PNAS

Supporting Information for

The *Holothuria leucospilota* genome elucidates sacrificial organ expulsion and bioadhesive trap enriched with amyloid-patterned proteins

Ting Chen[1,6,†,*], Chunhua Ren[1,6,†], Nai-Kei Wong[2,†], Aifen Yan[3,†], Caiyun Sun[4,†], Dingding Fan[5,†], Peng Luo[1,6], Xiao Jiang[1,6], Lvping Zhang[1,6], Yao Ruan[1,7], Jiaxi Li[3], Xiaofen Wu[1,7], Da Huo[1,7], Jiasheng Huang[1,7], Xiaomin Li[1,7], Feifei Wu[1,7], Zixuan E[1,7], Chuhang Cheng[1,8], Xin Zhang[1,7], Yanhong Wang[1,6], Chaoqun Hu[1,8,*]

* **Correspondence authors:**
T.C. (**Email:** chan1010@scsio.ac.cn; ORCID: 0000-0002-5777-909X);
C.H. (**Email:** hucq@scsio.ac.cn; ORCID: 0000-0003-1047-5598)

This PDF file includes:

Other supporting materials for this manuscript include the following:

Datasets S1 to S3
Movies S1 and S2

Supporting Information Text

Supplementary Materials and Methods

Sample collection

A female adult *H. leucospilota* specimen was collected from a shallow-water site near Yongxin island (112°20′E, 16°50′N), China. The rete mirabile was used for genome sequencing and construction of the Hi-C library. The tissue transcriptomic sequencing was performed alongside three additional female specimens for the body wall, muscle, oral tentacles, intestine, rete mirabile, transverse vessel, polian vesicle, respiratory tree, Cuvierian organ, coelomocytes and ovary and three male specimens for the testis. All animal specimens were used in accordance with research guidelines set up by the research ethics committee for animal experiments at the South China Sea Institute of Oceanology, Chinese Academy of Sciences.

Genome sequencing and assembly

Short reads used in genome size estimation and correction of genome assembly were obtained on an Illumina HiSeq X Ten platform (Illumina, San Diego, CA). A paired-end and mate-paired sequencing libraries was constructed with an insert length of 350 bp and sequenced with 150 bp pair-ended. Low quality and duplication reads were removed by using SOAPnuke1 v1.5.6. For SMRT sequencing, genomic DNA was fragmented by using 26 G Needle and 20 kb template was prepared using BluePippin Size-Selection System (PacBio, Menlo Park, CA) and sequenced by using the Pacbio Sequel I system (PacBio).

Genome size was initially estimated based on main k-mer depth from clean short reads (k = 17) by using kmerfreq (1). The genome size, heterozygosity and multiplicity were further calculated by using script genomescope.R of GenomeScope (2). Long reads generated on a PacBio Sequel platform were assembled into contigs by using the three assemblers: FALCON-unzip (3) v1.3.5, flye (4) v2.5 and WTDBG (5) v2 2.5. The assembly from WTDBG was used for curating heterozygous diploid genome assemblies by a Purge Haplotigs (6) pipeline.

Hi-C library were created as described previously (7). Purified DNA was fragmented with MboI restriction enzyme (NEB, Ipswich, MA) to a size of 300–500 bp and DNA ends were then repaired. Hi-C libraries were controlled for quality and sequenced on a BGI MGISEQ-2000 sequencer (BGI, Shenzhen, China). All clean data were mapped onto contigs by using BWA aliger (8) v0.7.16a-r1181. LACHESIS (9) was used for scaffolds *de novo* assemblies based on the locations of Hi-C reads on assembly contigs. Chromosome-scale scaffolds were adjusted manually with the help of JucieBox (10) on the basis of an interaction map created by Juicer (11). Illumina-derived short reads and Pacbio subreads (> 10 kb) were employed to correct for any remaining errors or gaps by using nextPolish (12) v1.3.0.

Short reads were generated by the Illumina platform onto the genome with BWA (13) v0.7.17-r1188 (-mem), which were then mapped and subjected to variant calling with SAMtools (14) v1.9 to evaluate the accuracy of the genome at single-base level. The set of single-copy orthologs of Benchmarking Universal Single-Copy Orthologs (BUSCO 3.0) (15) was searched against the assembled genome and gene sets.

Genome annotation

The gene structures were predicted by homologous comparison, RNA-seq prediction and *ab initio* prediction. The pipelines of Maker (16) v3.01.03 were utilized to integrate the three aspects of evidence to produce the final gene structures. As for homology prediction, protein sequences of *Homo sapiens, Acanthaster planci, Strongylocentrotus purpuratus* and *Apostichopus japonicas*, were concatenated to single files. For RNA-seq, paired-end clean reads from 47 RNA-seq libraries (SI Appendix, Dataset S1 I) from tissues were aligned to the *H. leucospilota* genome by using HISAT2 (17) and transcripts of each library were assembled and merged by using StringTie (17) v1.3.5. For *ab initio* prediction, AUGUSTUS (18) v3.2.3 and Fgenesh (19) in softbarry were employed. Proteins containing tandem repeats were detected by using TRF (20) by searching the coding sequence of gene stets.

Gene functions were annotated based on best-matched hits in SwissProt (21) by using BLAST 2.9.0+ (22) (-p blastp -e 1e-5). Gene motifs and domains were identified by InterProScan (23) v5.27-66.0 as protein sequences to be searched with the InterProScan software to identify molecular signatures. Gene ontology annotations were created based on InterPro2GO mapping results. Gene ontology enrichment were done by R package clusterProfiler (24).

Evolution analyses

To determine orthologues in comparative genomics, proteomes of 12 species including *H. sapiens*, *Gallus gallus*, *Xenopus tropicalis*, *Danio rerio*, *Ciona intestinalis*, *Branchiostoma lanceolatum*, *Ptychodera flava*, *Saccoglossus kowalevskii*, *S. purpuratus*, *A. planci*, *A. japonicas*, *Holothuria glaberrima* were retrieved for analysis run on OrthoFinder (25) v2.4.0. Based on orthologue relationships with human genes, coding sequences were aligned by using Prank (26) with a GUIDANCE (27) pipeline (codon model). Nucleic acid sites were selected to construct a phylogenetic tree by using RAxML (28). Divergence time estimation by approximate likelihood calculation was done by using mcmctree in paml (29) v4.9.2.

Gene family size changes were analyzed in an attempt to account for phylogenetic history and provide a statistical foundation for evolutionary inferences by using CAFE (30) v4.2 with a cutoff *p*-value of 0.01.

RNA sequencing and gene expression analyses

The RNA library preparations were sequenced on an Illumina Hi-seq platform, where 150 bp paired-end reads were generated. Clean data (clean reads) were obtained by using SOAPnuke (31) v1.5.6. Paired-end clean reads were aligned to the reference genome by using HISAT2 (17) v2.1.0. Transcripts were assembled and reads counts of each gene were calculated by using StringTie (17) v1.3.5. Counts per million mapped reads (CPM) was calculated and cross-sample normalization were done by using DESeq2 (32).

Histology

Quiescent CO tubules were collected by tissue dissection following anesthesia on ice, and elongated tubules were discharged by applying pressure on the sea cucumber body to induce tubules expulsion and allowed to attach naturally. Paraffin sectioning of quiescent and attached tubules were cut transversely into 4 μm thin and stained with hematoxylin and eosin (H/E). For SEM ultrastructures, specimens of quiescent tubules were attached to metallic stubs and observed and visualized by an S-3400N scanning electron microscope (Hitachi, Chiyoda, Japan) with an accelerating voltage of 30 kV in high vacuum mode. For TEM ultrastructures, specimens of quiescent tubules were cut to 60 nm thin on an EM UC7 Ultramicrotome (Leica) and observed and visualized by an HT7800 transmission electron microscope (Hitachi).

For F*IS*H, the 219-bp Hl-16258 and 213-bp Hl-21915 antisense cRNA probes (SI Appendix, Dataset S1 W) were labeled at the 5'-ends with the Cy5 and 6-FAM dyes, respectively. DAPI reagent was used for counter-staining of cell nuclei. The F*IS*H sections were then viewed and imaged by an LSM800 Confocal Laser Scanning Microcopy (Zeiss, Thornwood, NY). For immunofluorescence staining, the antigens for Hl-25083, Hl-25084, Hl-25088, Hl-30757, Hl-19376, Hl-19378 and Hl-25085 were generated by an *Escherichia coli* system. Polyclonal antibodies were purified from the serum of rabbits with antigen injection (Genecreate, Wuhan, China). The paraffin sections of CO were incubated with primary antibodies followed by FITC labeled secondary antibody and DAPI reagent for visualization of the endogenous proteins and the cell nuclei, respectively. Fluorescence images were acquired with Digital Eclipse C1 Microscope (Nikon, Tokyo, Japan).

For Congo red staining of amyloid fibrils, CO paraffin sections at 5 μm were deparaffinized and hydrated with distilled water, then stained in 1% Congo red solution (Sbjbio Life Sciences, Nanjing, China) for 30 min, and rinsed in deionized water for 5 min. Thereafter, the sections were counterstained with Gill's hematoxylin for 30 s, rinsed in water for 1 min, dipped in liquid ammonia for 30 s and rinsed in deionized water for 5 min. Amyloid signals were observed and visualized under a Leica DM-IRB light microscope (Leica).

Cuvierian organ component analyses

To calculate enrichment of tissue-specific genes, we use teGeneRetrieval (33) to define tissue-specific genes from TissueEnrich package. Here, a CPM matrix of 12 tissues was used with the following parameters: fold change threshold = 5; max number of tissues = 2; expressed gene threshold = 2. All tissue-enriched, group-enriched, tissue-enhanced genes of the tissues were merged as tissue-specific genes, and the high expression genes in CO was defined with the following threshold: CPM of CO tubules 20 folds greater than the median CPM of 12 tissues.

To detect tandem-repeated proteins, coding sequences of genes were called into Tandem Repeats Finder (20) v4.09.1 to find two or more adjacent, approximate copies of a pattern of nucleotides with the following parameters (2 5 7 80 10 50 2000). For genes of interest, protein

sequences were manually verified with online tools by RADAR.

AlphaFold (34) v2.1.1 (GitHub - Deepmind/Alphafold), an open-source code for AlphaFold, was used to predict a protein's 3D structure from its amino acid sequence. Five models were entered into CASP14 and full databases were employed. The project was performed by implementing commands on an A40 Nvidia GPU and a 72 core Xeon CPU with a 256-Gb memory server.

Measurement of intracellular Ca^{2+} levels, $[Ca^{2+}]_i$

Three double-stranded RNA (dsRNA) specifically targeting the transient receptor potential canonical (TRPC) Hl-16258 and Hl-21915 and the piezo-type mechanosensitive ion channel component 2 (PIEZ2) Hl-13522 were designed based on their coding sequences. dsRNAs were generated with a T7 RiboMAX Express RNAi System kit (Promega, Madison, WI, USA). In addition, a non-targeting dsRNA, which is specific to enhanced green fluorescent protein (EGFP), was used as a negative control. dsRNA quality was verified after annealing by gel electrophoresis. dsRNA sequences for Hl-21915, Hl-16258, Hl-13522 and EGFP are as shown in SI Appendix, Dataset S1 W. Concentrations of synthesized dsRNA were verified with a NanoDrop 2000 spectrophotometer (Thermo Scientific, Waltham, MA, USA). dsRNAs were then diluted to 600 ng/μL in 150 mM RNase-free saline solution (RFSS). In the negative control group, dsRNA solutions were replaced with RFSS containing no dsRNA. For *in vivo* dsRNA interference, a total of 150 sea cucumbers (weighing about 50 g each) were randomly assigned into 15 groups, each containing 10 individuals. Sea cucumbers in 3 groups (30 individuals) were employed for a treatment. In the blank control group, 100 μL of 150 mM RFSS was injected into each sea cucumber. In a negative control group, 100 μL of EGFP dsRNA solution which contained EGFP dsRNA was injected into each sea cucumber. In the three experimental groups, each cucumber was injected with 100 μL of dsRNA solution for TRPCs (Hl-21915 and Hl-16258) and PIEZ2 (Hl-13522), correspondingly. Samples of CO were collected to assess interference efficiency of dsRNAs by qPCR, with specific primers and reaction conditions as listed in SI Appendix, Dataset S1 X. At 24 h following injection, pressure on the middle-body were applied to induce Cuvierian organ expulsion. Behaviors were observed by visual inspection and recorded photographically within 1 min after stimulation.

Pressure-induced Cuvierian organ expulsion

A total of 450 sea cucumbers were randomly assigned into 45 tanks with 30‰ artificial seawater for acclimatization for 7 days prior to experiments. Three methods, including physical pressure (by squeezing), piercing and tactile stimulation (by stroking), were applied to test for any stimulatory effects on CO expulsion. In this case, 150 individuals in 15 tanks were used for a particular stimulation method. The stimulated sites of sea cucumbers included 5 parts, namely, the front end (oral tentacles), anterior-body, middle-body, posterior-body and rear end (anus). Biological replicates for each stimulated site for each stimulation method were 30 tested in 3 tanks. Behaviors of CO expulsion were observed by visual inspection and recorded within 1 min after stimulation.

Gene family analyses

Proteomes for *H. leucospilota* and other 12 genomes in typical deuterostome species were selected in this study to assign functions with SwissProt and InterProscan. TRP genes thereof were selected based on the keyword "transient receptor potential" in the SwissProt database. Specific classifications of TRPs including TRPC (canonical), TRPV (vanilloid), TRPM (melastatin), TRPA (ankyrin), TRPML (mucolipin), TRPAP (associated protein) and TRPP (polycystic) were determined by phylogenetical analysis. LGIC genes thereof were selected based on the keywords "IPR006029, neurotransmitter-gated ion-channel transmembrane domain" and "IPR006202, neurotransmitter-gated ion-channel ligand-binding domain" in the InterProscan database. Specific classifications of LGICs including neuronal acetylcholine receptor (nAchR), γ-aminobutyric acid receptor (GBR)/glycine receptor (GLR) and 5-hydroxytryptamine receptor (HTR) were determined by phylogenetical analysis. Multiple sequence alignment (MSA) for all gene families mentioned above were built with the MAFFT v7.487 aligner and their corresponding phylogeny was inferred with RAxML v8.2.12. Phylogenetic trees were constructed to classify gene families and visualized with Evolview (v3).

RNA interference (RNAi)

Double-stranded RNA (dsRNA) specifically targeting TRPCs (Hl-16258 and Hl-21915), PIEZ2 (Hl-13522) and EGFP (non-targeting negative control) were generated with a T7 RiboMAX Express

RNAi System kit (Promega, Madison, WI) and diluted to RNase-free saline solution (RFSS). For *in vivo* dsRNA interference, a total of 150 sea cucumbers (weighing about 50 g each) were randomly assigned into 15 groups, each containing 10 individuals. Sea cucumbers in 3 groups (30 individuals) were employed for a treatment. In the blank control group, 100 μL of 150 mM RFSS was injected into each sea cucumber. In a negative control group, 100 μL of EGFP dsRNA solution which contained EGFP dsRNA was injected into each sea cucumber. In the three experimental groups, each cucumber was injected with 100 μL of dsRNA solution for TRPCs (Hl-21915 and Hl-16258) and PIEZ2 (Hl-13522), correspondingly. Samples of CO were collected to assess interference efficiency of dsRNAs by qPCR (SI Appendix, Dataset S1 X). At 24 h following dsRNA injection, pressure on the middle-body were applied to induce CO expulsion. Behaviors were observed by visual inspection and recorded within 1 min after stimulation.

Inhibition of TRPC for Cuvierian organ expulsion

Inhibition of TRPC was performed by celomic injection of its blocker. Briefly, a total of 60 sea cucumbers were randomly assigned into 6 individual groups, each containing 10 biological replicates. Sea cucumbers in 3 groups (30 individuals) were employed for a treatment. SKF96365 (130495-35-1, InvivoChem, Libertyville, IL), a selective inhibitor of receptor-mediated and voltage-gated Ca^{2+} entry, were prepared in DMSO at a concentration of 10 mM. In the TRPC inhibitory group, sea cucumbers were injected with SKF-96365 in final concentration of 5 μM. In the control group, 100 μL of 150 mM PBS with 5 μM DMSO was injected into each sea cucumber. At 60 min following blocker injection, pressure on the middle-body were applied to induce CO expulsion. Behaviors were observed by visual inspection and recorded photographically within 1 min after stimulation.

Acetylcholine injection for Cuvierian organ expulsion

Pharmacological effects of acetylcholine on CO expulsion and other behaviors were detected following celomic injection. Briefly, 150 sea cucumbers were randomly assigned into 5 groups (30 individuals per group) in 15 200-L tanks (10 individuals per tank) with 30‰ artificial seawater for acclimatization for 7 days prior to experiments. Sea cucumbers were injected with acetylcholine (A2661-25G, Sigma-Aldrich, St. Louis, MO) in final concentrations of 10–1000 μM and nicotine (N-008-1ML, Sigma-Aldrich), the receptor agonist of nAChRs, in final concentration of 100 nM diluted in PBS, and PBS injection alone served as a control. Behaviors of body contraction, CO expulsion and non-CO organ expulsion were recorded upon their appearance, 0–16 h after injection.

Supplementary Figures

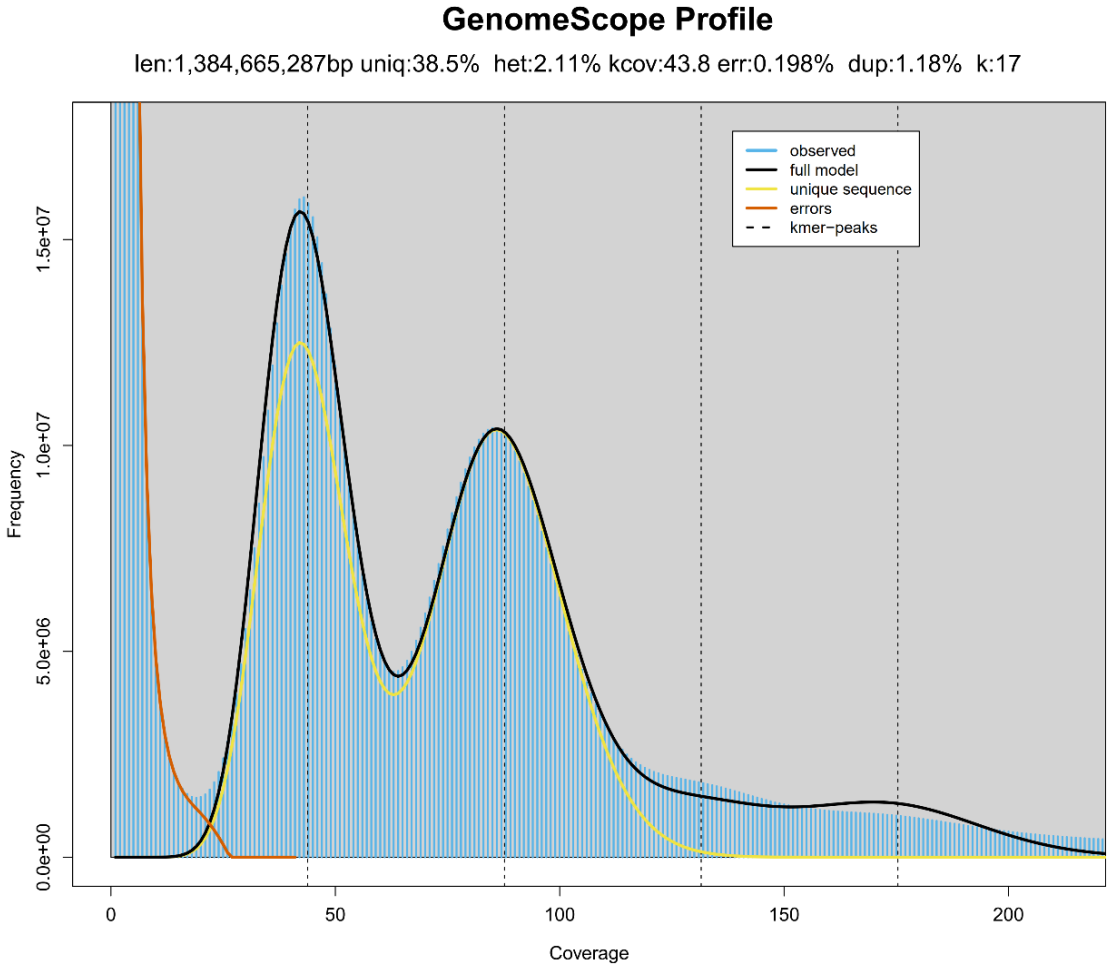

Fig. S1. Estimation of *H. leucospilota* genome size based on k-mer coverage distribution analyzed with Kmerfreq (k-mer size: 17 bp).

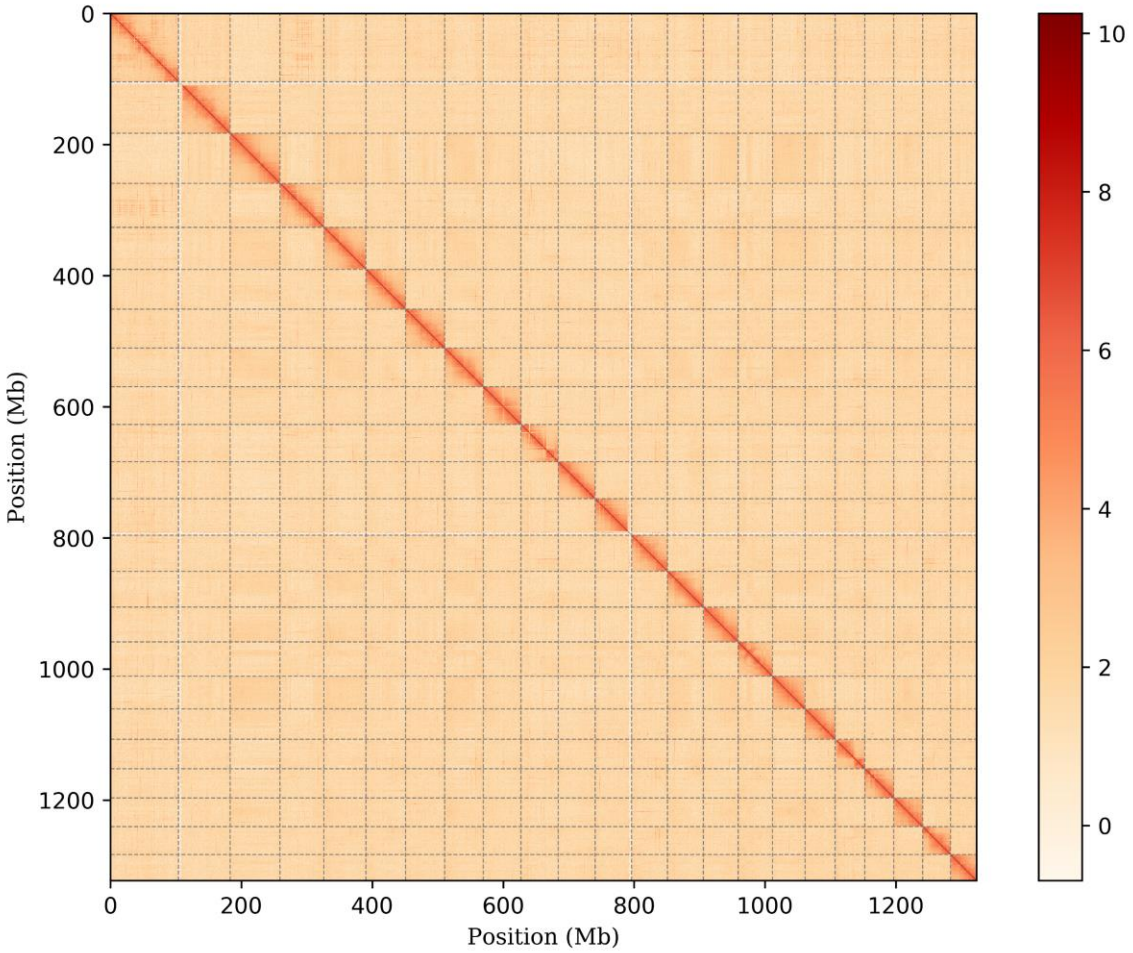

Fig. S2. Heatmap for Hi-C interactions, which shows the clustering of polished primary contigs into 23 sets of chromosome-scale scaffolds. Here, the sliding window size was set as 500 kb.

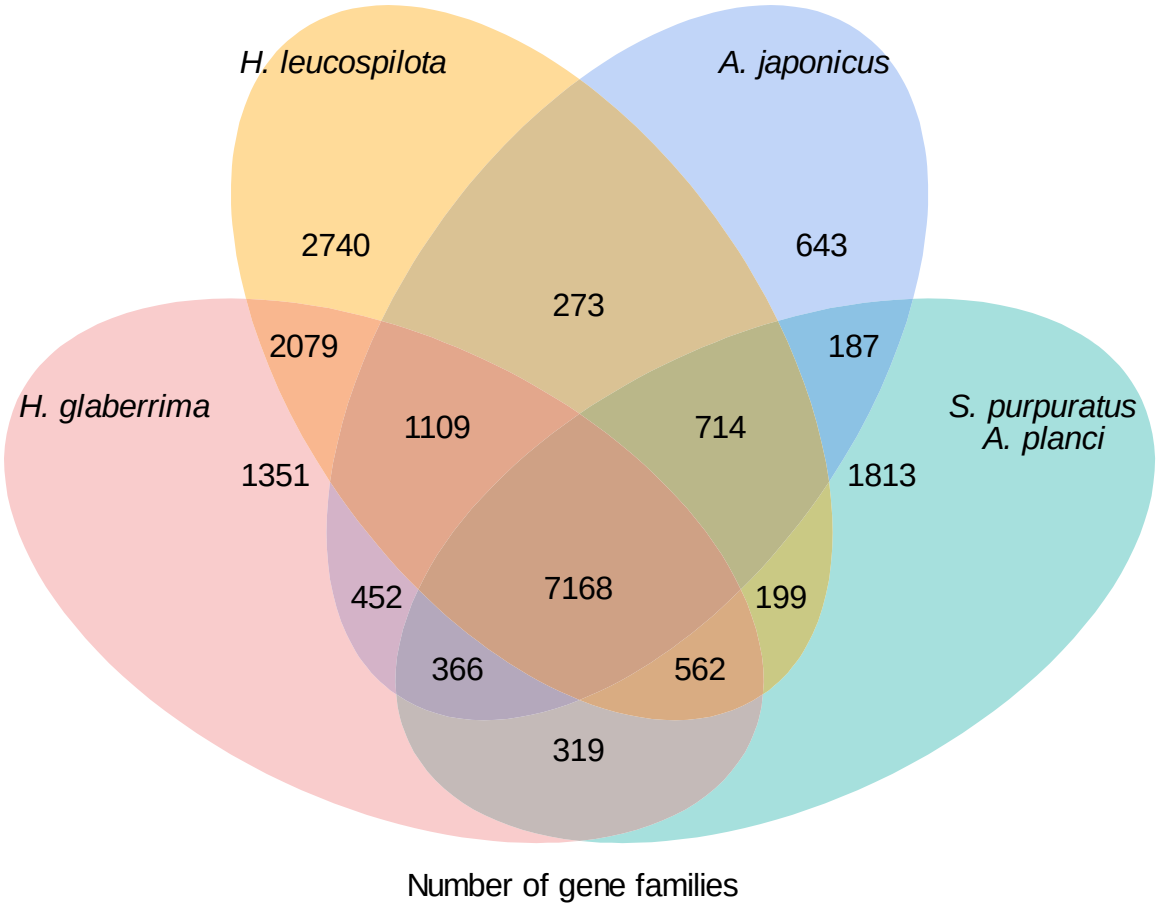

Fig. S3. Venn diagram for shared gene families in Echinodermata.

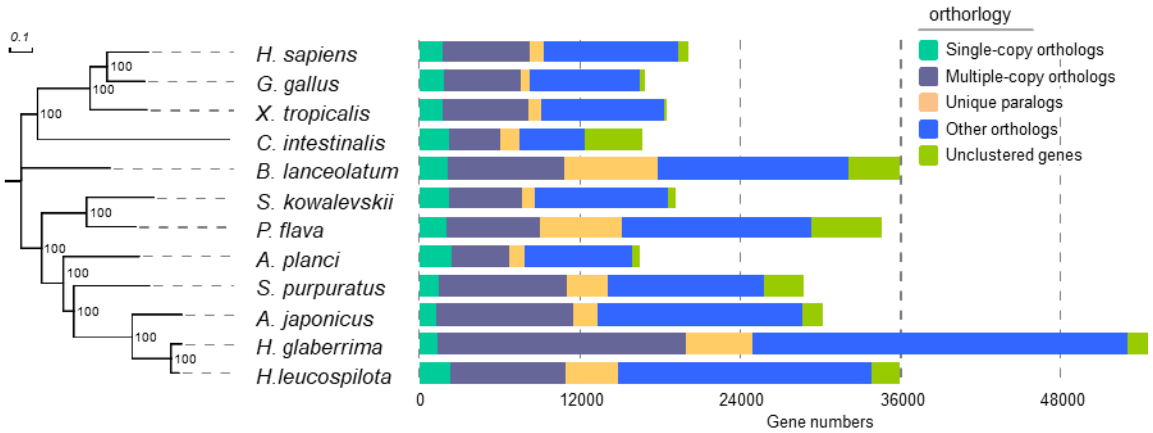

Fig. S4. Summary on relationships between orthologues of *H. leucospilota* and other animals in Deuterostomia. Single-copy orthologues mean that only one gene is present in a species within all 12 species. Multiple-copy orthologues mean that more than one homologous genes are present in a species among all 12 species. Unique paralogues mean that homologous genes exist only in a species but not in 11 other species. Other orthologues mean that the genes are present in some species but not in all 12 species. Unclustered genes mean that the genes do not form any observable clustering.

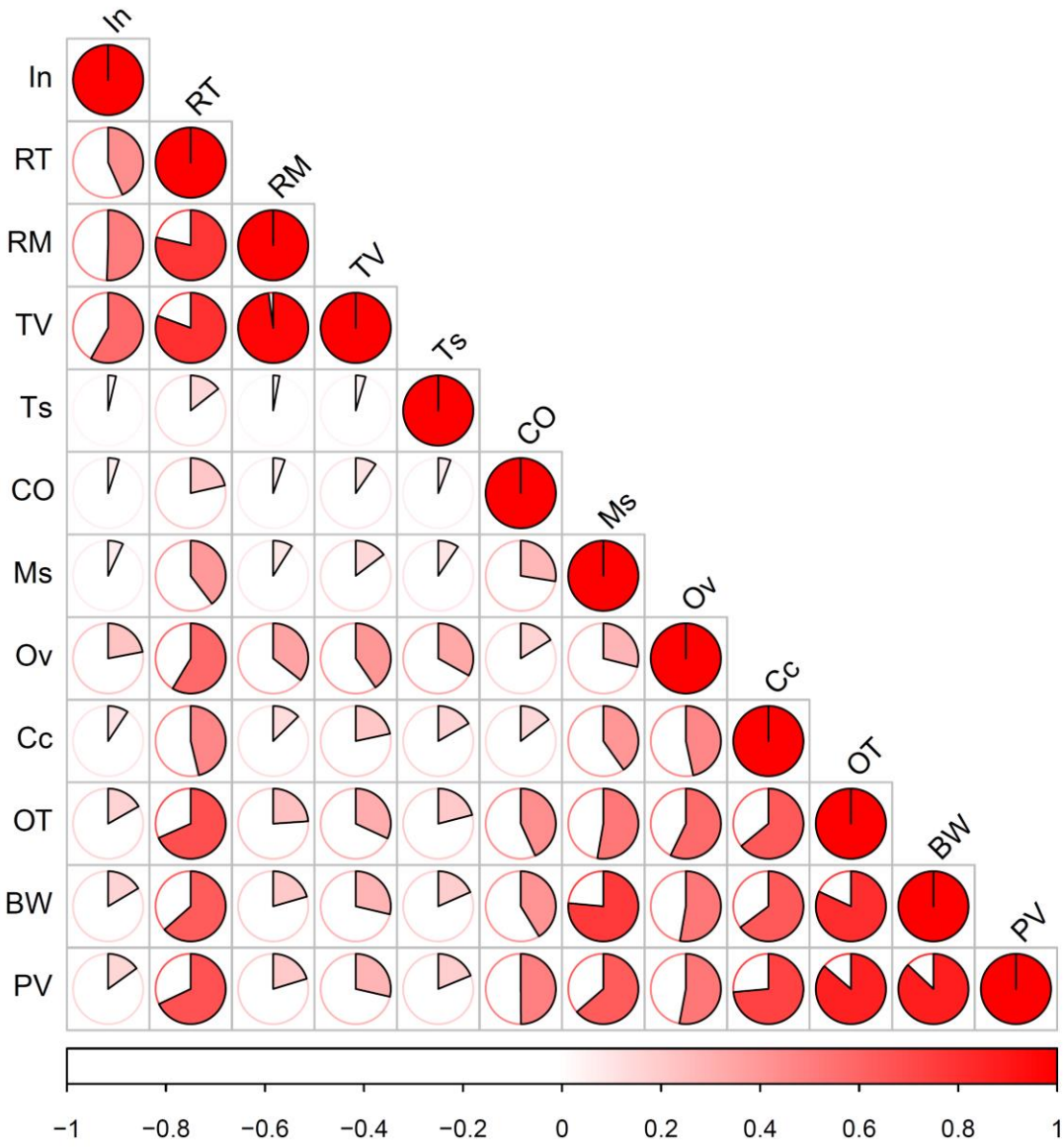

Fig. S5. Correlation for transcript expression all *H. leucospilota* genes in different tissues. Expression levels are represented as CPM. Selected tissues include the body wall (BW), Cuvierian organ (CO), coelomocytes (Cc), intestine (In), muscle (Ms), oral tentacles (OT), ovary (Ov), polian vesicle (PV), rete mirabile (RM), respiratory trees (RT), transverse vessel (TV) and testis (Ts).

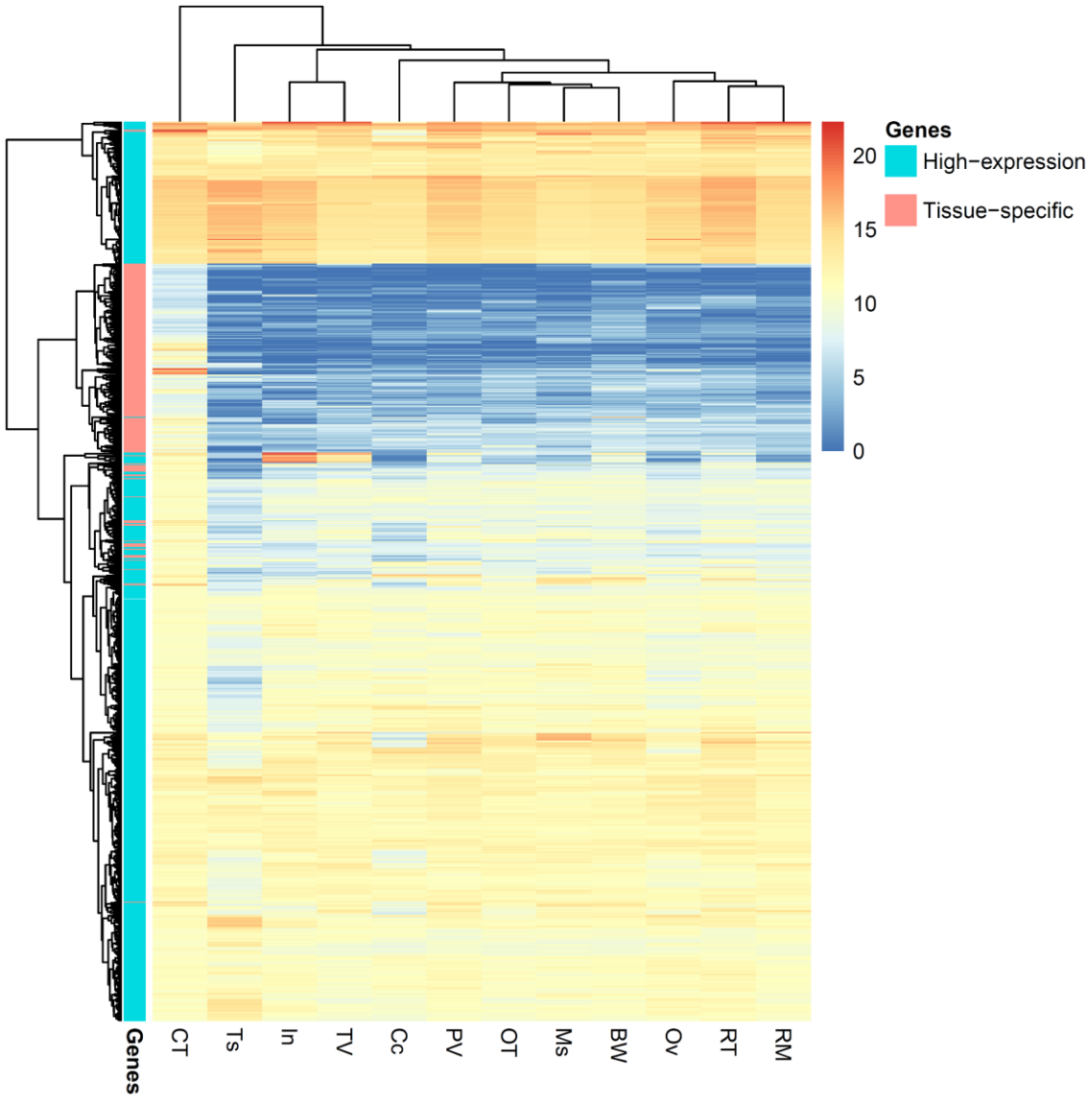

Fig. S6. Heatmap illustrating 611 tissue highly-expressed and 198 tissue specifically expressed genes in the Cuvierian organ. Expression levels are represented by CPM. Selected tissues include the body wall (BW), Cuvierian organ (CO), coelomocytes (Cc), intestine (In), muscle (Ms), oral tentacles (OT), ovary (Ov), polian vesicle (PV), rete mirabile (RM), respiratory trees (RT), transverse vessel (TV) and testis (Ts).

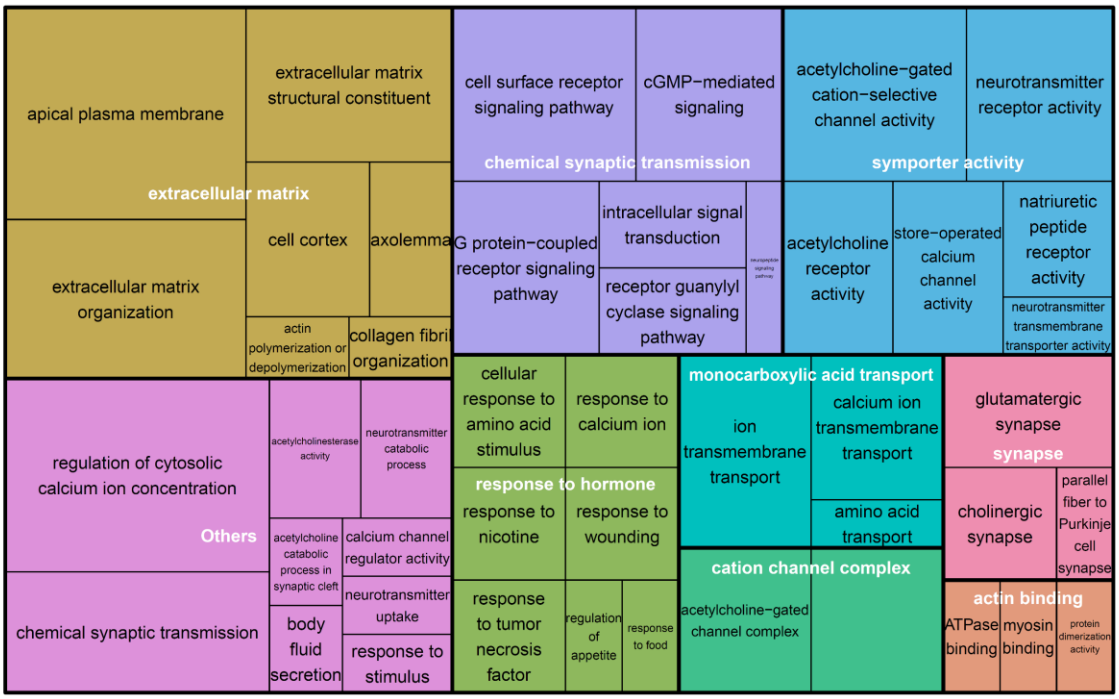

Fig. S7. Tree map visualize of gene ontology terms and summarizes for Cuvierian organ highly and specifically expressed genes using revigo.

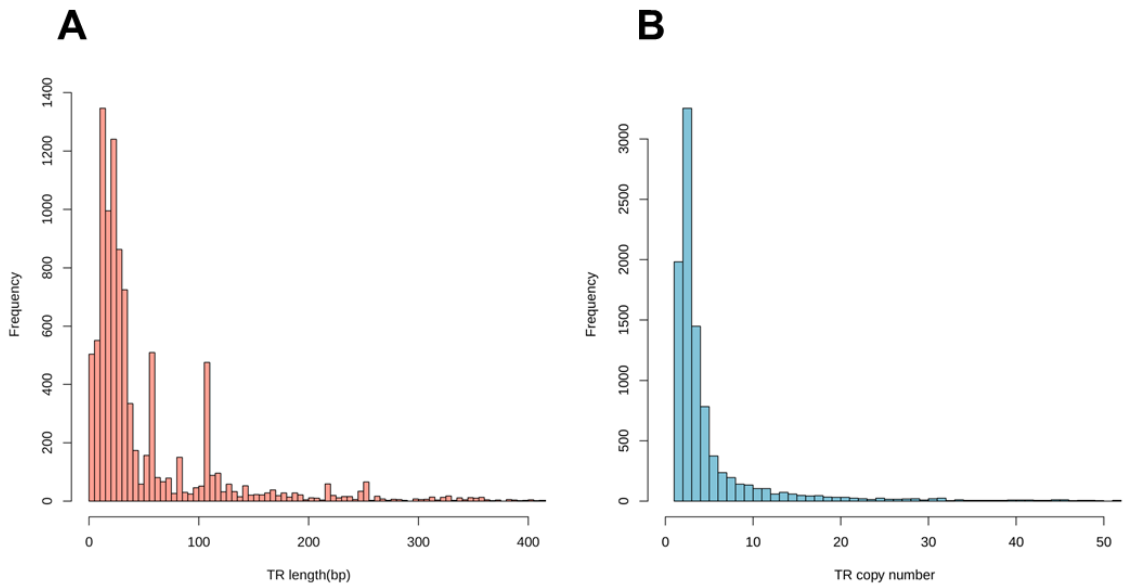

Fig. S8. Repeat lengths (**A**) and repeat numbers (**B**) of the tandem-repeats containing genes in *H. leucospilota*.

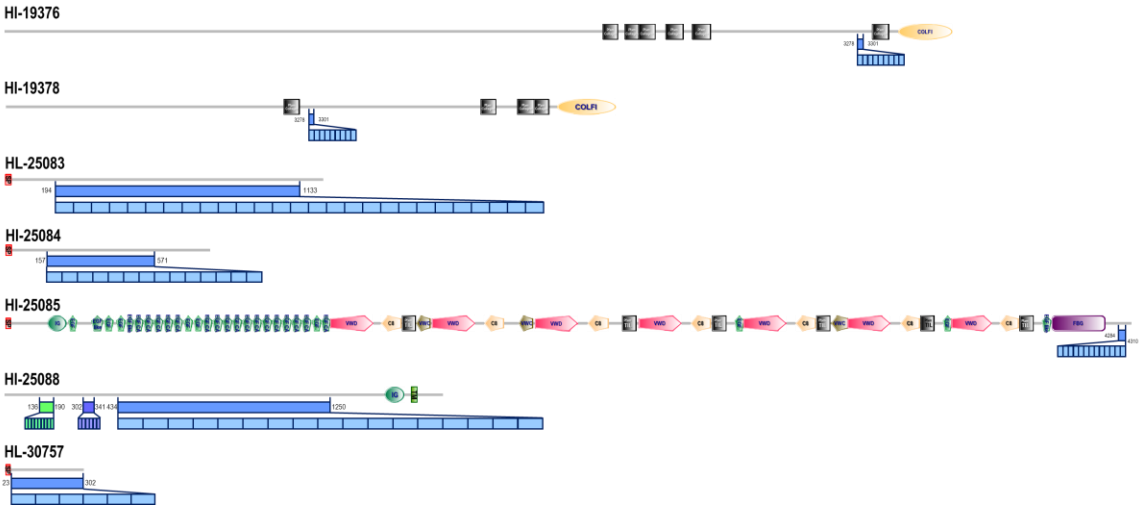

Fig. S9. Structural domains and repeats of 7 CO-highly and specifically expressed genes, namely: COL1A (HI-19376), COL1A (HI-19378), COOLAP124k (HI-25083), COOLAP73k (HI-25084), FCGBP (HI-25085), COOLAP178k (HI-25088) and COOLAP35k (HI-30757), as predicted by the Simple Modular Architecture Research Tool (SMART) program.

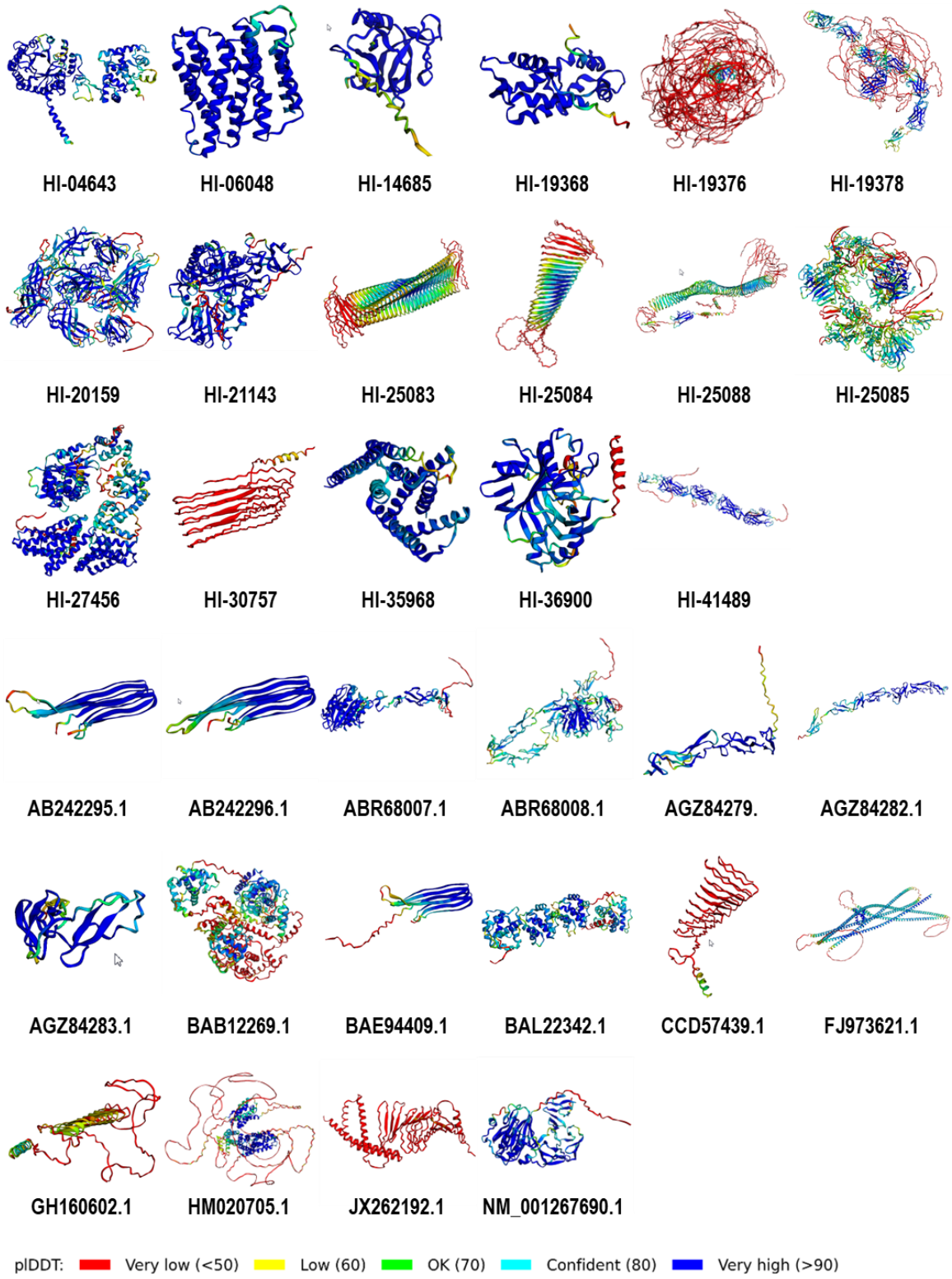

Fig. S10. Protein structures of CO-specific tandem-repeats containing protein in *H. leucospilota*, and adhesive proteins in other species, as predicted with AlphaFold.

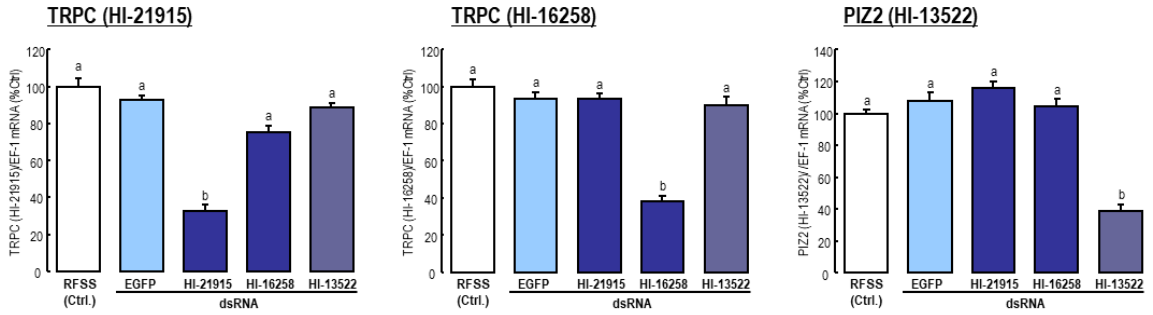

Fig. S11. Efficiencies of RNAi against TRPCs (HI-21915 and HI-16258) and PIZ2 (HI-13522) in the Cuvierian organ of *H. leucospilota* after injection of RFSS (blank control) or dsRNA of EGFP (negative control), TRPCs (HI-2915 and HI-16258) and PIZ2 (HI-13522) for 12 h. Identically assigned letters represent a comparable level of transcriptional expression ($p > 0.05$), while differently assigned letters represent significant difference in levels of transcriptional expression between two groups ($p < 0.05$, ANOVA followed by Fisher LSD test).

Legends for Supplementary Table Dataset

The following are the titles of the supplementary tables. Details of the supplementary tables are shown in the Excel file “Dataset S1.xlsx”.

Dataset S1 A (separate file). Summary of short reads sequencing on the Illumina HiSeq Xten platform for estimation of genome size and accuracy of the genome assembly.

Dataset S1 B (separate file). Summary of long reads sequencing by PacBio Sequel I system for contig assembly.

Dataset S1 C (separate file). Summary of Hi-C genome sequencing by BGI MGISEQ-2000 sequencer for chromosome construction.

Dataset S1 D (separate file). Assembly statistics of the *H. leucospilota* genome by using three assembler. The WTDGB assembler presents a good continuous assembly with a smaller contig number.

Dataset S1 E (separate file). Assembly statistics of the *H. leucospilota* chromosome-scale genome.

Dataset S1 F (separate file). Summary of short reads alignment to the final genome assembly. Here, the variations depth greater than 2-fold average sequence depth are removed in cases of errors produced by a repeat region.

Dataset S1 G (separate file). Benchmarking Universal Single-Copy Orthologs (BUSCOs) evaluates the genome completeness in the published Echinoderm genomes. The genome accession numbers of *Holothuria glaberrima*, *Apostichopus japonicas*, *Acanthaster planci*, *Strongylocentrotus purpuratus* are GCA_009936505.1, GCA_002754855.1, GCF_001949145.1 and GCF_000002235.3, respectively. The total gene number of single-copy ortholog in metazoa_odb9 set is 978.

Dataset S1 H (separate file). Lengths of the 23 assembled *H. leucospilota* chromosomes.

Dataset S1 I (separate file). Summary of RNAseq reads for 47 samples and their alignments onto the *H. leucospilota* genome. The reads for all libraries are paired-ends of 150 bp.

Dataset S1 J (separate file). Summary of gene sets of *H. leucospilota* and other 9 deuterostome animals.

Dataset S1 K (separate file). Summary functional assignments of *H. leucospilota* coding genes from known protein databases.

Dataset S1 L (separate file). Summary of repeat elements and their classification based on RepeatScout and Repbase.

Dataset S1 M (separate file). Expression of all genes in different tissues of *H. leucospilota* that are represented by CPM. The selected tissues include body wall (BW), Cuvierian organ (CO), coelomocytes (Cc), intestine (In), muscle (Ms), oral tentacles (OT), ovary (Ov), polian vesicle (PV), rete mirabile (RM), respiratory trees (RT), transverse vessel (TV) and testis (Ts).

Dataset S1 N (separate file). The Cuvierian organ highly expressed genes and their transcript levels in multiple tissues, as represented by CPM.

Dataset S1 O (separate file). The Cuvierian organ specifically expressed genes and their transcript levels in multiple tissues, as represented by CPM.

Dataset S1 P (separate file). The Cuvierian organ highly and specifically expressed genes and their transcript levels in multiple tissues, as represented by CPM.

Dataset S1 Q (separate file). The amino acid tandem-repeated sequences in the genes from the *H. leucospilota* genome, and their ranges in the genes, repeat lengths and repeat numbers.

Dataset S1 R (separate file). Expression levels of the tandem-repeats containing genes in multiple tissues and their relationship to the Cuvierian organ specifically and highly expressed genes. Ave Exp: gene average expression in 12 tissues; CO High Exp: Cuvierian organ highly expressed genes; CO Spe Exp: Cuvierian organ specifically expressed genes; CO High+Spe Exp: Cuvierian organ highly and specifically expressed genes.

Dataset S1 S (separate file). The *H. leucospilota* Cuvierian organ highly and specifically expressed genes occurring as potential orthologues in 12 species.

Dataset S1 T (separate file). Functional domains of the *H. leucospilota* Curvierian organ highly and specifically expressed genes.

Dataset S1 U (separate file). TRP superfamily genes in *H. leucospilota* and other 12 deuterostome animals.

Dataset S1 V (separate file). LGIC superfamily genes in *H. leucospilota* and other 12 deuterostome animals.

Dataset S1 W (separate file). cRNA probes and dsRNA sequences used in this study.

Dataset S1 X (separate file). Primers and qPCR conditions used in this study.

Legends for Phylogenetic Tree Datasets

Phylogenetic Tree Data S2 (separate file). Detailed phylogenetic tree of TRP superfamily genes in *H. leucospilota* and other 12 deuterostome animals. HOMsap: *Homo sapiens*; GALgal: *Gallus gallus*; XENtro: *Xenopus tropicalis*; DANrer: *Danio rerio*; CIOint: *Ciona intestinalis*; BRAlan: *Branchiostoma lanceolatum*; PTYfla: *Ptychodera flava*; SACkow: *Saccoglossus kowalevskii*; STRpur: *Strongylocentrotus purpuratus*; ACApla: *Acanthaster planci*; APOjap: *Apostichopus japonicas*; HOLgla: *Holothuria glaberrima*; HOLleu: *Holothuria leucospilota*.

Phylogenetic Tree Data S3 (separate file). Detailed phylogenetic tree of LGIC superfamily genes in H. leucospilota and other 12 deuterostome animals. HOMsap: *Homo sapiens*; GALgal: *Gallus gallus*; XENtro: *Xenopus tropicalis*; DANrer: *Danio rerio*; CIOint: *Ciona intestinalis*; BRAlan: *Branchiostoma lanceolatum*; PTYfla: *Ptychodera flava*; SACkow: *Saccoglossus kowalevskii*; STRpur: *Strongylocentrotus purpuratus*; ACApla: *Acanthaster planci*; APOjap: *Apostichopus japonicas*; HOLgla: *Holothuria glaberrima*; HOLleu: *Holothuria leucospilota*.

Legends for Movies

Movie S1 (separate file). *H. leucospilota* expel Cuvierian tubules toward a crab and the crab was tied up.

Movie S2 (separate file). Biochemical properties as assessed in terms of adhesiveness and tenacity of *H. leucospilota* Cuvierian tubules.

SI References

1. B. Liu *et al.*, Estimation of genomic characteristics by analyzing k-mer frequency in de novo genome projects. *arXiv preprint arXiv:1308.2012*, (2013).
2. T. R. Ranallo-Benavidez, K. S. Jaron, M. C. Schatz, GenomeScope 2.0 and Smudgeplot for reference-free profiling of polyploid genomes. *Nat. Commun.* **11**, 1432 (2020).
3. M. Pendleton *et al.*, Assembly and diploid architecture of an individual human genome via single-molecule technologies. *Nat. Methods.* **12**, 780-786 (2015).
4. M. Kolmogorov, J. Yuan, Y. Lin, P. A. Pevzner, Assembly of long, error-prone reads using repeat graphs. *Nat. Biotechnol.* **37**, 540-546 (2019).
5. J. Ruan, H. Li, Fast and accurate long-read assembly with wtdbg2. *Nat. Methods.* **17**, 155-158 (2020).
6. M. J. Roach, S. A. Schmidt, A. R. Borneman, Purge Haplotigs: allelic contig reassignment for third-gen diploid genome assemblies. *BMC bioinformatics* **19**, 460 (2018).
7. G. Gong *et al.*, Chromosomal-level assembly of yellow catfish genome using third-generation DNA sequencing and Hi-C analysis. *GigaScience* **7**, 1-9 (2018).
8. H. Li, R. Durbin, Fast and accurate long-read alignment with Burrows-Wheeler transform. *Bioinformatics* **26**, 589-595 (2010).
9. J. N. Burton *et al.*, Chromosome-scale scaffolding of de novo genome assemblies based on chromatin interactions. *Nat. Biotechnol.* **31**, 1119-1125 (2013).
10. N. C. Durand *et al.*, Juicebox provides a visualization system for Hi-C contact maps with unlimited zoom. *Cell Syst.* **3**, 99-101 (2016).
11. N. C. Durand *et al.*, Juicer provides a one-click system for analyzing loop-resolution Hi-C experiments. *Cell Syst.* **3**, 95-98 (2016).
12. J. Hu, J. Fan, Z. Sun, S. Liu, NextPolish: a fast and efficient genome polishing tool for long-read assembly. *Bioinformatics* **36**, 2253-2255 (2020).
13. H. Li, R. Durbin, Fast and accurate short read alignment with Burrows–Wheeler transform. *Bioinformatics* **25**, 1754-1760 (2009).
14. H. Li *et al.*, The sequence alignment/map format and SAMtools. *Bioinformatics* **25**, 2078-2079 (2009).
15. F. A. Simão, R. M. Waterhouse, P. Ioannidis, E. V. Kriventseva, E. M. Zdobnov, BUSCO: assessing genome assembly and annotation completeness with single-copy orthologs. *Bioinformatics* **31**, 3210-3212 (2015).
16. B. L. Cantarel *et al.*, MAKER: an easy-to-use annotation pipeline designed for emerging model organism genomes. *Genome Res.* **18**, 188-196 (2008).
17. M. Pertea, D. Kim, G. M. Pertea, J. T. Leek, S. L. Salzberg, Transcript-level expression analysis of RNA-seq experiments with HISAT, StringTie and Ballgown. *Nat. Protoc.* **11**, 1650-1667 (2016).
18. M. Stanke, S. J. B. Waack, Gene prediction with a hidden Markov model and a new intron submodel. *Bioinformatics* **19**, ii215-ii225 (2003).
19. S. L. Zhang, D. F. Li, G. S. Zhang, J. W. Wang, N. Na, The prediction of rice gene by Fgenesh. *Agr. Sci. China* **7**, 387-394 (2008).
20. G. Benson, Tandem repeats finder: a program to analyze DNA sequences. *Nucleic Acids Res.* **27**, 573-580 (1999).
21. B. Boeckmann *et al.*, The SWISS-PROT protein knowledgebase and its supplement TrEMBL in 2003. *Nucleic Acids Res.* **31**, 365-370 (2003).
22. S. F. Altschul, W. Gish, W. Miller, E. W. Myers, D. Lipman, Basic local alignment search tool. *J. Mol. Biol.* **215**, 403-410 (1990).
23. E. Quevillon *et al.*, InterProScan: protein domains identifier. **33**, *Nucleic Acids Res*. W116-W120 (2005).
24. G. Yu, L. G. Wang, Y. Han, Q. He, clusterProfiler: an R package for comparing biological themes among gene clusters. *J. Mol. Biol.* **16**, 284-287 (2012).
25. D. M. Emms, S. Kelly, OrthoFinder: phylogenetic orthology inference for comparative genomics. *Genome Biol.* **20**, 1-14 (2019).
26. A. Löytynoja, N. Goldman, webPRANK: a phylogeny-aware multiple sequence aligner with interactive alignment browser. *BMC Bioinformatics* **11**, 579 (2010).

27. O. Penn *et al.*, GUIDANCE: a web server for assessing alignment confidence scores. **38**, *Nucleic Acids Res.* W23-W28 (2010).
28. A. Stamatakis, RAxML version 8: a tool for phylogenetic analysis and post-analysis of large phylogenies. *Bioinformatics* **30**, 1312-1313 (2014).
29. Z. Yang, PAML 4: Phylogenetic Analysis by Maximum Likelihood. *Mol. Biol. Evol.* **24**, 1586-1591 (2007).
30. T. De Bie, N. Cristianini, J. P. Demuth, M. W. Hahn, CAFE: a computational tool for the study of gene family evolution. *Bioinformatics* **22**, 1269-1271 (2006).
31. Y. Chen *et al.*, SOAPnuke: a MapReduce acceleration-supported software for integrated quality control and preprocessing of high-throughput sequencing data. *GigaScience* **7**, 1-6 (2018).
32. M. I. Love, W. Huber, S. Anders, Moderated estimation of fold change and dispersion for RNA-seq data with DESeq2. *Genome Biol.* **15**, 550 (2014).
33. A. Jain, G. Tuteja, TissueEnrich: Tissue-specific gene enrichment analysis. *Bioinformatics* **35**, 1966-1967 (2019).
34. J. Jumper *et al.*, Highly accurate protein structure prediction with AlphaFold. *Nature* **596**, 583-589 (2021).