RNeasy Plus Kit following the manufacturer's protocols. Before library preparation, we performed DNase treatment followed by AMPure bead cleanup and QIAGEN FastSelect HMR rRNA depletion. Eight RNA-seq libraries (4 types of tissues x 2 individuals) were prepared using the NEBNext Ultra II RNA Library Prep Kit following the manufacturer's protocols. These libraries were sequenced on the NovaSeq 6000 platform in a 2 × 150 bp configuration to generate approximately 57 million reads per library.

***De novo* assembly.** Genome assembly was conducted using 50.2 Gb of PacBio Circular Consensus Sequencing (CCS) reads as input for Hifiasm v0.15.4-r347[24], with default parameters, as outlined in Table 1. The resultant primary contigs underwent contamination screening using BLAST[25] against the complete NCBI nucleotide (nt) database, encompassing all GenBank and RefSeq sequences. The output was processed using BlobTools v1.1.1[26] to identify and eliminate putative contaminant sequences, generating a filtered assembly. Subsequently, haplotigs and overlapping contigs were removed using purge_dups v1.2.5[27], resulting in the final purged assembly.

Assembly	Software and any non-default options	Version
K-mer counting	Meryl (k = 21)	1.4.1
Estimation of genome size	GenomeScope	2.0
De novo assembly (contiging)	HiFiasm (default parameters)	0.15.4-r347
Scaffolding		
Omni-C data alignment	bwa	0.7.19
Omni-C Scaffolding	HiRise	1.0
Short-read alignment	BWA-MEM (-5SP)	0.7.17-r1188
SAM/BAM processing	samtools	1.11
SAM/BAM filtering	pairtools	0.3.0
Pairs indexing	pairix	0.3.7
Matrix generation	cooler	0.8.10
Contact map visualization	In-house pipeline (Dovetail Genomics)	
Genome quality assessment		
Basic assembly metrics	QUAST (--est-ref-size)	5.0.2
Assembly completeness	BUSCO (eukaryota_odb10)	5.0.0
	Merqury	1.3
Assembly metrics visualization	BlobToolKit	2.6.3
Contamination screening		
Local alignment tool	BLAST (-db nt)	2.9
General contamination screening	BlobToolKit (PacBIo HiFi Coverage, NCBI Taxa ID = 200383, BUSCODB = Eukaryota)	1.1.1
Gene Prediction		
RNAseq aligner	STAR	2.7
Ab initio gene prediction model	SNAP	v2006-07-28
Ab initio gene prediction model	AUGUSTUS	2.5.5
Peptide evidence	MAKER	3.01.04
Gene model prediction refinement	PASA	v2.5.2
tRNA prediction	tRNAscan-SE	v2.05
Functional annotation	BLASTP: UniProtKB/Swiss-Prot database	
Mitochondrial assembly		
Mitochondrial genome assembly	MitoHiFi (-o 9) Reference genome: Isostichopus badionotus	3.0.0
Synteny		
Genome-Genome alignment	ntSynt (-d 15)	1.0.2
Synteny visualization	ntSynt-viz	1.0.0

Table 1. Assembly pipeline and software used. Software citations are listed in the text.

Scaffolding the assembly with HiRise. We scaffolded the *I. fuscus* genome assembly using HiRise, a pipeline designed specifically for using proximity ligation data to scaffold genome assemblies[28], with default parameters. The long-read *de novo* assembly (Hifiasm) and Dovetail Omni-C library reads were used as input data for HiRise. Dovetail Omni-C library sequences were aligned to the draft input assembly using bwa[29]. The positions of Dovetail Omni-C read pairs mapped within draft scaffolds were analyzed by HiRise to produce a likelihood model for the genomic distance between read pairs. This model was used to identify and break putative misjoins, score prospective joins, and make joins above a threshold.

We then analyzed the scaffolded genome assembly of *I. fuscus* to *de novo* identify and classify repeat families using the software package RepeatModeler2 v2.0.1[30]. This software uses the programs RECON v1.08[31] and RepeatScout v1.0.6[32] for *de novo* identification of repeats within the genome, and the resulting custom repeat library was used to discover, identify, and mask the repeats in the assembly file using RepeatMasker v4.1.0 (https://www.repeatmasker.org).

Finally, the output scaffolded assembly was processed using the NCBI Foreign Contamination Screen (FCS), which utilizes the genome cross-species aligner (GX) to identify and remove contaminant sequences[33] in our assembly before submission to GenBank.

AB Initio genome annotation. Evidence-guided structural annotation combined ab initio prediction with splice-aware RNA-seq alignments and protein homology on the HiRise-scaffolded assembly, prioritizing conserved exon–intron signals that are robust within species and not expected to depend on geographic provenance of RNA donors. Coding sequences from other echinoderm species with high-quality reference genomes (*Apostichopus japonicus*[21], *Asterias rubens*[34], *Holothuria glaberrima*[35], and *Strongylocentrotus purpuratus*[36]) were used to train the initial ab initio gene prediction model for *I. fuscus* using AUGUSTUS v2.5.5[37]. Six rounds of prediction optimization were done using the software package provided by AUGUSTUS. The same coding sequences were also used to train a separate ab initio model for *I. fuscus* using SNAP v2006-07-28[38]. The RNAseq

reads we generated from *I. fuscus* were then mapped to the genome assembly using the STAR aligner v2.7[39] with intron hints generated with the bam2hints tools within the AUGUSTUS software. SNAP and AUGUSTUS (with intron-exon boundary hints provided from RNA-Seq) were then used to predict genes in the repeat-masked reference genome. To help guide the prediction process, Swiss-Prot peptide sequences from the UniProt database were downloaded and used in conjunction with the protein sequences from *A. japonicus*, *A. rubens*, *H. glaberrima*, and *S. purpuratus* to generate peptide evidence in the MAKER pipeline[40]. Only genes predicted by the SNAP and AUGUSTUS software were retained in the final gene sets. The gene models were further refined through two rounds of PASA v2.5.2[41], incorporating RNA-seq assemblies to improve exon-intron structures and identify alternative isoforms. Functional annotations were assigned using BLASTP[25] (E-value < 1e-5) against the UniProtKB/Swiss-Prot database[42]. Transfer RNAs (tRNA) were predicted using the software tRNAscan-SE v2.05[43].

Genome quality assessment. To assess genome characteristics, we generated k-mer counts (k = 21) from the PacBio HiFi reads using Meryl[44] with default parameters. The k-mer counts were then used in GenomeScope 2.0[45], using default parameters to estimate genome size, heterozygosity, and repeat content. To obtain general contiguity metrics, we ran QUAST[46]. To evaluate genome quality and functional completeness, we used BUSCO v5.5.0[47] with eukaryota_odb10 (Creation date: 2024-01-08, number of genomes: 70, number of BUSCOs: 255).

Mitochondrial genome assembly. We assembled the mitochondrial genome of *I. fuscus* from the PacBio CCS reads using the reference-guided pipeline MitoHiFi v3.2.2 with MitoFinder v1.4.2 for annotation[48,49]. The mitochondrial sequence of the congener *Isostichopus badionotus* (GenBank: IMZ188901.1) was used as the starting sequence. After completing the nuclear genome assembly, we searched for matches of the resulting mitochondrial assembly sequence in the nuclear genome using BLAST+[50]. We filtered out contigs and scaffolds from the nuclear genome with a sequence identity greater than 99% and a size smaller than the mitochondrial assembly sequence. No other manual curation was performed on the mitochondrial genome. The final mitochondrial sequence was 16,300 bp long, with base composition of A = 31.7%, C = 24.2%, G = 15.9%, and T = 28.2%. It consists of 2 non-coding RNAs, 22 unique transfer RNAs and 13 protein-coding genes. The mitochondrial genome assembly is available on GenBank (see Table 2 and the Data records section for details).

Synteny analysis. We compared our final *I. fuscus* genome assembly to the two closely related chromosome-level assemblies available for the sea cucumbers *A. japonicus*[21] and *Stichopus monotuberculatus*[20]. To assess homology of the *I. fuscus* scaffolds with the 23 chromosomes identified for *A. japonicus* and *S. monotuberculatus* genomes, we aligned the 23 largest *I. fuscus* scaffolds to the *A. japonicus* and *S. monotuberculatus* chromosomes using the program ntSynt v1.0.2[51] and visualized the alignment using the program ntSynt-viz v1.0.0[52].

Data Records

The genomic PacBio sequencing data were deposited in the SRA at NCBI under accession numbers SRR34364192 and SRR34364193[53].

The transcriptomic sequencing data were deposited in the SRA at NCBI under accession numbers SRR34366410, SRR34366411, SRR34366412, SRR34366413, SRR34366414, SRR34366415, SRR34366416, and SRR34366417[41].

The Omni-C sequencing data were deposited in the SRA at NCBI under accession number SRR34364191[41].

The final chromosome assembly (accession: GCA_051397855.1[54]) and genome annotation files are available at NCBI under BioProject number PRJNA1284402[55].

The final mitochondrial genome assembly and annotation files are available at GenBank under the accession number (accession: PV946926[56]).

Technical Validation

The final assembly comprises 1,102 scaffolds spanning 897.75 Mb, with a scaffold N50 of 36.9 Mb and the largest scaffold of 57.13 Mb. Using the eukaryota_odb10 lineage dataset (number of species: 70, number of BUSCOs: 255), the estimated BUSCO completeness score corresponds to 98.5% (including Duplicated BUSCOs; Table 2), indicating a high-quality assembly (Fig. 2b). The genome exhibited high contiguity, as 94.8% of its total sequence length was found in the largest 23 scaffolds. Analyses using RepeatModeler and GenomeScope indicated that the genome comprised 38.27% and 30.8% repetitive sequences, respectively. Multiple lines of evidence supported the hypothesis that, similar to other sea cucumbers[37,38], the nuclear genome is organized into 23 chromosomes. First, the contact map indicated that the vast majority of Omni-C reads mapped to the genome were arranged into 23 large scaffolds (Fig. 3). Further, both alignments and synteny analysis showed that the largest 23 scaffolds in *I. fuscus* corresponded closely with the 23 chromosomes of *A. japonicus* and *S. monotuberculatus* (Fig. 4).

Standard QC metrics (91.0% BUSCO completeness, 70.6% strong evidence support) also confirm high annotation quality, even though the annotation was partly based on RNA-seq evidence from a different geographic location from where the individual used for assembling the genome sequence was sampled. Using conspecific RNA-seq generated from different localities than the reference genome individual is a common practice in evidence-guided annotation frameworks, which rely on splice-junction and coding signals rather than geographic provenance[26]. The annotation pipeline predicted 36,188 genes (35,811 protein-coding and 377 tRNA) in the *I. fuscus* genome, with an average length of 1,353 bp (Table 2). The BUSCO evaluation showed that these predicted genes cover 97.3% of eukaryota_odb10 genes (91.0% complete, 6.3% fragmented; n = 255), demonstrating comprehensive gene content. The annotation edit distance (AED) revealed that a substantial portion of gene annotations are well-supported, with 70.6% (n = 140,157) assigned functional annotations based on homology to Swiss-Prot. Additionally, 377 transfer RNA genes were detected, representing tRNAs for all 20 standard amino acids, two selenocysteine tRNAs, and four stop codon suppressors (documented as tRNA-Leu).

Bio projects & vouchers						
Species NCBI BioProject	PRJNA1284402: *Isostichopus fuscus* isolate:PM-IF020 Assembly (TaxID: 200383)					
NCBI BioSample	SAMN49722880: GCCDRS-3457 (TaxID: 200383)					
Specimen identification	Charles Darwin Research Foundation for the Galapagos Islands					
NCBI genome accessions						
Assembly accession	JBPNYN000000000					
Genome sequence						
PacBio HiFi reads	Run	2 PACBIO 8 M SMRT (Sequel II) 50.2 Gb				
	Accession	SRR34364192 and SRR34364193				
Omni-C Illumina reads	Run	1 ILLUMINA (Illumina HiSeqX) 81.6 M reads				
	Accession	SRR34364191				
Genome assembly quality metrics						
HiFi Read coverage	~56X					
Number of contigs	1,160					
Contig N50 (Mbp)	21.6					
Longest contigs (Mbp)	50.58					
Number of scaffolds	1115					
Scaffold N50 (Mbp)	36.9					
Largest scaffold (Mbp)	57.13					
Size of final assembly (Mbp)	898.25					
k-mer completeness	99.94					
BUSCO - Genome assembly (eukaryota_odb10) n = 255		**Complete**	**Single-Copy**	**Duplicated**	**Fragmented**	**Missing**
		98.5%	91.8%	6.7%	1.6%	0.1%
Organelles (complete mitochondrial sequence)	Size (bp)	16,300				
	Accession	PV946926				
Gene annotation quality metrics						
Total genes	36,188					
Protein-coding	35,811					
tRNA genes	377					
mRNAs	36,522					
CDS	198,510					
Functional annotation	140,157					
Hypothetical proteins	58,353					
tRNA features	377					
Selenocysteine (tRNA-Sec)	2					
Suppressor tRNAs	4					
BUSCO - Gene annotation (eukaryota_odb10) n = 255		**Complete**	**Single-Copy**	**Duplicated**	**Fragmented**	**Missing**
		91.0%	89.0%	2.0%	6.3%	2.7%

Table 2. Sequencing and assembly statistics, and accession numbers.

Compared to other chromosome-level assemblies of other sea cucumber species in the *Stichopodidae* family, the assembled genome size of *I. fuscus* (897.75 Mb) is slightly larger. However, the length of our assembled genome is similar to the genome size estimated by the K-mer-based analysis of the PacBio CCS reads using Genomescope 2.0, which estimated a genome size of 0.9 Gb, a 0.248% sequencing error rate, and a heterozygosity rate of 1.94% (Fig. 2a). The genome of *A. japonicus*, arguably the most economically significant sea cucumber globally, has an assembled genome size of 671.60 Mb[37], while the size of the genome assembly for the tropical sea cucumbers *S. monotuberculatus* and *Stichopus fusiformiossa* are 810.54 Mb and 766.55 Mb, respectively[20,22]. The chromosome-level genome assembly presented here shows excellent completeness, with a BUSCO score of 98.5%, and a high level of contiguity, as reflected by a scaffold N50 of 36.9 Mb. *A. japonicus* exhibited a slightly lower completeness with a BUSCO score of 97.2% and a lower scaffold N50 of 29.65 Mb[21], while *S. monotuberculatus* displayed a comparable scaffold N50 value of 35.36 Mb and a BUSCO completeness score of 97.8%[20]. *S. fusiformiossa* showed a comparable BUSCO completeness score of 98.1% but a lower scaffold N50 of 31.8 Mb[22]. However, all four *Stichopodidae* genomes are notably smaller than the chromosome-level assemblies of more distantly related sea cucumber species from the *Holothuriidae* (*Holothuria scabra*, 1.19 Gb[17]; *Holothuria leucospilota*, 1.39 Gb[18]), the *Chiridotidae* (*Chiridota heheva*, 1.43 Gb[16]), and the *Cucumariidae* (*Colochirus anceps*, 2.24 Gb[19]). Notably, the chromosome number suggested by our assembly of *I. fuscus* (n = 23) matches the chromosome number described for all these sea cucumber species: *C. heheva*[16], *H. scabra*[17], *H. leucospilota*[18], *C. anceps*[19], *S. monotuberculatus*[20], *A. japonicus*[21], and *S. fusiformiossa*[22]. Accordingly, all quality metrics suggest that our generated reference genome for *I. fuscus* is a high-quality resource for future research.

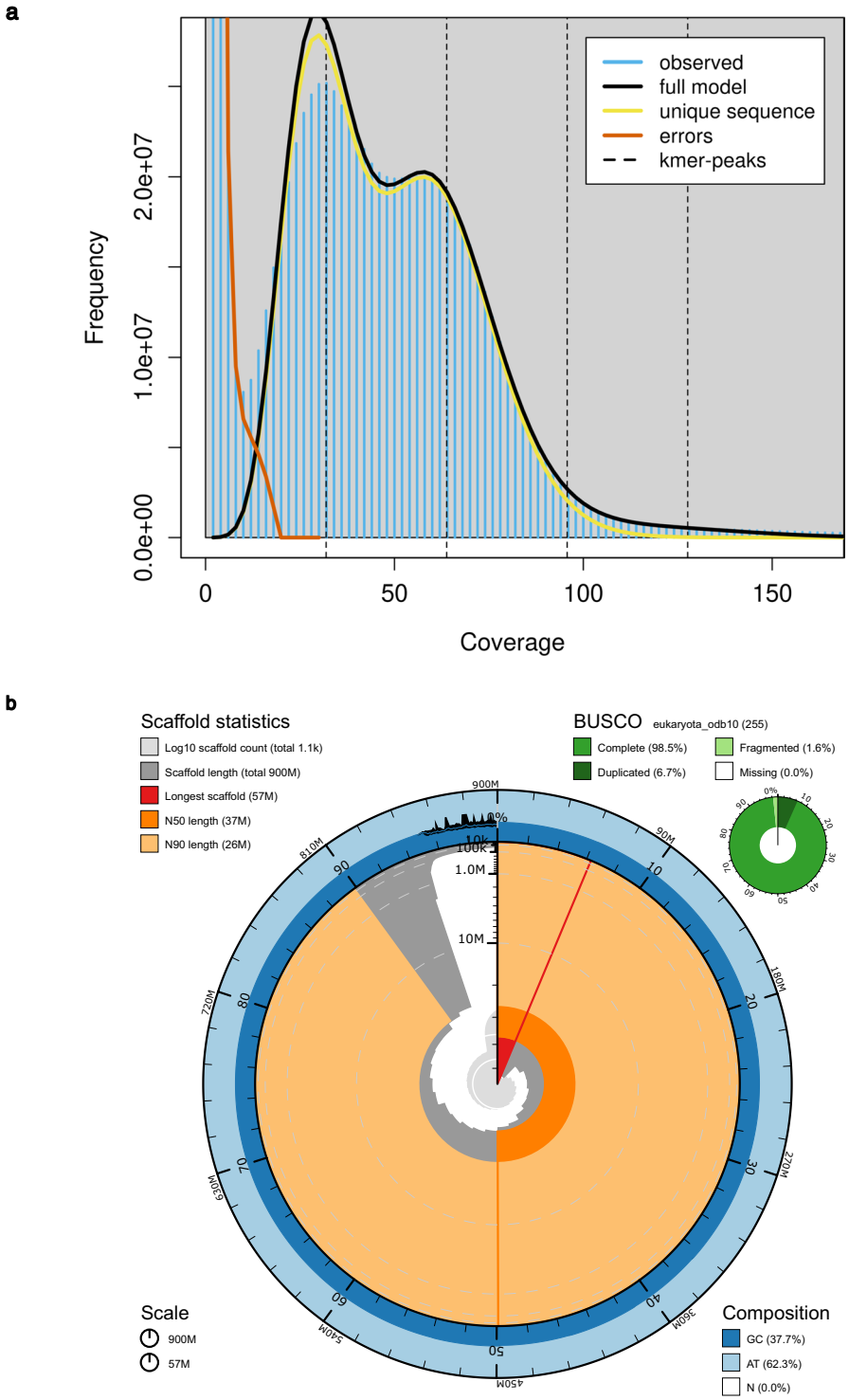

Fig. 2 Visual overview of *I. fuscus* genome assembly metrics. (**a**) K-mer spectrum output generated from PacBio CCSS data without adapters using GenomeScope 2.0. (**b**) BlobToolKit Snail plot showing a graphical representation of the quality metrics presented in Table 2 for the *I. fuscus* assembly and BUSCO assessment results based on the eukaryota_odb10 set of orthologous genes (n = 255). The plot circle represents the full size of the assembly. From the inside to the outside, the central plot covers length-related metrics. The red line shows the longest scaffold; others are sorted by size, arranged clockwise around the plot, and drawn in gray from the center outward. Dark and light orange arcs show scaffold N50 and N90 values. The central light gray spiral depicts the cumulative scaffold count, with a white line marking each magnitude. White areas show the proportion of Ns in the assembly. The dark and light blue areas around display average, maximum, and minimum GC versus AT content at 0.1% intervals.

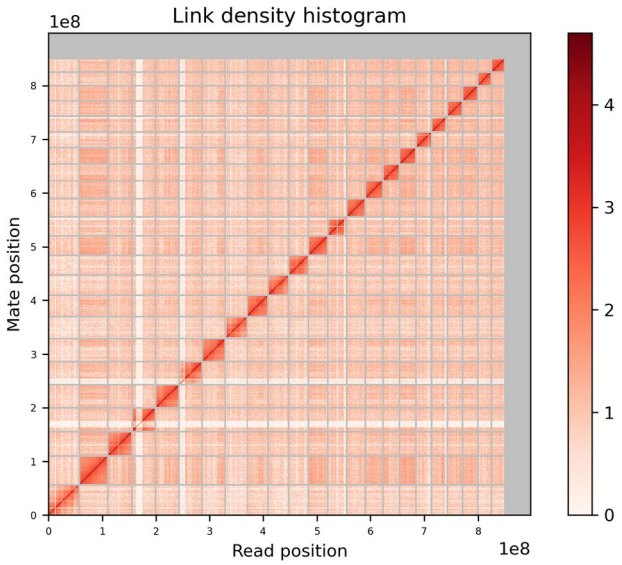

Fig. 3 Omni-C contact map for the scaffold-level genome assembly of *I. fuscus*. Each pixel represents a pair of read alignments from proximity-ligated DNA fragments, with the x- and y-axes showing the genomic positions of each read in the pair. Darker regions indicate higher contact frequency. Scaffolds are arranged by size, and the diagonal reflects frequent contacts within scaffolds. Only scaffolds larger than 1 Mb are shown.

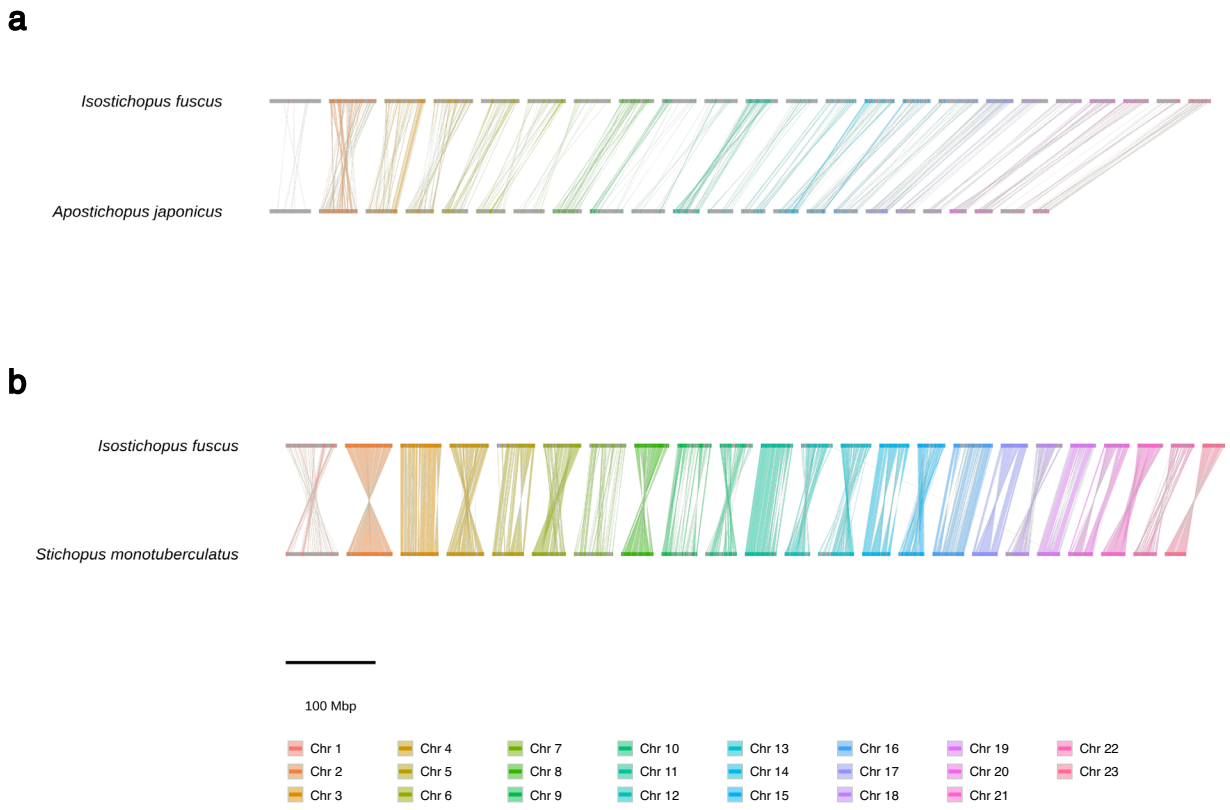

Fig. 4 Chromosome-level synteny among three sea cucumber species. Synteny relationships are shown between the chromosomes of (**a**) *A. japonicus* and *I. fuscus*; and (**b**) *S. monotuberculatus* and *I. fuscus*. Colored lines connect homologous chromosomal regions across species, with each chromosome represented by a distinct color.

Data availability

The dataset has been deposited in GenBank under accession PRJNA1284402.

Code availability

No custom scripts or code were utilized.

Received: 21 July 2025; Accepted: 5 December 2025;
Published online: 17 December 2025

References

1. Toral-Granda, V. Galapagos Islands: a hotspot of sea cucumber fisheries in Latin America and the Caribbean. **24**.
2. Purcell, S. W., Conand, C., Uthicke, S. & Byrne, M. Ecological roles of exploited sea cucumbers. *Oceanography and marine biology* **54**, 367–386 (2016).
3. Toral-Granda, M. V. & Martínez, P. C. Reproductive biology and population structure of the sea cucumber Isostichopus fuscus (Ludwig, 1875) (Holothuroidea) in Caamaño, Galápagos Islands, Ecuador. *Marine Biology* **151**, 2091–2098 (2007).
4. Uthicke, S. Nutrient regeneration by abundant coral reef holothurians. *Journal of Experimental Marine Biology and Ecology* **265**, 153–170 (2001).
5. Pan, W. *et al*. Sea cucumbers and their symbiotic microbiome have evolved to feed on seabed sediments. *Nat Commun* **15**, 8825 (2024).
6. Clements, C. S., Pratte, Z. A., Stewart, F. J. & Hay, M. E. Removal of detritivore sea cucumbers from reefs increases coral disease. *Nature Communications* **15**, 1338 (2024).
7. Ramírez-González, J., Moity, N., Andrade-Vera, S. & Reyes, H. Overexploitation and More Than a Decade of Failed Management Leads to No Recovery of the Galápagos Sea Cucumber Fishery. *Frontiers in Marine Science* **7**, (2020).
8. Ruiz, D. J. & Wolff, M. The Bolivar Channel Ecosystem of the Galapagos Marine Reserve: Energy flow structure and role of keystone groups. *Journal of Sea Research* **66**, 123–134 (2011).
9. Hearn, A. & Martinez, P. Veronica Toral-Granda, M., Murillo, J. C. & Polovina, J. Population dynamics of the exploited sea cucumber Isostichopus fuscus in the western Galapagos Islands, Ecuador. *Fisheries Oceanography* **14**, 377–385 (2005).
10. Wolff, M., Schuhbauer, A. & Castrejón, M. A revised strategy for the monitoring and management of the Galapagos sea cucumber *Isostichopus fuscus* (Aspidochirotida: Stichopodidae). *Revista de Biología Tropical* **60**, (2012).
11. Toral-Granda, V. *et al*. Sea cucumbers. A global review on fishery and trade. *SPC Beche-de-mer Information Bulletin* **28**, 4–6 (2008).
12. Toral-Granda, M. V. Requiem for the Galápagos sea cucumber fishery? 4.
13. Mercier, A. *et al*. *Isostichopus Fuscus, Brown Sea Cucumber*. https://doi.org/10.2305/IUCN.UK.2013-1.RLTS.T180373A1621878.en (2013).
14. CITES. Checklist of CITES species. https://checklist.cites.org/#/en (2020).
15. WWF. Entra en vigencia el Calendario Pesquero 2016–2021 para el desarrollo sostenible de las pesquerías en Galápagos | WWF. https://wwf.panda.org/wwf_news/?282870/Entra-en-vigencia-el-Calendario-Pesquero-2016-2021-para-el-desarrollo-sostenible-de-las-pesqueras-en-Galpagos (2016).
16. Pu, Y., Zhou, Y., Liu, J. & Zhang, H. A high-quality chromosomal genome assembly of the sea cucumber Chiridota heheva and its hydrothermal adaptation. *GigaScience* **13**, giad107 (2024).
17. Zhong, S. *et al*. Chromosomal-level genome assembly and annotation of the tropical sea cucumber Holothuria scabra. *Sci Data* **11**, 474 (2024).
18. Chen, T. *et al*. The Holothuria leucospilota genome elucidates sacrificial organ expulsion and bioadhesive trap enriched with amyloid-patterned proteins. *Proceedings of the National Academy of Sciences* **120**, e2213512120 (2023).
19. Jiang, C. *et al*. Chromosome-level genome assembly of the sea cucumber, Colochirus anceps. *Scientific Data* **12**, 1643 (2025).
20. Chen, T. *et al*. Chromosome-level genome assembly and annotation of the tropical sea cucumber Stichopus monotuberculatus. *Sci Data* **11**, 1245 (2024).
21. Sun, L., Jiang, C., Su, F., Cui, W. & Yang, H. Chromosome-level genome assembly of the sea cucumber Apostichopus japonicus. *Sci Data* **10**, 454 (2023).
22. Lau, N.-S. *et al*. A chromosomal-level genome assembly of the tropical sea cucumber Stichopus fusiformiossa. *Sci Data* **12**, 973 (2025).
23. Jacobs, D. K. *et al*. Reference genome for the endangered, genetically subdivided, northern tidewater goby, Eucyclogobius newberryi. *J Hered* **116**, 170–178 (2025).
24. Cheng, H., Concepcion, G. T., Feng, X., Zhang, H. & Li, H. Haplotype-resolved de novo assembly using phased assembly graphs with hifiasm. *Nat Methods* **18**, 170–175 (2021).
25. Altschul, S. F. *et al*. Gapped BLAST and PSI-BLAST: a new generation of protein database search programs. *Nucleic Acids Res* **25**, 3389–3402 (1997).
26. Laetsch, D. R. & Blaxter, M. L. BlobTools: Interrogation of genome assemblies. *F1000Res* **6**, 1287 (2017).
27. Guan, D. *et al*. Identifying and removing haplotypic duplication in primary genome assemblies. *Bioinformatics* **36**, 2896–2898 (2020).
28. Putnam, N. H. *et al*. Chromosome-scale shotgun assembly using an *in vitro* method for long-range linkage. *Genome Res* **26**, 342–350 (2016).
29. Li, H. & Durbin, R. Fast and accurate short read alignment with Burrows–Wheeler transform. *Bioinformatics* **25**, 1754–1760 (2009).
30. Flynn, J. M. *et al*. RepeatModeler2 for automated genomic discovery of transposable element families. *Proceedings of the National Academy of Sciences* **117**, 9451–9457 (2020).
31. Bao, Z. & Eddy, S. R. Automated De Novo Identification of Repeat Sequence Families in Sequenced Genomes. *Genome Res.* **12**, 1269–1276 (2002).
32. Price, A. L., Jones, N. C. & Pevzner, P. A. De novo identification of repeat families in large genomes. *Bioinformatics* **21**, i351–i358 (2005).
33. Astashyn, A. *et al*. Rapid and sensitive detection of genome contamination at scale with FCS-GX. *Genome Biology* **25**, 60 (2024).
34. NCBI GenBank https://www.ncbi.nlm.nih.gov/datasets/genome/GCF_902459465.1. (2020).
35. Medina-Feliciano, J. G., Pirro, S., García-Arrarás, J. E., Mashanov, V. & Ryan, J. F. Draft Genome of the Sea Cucumber Holothuria glaberrima, a Model for the Study of Regeneration. *Front Mar Sci* **8**, 603410 (2021).
36. NCBI GenBank https://www.ncbi.nlm.nih.gov/datasets/genome/GCF_000002235.5. (2019).
37. Stanke, M., Diekhans, M., Baertsch, R. & Haussler, D. Using native and syntenically mapped cDNA alignments to improve de novo gene finding. *Bioinformatics* **24**, 637–644 (2008).
38. Korf, I. Gene finding in novel genomes. *BMC Bioinformatics* **5**, 59 (2004).
39. Dobin, A. *et al*. STAR: ultrafast universal RNA-seq aligner. *Bioinformatics* **29**, 15–21 (2013).
40. Cantarel, B. L. *et al*. MAKER: An easy-to-use annotation pipeline designed for emerging model organism genomes. *Genome Res.* **18**, 188–196 (2008).
41. Haas, B. J. *et al*. Improving the Arabidopsis genome annotation using maximal transcript alignment assemblies. *Nucleic Acids Res* **31**, 5654–5666 (2003).
42. The UniProt Consortium. UniProt: the universal protein knowledgebase. *Nucleic Acids Res* **45**, D158–D169 (2017).

43. Chan, P. P., Lin, B. Y., Mak, A. J. & Lowe, T. M. tRNAscan-SE 2.0: improved detection and functional classification of transfer RNA genes. *Nucleic Acids Research* **49**, 9077–9096 (2021).
44. Rhie, A., Walenz, B. P., Koren, S. & Phillippy, A. M. Merqury: reference-free quality, completeness, and phasing assessment for genome assemblies. *Genome Biology* **21**, 245 (2020).
45. Ranallo-Benavidez, T. R., Jaron, K. S. & Schatz, M. C. GenomeScope 2.0 and Smudgeplot for reference-free profiling of polyploid genomes. *Nat Commun* **11**, 1432 (2020).
46. Gurevich, A., Saveliev, V., Vyahhi, N. & Tesler, G. QUAST: quality assessment tool for genome assemblies. *Bioinformatics* **29**, 1072–1075 (2013).
47. Manni, M., Berkeley, M. R., Seppey, M., Simão, F. A. & Zdobnov, E. M. BUSCO Update: Novel and Streamlined Workflows along with Broader and Deeper Phylogenetic Coverage for Scoring of Eukaryotic, Prokaryotic, and Viral Genomes. *Molecular Biology and Evolution* **38**, 4647–4654 (2021).
48. Allio, R. *et al*. MitoFinder: Efficient automated large-scale extraction of mitogenomic data in target enrichment phylogenomics. *Molecular Ecology Resources* **20**, 892–905 (2020).
49. Uliano-Silva, M. *et al*. MitoHiFi: a python pipeline for mitochondrial genome assembly from PacBio high fidelity reads. *BMC Bioinformatics* **24**, 288 (2023).
50. Camacho, C. *et al*. BLAST+: architecture and applications. *BMC Bioinformatics* **10**, 421 (2009).
51. Coombe, L., Kazemi, P., Wong, J., Birol, I. & Warren, R. L. Multi-genome synteny detection using minimizer graph mappings. 2024.02.07.579356 Preprint at https://doi.org/10.1101/2024.02.07.579356 (2024).
52. Coombe, L., Warren, R. L. & Birol, I. ntSynt-viz: Visualizing synteny patterns across multiple genomes. 2025.01.15.633221 Preprint at https://doi.org/10.1101/2025.01.15.633221 (2025).
53. *NCBI Sequence Read Archive* https://identifiers.org/ncbi/insdc.sra:SRP598206 (2025).
54. *NCBI GenBank* https://identifiers.org/ncbi/insdc.gca:GCA_051397855.1 (2025).
55. *NCBI GenBank* https://identifiers.org/ncbi/bioproject:PRJNA1284402 (2025).
56. *NCBI GenBank* https://identifiers.org/ncbi/insdc:PV946926.1 (2025).

Acknowledgements

All DNA and RNA extraction, library preparation, sequencing, and assembly were performed by Dovetail Genomics (Cantata Bio. LLC), and we are grateful to our project manager, Jordan Zhang, and their other staff for support and guidance throughout the process. This project was funded through a grant from the Revive and Restore Catalyst Science Fund to N.O.T., and we are thankful for their support. Samples were collected under the Genetic Research Permit from the Ecuadorian Government: "ECOLOGÍA. EVALUACIÓN Y MANEJO DE PESQUERÍAS - PASOS HACIA LA SOSTENIBILIDAD" MAAE-DNB-CM-2021-0202 and CITES export permits N° 1752606 and N° 1752036. We thank the Galapagos National Park, MAATE, and the Charles Darwin Foundation for their institutional support. This publication is contribution number 2744 of the Charles Darwin Foundation for the Galapagos Islands. J. Ortiz was supported by a fellowship from the SENESCYT. Local artisanal fishers helped collect samples. We are grateful to Captain Wilson Gustavo Rivadeneira Arias and his crew from the fishing boat "Pesquero Piquero" and the GNPD captain and crew of the "Molme" vessel.

Author contributions

Jaime D. Ortiz-Pachar and Nina Overgaard Therkildsen conceived the research project and wrote the manuscript. Jaime D. Ortiz-Pachar performed the post-processing bioinformatics analysis, mitochondrion assembly and annotation, and quality validation. Jorge Ramírez-González, Nicolas Moity, Solange Andrade-Vera, and Wilson D. Rivadeneira collected the samples and obtained all the necessary research, export, and CITES permits. All authors approved the final submission.

Competing interests

The authors declare no competing interests.

Additional information

Correspondence and requests for materials should be addressed to J.D.O.-P.