scientific data

Check for updates

OPEN

DATA DESCRIPTOR

Chromosome-level genome assembly and annotation of the tropical sea cucumber *Holothuria fuscocinerea*

Xuan Wang[1,2,7], Qianying Huang[3,7], Zhou Qin[1,2], Dingding Fan[1], Chunhua Ren[1], Wenjie Pan[1,2], Jiasheng Huang[1,2], Zepeng Zhang[4], Hua Ge[3], Jingxuan Liang[1,2], Jianfeng Xu[1,2], Yi Zhang[1,5], Peng Luo[1], Xiao Jiang[1], Lina Sun[6], Hongyan Sun[4], Chaoqun Hu[1], Aifen Yan[3] ✉ & Ting Chen[1] ✉

By the integration of MGI short-read sequencing, PacBio HiFi long-read sequencing, and Hi-C techniques, we constructed a chromosome-scale genome assembly for *Holothuria fuscocinerea*, a highly abundant and widely distributed sea cucumber species in the family Holothuriidae. The assembly was anchored into 23 pseudochromosomes with a total size of 1.556 Gb after redundancy removal, demonstrating contig and scaffold N50 values of 5.18 Mb and 63.82 Mb, respectively, indicative of high contiguity. Genome quality assessment using Merqury and BUSCO analyses yielded a QV score of 60.4087 and a BUSCO completeness rate of 97.8%, confirming high assembly quality with minimal gaps. Among the 29,878 protein coding genes, 95 percent obtained functional annotations from at least one database. This high-resolution genome resource is expected to advance research on *H. fuscocinerea* and the Holothuroidea class, providing insights for population management, evolutionary studies, and genetics research.

Background & Summary

The class Holothuroidea, or sea cucumbers, as both diverse and abundant member of the phylum Echinodermata, are globally distributed across marine benthic habitats[1]. Among the 1,693 currently accepted extant species, around 800 species inhabit the shallow regions of the Indo-West-Pacific Ocean, where the highest species diversity occurs[2]. Commercially valuable species predominantly belong to the families Holothuriidae and Stichopodidae, whose dried body walls are consumed for either food or traditional medicine[3]. Sea cucumbers enhance coastal ecosystem health (e.g., coral reefs) through ecological functions including sediment bioturbation, nitrogen cycling, seawater alkalinity regulation, ocean acidification buffering, symbiont biodiversity maintenance, and trophic energy transfer[4]. However, despite intensive exploitation and slow population recovery rates under escalating demand, critical knowledge gaps persist regarding commercial species' population structures, community composition, and substrate-food interactions, causing difficulties in formulating conservation strategies[5]. In overharvested regions, population restoration through release of artificially cultured juveniles has become essential[6,7]. Despite ongoing translocation of hatchery and wild individuals across regions, the extent of genomic alterations or admixture introduced by these practices hasn't yet been evaluated.

[1]Key Laboratory of Breeding Biotechnology and Sustainable Aquaculture, Laboratory of Tropical Marine Bioresources and Ecology, Guangdong Provincial Key Laboratory of Applied Marine Biology, South China Sea Institute of Oceanology, Chinese Academy of Sciences, Guangzhou, 510301, China. [2]University of Chinese Academy of Sciences, Beijing, 100049, China. [3]School of Medicine, Foshan University, Foshan, 528225, China. [4]Guangdong Laboratory for Lingnan Modern Agriculture, College of Marine Sciences, South China Agricultural University, Guangzhou, 510642, China. [5]College of Fisheries, Huazhong Agricultural University, Wuhan, 430070, China. [6]Key Laboratory of Breeding Biotechnology and Sustainable Aquaculture, Institute of Oceanology, Chinese Academy of Sciences, Qingdao, 266071, China. [7]These authors contributed equally: Xuan Wang, Qianying Huang. ✉e-mail: yanaifen@fosu.edu.cn; chan1010@scsio.ac.cn

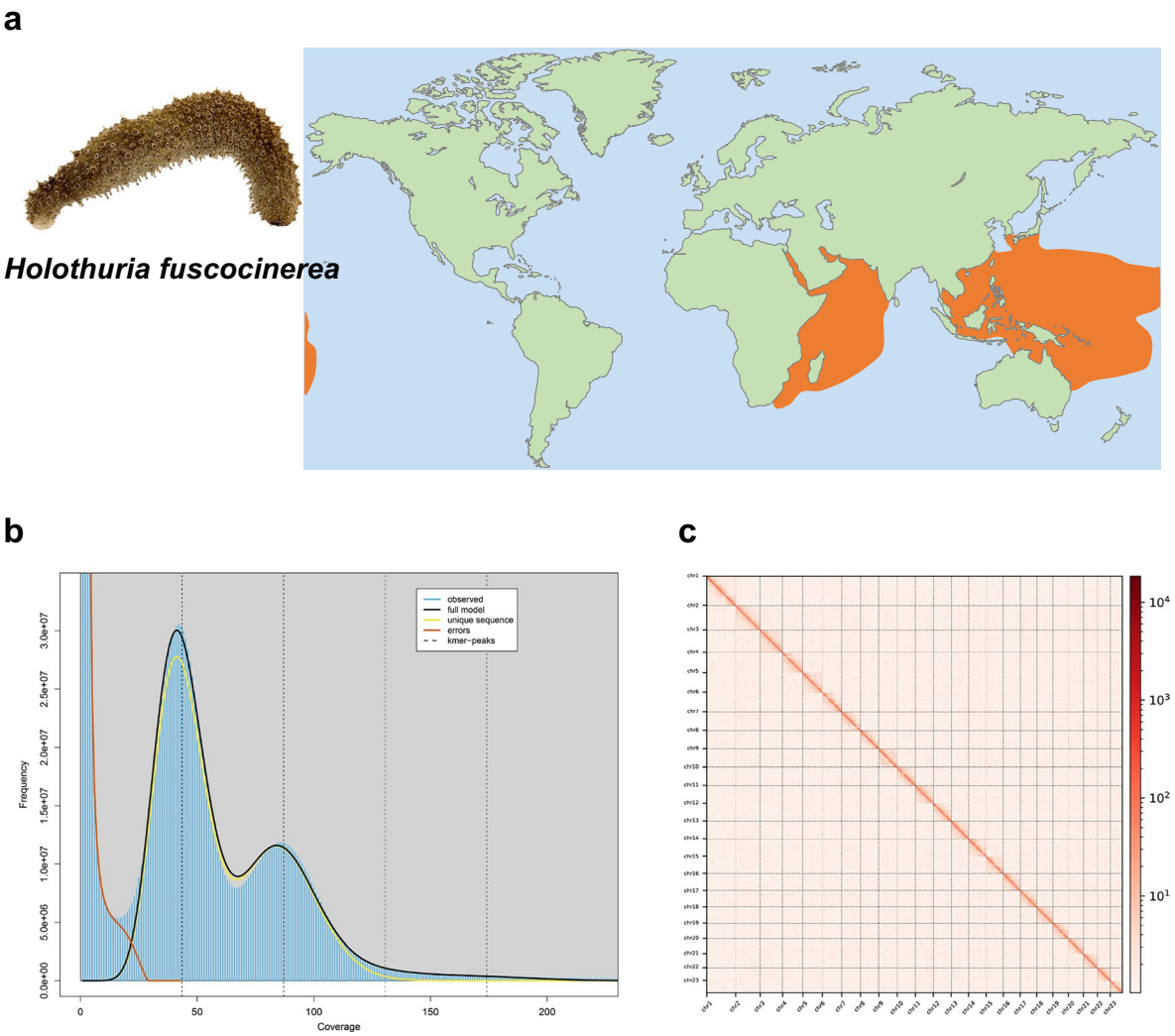

Fig. 1 Basic attributes of the *H. fuscocinerea* genome assembly. (**a**) Illustration of *H. fuscocinerea* with its natural range of distribution; (**b**) Histogram of 21-mer count distribution of the *H. fuscocinerea* genome; (**c**) Hi-C interaction map of the *H. fuscocinerea* genome, visualized using Juicebox[77].

Holothuria fuscocinerea, a commercially harvested species in the family Holothuriidae, is distributed across tropical and subtropical reef habitats in the Pacific and Indian Oceans, ranging from the Red Sea and East African coasts eastward to China, southern Japan, northern Australia, and central Pacific islands (Fig. 1a)[8,9]. This species reaches an average mature length of 50 cm, exhibiting an oval-cylindrical morphology with a flattened bottom and rounded tip[10]. The upper surface displays a rough, brown-grey texture frequently coated with sand grains, potentially serving as an adaptation for sediment camouflage[11]. Like many other species in family Holothuriidae, *H. fuscocinerea* possesses Cuvierian tubules—defensive organs that are expelled when threatened[9,12]. Having sequenced the mitochondrial genome, phylogenetic analyses of *H. fuscocinerea* indicate its closest relatives are either *Holothuria leucospilota* or *Holothuria polii*, though this phylogenetic relationship requires verification through sequencing of more species[13,14]. Although possessing limited commercial value, *H. fuscocinerea* is projected to become a target fishery species following depletion of higher-value commercial species[8,9]. This transition may occur imminently, given that overexploitation has caused population declines in three-quarters of global commercial sea cucumber species, largely reducing potential profits generated from them[15]. Besides, three new triterpene glycosides with demonstrated cytotoxicity, antimicrobial activity, and immunomodulatory potential have been isolated from *H. fuscocinerea*, establishing their pharmacological relevance[16–18]. Knowledge for this species at the organismal level remains limited, with research focused primarily on regional abundance[19,20], molecular identification[21], polysaccharide composition[22], predator-prey interaction[23], feeding mode[20,24] and bioactive compounds[16–18,25,26]. The lack of whole genome assembly and tissue transcriptome of *H. fuscocinerea* hinders the implementation of mechanical studies of development, genetics and behaviours, leaving its phylogeny uncertain as inferred only from the mitochondrial genomic data[13,14].

By integrating the clean data obtained from MGI short-read (100.3 ×), PacBio HiFi long-read (30.54 ×) and Hi-C (89×) sequencing, our study generated a chromosome-scale genome assembly for *H. fuscocinerea*,

Libraries	Insert size (bp)	Clean reads	Clean bases (Gb)	Read Length (bp)	Coverage	Q20	Q30
MGI	350	519,887,557	155.97	150	100.3	98.84%: 97.86%	96.49%: 93.64%
Pacbio HiFI	20,000	2,633,444	47.5	N50: 18,099	30.54	—	—
Hi-C	350	923,272,938	138.4	150	89	96.75%: 95.95%	92.03%: 90.94%
RNA Seq	—	1,209,843,921	362.95	150	—	—	—

Table 1. The sequencing reads and their characteristics obtained from each sequencing. For each Q20/Q30 metric, the value on the left corresponds to the percentage of raw reads bases and the value on the right corresponds to the percentage of clean reads bases after filtering.

organized into 23 pseudochromosomes. The final assembly spans 1.556 Gb with a scaffold N50 value of 63.82 Mb, respectively. Analysis of telomeric and centromeric repeats further supported the high structural completeness of the assembly. BUSCO analysis validated 96.1% complete BUSCOs, including 90.8% single-copy and 5.3% duplicated orthologs, with 1.7% fragmented and 2.2% missing BUSCOs through conserved ortholog alignment. We annotated 29,878 protein coding genes, with 28,110 (95%) receiving functional assignments from at least one database. In addition to the nuclear genome, we also recovered the complete mitochondrial genome using PacBio HiFi reads, providing a complementary mitogenome reference for taxonomic authentication and comparative analyses. Our study provides the genomic resource required for population monitoring, genetic diversity investigation, and broodstock management of *H. fuscocinerea*. By comparing with an increasing number of available holothurian genomes[27–31], this genome may elucidate common gene patterns underlying critical holothurian specializations such as mutable connective tissues and oral tentacles, which shape their ecological niche through evolution. It may also help clarify the ancestral genome arrangement of the genus *Holothuria*, thus revealing the driving forces of diversification within this ancient and complex lineage[32].

Methods

Sample collection and nucleic acid extraction. A female *H. fuscocinerea* specimen was collected at Tanmen Port (19.33°N, 110.49°E), Qionghai City, Hainan Province, China and dissected to obtain sequencing samples. The gonads were rinsed with phosphate-buffered saline (PBS) and frozen in liquid nitrogen immediately. Genomic DNA was extracted using the QIAamp DNA Mini Kit (QIAGEN, Hilden, Germany) for both long-read and short-read whole-genome sequencing. Ten tissues, namely, the skin, intestine, Polian vesicle, muscle, oral tentacle, Cuvierian organ, rete mirabile, body wall, gonad, and respiratory tree were excised with scissors and forceps, rinsed with PBS, and snap frozen with liquid nitrogen. Total RNA was extracted from these tissues using the RNAprep Pure Plant Plus Kit (Tiangen Biotech Co. Ltd., Beijing, China) for transcriptomic sequencing.

Library preparation and sequencing. Short-read sequencing was performed using the MGI DNBSEQ-T7 platform (MGI Tech Co., Shenzhen, China) with a PE150 read configuration. After the high-quality DNA passed purity, concentration and integrity examinations, enzymatic fragmentation and sequencing adapter ligation were conducted with the MGIEasy FS DNA Library Prep Set (MGI, Shenzhen, China) to construct a pair-end library of 350 bp insert size. Enzymatic fragmentation was performed using the Fragmentation Enzyme Mix supplied with the MGIEasy FS DNA Library Prep Set. Three MGI short-read libraries were prepared, but only one high-quality library was used and pooled with unrelated libraries during sequencing. DNA nanoballs (DNBs) were generated through rolling circle replication for subsequent sequencing. The obtained raw data was assessed with FastQC (v0.12.1)[33] and filtered with SOAPnuke (v2.1.4)[34]. A total of 155.97 Gb (100.3×) clean reads were kept for downstream analysis (Table 1; for genome size estimation, see below).

Long-read sequencing was conducted using the PacBio Revio platform (Pacific Biosciences, Menlo Park, USA). High-quality DNA was sheared to approximately 20 kb using the Covaris ultrasonic disruptor (Woburn, MA, USA), then A-tailing and adapter ligation were performed with the SMRTbell Prep Kit 3.0 (PacBio, Menlo Park, CA, USA). After size selection with BluePippin for approximately 20 kb fragments, polymerase binding to the constructed library was performed using the Sequel Binding Kit 3.2 (PacBio, Menlo Park, CA, USA). The library was loaded onto SMRT Cell 8 M and sequenced with a 15-hour movie time. Raw data were converted to Circular Consensus Sequencing (CCS) reads with SMRT Link v13.1 (Pacific Biosciences) and filtered at QV 20[35]. A total of 47.5 Gb (30.54×) clean reads were retained for further analysis (Table 1).

Hi-C sequencing was conducted using the MGI DNBSEQ-T7 platform with PE150 configuration. The gonad sample was fixed with formaldehyde and digested with DpnII enzyme (NEB, Ipswich, MA, USA). DNA was extracted and fragmented to approximately 350 bp fragments after biotin-based end polishing and ligation of spatially adjacent chromatin fragments. These fragments were captured using streptavidin magnetic beads for library construction. The library was sequenced, and raw data were assessed with FastQC (v0.12.1). A total of 138.4 Gb (89×) clean reads were retained (Table 1).

Transcriptomic sequencing was likewise conducted using the MGI DNBSEQ-T7 platform. Total RNA was fragmented, reverse-transcribed, and ligated to adapters using the MGIEasy RNA Library Prep Set (MGI, Shenzhen, China). The constructed libraries were pooled after indexing with sample-specific barcoded primers and sequenced to obtain raw reads from 10 tissues. Clean reads of 362.95 Gb in total were acquired.

Genome assessment and assembly. To assess genome size, heterozygosity, and repeat content, k-mer analysis was performed. K-mers were counted from raw reads using Jellyfish (v2.3.0)[36] with 10 threads and both

K-mer length	K-mer number	K-mer depth	Genome size	Heterozygous ratio (%)	Repeat (%)
21	54,250,555,200	43.5	1,247,369,200	2.99124	34.1

Table 2. *H. fuscocinerea* genomic statistics obtained from K-mer and GenomeScope analyses.

strands considered, while k-mer and hash sizes were set to 21 and 100 million, respectively (Fig. 1b; Table 2). The resulting k-mer histogram (k = 21) was processed with GenomeScope v2.0.0[37], under the assumption of a diploid genome. After removing erroneous-frequency k-mers, 54,250,555,200 k-mers exhibiting a main peak at depth 43.5 were retained (Table 2). GenomeScope analysis estimated a *H. fuscocinerea* genome size of ~1.247 Gb, with a heterozygosity of 2.99% and repetitive sequence component of 34.09%.

Using Hifiasm (v0.19.8-r603)[38], PacBio long-read data was assembled into a draft genome based on overlapping reads after all-versus-all reads comparison and self-correction. MGI short-read data was used to correct residual errors using Bwa (v0.7.17-r1188)[39] and Pilon (v1.23)[40], followed by Purge_dups (v1.2.5)[41] for haplotig reduction and heterozygosity resolution. The resulting 541 contigs were assigned and oriented with Hi-C reads using the software ALLHiC (v1.1)[41], producing a chromosome-level assembly with 23 pseudochromosomes at 94.82% anchoring rate (Fig. 1c). We created a plot displaying chromosomal structure with Circos (v0.69-8)[42], which facilitated the interpretation of various genomic features and their distribution along the chromosomes (Fig. 2; Table 3). This visualization included the following tracks from inner to outer circles: genic regions, repeat sequences, SNP percentage, NGS sequencing depth, and chromosomes. At a 500-kb window scale, genic regions (exons + introns) covered on average 60% of each window, while repeat sequences occupied about 20%. The SNP percentage equals to the fraction of heterozygous SNPs per window derived from MGI short-read mapping and variant calling, with a genome-wide mean of 0.019%. The NGS sequencing depth equals to the local coverage of the MGI short reads, with an average depth of 173× in each window. The final assembly comprises 64 scaffolds (1.556 Gb total) of 23 anchored pseudochromosomes and 41 unanchored scaffolds,with 63.82 Mb scaffold N50 (Table 4).

The mitochondrial genome was not recovered from the initial nuclear genome assembly. Following the reviewer's suggestion, we assembled the mitochondrial genome using mitoHiFi (v3.2.3)[43], which reconstructs circular mitochondrial genomes directly from PacBio HiFi reads. The mitochondrial genome of Holothuria spinifera strain S-4 (GenBank accession: NC_046508) was used as the reference sequence. The resulting mitogenome was validated by BLAST searches against the NCBI nucleotide (nt) database, which confirmed that it belongs to *Holothuria fuscocinerea*.

Repetitive element annotation. Transposable element annotations were achieved employing RepeatMasker (v4.09, plus RepBase 20181026) and EDTA[44], a whole-genome *de-novo* TE annotation pipeline. LTR retrotransposons were identified with LTRharvest (LTR_FINDER alternative)[45] and filtered with LTR_retriever[46], while other TEs were detected by TIR-Learner[47], HelitronScanner[48], and GRF[49], which were further supplemented with RepeatModeler[50]. Prior to removing redundant results from the three methods, a nucleotide coding sequence (CDS) reference database was applied to minimize erroneous gene fragment annotations in the TE library. It is determined that the genome contains 49.59% repetitive elements, with DNA transposons (19.72%) constituting the major component, alongside simple repeats (2.20%), LINEs (4.52%), and LTRs (0.67%) (Fig. 3a; Table 5).

Noncoding RNA annotation. Under default configurations, Barrnap (v0.9)[51] was used to scan ribosomal RNAs (rRNAs) and tRNAscan-SE (v2.0.11)[52] was applied to find transfer RNAs (tRNAs). The Rfam database (v14.8)[53] was employed to detect microRNAs (miRNAs) and small nuclear RNAs (snRNAs) based on sequence similarity, while subsequent identification was performed with Infernal (v1.1.4)[54] through alignment based on covariance models. In summary, the genome contained 586 rRNAs, 31,158 tRNAs, 70 miRNAs and 919 snRNAs (Fig. 3b; Table 6).

Gene prediction and functional annotation. The RNAseq data was applied to enhance the prediction of exon–intron boundaries and start codons, and MAKER3 (v3.01.03)[55] was used to combine annotations from different modes. Specifically, HISAT2 (v2.2.1)[56], a splice-aware aligner optimized for fast and accurate mapping of transcriptomic data, was used to map 362.95 Gb of clean reads obtained from 28 samples onto the reference genome, followed by assembly of transcripts with StringTie (v2.1.7) (Table 7)[57]. Annotation of the genome by homology was then performed by matching the *H. fuscocinerea* reference genome with sequences from 9 high-quality protein datasets, namely the *Lytechinus pictus* (GCF_015342785.2)[58], *Patiria miniata* (GCF_015706575.1)[59], *Anneissia japonica* (GCF_011630105.1)[59], *Asterias rubens* (GCF_902459465.1)[59], *Acanthaster planci* (GCF_001949145.1)[60], *Lytechinus variegatus* (GCF_018143015.1)[61], *Strongylocentrotus purpuratus* (GCF_000002235.5)[62], *Holothuria leucospilota* (GCA_029531755.1)[28] and *Apostichopus japonicus* (GCA_002754855.1)[27].

Gene prediction and annotation were performed using the BRAKER3 pipeline (v3.0.3)[63], which integrates GeneMark-ETP+ and AUGUSTUS[64] to predict gene models in an unsupervised, evidence-guided manner. This approach ensures the annotation of both well-supported coding regions and novel gene models by leveraging RNA-seq expression profiles and protein orthologs simultaneously. The results were merged into a final gene collection consisting of 29,878 genes, and 57.3% of the genes were approved by both evidence with ≥50% coverage of coding region (Fig. 4a). Gene sets of three other members of Holothuroidea, namely *H. leucospilota*[28], *A. japonicus*[27], and *Stichopus monotuberculatus*[30], were compared with the predicted *H. fuscocinerea* collection

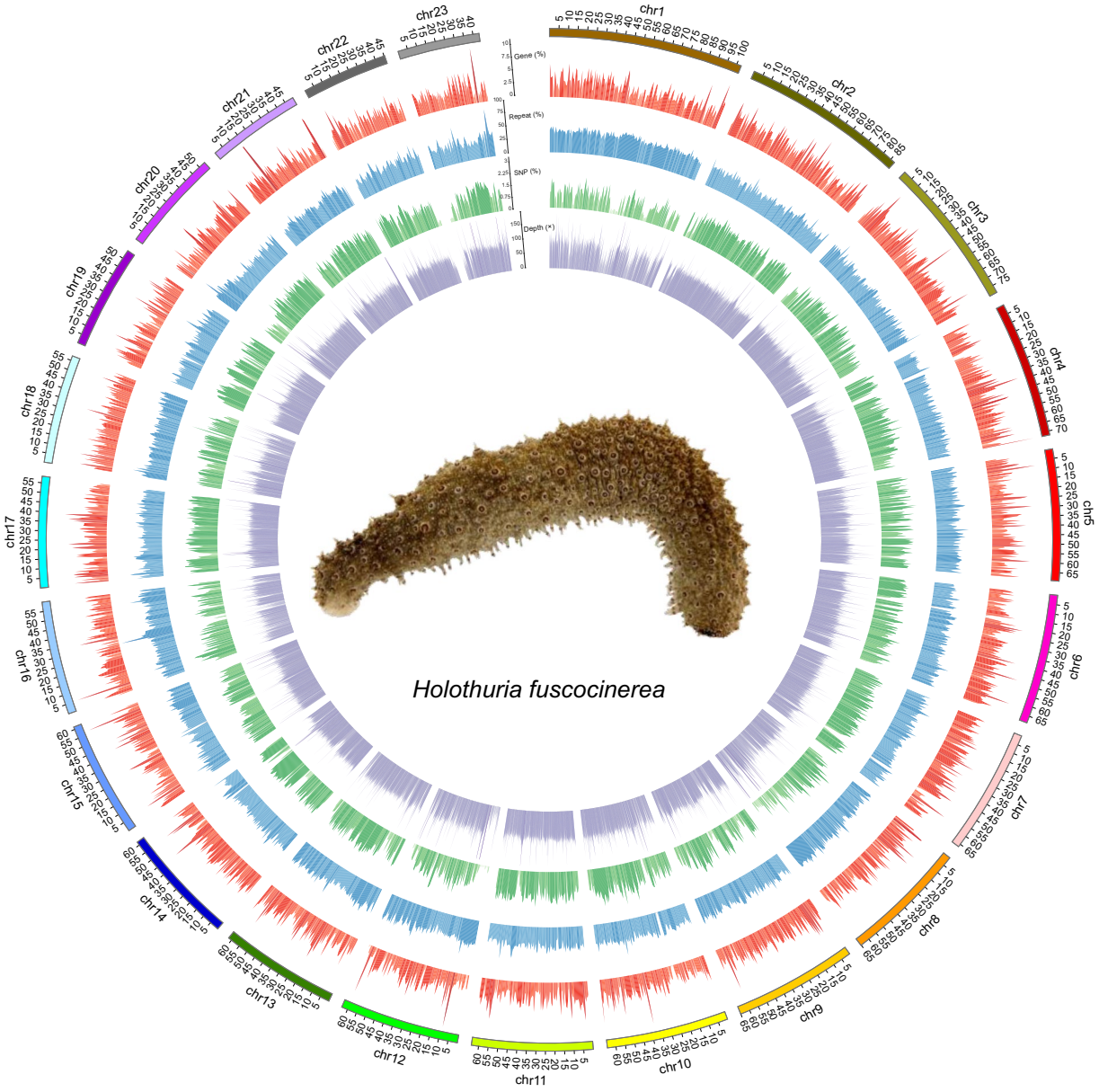

Fig. 2 Circos plot depicting the genomic features of *H. fuscocinerea*. The outermost circle exhibits the pseudochromosomes with marked x-axis showing their length. The inner circles from inner to outer are genic regions, repeat sequences, SNP percentage, and NGS sequencing depth. For each track, the y-axis from 0 to the maximum indicates respective observed ranges. Average values for these features are 60%, 20%, 0.019%, and 173x, respectively. Each feature employed a window size of 500 kb.

in terms of gene amount and detailed attributes (Table 8). The average transcript length of *H. fuscocinerea* was shown to be notably longer due to the addition of UTR regions to its transcripts by the prediction method. BUSCO analysis was then used to assess the quality of the predicted gene collection. The result showed that our collection contained homologs corresponding to 97.8% of the 954 metazoan conserved genes from 65 genomes (Table 9), while fragmented BUSCOs only made up 1.7%. Thus, the 96.1% complete BUSCO results confirm the general completeness of protein-coding genes uncovered in the *H. fuscocinerea* genome.

The functions and pathways of coding genes from the collection were further annotated using BLAST (v2.12.0+, e-value 1e^-5)[65], by aligning their protein sequences with data retrieved from UniProtKB/Swiss-Prot[66], UniProt[67] and KEGG[68]. InterProScan (v5.56)[69] was also applied to assign GO terms[70], in addition to the prediction of gene families, domains and active sites. Among these genes, 95% received annotation from at least one database (Fig. 4b; Table 10).

Telomere and centromere identification. To evaluate the structural completeness of the assembled genome and to identify putative telomeric and centromeric regions, we analyzed all 23 pseudochromosomes using tidk (v0.2.31)[71] and quarTeT (v1.2.5r4)[72]. Telomeric repeat identification was performed with tidk, which

Chromosome ID	scaffolded contigs	Length (bp)	GC Content	MGI Base Coverage	HIFI Base Coverage
Chr1	61	102,481,944	0.3925	82.5487	26.75
Chr2	10	87,536,712	0.3899	100.5617	30.21
Chr3	26	79,194,133	0.3905	100.6603	30.25
Chr4	12	72,173,421	0.3939	100.4911	30.41
Chr5	25	69,305,304	0.3916	96.2229	29.09
Chr6	20	68,610,530	0.3965	99.5423	28.75
Chr7	13	66,539,636	0.3935	97.1762	30.20
Chr8	46	65,158,882	0.3902	92.7341	27.96
Chr9	15	65,049,147	0.3908	100.9037	29.83
Chr10	19	64,527,398	0.3908	97.8433	30.26
Chr11	11	63,820,209	0.3918	102.0026	29.97
Chr12	59	63,236,084	0.3915	87.6946	26.43
Chr13	21	61,883,644	0.3914	92.0205	28.67
Chr14	9	61,823,350	0.3917	100.9595	30.34
Chr15	12	61,340,513	0.3950	100.5393	29.60
Chr16	28	59,983,962	0.3918	84.7402	28.18
Chr17	19	58,442,115	0.3900	101.0145	29.82
Chr18	19	57,425,682	0.3928	95.2754	28.05
Chr19	14	54,745,315	0.3928	99.0963	29.71
Chr20	30	53,662,173	0.3912	98.7249	29.62
Chr21	11	49,797,694	0.3927	99.4928	29.58
Chr22	9	45,330,500	0.3915	101.6151	30.24
Chr23	11	43,245,116	0.3910	100.5426	28.97
unanchored	41	80,589,354	—	—	—
Total	541	1555902818	—	—	—

Table 3. Details of the 23 Hi-C sequencing based chromosomal-level assemblies.

	Scaffold Contig	
	Length(bp)	**Length(bp)**
Max length	102481944	20657421
N10	87536712	13128536
N20	72173421	9421218
N30	68610530	8130017
N40	65049147	6436443
N50	**63820209**	**5179660**
N60	61823350	4362018
N70	59983962	3216522
N80	54745315	2409083
N90	45330500	1522653
Total length	1555902818	1555855118
	Scaffold Contig	
	Number	**Number**
Max length		
L10	2	10
L20	4	24
L30	6	42
L40	9	64
L50	**11**	**90**
L60	14	123
L70	16	164
L80	19	221
L90	22	300
Number > = 100 bp	64	541
Number > = 2000 bp	64	541

Table 4. Summary of the assembled *H. fuscocinerea* genome after the removal of redundant sequences.

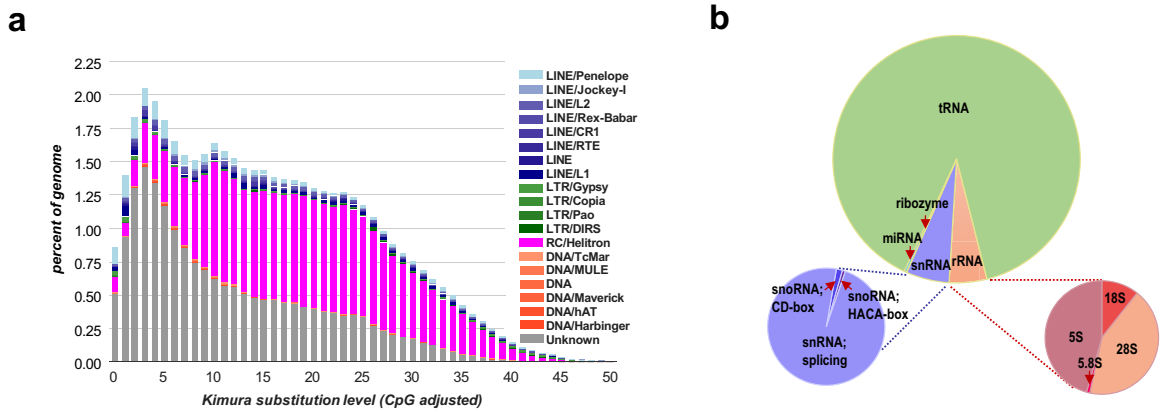

Fig. 3 Results of non-coding elements and RNA prediction. (**a**) Classification of annotated repetitive elements with Kimura substitution levels displayed. The x-axis shows Kimura divergence levels, which reflects the approximate period of TE insertions. Lower divergence hints recent insertions, while higher divergence indicates paleo insertions. The y-axis gives the genome proportion from each TE family at every divergence level. This plot implies the expansions of LINEs, LTR elements, RCs, and DNA transposons throughout time in the *H. fuscocinerea* genome; (**b**) Classification of annotated noncoding RNAs.

TE subtype	EDTA + Repeat Masker		EDTA		Repeat Masker	
	Length (bp)	length/genome	Length (bp)	length/genome	Length (bp)	length/genome
DNA	245,040,454	19.72	244,704,019	19.69	388,040	0.03
LINE	56,184,204	4.52	54,141,428	4.36	4,301,695	0.35
LTR	8,333,561	0.67	8,174,474	0.66	288,544	0.02
Low_complexity	3,055,729	0.25	0	0.00	3,055,729	0.25
SINE	1,083,380	0.09	0	0.00	1,083,380	0.09
Simple_repeat	27,290,623	2.20	0	0.00	27,290,623	2.20
Unknown	294,364,283	23.69	294,360,033	23.69	5,456	0.00
Total	616,319,292	49.59	596,560,351	48.00	36,481,589	2.94

Table 5. List of repetitive elements obtained through repeat annotation.

	Type	Copy number	Total length (bp)	Percentage (%)
	miRNA	70	7,738	0.0006%
	tRNA	31,158	2,318,302	0.1865%
rRNA	rRNA	586	127,476	0.0103%
	18S	10	13,413	0.0011%
	28S	28	54,755	0.0044%
	5.8S	6	940	0.0001%
	5S	542	58,368	0.0047%
snRNA	snRNA	919	146,406	0.0118%
	snoRNA; CD-box	22	2,379	0.0002%
	snoRNA; HACA-box	7	1,340	0.0001%
	snoRNA; scaRNA	1	133	0.0000%
	snRNA; splicing	889	142554	0.0115%
	IncRNA	2	254	0.0000%
ribozyme		3	849	0.0001%

Table 6. List of non-coding RNAs obtained through annotation.

detected the canonical echinoderm telomeric motif "AACCCT" across the genome. Among the 23 pseudochromosomes, 17 contained telomeric sequences, and 10 pseudochromosomes exhibited telomeric repeats at both ends, suggesting a high degree of assembly continuity approaching near telomere-to-telomere (T2T) quality (Fig. 5; Fig. 6).

Putative centromeric regions were identified using the CentroMiner module of quarTeT. The analysis detected centromeric signatures on nine pseudochromosomes: chr4, chr6, chr7, chr10, chr12, chr14, chr15, chr16, and chr19.

Sample	Tissue	Clean reads	Clean bases	Q20 bases	Q30 bases
BW-8-1	Body Wall	34,384,163	10,315,248,900	98.90%: 98.33%	96.59%: 95.87%
BW-8-2	Body Wall	33,385,016	10,015,504,800	99.10%: 98.59%	97.07%: 96.33%
CT-1-2	Cuvierian Organ	38,317,795	11,495,338,500	99.22%: 99.66%	97.10%: 98.69%
CT-3-1	Cuvierian Organ	39,124,760	11,737,428,000	99.29%: 99.66%	97.35%: 98.66%
CT-4-1	Cuvierian Organ	35,769,453	10,730,835,900	99.28%: 99.67%	97.27%: 98.72%
IN-2-1	Intestine	49,969,090	14,990,727,000	99.31%: 99.64%	97.40%: 98.59%
IN-3-1	Intestine	49,133,968	14,740,190,400	99.24%: 99.59%	97.12%: 98.34%
IN-4-1	Intestine	50,653,582	15,196,074,600	99.35%: 99.61%	97.53%: 98.45%
MS-3-1	Muscle	53,913,944	16,174,183,200	99.03%: 99.56%	96.37%: 98.18%
MS-8-1	Muscle	45,990,549	13,797,164,700	99.35%: 99.55%	97.52%: 98.18%
OA-2-1	Ovary	32,522,315	9,756,694,500	99.28%: 99.68%	97.28%: 98.74%
OA-3-3	Ovary	39,012,042	11,703,612,600	99.35%: 99.73%	97.54%: 98.93%
OT-1-1	Oral Tentacles	49,141,990	14,742,597,000	99.39%: 99.58%	97.69%: 98.32%
OT-2-1	Oral Tentacles	40,831,306	12,249,391,800	98.79%: 98.47%	96.24%: 96.18%
OT-3-1	Oral Tentacles	42,400,416	12,720,124,800	99.23%: 99.61%	97.12%: 98.44%
PV-3-1	Polian Vesicles	32,994,916	9,898,474,800	99.29%: 99.68%	97.34%: 98.76%
PV-6-1	Polian Vesicles	37,316,538	11,194,961,400	99.34%: 99.60%	97.52%: 98.43%
PV-8-1	Polian Vesicles	41,250,982	12,375,294,600	99.30%: 99.59%	97.34%: 98.41%
RM-1-1	Rete Mirabile	42,809,469	12,842,840,700	98.91%: 98.11%	96.59%: 95.35%
RM-3-1	Rete Mirabile	38,074,253	11,422,275,900	98.97%: 98.52%	96.80%: 96.34%
RM-4-1	Rete Mirabile	39,517,117	11,855,135,100	98.81%: 98.30%	96.28%: 95.77%
RT-1-1	Respiratory Tree	48,709,434	14,612,830,200	98.83%: 97.93%	96.37%: 94.99%
RT-4-1	Respiratory Tree	44,316,868	13,295,060,400	98.71%: 97.83%	96.02%: 94.72%
RT-6-1	Respiratory Tree	55,761,211	16,728,363,300	98.91%: 98.03%	96.58%: 95.18%
SK-4-1	Skin	40,520,645	12,156,193,500	99.40%: 99.67%	97.75%: 98.71%
SK-4-2	Skin	37,532,130	11,259,639,000	99.27%: 99.64%	97.30%: 98.61%
SK-7-1	Skin	44,086,827	13,226,048,100	99.33%: 99.64%	97.48%: 98.60%
TS-4-1	Testis	72,403,142	21,720,942,600	98.93%: 97.94%	96.60%: 94.88%
Total		**1209843921**	**362953176300**	—	—

Table 7. Data generated from the transcriptomic sequencing of tissue samples. For each Q20/Q30 metric, the value on the left corresponds to the percentage of raw reads bases and the value on the right corresponds to the percentage of clean reads bases after filtering.

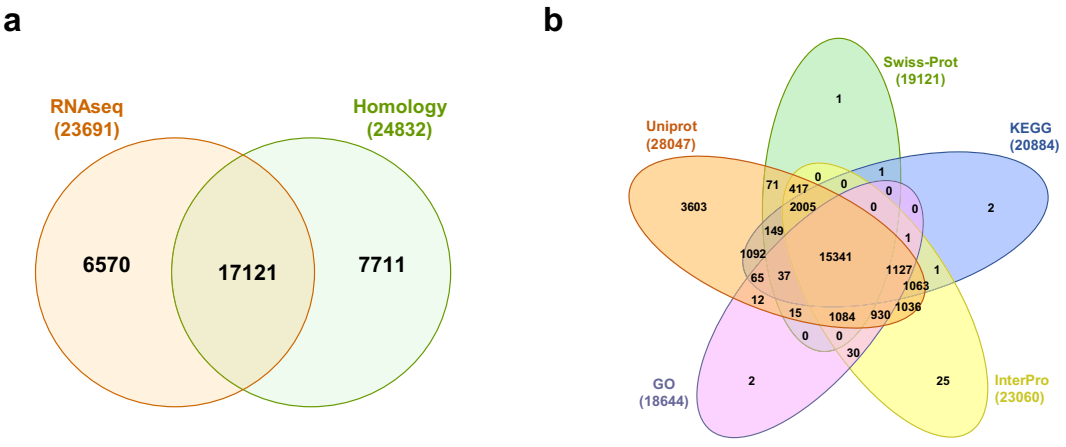

Fig. 4 Results of gene prediction and various annotations. (**a**) Venn Diagram of genes predicted by RNAseq data and homology. (**b**) Venn Diagram of genes annotated with Swiss-Prot, Uniprot, KEGG, GO and InterProScan.

Data Records

The following sites store all the data generated in this study available for downloading. The BioProject PRJNA1250249[73] of NCBI has been created to store raw data acquired in various sequencings and the final *H. fuscocinerea* genome assembled into 23 chromosomes. A complete one-to-one correspondence between each sequencing library and its SRA run accession is provided as follows. The PacBio HiFi long-read library is available

Species	Number of genes	Single exon gene	Average transcript length (bp)	Average CDS length (bp)	Average exons per gene	Average exon length (bp)	Average intron length (bp)
Holothuria fuscocinerea	29,878	13.87%	50,219	1,764	8	219	3,309
Holothuria leucospilota	37,209	25.02%	18,434	1,350	6	240	2,960
Apostichopus japonicus	19,828	11.60%	14,590	1,573	8	195	1,845
Stichopus monotuberculatus	29,596	11.67%	16,701	1,384	7	201	2,251

Table 8. The *Holothuria fuscocinerea* assembly in comparison with three other available Holothuroidea genomes in terms of genes and gene elements.

Feature	*H. fuscocinerea*	
	Orthologs	Percent (%)
Complete BUSCOs (C)	917	96.1
Complete and single-copy BUSCOs (S)	866	90.8
Complete and duplicated BUSCOs (D)	51	5.3
Fragmented BUSCOs (F)	16	1.7
Missing BUSCOs (M)	21	2.2
Total BUSCO groups searched	954	100.0

Table 9. Detailed results obtained from the BUSCO analysis of the annotated genes.

Feature	Functionally annotated genes	Percent of all genes (%)
UniProt	28,047	93%
UniProtKB/Swiss-Prot KEGG	19,121 20,884	63% 69%
InterProscan	23,060	77%
GeneOntology	18,644	62%
Total	28,110	95%

Table 10. Number of coding genes identified by each method of functional annotation.

under accession SRR33220574 (SRX28467615), the MGI short-read whole-genome sequencing (WGS) library is deposited under SRR33220573 (SRX28467616), and the Hi-C library used for chromosome-level scaffolding is archived under SRR33220562 (SRX28467627). A total of 28 RNA-seq libraries derived from ten tissues were represented by their unique SRA run accessions. RNA-seq libraries for intestines are SRR33220544 (SRX28467645), SRR33220545 (SRX28467644) and SRR33220572 (SRX28467617). RNA-seq libraries for Cuvierian organs are SRR33220546 (SRX28467643), SRR33220547 (SRX28467642) and SRR33220548 (SRX28467641). RNA-seq libraries for body walls are SRR33220549 (SRX28467640) and SRR33220551 (SRX28467638). The RNA-seq library for the testis is SRR33220550 (SRX28467639). RNA-seq libraries for skins are SRR33220552 (SRX28467637), SRR33220553 (SRX28467636) and SRR33220554 (SRX28467635). RNA-seq libraries for respiratory trees are SRR33220555 (SRX28467634), SRR33220556 (SRX28467633) and SRR33220557 (SRX28467632). RNA-seq libraries for rete mirabilia are SRR33220558 (SRX28467631), SRR33220559 (SRX28467630) and SRR33220560 (SRX28467629). RNA-seq libraries for polian vesicles are SRR33220563 (SRX28467626), SRR33220564 (SRX28467625) and SRR33220561 (SRX28467628). RNA-seq libraries for ovaries are SRR33220569 (SRX28467620) and SRR33220568 (SRX28467621). RNA-seq libraries for oral tentacles are SRR33220567 (SRX28467622), SRR33220566 (SRX28467623) and SRR33220565 (SRX28467624). RNA-seq libraries for muscles are SRR33220571 (SRX28467618), SRR33220570 (SRX28467619) and SRR33220552 (SRX28467637). Additionally, the accession number JBNFMV000000000[74] of DDBJ/ENA/GenBank store the genome assembly. The mitochondrial genome generated in this study has also been deposited in GenBank. The Figshare database[75] hosts all data generated from genome assembly, gene annotation and functional annotations at the link: https://figshare.com/s/6c5acac389a19d8c8386.

Technical Validation

Qualification of Nucleic acid. Agarose gel electrophoresis (0.75%) was performed to determine the degradation level of DNA and RNA samples, followed by purity measurement with NanoDrop One spectrophotometer (Thermo Fisher Scientific, Waltham, MA, USA) and concentration assessment with Qubit 3.0 Fluorometer (Life Technologies, Carlsbad, CA, USA). DNA samples and RNA samples with concentrations above 10 ng/μL were allowed to proceed to library preparation.

Qualification of Genome assembly and Annotation. The k-mer coverage and QV of the final genome assembly was assessed using Merqury[76]. By running the script "best_k.sh", the best-fit length of k-mer was determined to be 19. The k-mer amount of the short-read sequencing data was then estimated with Meryl using default

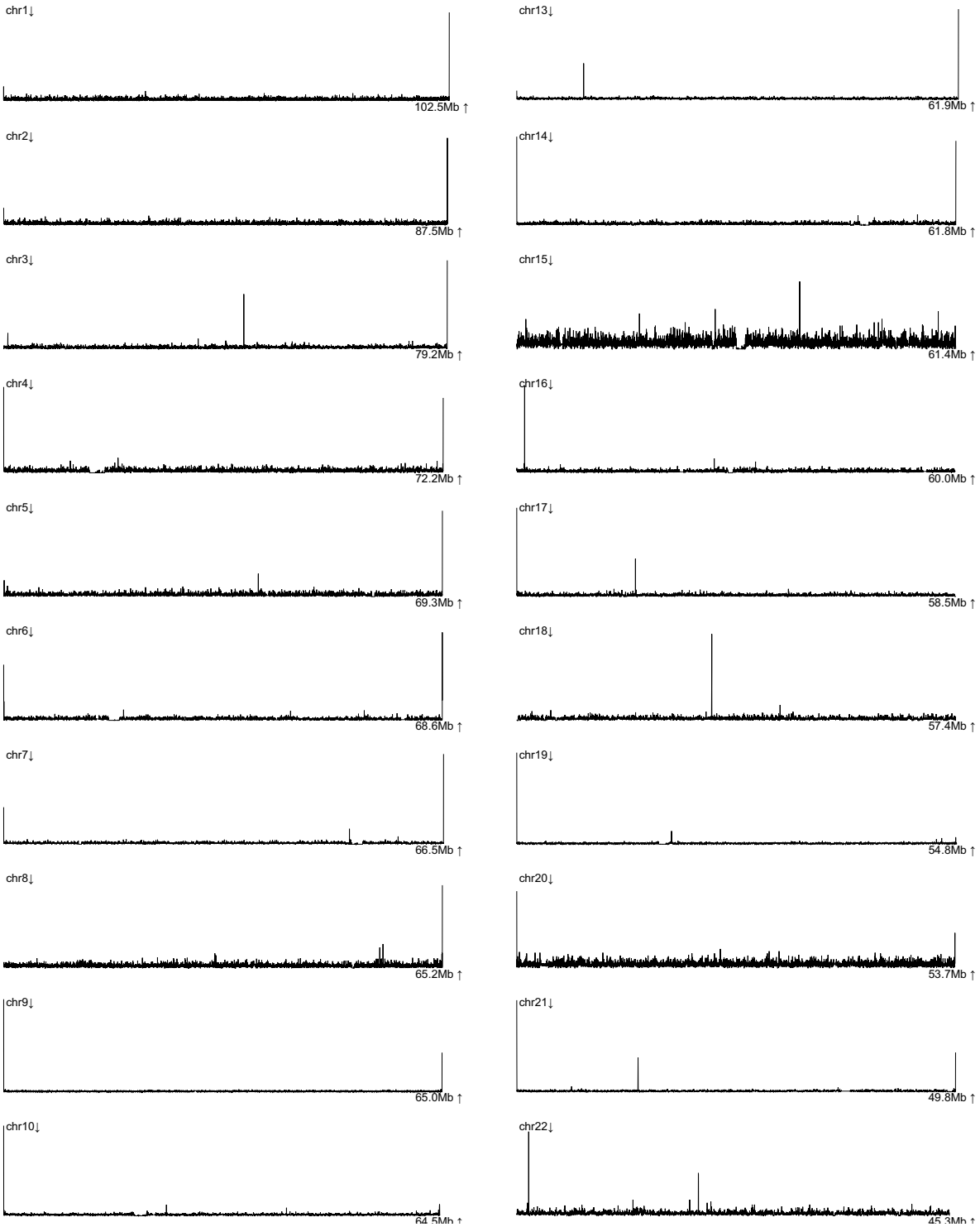

Fig. 5 Telomeric repeat motif AACCCT distribution discovered with Tidk.

parameters, which together with k-mer amount of the assembly was processed using Merqury for QV value. This calculation determined the k-mer completeness to be 72.0596% and QV to be 60.4087.

BUSCO analysis, employing 65 genomes consisting of 954 conserved genes from the data collection metazoa_odb10, was also applied to estimate genomic coverage (Table 11). A total of 97.8% BUSCOs were found in the genome, of which only 3.1% were fragmented. This high complete BUSCOs rate of 94.7% implied that the genome was well-assembled with minimal gaps. It's slightly lower than the 96.1% complete BUSCOs rate of the annotated gene set, since BUSCO analysis of the genome is based on de novo prediction of genes that may involve skipping of weakly predicted loci, while BUSCO analysis of the gene set directly evaluates already predicted gene models. Thus, the evidence confirms a structurally reliable genome suitable for future studies on *H. fuscocinerea* population genetics, resource management, and artificial breeding.

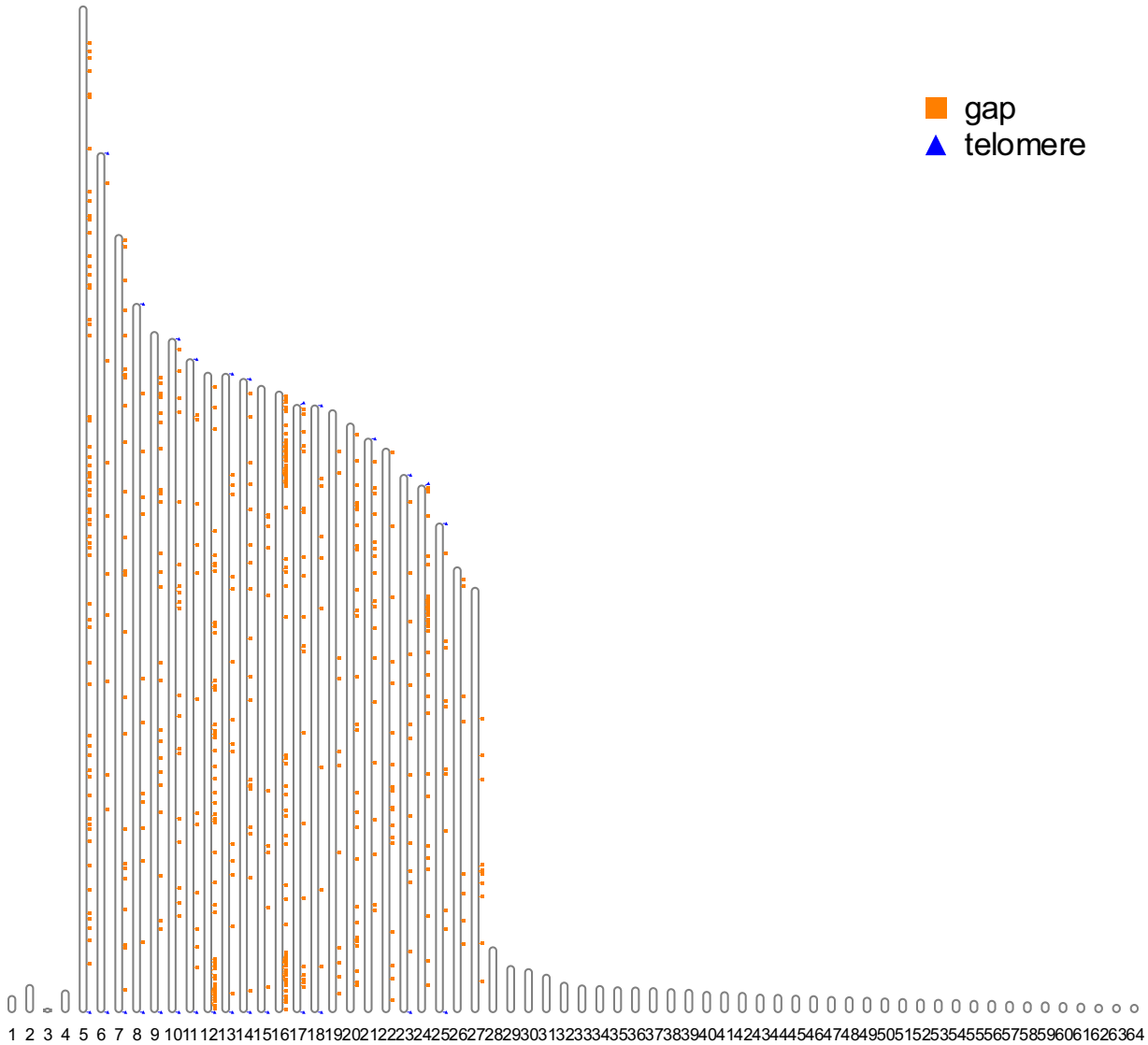

Fig. 6 Telomeric repeat chromosomal distribution identified with QuarTeT.

Feature	*H. fuscocinerea*	
	Orthologs	**Percent (%)**
Complete BUSCOs (C)	903	94.7
Complete and single-copy BUSCOs (S)	862	90.4
Complete and duplicated BUSCOs (D)	41	4.3
Fragmented BUSCOs (F)	30	3.1
Missing BUSCOs (M)	21	2.2
Total BUSCO groups searched	954	100.0

Table 11. Detailed results obtained from the BUSCO analysis of the assembled genome.

Data availability

The following sites store all the data generated in this study available for downloading. The BioProject PRJNA1250249[73] of NCBI has been created to store raw data acquired in various sequencings and the final *H. fuscocinerea* genome assembled into 23 chromosomes. Additionally, the accession number JBNFMV000000000[74] of DDBJ/ENA/GenBank store the genome assembly. The Figshare database[75] hosts all data generated from genome assembly, gene annotation and functional annotations at the link: https://figshare.com/s/6c5acac389a19d8c8386.

Code availability

The following packages, under their established procedures and guidelines, were applied to compute and analyze the data in this study. The workflows and scripts were not altered unless otherwise specified. No custom code was used in this study for data processing, curation, or validation. All analyses were performed using publicly available tools as described in the Methods section. The detailed list of parameters of each software is as follows:

SOAPnuke (v2.1.4): Applied to remove low-quality reads from MGI sequencing raw reads using built-in configurations.
SMRT Link (v13.1): Applied to transform and filter PacBio sequencing raw data using built-in configurations.
Jellyfish (v2.3.0): Applied to count K-mers for assessment of genome size and heterozygosity with a K-mer size of 21.
GenomeScope (v2.0.0): Applied to process the K-mer histogram for the assessment of genome size, heterozygosity and repetitive component using built-in configurations.
Hifiasm (v0.19.8-r603): Applied to assemble the PacBio HiFi data after reads comparison and self-correct using built-in configurations.
Bwa (v0.7.17-r1188): Applied to map the MGI short read data onto the draft assembly using built-in configurations.
Pilon (v1.23): Applied to correct residual errors with Bwa alignment result using built-in configurations.
Purge_dups (v1.2.5): Applied to reduce redundant haplotigs and determine heterozygosity for the draft genome under a configuration of -j 80 -s 80.
ALLHiC (v1.1): Applied to assign and orient scaffolds using Hi-C reads into chromosome-level assemblies.
Merqury (v1.3): Applied to assess k-mer coverage and QV value for the qualification of the assembled genome using best-fit K-mer = 19.
BUSCO (v5.7.1): Applied to estimate genomic coverage using the metazoa_odb10 data collection.
Circos (v0.69): Applied to display chromosomal structure and visualize the distribution of gene regions, repeat sequences, SNP percentage and NGS sequencing depth.
RepeatMasker (v4.09): Applied to annotate transposable elements using built-in configurations.
EDTA: Applied to annotate de-novo transposable elements using built-in configurations.
Barrnap (v0.9): Applied to identify ribosomal RNAs (rRNAs) using built-in configurations.
tRNAscan-SE (v2.0.11): Applied to search for transfer RNAs (tRNAs) sing built-in configurations.
Infernal (v1.1.4): Applied to identify microRNAs (miRNAs) and small nuclear RNAs (snRNAs) using built-in configurations.
Braker (v3.0.8): Applied to integrate gene prediction results with 9 selected proteomes and RNAseq reads from tissues with parameters set to gff3, threads 48, prot_seq = pep.fasta, bam = bams and UTR = on.
HISAT2 (v2.2.1): Applied to map transcriptomic data for genome annotation using built-in configurations.
StringTie (v2.1.7): Applied to assemble the transcripts for the prediction of gene structures using built-in configurations.
MAKER3 (v3.01.03): Applied to combine outputs from various prediction modes into the final gene collection using built-in configurations.
BLAST (v2.11.0+): Applied for synteny analysis and functional annotation of predicted genes with blastp module set to e-value = 1e^-5.

Received: 20 June 2025; Accepted: 9 January 2026;
Published online: 24 January 2026

References

1. Conand, C. Expansion of global sea cucumber fisheries buoys exports. *Revista de Biología Tropical* **65**, S1–S10–S11–S10 (2017).
2. Miller, A. K. *et al*. Molecular phylogeny of extant Holothuroidea (Echinodermata). *Molecular Phylogenetics and Evolution* **111**, 110–131 (2017).
3. Conand, C. Tropical sea cucumber fisheries: Changes during the last decade. *Mar Pollut Bull* **133**, 590–594 (2018).
4. Purcell S., Conand C., Uthicke S., Byrne M. Ecological Roles of Exploited Sea Cucumbers. In: *Oceanography and Marine Biology An Annual Review V54*) (2016).
5. Hamamoto, K., Poliseno, A., Soliman, T. & Reimer, J. D. Shallow epifaunal sea cucumber densities and their relationship with the benthic community in the Okinawa Islands. *PeerJ* **10**, e14181 (2022).
6. Zixuan, E. *et al*. Applications of Environmental DNA (eDNA) in Monitoring the Endangered Status and Evaluating the Stock Enhancement Effect of Tropical Sea Cucumber Holothuria Scabra. *Marine biotechnology (New York, NY)* **25**, 778–789 (2023).
7. Huang, W. *et al*. Spawning, larval development and juvenile growth of the tropical sea cucumber Holothuria leucospilota. *Aquaculture* **488**, 22–29 (2018).
8. Purcell, S. W. *et al*. *Commercially important sea cucumbers of the world*. *Food & Agriculture Org.*, (2023).
9. Conand C., Gamboa, R. & Purcell, S. Holothuria fuscocinerea. The IUCN Red List of Threatened Species 2013: e.T180494A1638255.) (2013).
10. G P. Holothuria (Stauropora) fuscocinerea Jaeger, 1833..). *World Register of Marine Species* (2018).
11. Dermawan A. Pedoman umun identifikasi dan monitoring Teripang. *Direktorat Konservasi dan Keanekaragaman Hayati Laut Jakarta*, 8-56 (2015).
12. Demeuldre, M. *et al*. Mechanical adaptability of sea cucumber Cuvierian tubules involves a mutable collagenous tissue. *J Exp Biol* **220**, 2108–2119 (2017).
13. Cheng, C. *et al*. Characterization and phylogenetic analysis of the complete mitochondrial genome of a tropical sea cucumber, Holothuria fuscocinerea. *Mitochondrial DNA Part B* **5**, 2677–2678 (2020).
14. Sun, Y. *et al*. The complete mitochondrial genome of Holothuria fuscocinerea (Jaeger,1833). *Mitochondrial DNA B Resour* **5**, 33–34 (2019).
15. Anderson, S. C., Flemming, J. M., Watson, R. & Lotze, H. K. Serial exploitation of global sea cucumber fisheries. *Fish and Fisheries* **12**, 317–339 (2011).

16. Zhang, S.-Y., Yi, Y.-H. & Tang, H.-F. Bioactive triterpene glycosides from the sea cucumber Holothuria fuscocinerea. *Journal of natural products* **69**, 1492–1495 (2006).
17. Cayabo, G. D. B. & Mabuhay-Omar, J. A. Antibacterial potential of crude extracts from sea cucumber Holothuria fuscoscinerea Jaeger, 1833. *The Palawan Scientist* **8**, 1–1 (2016).
18. Siregar, M. S. A., Bachtiar, E., Nurhayati, A. & Lewaru, M. W. Antioxidant Activity of Gamat (Stichopus variegatus) and Milk Sea Cucumbers (Holothuria fuscocinerea) from the Thousand Islands National Park Waters. *Journal of Aquaculture and Fish Health* **12**, 390–404 (2023).
19. Komala R., Miarsyah M. & Wulaningsih R. Holothuroidea as a constituent of benthic communities in the seagrass ecosystems at Bira Island Islands. In: *IOP Conference Series: Earth and Environmental Science*). IOP Publishing (2019).
20. Arriesgado, E. *et al*. Diversity and abundance of sea cucumbers in selected areas of Mindanao, Philippines. *Philippine Journal of Science* **151**, 863–877 (2022).
21. Patantis, G., Dewi, A. S., Fawzya, Y. N. & Nursid, M. Identification of Beche-de-mers from Indonesia by molecular approach. *Biodiversitas Journal of Biological Diversity* **20**, 537–543 (2019).
22. Bilan, M. I. *et al*. The Structure of Sulfated Polysaccharides from the Sea Cucumber Holothuria (Stauropora) fuscocinerea. *Russian Journal of Bioorganic Chemistry* **49**, 758–767 (2023).
23. Morton, B. Aspects of predation by Tonna zonatum (Prosobranchia: Tonnoidea) feeding on holothurians in Hong Kong. *Journal of molluscan studies* **57**, 11–19 (1991).
24. López-Pérez, A., Pérez-López, J., Granja-Fernández, R., Calva-Benítez, L.-G. & Maya-Alvarado, B. Feeding habits of Holothuria (Stauropora) fuscocinerea (Echinodermata: Holothuroidea) in a Mexican Pacific reef. *Revista de Biología Tropical* **69**, 66–79 (2021).
25. Bondoc, K. G. V., Lee, H., Cruz, L. J., Lebrilla, C. B. & Juinio-Meñez, M. A. Chemical fingerprinting and phylogenetic mapping of saponin congeners from three tropical holothurian sea cucumbers. *Comparative Biochemistry and Physiology Part B: Biochemistry and Molecular Biology* **166**, 182–193 (2013).
26. Pangestuti R. *et al*. Free Radical Scavenging Activity of Selected Sea Cucumber Species from Mataram-Lombok, Indonesia. *Jurnal Teknologi* **78** (2016).
27. Sun, L., Jiang, C., Su, F., Cui, W. & Yang, H. Chromosome-level genome assembly of the sea cucumber Apostichopus japonicus. *Scientific Data* **10**, 454 (2023).
28. 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).
29. Zhong, S. *et al*. Chromosomal-level genome assembly and annotation of the tropical sea cucumber Holothuria scabra. *Scientific Data* **11**, 474 (2024).
30. Chen, T. *et al*. Chromosome-level genome assembly and annotation of the tropical sea cucumber Stichopus monotuberculatus. *Scientific Data* **11**, 1245 (2024).
31. 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. *Frontiers in Marine Science* **8**, 603410 (2021).
32. Borrero-Perez, G. H., Gomez-Zurita, J., Gonzalez-Wanguemert, M., Marcos, C. & Perez-Ruzafa, A. Molecular systematics of the genus Holothuria in the Mediterranean and Northeastern Atlantic and a molecular clock for the diversification of the Holothuriidae (Echinodermata: Holothuroidea). *Mol Phylogenet Evol* **57**, 899–906 (2010).
33. Andrews S. FastQC: a quality control tool for high throughput sequence data. 2010) (2017).
34. Chen, Y. *et al*. SOAPnuke: a MapReduce acceleration-supported software for integrated quality control and preprocessing of high-throughput sequencing data. *Gigascience* **7**, gix120 (2018).
35. Wenger, A. M. *et al*. Accurate circular consensus long-read sequencing improves variant detection and assembly of a human genome. *Nature biotechnology* **37**, 1155–1162 (2019).
36. Marçais, G. & Kingsford, C. A fast, lock-free approach for efficient parallel counting of occurrences of k-mers. *Bioinformatics* **27**, 764–770 (2011).
37. Vurture, G. W. *et al*. GenomeScope: fast reference-free genome profiling from short reads. *Bioinformatics* **33**, 2202–2204 (2017).
38. Cheng, H., Concepcion, G. T., Feng, X., Zhang, H. & Li, H. Haplotype-resolved de novo assembly using phased assembly graphs with hifiasm. *Nature methods* **18**, 170–175 (2021).
39. Li, H. & Durbin, R. Fast and accurate short read alignment with Burrows–Wheeler transform. *bioinformatics* **25**, 1754–1760 (2009).
40. Walker, B. J. *et al*. Pilon: an integrated tool for comprehensive microbial variant detection and genome assembly improvement. *PloS one* **9**, e112963 (2014).
41. Guan, D. *et al*. Identifying and removing haplotypic duplication in primary genome assemblies. *Bioinformatics* **36**, 2896–2898 (2020).
42. Krzywinski, M. *et al*. Circos: an information aesthetic for comparative genomics. *Genome research* **19**, 1639–1645 (2009).
43. Uliano-Silva, M. *et al*. MitoHiFi: a python pipeline for mitochondrial genome assembly from PacBio high fidelity reads. *BMC bioinformatics* **24**, 288 (2023).
44. Ou, S. *et al*. Benchmarking transposable element annotation methods for creation of a streamlined, comprehensive pipeline. *Genome biology* **20**, 1–18 (2019).
45. Ellinghaus, D., Kurtz, S. & Willhoeft, U. LTRharvest, an efficient and flexible software for de novo detection of LTR retrotransposons. *BMC bioinformatics* **9**, 18 (2008).
46. Ou, S. & Jiang, N. LTR_retriever: a highly accurate and sensitive program for identification of long terminal repeat retrotransposons. *Plant physiology* **176**, 1410–1422 (2018).
47. Su, W., Gu, X. & Peterson, T. TIR-Learner, a new ensemble method for TIR transposable element annotation, provides evidence for abundant new transposable elements in the maize genome. *Molecular plant* **12**, 447–460 (2019).
48. Xiong, W., He, L., Lai, J., Dooner, H. K. & Du, C. HelitronScanner uncovers a large overlooked cache of Helitron transposons in many plant genomes. *Proceedings of the National Academy of Sciences* **111**, 10263–10268 (2014).
49. Shi, J. & Liang, C. Generic repeat finder: a high-sensitivity tool for genome-wide de novo repeat detection. *Plant physiology* **180**, 1803–1815 (2019).
50. Smit A. F. A. HRGP. RepeatModeler Open-1.0. 2008–2015. (ed Biology IfS) (2018).
51. Aylward, F. Introduction to Prokaryotic gene prediction (CDS and rRNA) V. 2. *BMC Bioinformatics* **11**, 1 (2010).
52. Lowe, T. M. & Eddy, S. R. tRNAscan-SE: a program for improved detection of transfer RNA genes in genomic sequence. *Nucleic acids research* **25**, 955–964 (1997).
53. Kalvari, I. *et al*. Rfam 13.0: shifting to a genome-centric resource for non-coding RNA families. *Nucleic acids research* **46**, D335–D342 (2018).
54. Nawrocki, E. P. & Eddy, S. R. Infernal 1.1: 100-fold faster RNA homology searches. *Bioinformatics* **29**, 2933–2935 (2013).
55. Cantarel, B. L. *et al*. MAKER: an easy-to-use annotation pipeline designed for emerging model organism genomes. *Genome research* **18**, 188–196 (2008).
56. Kim, D., Paggi, J. M., Park, C., Bennett, C. & Salzberg, S. L. Graph-based genome alignment and genotyping with HISAT2 and HISAT-genotype. *Nature biotechnology* **37**, 907–915 (2019).
57. Pertea, M. *et al*. StringTie enables improved reconstruction of a transcriptome from RNA-seq reads. *Nature biotechnology* **33**, 290–295 (2015).

58. Warner, J. F. *et al*. Chromosomal-level genome assembly of the painted sea urchin Lytechinus pictus: a genetically enabled model system for cell biology and embryonic development. *Genome Biology and Evolution* **13**, evab061 (2021).
59. Telmer, C. A. *et al*. Echinobase: a resource to support the echinoderm research community. *Genetics* **227**, iyae002 (2024).
60. Hall, M. R. *et al*. The crown-of-thorns starfish genome as a guide for biocontrol of this coral reef pest. *Nature* **544**, 231–234 (2017).
61. Davidson, P. L. *et al*. Chromosomal-level genome assembly of the sea urchin Lytechinus variegatus substantially improves functional genomic analyses. *Genome Biology and Evolution* **12**, 1080–1086 (2020).
62. Sodergren, E. *et al*. The genome of the sea urchin Strongylocentrotus purpuratus. *Science* **314**, 941–952 (2006).
63. Brůna, T., Hoff, K. J., Lomsadze, A., Stanke, M. & Borodovsky, M. BRAKER2: automatic eukaryotic genome annotation with GeneMark-EP+ and AUGUSTUS supported by a protein database. *NAR genomics and bioinformatics* **3**, lqaa108 (2021).
64. Nachtweide S., Stanke M. Multi-genome annotation with AUGUSTUS. In: *Gene prediction: methods and protocols*). Springer (2019).
65. Stover, N. A. & Cavalcanti, A. R. Using NCBI BLAST. *Current Protocols Essential Laboratory Techniques* **14**, 11.11. 11–11.11. 34 (2017).
66. Boutet E. *et al*. UniProtKB/Swiss-Prot, the manually annotated section of the UniProt KnowledgeBase: how to use the entry view. *Plant bioinformatics: methods and protocols*, 23-54 (2016).
67. UniProt Consortium, T. UniProt: the universal protein knowledgebase. *Nucleic acids research* **46**, 2699–2699 (2018).
68. Kanehisa, M., Sato, Y. & Morishima, K. BlastKOALA and GhostKOALA: KEGG tools for functional characterization of genome and metagenome sequences. *Journal of molecular biology* **428**, 726–731 (2016).
69. Mitchell, A. L. *et al*. InterPro in 2019: improving coverage, classification and access to protein sequence annotations. *Nucleic acids research* **47**, D351–D360 (2019).
70. Consortium GO. The gene ontology resource: 20 years and still GOing strong. *Nucleic acids research* **47**, D330–D338 (2019).
71. Brown, M. R., Manuel Gonzalez de La Rosa, P. & Blaxter, M. Tidk: a toolkit to rapidly identify telomeric repeats from genomic datasets. *Bioinformatics* **41**, btaf049 (2025).
72. Lin, Y. *et al*. quarTeT: a telomere-to-telomere toolkit for gap-free genome assembly and centromeric repeat identification. *Horticulture research* **10**, uhad127 (2023).
73. *NCBI Sequence Read Archive* https://identifiers.org/ncbi/insdc.sra:SRP579817 (2025).
74. T C. Holothuria fuscocinerea isolate:south_sea_zhhs | cultivar:nanhai Genome sequencing and assembly). GenBank. https://identifiers.org/ncbi/insdc:JBNFMV000000000 (2025).
75. T C. Chromosome-level genome assembly and annotation of the tropical sea cucumber Holothuria fuscocinerea). *Figshare*. https://doi.org/10.6084/m9.figshare.28831169 (2025).
76. Rhie, A., Walenz, B. P., Koren, S. & Phillippy, A. M. Merqury: reference-free quality, completeness, and phasing assessment for genome assemblies. *Genome biology* **21**, 1–27 (2020).
77. Durand, N. C. *et al*. Juicebox provides a visualization system for Hi-C contact maps with unlimited zoom. *Cell systems* **3**, 99–101 (2016).

Acknowledgements

This study was graciously supported by grants from the National Natural Science Foundation of China (42176132 and 32573487 to T.C., and W2512089 to L.S. and A.Y.), the Research on breeding technology of candidate species for Guangdong modern marine ranching (2024-MRB-00-001 to T.C.), the Science and Technology Program of Nansha District (NSJL202103 to C.H.),, the National Key R & D Program of China (2022YFD2401301 to C.H.), the Guangdong Province Project (2024A1515011418 to T.C., 2024A1515010899 to X.J.).

Author contributions

Chunhua Ren, Chaoqun Hu, Aifen Yan, and Ting Chen planned and conceptualized the research. Xuan Wang, Qianying Huang, Zhou Qin, Wenjie Pan, Jiasheng Huang, Zepeng Zhang, Hua Ge, Jingxuan Liang, Jianfeng Xu, and Yi Zhang acquired and processed the samples. Zhou Qin and Dingding Fan constructed the genome and performed annotations. Xuan Wang, Qianying Huang, Zhou Qin and Dingding Fan analysed gene functions. Xuan Wang, Qianying Huang, Zhou Qin, Dingding Fan and Ting Chen conducted bioinformatic analyses. Chunhua Ren, Peng Luo, Xiao Jiang, Lina Sun, Hongyan Sun, Chaoqun Hu, Aifen Yan, and Ting Chen offered experimental materials and computational resources. Xuan Wang, Qianying Huang, Dingding Fan, Aifen Yan and Ting Chen composed the manuscript. Chunhua Ren, Chaoqun Hu, Aifen Yan, and Ting Chen carried out revisions. All authors have reviewed and consented to the final version of the manuscript.

Competing interests

The authors declare no competing interests.

Additional information

Correspondence and requests for materials should be addressed to A.Y. or T.C.

Reprints and permissions information is available at www.nature.com/reprints.

Publisher's note Springer Nature remains neutral with regard to jurisdictional claims in published maps and institutional affiliations.