scientific data

Check for updates

OPEN

DATA DESCRIPTOR

A chromosomal-level genome assembly of the tropical sea cucumber *Stichopus fusiformiossa*

Nyok-Sean Lau[1], Noorizan Miswan[1], Nurlina Rosli[2], Sixu Gong[1], Izzat Irfan[3], Mathan Ganasan[3], Sau Pinn Woo[3] & Annette Jaya-Ram[3] ✉

The genus *Stichopus* comprises sea cucumber species that are ecologically and economically valuable, known for their roles in nutrient cycling and as sources of bioactive compounds in traditional medicine. Genomic insights are essential for understanding the molecular pathways and biological functions underlying their adaptation, resilience, and ecological significance. We present a chromosome-level genome assembly for the tropical sea cucumber *Stichopus fusiformiossa*, constructed using Nanopore long-read and Hi-C sequencing. The genome assembly spans 766.55 Mb, with a scaffold N50 of 31.80 Mb, and is anchored to 23 pseudo-chromosomes. Genome completeness analysis with BUSCO revealed 98.1% of conserved metazoan genes, reflecting the assembly's quality. Nanopore Iso-Seq analysis of muscle, intestine, and respiratory tree samples generated 10,419-35,371 non-redundant full-length isoforms, which were integrated into genome annotation, enabling the annotation of 27,991 protein-coding genes in the genome. This high-quality genome assembly and Iso-Seq transcriptome data provide a valuable resource for advancing genetic and functional studies in *S. fusiformiossa*, supporting future research into its unique biological traits and potential applications.

Background & Summary

Sea cucumbers, belonging to the class Holothuroidea, are marine invertebrates within the phylum Echinodermata. Unlike other echinoderms such as sea urchins and sea stars, holothurians exhibit a worm-like body shape and lack the classic five-pointed radial symmetry[1]. Their skeletons are composed of microscopic calcified ossicles embedded in their tissues, and they inhabit diverse marine environments, from shallow littoral zones to the deep sea[2,3]. Ecologically, sea cucumbers play vital roles in marine ecosystems, including bioturbation, nutrient cycling, and organic matter processing[4]. For example, *Holothuria atra* bioturbates an estimated 64,000 metric tonnes of sediment annually in the Heron Island Reef, underscoring their ecological significance[5].

Nearly 2,000 species of sea cucumbers have been recorded, with the Asia-Pacific region harbouring the most diverse populations[2,6]. However, many species face overexploitation due to high commercial demand, particularly in Asian markets where they are valued for their nutritional and medicinal properties[7]. Given the increasing pressure on wild populations, genomic and transcriptomic studies are crucial for species conservation, aquaculture development, and understanding their adaptive traits. Previous genomic studies have uncovered significant genetic diversity among sea cucumber species, reflecting their adaptations to a wide range of marine environments[8–13]. For instance, *Apostichopus japonicus* possesses unique gene families associated with intestinal regeneration, a trait of interest for biomedical research[13]. Similarly, *Chiridota heheva*, a species inhabiting deep-sea hydrothermal vents, has evolved specialised iron homeostasis mechanisms, including unique ion-binding and transport genes, to survive in extreme conditions[12]. These findings highlight the potential of sea cucumber genomes to reveal insights into both evolutionary adaptations and practical applications.

The genus *Stichopus* represents an ecologically and economically significant group of tropical sea cucumbers[14,15]. Among its recently described members is *Stichopus fusiformiossa*, identified by Woo *et al.* (2015) from sandy substrates in shallow reef waters near Pulau Songsong, Kedah, Malaysia, within the Straits of Malacca[16]. The species name is derived from the Latin *fusiformis* (fusiform) and *ossa* (bone), reflecting its distinctive fusiform-like ossicles in the tentacles—a trait unique within its genus. Morphologically similar to *Stichopus*

[1]Centre for Chemical Biology, Universiti Sains Malaysia, 11900, Bayan Lepas, Penang, Malaysia. [2]School of Biological Sciences, Universiti Sains Malaysia, 11800, USM, Penang, Malaysia. [3]Centre for Marine and Coastal Studies, Universiti Sains Malaysia, 11800, USM, Penang, Malaysia. ✉e-mail: annettejr@usm.my

Library	No. of reads	Total bases (Gb)	Average read length (bp)	Sequence coverage (X)
Nanopore	94,314,232	50.28	2,398	65.59
Hi-C	532,616,718	79.89	150	104.22

Table 1. Sequencing statistics for Nanopore and Hi-C reads.

Feature	Flye	Canu	NextDenovo	Wtdbg2
Assembly size (Mb)	772.69	875.89	611.86	580.97
No. of contigs	21,516	23,686	5,776	10,458
Contigs N50 (Kb)	245.70	100.03	144.93	111.26
Longest contig (Kb)	1,669.75	1,219.99	2,521.46	815.56
GC content (%)	37.73	37.75	37.76	37.74

Table 2. Comparison of draft genome assemblies for *S. fusiformiossa* generated using different software.

horrens, *S. fusiformiossa* exhibits a brownish-grey body interspersed with irregular black and white papillae and can reach up to 30 cm in length[16]. *S. fusiformiossa* have been harvested from the wild for the production of pharmaceutical and traditional medicine industry in Malaysia. Growing demand and overharvesting have driven efforts in breeding and aquaculture development for *Stichopus* species, including attempts to induce spawning of *S. fusiformiossa* under controlled conditions[17]. Production efforts are critical to ensure the conservation of species in the natural environment and the sustainability of the pharmaceutical industry which relies on it. Beyond aquaculture, research on this genus has expanded to taxonomy, population genetics, and gut microbiota, providing insights essential for sustainable management and conservation. While genomic, mitochondrial, transcriptomic, and proteomic data are available for *Stichopus* species such as *Stichopus monotuberculatus, Stichopus horrens*, and *Stichopus ocellatus*[9,18–20], genetic resources for *S. fusiformiossa* remain scarce.

This study presents a chromosome-level genome assembly for *S. fusiformiossa*, generated using Nanopore long-read and Hi-C sequencing data. This resource, complemented by full-length isoform sequencing (Iso-Seq) of muscle, intestine, and respiratory tree tissues, provides a foundation for understanding the genetic basis of *S. fusiformiossa*'s unique traits and adaptations. The genome assembly and annotation will support future research on sea cucumber biology, conservation, and sustainable aquaculture practices.

Methods

Sampling and genome sequencing. *S. fusiformiossa* were maintained at the Centre for Marine and Coastal Studies, Universiti Sains Malaysia. Specimens were housed in a 500 L tank with continuous aeration at a temperature of 28–29 °C[17]. Healthy adult specimens were selected for genome and isoform sequencing. Specimens were dissected and tissues were stored in RNAlater (Thermo Fisher Scientific, USA) at −20 °C. High molecular weight genomic DNA was extracted from muscle tissue using the Qiagen Genomic-tip 20/G kit (Qiagen, Germany) following the manufacturer's protocol. DNA quality and quantity were assessed by 1% agarose gel electrophoresis and Qubit fluorometry (Invitrogen, USA). For Nanopore library construction, DNA fragments larger than 20 kb were first size selected using the BluePippin system (Sage Science, USA). Subsequently, the library was prepared with the Ligation Sequencing Kit (SQK-LSK109, Oxford Nanopore Technologies, UK) and sequenced on a flow cell using the Nanopore PromethION sequencer. This sequencing yielded a total of 50.28 Gb of data, providing approximately 65.59-fold coverage of the genome (Table 1). A high-throughput chromatin conformation capture (Hi-C) library was constructed following standard protocol[20]: briefly, the homogenised muscle tissue was crosslinked with formaldehyde, digested with the *Dpn II* restriction enzyme, and the 5′ overhangs were biotinylated. The DNA was then blunt-end ligated and sheared into fragments of 300–700 bp in size. The final Hi-C library was sequenced on the Illumina NovaSeq 6000 platform with 150-bp paired-end reads.

Genome assembly. Nanopore long-read sequencing data were used to generate a contig-level assembly of the *S. fusiformiossa* genome. The *k*-mer distribution of the error-corrected Nanopore reads was analysed using Jellyfish (v2.3.1) to estimate the genome size. The genome size of *S. fusiformiossa* was estimated using GenomeScope (v2.0) based on 21-mer at 674.63 Mb with a heterozygosity rate of 0.80%. Preliminary assemblies were constructed with Flye (v2.9.5)[21], Canu (v2.3)[22], NextDenovo (v2.5.2)[23], and wtdbg (v2.5)[24]. Among these, the Flye assembly was selected for further analysis due to its best assembly metrics (Table 2, Fig. 1). The Flye-generated assembly comprised 21,516 contigs, with a total assembly size of 772.69 Mb, a contig N50 length of 245.70 Kb, and coverage of 92.1% of metazoan genes based on Benchmarking Universal Single-Copy Orthologs (BUSCO) (v5.8.2)[25] analysis. To improve base-level accuracy, the genome assembly was polished using Nanopore reads with Medaka (v2.0.1). Heterozygous redundant sequences in the assembly were removed using Purge_haplotigs (v1.1.3)[26]. A total of 79.89 Gb of Hi-C reads were sequenced and mapped to the contig-level assembly using the Juicer pipeline (v1.6)[27] with BWA (v0.7.18). The mapping resulted in 59.21% of unique alignments, of which 35.54% were valid contacts (Table 3). The contig-level assembly was then scaffolded and organised into pseudo-chromosomes using the 3d-dna pipeline (v201008)[28] (Fig. 2a,b). More than 91% of the genome assembly was anchored to the 23 largest scaffolds, which likely represent the 23 chromosomes of the *S. fusiformiossa* genome (Table 4). The final genome assembly was 766.55 Mb in size, containing 8,723 scaffolds with N50 of 31.80 Mb (Table 5). The Hi-C scaffolding assembled 23 pseudo-chromosomes in *S. fusiformiossa*, consistent with

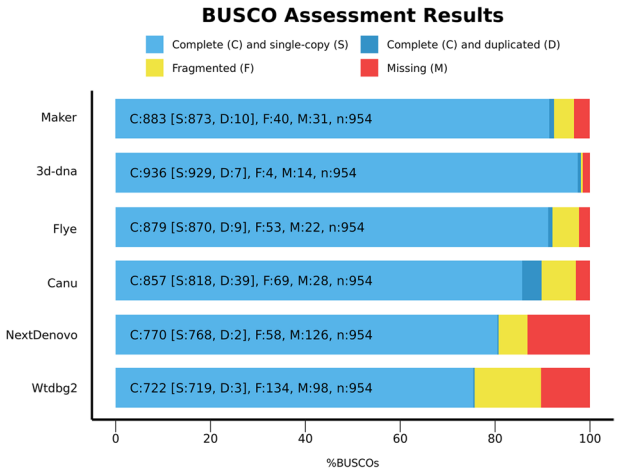

Fig. 1 Assessment of genome or gene model completeness using BUSCO software. Results shown are for gene models annotated with Maker, the final genome assembly generated using 3d-dna, and draft genome assemblies produced with Canu, NextDenovo, and Wtdbg2.

Type	Number	Ratio (%)
Sequenced read pairs	266,308,359	100
Mapped read pairs	247,647,311	92.99
Normal paired	206,022,162	77.36
Unique reads	157,683,681	59.21
Valid interaction pairs	94,654,880	35.54

Table 3. Hi-C read mapping for the *S. fusiformiossa* genome.

the chromosome number reported for another *Stichopus* species, *S. monotuberculatus*[1]. Further synteny analysis using MCScanX (v1.0.0) revealed conserved macrosynteny between the two species, particularly between chromosome pairs such as Sf3–Sm4, Sf4–Sm5, Sf5–Sm8, Sf6–Sm9, and Sf7–Sm10 (Fig. 3a).

RNA sequencing and analysis. Total RNA was extracted from muscle, intestine and respiratory tree samples using the RNeasy Mini Kit (Qiagen) according to the manufacturer's instructions. RNA concentration was quantified with a Qubit fluorometer, and RNA integrity was assessed using an Agilent 2100 Bioanalyzer (Agilent Technologies, USA). Full-length cDNA libraries were prepared from the extracted RNA using the cDNA-PCR Sequencing Kit (SQK-PCS109, Oxford Nanopore Technologies), and the cDNA was PCR amplified with specific barcoded adapters from the PCR Barcoding Kit (SQKPBK004, Oxford Nanopore Technologies). The prepared libraries were sequenced on a PromethION platform. Raw reads generated by Nanopore sequencing were pre-processed to remove reads with a quality score below 10. The total size of clean reads obtained ranged from 10.41 to 17.08 Gb across the samples (Table 6). Reads passing the quality filter were processed with Pychopper (v2.4.0) using parameters '-Q7 -z 50'. Pychopper identified full-length cDNA sequences while removing the adapter, barcode, and primer sequences, resulting in 10,773,413, 11,051,406, and 23,462,542 full-length reads for intestine, respiratory tree, and muscle samples, respectively. The full-length sequences were further filtered using NanoFilt (v2.8.0) to exclude sequences with a quality score below 7 or a length shorter than 150 bp. The filtered full-length sequences were then analysed with Pinfish (v0.1.0) to construct a non-redundant transcript set. In Pinfish, the sequences were first mapped to the reference genome using Minimap2 (v2.17-r941), followed by clustering, redundancy removal, and polishing steps to produce a final set of consensus sequences. The number of non-redundant full-length isoforms ranged from 10,419 for the respiratory tree sample to 35,371 for the muscle sample, with muscle showing the highest transcript diversity.

Genome annotation. The MAKER (v3.01.03)[29] annotation pipeline, which integrates *ab initio* gene prediction, homology-based prediction, and transcript prediction, was used to predict consensus protein-coding genes for the *S. fusiformiossa* genome. The *ab initio* gene prediction was performed using Augustus (v3.2.2)[30], Snap (v2013-11-29)[31], and GeneMark (v3.61)[32]. For the homology-based prediction, protein sequences from closely related species including *S. monotuberculatus*, *Apostichopus japonicus*, and *Holothuria leucospilota* were used as protein support. For transcript prediction, Iso-Seq transcripts from muscle, intestine and respiratory tree samples were used as expressed sequence tag (EST) evidence. The annotation process resulted in the identification of 27,991 protein-coding genes in the *S. fusiformiossa* genome. The predicted genes had an average transcript length of 11,134 bp and an average coding sequence length of 1,429 bp. Functional annotation of the predicted protein-coding genes was performed using Diamond (v2.1.9) against the NCBI-NR, Swiss-Prot and TrEMBL

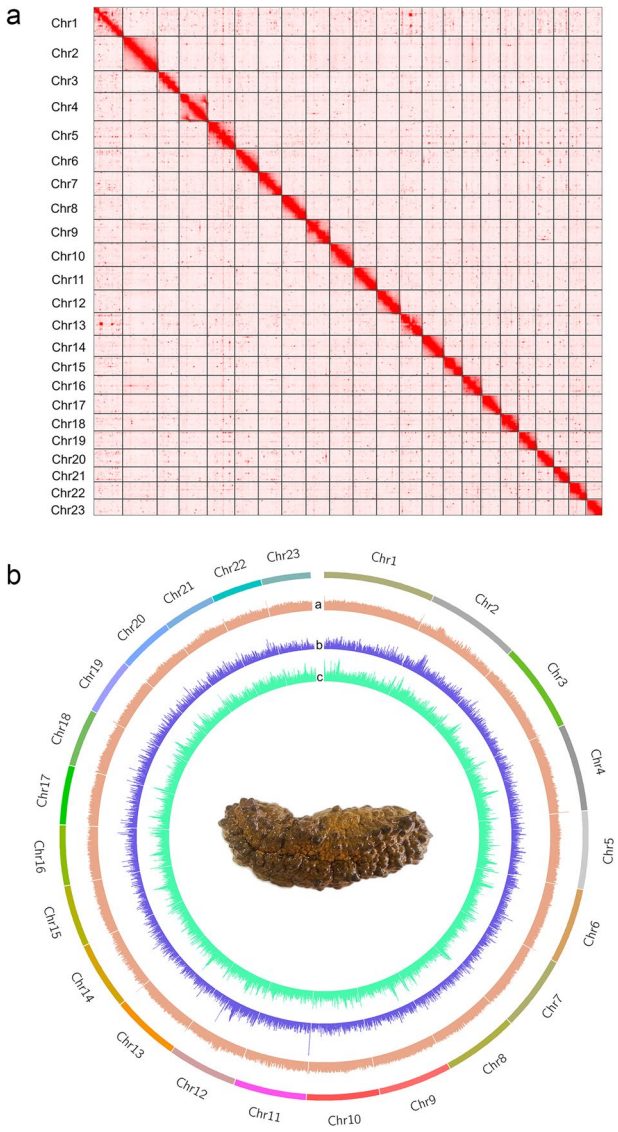

Fig. 2 Genomic organisation of *S. fusiformiossa*. (**a**) Circos plot showing the chromosomal organisation of *S. fusiformiossa*. The outermost ring represents the chromosomes. From outside to inside, the tracks show: GC content (track a), gene density (track b), and repeat density (track c), calculated using a 100 Kb sliding window. (**b**) Hi-C contact map of the *S. fusiformiossa* genome assembly. The Hi-C map was visualised with Juicebox (v2.17.00) at a 100 Kb resolution.

databases at an e-value cutoff of 1e-5. Additionally, gene functions were annotated using eggNOG-mapper (v2.1.12) and BlastKOALA annotation server.

Repeats in the *S. fusiformiossa* genome assembly were annotated using RepeatMasker (v4.1.0)[33] with a combined homology search against the Dfam and Repbase databases, along with a custom *de novo* repeat library. The *de novo* repeat library was generated using the EDTA pipeline (v2.2.2)[34], which incorporates multiple tools including LTRharvest (v1.1), LTR_FINDER (v1.1), LTR_retriever (v3.0.1), GRF (v1.0.2), TIR-Learner (v3.0.5), HelitronScanner (v1.10), and RepeatModeler (v2.0.6). This comprehensive approach identified a total of 195.34 Mb of repeat sequences, accounting for 25.49% of the genome assembly (Table 7). The proportion of repetitive sequences in the *S. fusiformiossa* genome was comparable to *S. monotuberculatus* (25.02%)[1] but lower than other echinoderms such as *A. japonicus* (47.33%)[35] and *Holothuria scabra* (52.31%)[36]. DNA transposable elements were the most abundant class of repeats, constituting 19.11% of the genome, followed by long interspersed nuclear elements (LINEs) at 3.03%, long terminal repeats (LTRs) at 2.92%, and short interspersed nuclear elements (SINEs) at 0.01%. Noncoding RNAs in the *S. fusiformiossa* genome were identified using INFERNAL (v1.1.4)[37] with the Rfam database. This analysis revealed various noncoding RNA types, including 73 microRNAs (miRNAs), 364 transfer RNAs (tRNAs), 71 ribosomal RNAs (rRNAs), 49 spliceosomal RNAs (snRNAs), and 34 small nucleolar RNAs (snoRNAs) (Table 8).

Pseudomolecule	Length (bp)	Repeat length (bp)	Repeat (%)
Chr1	48,857,395	9,747,999	20.0
Chr2	41,251,824	8,779,070	21.3
Chr3	39,124,030	7,692,063	19.7
Chr4	38,366,664	8,272,663	21.6
Chr5	34,109,107	7,164,286	21.0
Chr6	33,304,964	6,644,621	20.0
Chr7	32,618,788	6,753,298	20.7
Chr8	32,583,249	6,283,257	19.3
Chr9	32,113,486	6,528,594	20.3
Chr10	31,835,754	6,276,039	19.7
Chr11	31,798,684	5,918,966	18.6
Chr12	30,232,459	6,386,642	21.1
Chr13	29,800,220	6,038,968	20.3
Chr14	29,645,807	5,581,240	18.8
Chr15	26,936,153	5,452,733	20.2
Chr16	26,285,383	5,355,063	20.4
Chr17	25,863,859	4,814,960	18.6
Chr18	25,253,698	4,898,247	19.4
Chr19	24,644,399	4,978,005	20.2
Chr20	24,107,996	4,687,061	19.4
Chr21	22,839,700	4,242,876	18.6
Chr22	22,151,553	4,493,561	20.3
Chr23	21,457,154	4,008,174	18.7
Unanchored	61,369,400	16,881,183	27.5
Total	766,551,726	195,339,938	25.5

Table 4. Summary of the *S. fusiformiossa* Hi-C assisted genome assembly.

Feature	Value
Assembly size (Mb)	766.55
No. of scaffolds	8,723
Scaffolds N50 (Mb)	31.80
Longest scaffold (Mb)	48.86
GC content (%)	37.64
No. of gene models	27,991
Average transcript length (bp)	11,134
Average coding sequence length (bp)	1,429
Average no. of exons per transcript	7.26
Average exon length (bp)	197
Average intron length (bp)	1511
No. of annotated genes	24,471
No. of genes annotated to NR	24,178
No. of genes annotated to Swiss-Prot	13,986
No. of genes annotated to TremBL	24,217
No. of genes annotated to eggNOG	13,825
No. of genes annotated to KEGG	12,942

Table 5. Summary statistics of the final *S. fusiformiossa* genome assembly and annotation.

Data Records

All the genome sequencing data of Nanopore and Illumina Hi-C, as well as Nanopore Iso-Seq data obtained in this study have been deposited in the NCBI Sequence Read Archive database with the accession number SRR32120561[38], SRX27455368[39], SRR32120645-SRR32120647[40–42]. The final genome assembly had been deposited in the NCBI genome database under the accession number JBLGPS000000000[43]. The genome annotation files including gene CDS and protein data have been deposited in Figshare[44].

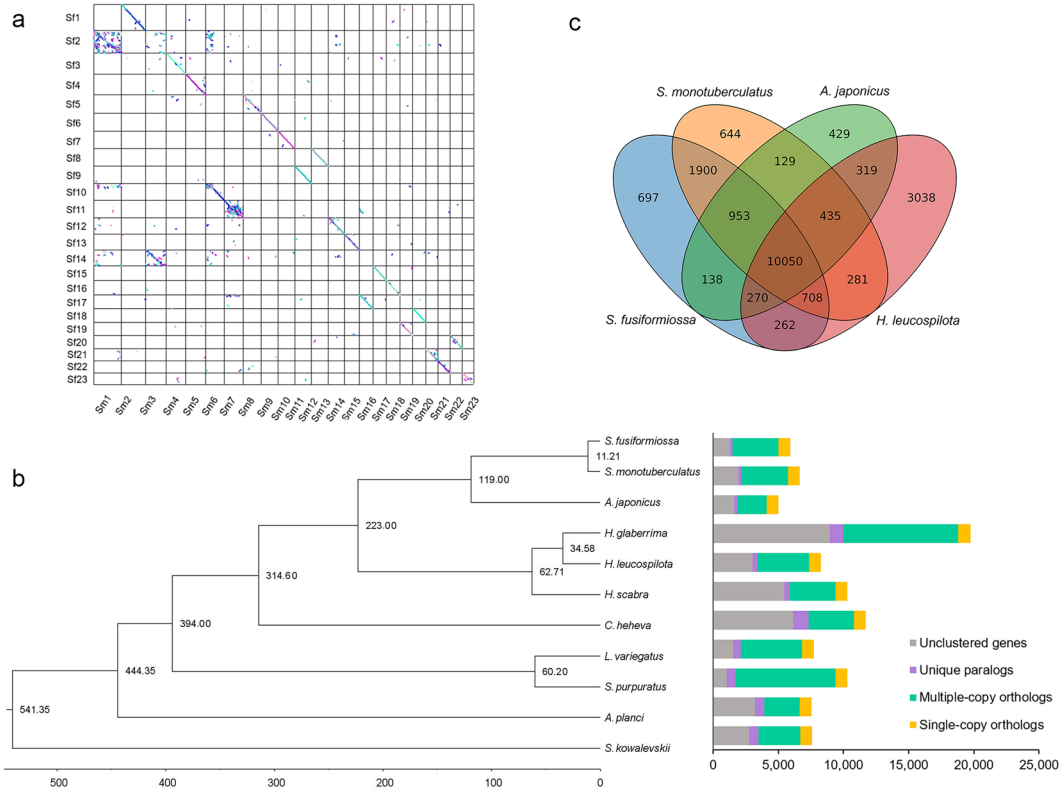

Fig. 3 Comparative genomic analyses of *S. fusiformiossa* and related species. (**a**) Dot plot showing synteny between chromosomes of *S. fusiformiossa* (Sf) and *S. monotuberculatus* (Sm). Each dot represents a collinear gene pair, and diagonal patterns indicate conserved gene order. (**b**) Right: Maximum likelihood phylogenetic tree constructed using single-copy orthologs, inferred with the LG + F + I + R9 model, and 1,000 ultrafast bootstrap replicates. Node labels indicate estimated divergence times (Mya) and the scale bar at the bottom represents divergence times in Mya. Left: Distribution of ortholog and paralog categories across 11 metazoan genomes: unclustered genes (grey), unique paralogs (purple), multiple-copy orthologs (green), and single-copy orthologs (orange). This analysis depicts the evolutionary relationships of genes, classifying them based on duplication and speciation events. (**c**) Venn diagram illustrating the number of shared and unique gene families among four sea cucumber species: *Stichopus monotuberculatus*, *Apostichopus japonicus*, *Stichopus fusiformiossa*, and *Holothuria leucospilota*.

Feature	Muscle	Intestine	Respiratory tree
Total reads	26,244,413	13,736,743	14,222,920
Average read length (bp)	650.77	757.82	737.8
Total read length (Gb)	17.08	10.41	10.49
Full length reads	23,462,542	10,773,413	11,051,406
Average full-length read length (bp)	472.85	566.06	590.06
Total full-length read length (Gb)	11.09	6.10	6.52
Non-redundant isoforms	35,371	25,817	10,419
Average isoform length (bp)	274	379	301

Table 6. Summary of Nanopore Isoform sequencing statistics for *S. fusiformiossa* muscle, intestine and respiratory tree samples.

Technical Validation

Assessment of genome assembly and gene annotation. The completeness of the *S. fusiformiossa* genome assembly was evaluated using BUSCO against the conserved metazoan gene set (metazoan_odb 10) (Fig. 1). The genome assembly was found to have a high level of completeness (98.1%). Of the 954 single-copy orthologs, 97.4% were complete and single-copy, 0.7% were complete and duplicated, 0.4% were fragmented, and 1.5% were missing. To assess the completeness of the *S. fusiformiossa* gene models, BUSCO was also applied to its gene annotation using the same metazoan dataset. This analysis indicated that 92.6% of the conserved single-copy orthologous genes were complete, while 4.2% were fragmented and 3.2% were missing. Nanopore reads

Type	Copy number	Length (bp)	% in genome
DNA transposon	896,469	146,449,920	19.11
Retrotransposon	149,556	45,652,237	5.96
LINE	58,577	23,213,805	3.03
SINE	617	57,806	0.01
LTR	90,362	22,380,626	2.92
Unclassified	27,721	3,237,781	0.42
Total	1,073,746	195,339,938	25.49

Table 7. Repeat content and classification in the *S. fusiformiossa* genome.

Type	Copy number	Total length (bp)
miRNA	73	6,945
tRNA	364	26,829
rRNA	71	17,335
snRNA	49	6,506
snoRNA	34	4,430

Table 8. Summary of non-coding RNA in the *S. fusiformiossa* genome.

were mapped to the genome assembly using Minimap2 (v2.28), with 95.80% of reads aligning to the genome. This mapping yielded 99.82% genome coverage ($\geq 1\times$) and an average sequencing depth of 60.98×. Functional annotation of the *S. fusiformiossa* genes showed that 87.42% of the annotated genes were assigned functional information by at least one database, including NCBI NR, Swiss-Prot, TrEMBL, eggNOG, and KEGG.

Gene family and phylogenomic analyses. Gene family analysis was conducted using OrthoFinder (v2.5.5)[45] to examine the orthologous genes shared between *S. fusiformiossa* and other related species including *Acanthaste planci*[46], *A. japonicus*[47], *Chiridota heheva*[48], *Lytechinus variegatus*[49], *Holothuria glaberrima*[50], *H. leucospilota*[51], *H. scabra*[52], *S. monotuberculatus*[53], *Strongylocentrotus purpuratus*[54], and *Saccoglossus kowalevskii*[55]. Among the protein-coding genes of *S. fusiformiossa*, a total of 26,666 genes were grouped into orthologous gene clusters, with 484 classified as species-specific genes (Fig. 3b). Additionally, 6,364 orthogroups were shared among the Echinodermata species analysed in this study, including 914 single-copy orthogroups. Multiple sequence alignment of these single-copy orthogroups was performed using MAFFT (v7.520). A phylogenomic tree was constructed using IQ-TREE (v2.4.0)[56] with the best-fit substitution model determined by ModelFinder based on the Bayesian Information Criterion (BIC). The maximum likelihood tree was inferred using the LG + F + I + R9 model with 1,000 ultrafast bootstrap replicates, with *S. kowalevskii* as the outgroup. The divergence times were estimated using the least square dating method[57] implemented in IQ-TREE (since v2.0.3) and we calibrated the model using the divergence times between *A. japonicus* and *S. monotuberculatus* (119 Mya), *A. japonicus* and *H. scabra* (233 Mya), and *A. japonicus* and *S. purpuratus* (394 Mya) obtained from the TimeTree database. The Venn diagram illustrates the shared and unique orthogroups among *S. monotuberculatus*, *A. japonicus*, *S. fusiformiossa*, and *H. leucospilota* (Fig. 3c). *S. fusiformiossa* shares a substantial number of orthogroups with *S. monotuberculatus*, indicating their close evolutionary relationship. The phylogenetic tree further supports this, clustering *S. fusiformiossa* with *S. monotuberculatus*, while *A. japonicus* diverged from their common ancestor approximately 119 million years ago (mya) (Fig. 3b).

Code availability

No custom scripts were used in this study. All data processing commands and pipelines followed the manuals or protocols provided by the respective bioinformatics software. Details of the parameters and settings are described in the Methods section.

Received: 6 March 2025; Accepted: 2 June 2025;
Published online: 10 June 2025

References

1. Stricker, S. A. The fine structure and development of calcified skeletal elements in the body wall of holothurian echinoderms. *J. Morphol.* **188**, 273–288 (1986).
2. Mercier, A., Gebruk, A., Kremenetskaia, A. & Hamel, J.-F. Chapter 1 - An overview of taxonomic and morphological diversity in sea cucumbers (Holothuroidea: Echinodermata). in *The World of Sea Cucumbers* (eds. Mercier, A., Hamel, J.-F., Suhrbier, A. D. & Pearce, C. M.) 3–15 (Academic Press, 2024).
3. Gallo, N. D. Submersible- and lander-observed community patterns in the Mariana and New Britain trenches: Influence of productivity and depth on epibenthic and scavenging communities. *Deep-Sea Res. I: Oceanogr. Res. Pap.* **99**, 119–133 (2015).
4. Purcell, S., Conand, C., Uthicke, S. & Byrne, M. Ecological roles of exploited sea cucumbers. *Oceanography and Marine Biology* **54**, 367–386 (2016).
5. Williamson, J. E., Duce, S., Joyce, K. E. & Raoult, V. Putting sea cucumbers on the map: projected holothurian bioturbation rates on a coral reef scale. *Coral Reefs* **40**, 559–569 (2021).

6. Abdul Kadar, N. B., Rosli, N. S. B. & Muzli, S. S. B. Sea cucumber (Echinodermata: Holothuroidea) species diversity on the West Coast of Sabah, Malaysia. in *Marine Biotechnology: Applications in Food, Drugs and Energy* (eds. Shah, M. D., Ransangan, J. & Venmathi Maran, B. A.) 149–169 (Springer Nature, Singapore, 2023).
7. Bordbar, S., Anwar, F. & Saari, N. High-value components and bioactives from sea cucumbers for functional foods—A Review. *Mar Drugs* **9**, 1761–1805 (2011).
8. Zhong, S. *et al*. The draft genome of the tropical sea cucumber *Stichopus monotuberculatus* (Echinodermata, Stichopodidae) reveals critical genes in fucosylated chondroitin sulfates biosynthetic pathway. *Front. Genet*. **14** (2023).
9. Chen, T. *et al*. Chromosome-level genome assembly and annotation of the tropical sea cucumber *Stichopus monotuberculatus*. *Sci Data* **11**, 1245 (2024).
10. Shao, G. *et al*. The genome of a hadal sea cucumber reveals novel adaptive strategies to deep-sea environments. *iScience* **25** (2022).
11. Chen, T. *et al*. The *Holothuria leucospilota* genome elucidates sacrificial organ expulsion and bioadhesive trap enriched with amyloid-patterned proteins. *Proc. Natl. Acad. Sci. USA*. **120**, e2213512120 (2023).
12. 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).
13. Zhang, X. *et al*. The sea cucumber genome provides insights into morphological evolution and visceral regeneration. *PLOS Biology* **15**, e2003790 (2017).
14. Sulardiono, B., Hartoko, A., Aini, A. N., Wulandari, D. & Budiharjo, A. Genetic diversity of commercial sea cucumbers *Stichopus* (Echinoderm: Stichopodidae) based on DNA barcoding in Karimunjawa, Indonesia. *Biodiversitas* **23** (2022).
15. Kamarudin, K. R. & Rehan, M. M. Morphological and molecular identification of *Holothuria* (Merthensiothuria) *leucospilota* and *Stichopus horrens* from Pangkor Island, Malaysia. *Trop Life Sci Res* **26**, 87–99 (2015).
16. Woo, S. P. Sea cucumbers of the genus *Stichopus* Brandt, 1835 (Holothuroidea, Stichopodidae) in Straits of Malacca with description of a new species. *ZooKeys* **545**, 1–26 (2015).
17. Cheong, K.-M. *et al*. Induced spawning of tropical sea cucumber, *Stichopus fusiformiossa* by neurotransmitters: Challenges and way forward. *Aquac. Rep.* **39**, 102464 (2024).
18. Wei, S. S. *et al*. Gut microbiome and metabolome of sea cucumber (*Stichopus ocellatus*) as putative markers for monitoring the marine sediment pollution in Pahang, Malaysia. *Mar Pollut Bull* **182**, 114022 (2022).
19. Fan, S., Hu, C., Wen, J. & Zhang, L. Characterization of mitochondrial genome of sea cucumber *Stichopus horrens*: A novel gene arrangement in Holothuroidea. *Sci. China Life Sci.* **54**, 434–441 (2011).
20. Belton, J.-M. *et al*. Hi-C: a comprehensive technique to capture the conformation of genomes. *Methods* **58**, 268–276 (2012).
21. Kolmogorov, M., Yuan, J., Lin, Y. & Pevzner, P. A. Assembly of long, error-prone reads using repeat graphs. *Nat. Biotechnol.* **37**, 540–546 (2019).
22. Koren, S. *et al*. Canu: scalable and accurate long-read assembly via adaptive *k* -mer weighting and repeat separation. *Genome Res* **27**, 722–736 (2017).
23. Hu, J. *et al*. NextDenovo: an efficient error correction and accurate assembly tool for noisy long reads. *Genome Biology* **25**, 107 (2024).
24. Ruan, J. & Li, H. Fast and accurate long-read assembly with wtdbg2. *Nat Methods* **17**, 155–158 (2020).
25. 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).
26. Roach, M. J., Schmidt, S. A. & Borneman, A. R. Purge Haplotigs: allelic contig reassignment for third-gen diploid genome assemblies. *BMC Bioinformatics* **19**, 460 (2018).
27. Durand, N. C. *et al*. Juicer provides a one-click system for analyzing loop-resolution Hi-C experiments. *Cell Syst* **3**, 95–98 (2016).
28. Dudchenko, O. *et al*. *De novo* assembly of the *Aedes aegypti* genome using Hi-C yields chromosome-length scaffolds. *Science* **356**, 92–95 (2017).
29. Cantarel, B. L. *et al*. MAKER: an easy-to-use annotation pipeline designed for emerging model organism genomes. *Genome Res* **18**, 188–196 (2008).
30. Stanke, M. *et al*. AUGUSTUS: *ab initio* prediction of alternative transcripts. *Nucleic Acids Res* **34**, W435–439 (2006).
31. Korf, I. Gene finding in novel genomes. *BMC Bioinformatics* **5**, 59 (2004).
32. Borodovsky, M. & Lomsadze, A. Eukaryotic gene prediction using GeneMark.hmm-E and GeneMark-ES. *Curr Protoc Bioinformatics* CHAPTER, Unit-4.610 (2011).
33. Tarailo-Graovac, M. & Chen, N. Using RepeatMasker to identify repetitive elements in genomic sequences. *Curr Protoc Bioinformatics* Chapter 4, 4.10.1–4.10.14 (2009).
34. Ou, S. *et al*. Benchmarking transposable element annotation methods for creation of a streamlined, comprehensive pipeline. *Genome Biology* **20**, 275 (2019).
35. Su, F. *et al*. Understanding gene regulation during the development of the sea cucumber *Apostichopus japonicus* using comparative transcriptomics. *Front. Mar. Sci.* **10** (2023).
36. Zhong, S. *et al*. Chromosomal-level genome assembly and annotation of the tropical sea cucumber *Holothuria scabra*. *Sci Data* **11**, 474 (2024).
37. Nawrocki, E. P. & Eddy, S. R. Infernal 1.1: 100-fold faster RNA homology searches. *Bioinformatics* **29**, 2933–2935 (2013).
38. *NCBI Sequence Read Archive* https://identifiers.org/ncbi/insdc.sra:SRR32120561 (2025).
39. *NCBI Sequence Read Archive* https://identifiers.org/ncbi/insdc.sra:SRX27455368 (2025).
40. *NCBI Sequence Read Archive* https://identifiers.org/ncbi/insdc.sra:SRR32120645 (2025).
41. *NCBI Sequence Read Archive* https://identifiers.org/ncbi/insdc.sra:SRR32120646 (2025).
42. *NCBI Sequence Read Archive* https://identifiers.org/ncbi/insdc.sra:SRR32120647 (2025).
43. *NCBI GenBank* https://identifiers.org/ncbi/insdc:JBLGPS000000000 (2025).
44. Lau, N.-S. A chromosomal-level genome assembly of the tropical sea cucumber *Stichopus fusiformiossa*. *Figshare* https://doi.org/10.6084/m9.figshare.28545557 (2025).
45. Emms, D. M. & Kelly, S. OrthoFinder: phylogenetic orthology inference for comparative genomics. *Genome Biol* **20**, 238 (2019).
46. *NCBI GenBank* https://identifiers.org/ncbi/insdc.gca:GCA_001949145.1 (2016).
47. *NCBI GenBank* https://identifiers.org/ncbi/insdc.gca:GCA_037975245.1 (2024).
48. *NCBI GenBank* https://identifiers.org/ncbi/insdc.gca:GCA_020152595.1 (2021).
49. *NCBI GenBank* https://identifiers.org/ncbi/insdc.gca:GCA_018143015.1 (2021).
50. *NCBI GenBank* https://identifiers.org/ncbi/insdc.gca:GCA_009936505.3 (2024).
51. *NCBI GenBank* https://identifiers.org/ncbi/insdc.gca:GCA_029531755.1 (2023).
52. *NCBI GenBank* https://identifiers.org/ncbi/insdc.gca:GCA_037179385.1 (2024).
53. *NCBI GenBank* https://identifiers.org/ncbi/insdc.gca:GCA_041315955.1 (2024).
54. *NCBI GenBank* https://identifiers.org/ncbi/insdc.gca:GCA_000002235.4 (2019).
55. *NCBI GenBank* https://identifiers.org/ncbi/insdc.gca:GCA_000003605.1 (2009).
56. Minh, B. Q. *et al*. IQ-TREE 2: New models and efficient methods for phylogenetic inference in the genomic era. *Molecular Biology and Evolution* **37**, 1530–1534 (2020).
57. To, T.-H., Jung, M., Lycett, S. & Gascuel, O. Fast dating using least-squares criteria and algorithms. *Syst Biol* **65**, 82–97 (2016).

Acknowledgements
This study was supported by grant from the Ministry of Higher Education Malaysia (Kementerian Pendidikan Tinggi Malaysia), Transdisciplinary Research Grant Scheme (TRGS/1/2020/USM/02/2/2).

Author contributions
N.-S.L., S.P.W. and A.J.-R. conceived and designed this study. N.M., I.I., M.G. and S.G. collected the samples. N.-S.L. performed the bioinformatics analysis. N.-S.L., N.R. and A.J.-R. wrote the manuscript. All authors reviewed and approved the final manuscript.

Competing interests
The authors declare no competing interests.

Additional information
Correspondence and requests for materials should be addressed to A.J.-R.

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.