scientific data

OPEN

DATA DESCRIPTOR

An Improved Chromosome-Level Genome Assembly and Comprehensive Annotation of the Model Ascidian *Ciona savignyi*

Xuena Huang[1] & Aibin Zhan[1,2] ✉

Ciona savignyi **is an important model organism across multiple disciplines, including developmental biology, ecology, evolution, and invasion science, and has become increasingly valuable for elucidating the genomic basis of adaptation in a rapidly changing world. Here, we present an improved chromosome-level genome assembly of *C. savignyi*, generated using a combination of Illumina short-read sequencing, Oxford Nanopore long-read sequencing, and High-Throughput Chromosome Conformation Capture (Hi-C). The final assembly spans 200.63 Mb, with a contig N50 of 1.66 Mb and a scaffold N50 of 14.37 Mb, successfully anchored to 13 chromosomes. Interestingly, despite the compact genome, approximately 45% of the genome consists of repetitive sequences. We predicted 15,327 protein-coding genes, of which 96.2% were functionally annotated. This assembly substantially improves the quality and completeness relative to the previous version and represents one of the most contiguous ascidian genomes assembled to date. This genomic resource provides an essential resource for future research using *C. savignyi* as a model in multiple disciplines.**

Background & Summary

Tunicates (subphylum Urochordata), the closest living invertebrate relatives of vertebrates, occupy a key comparative position for reconstructing the evolutionary origins of vertebrate traits, such as centralized nervous system development, neural crest and placode formation, complex gene regulatory networks, and the establishment of the notochord and dorsal nerve cord[1–5]. Among them, ascidians (class Ascidiacea) serve as versatile model organisms across developmental biology, multi-omics, evolutionary biology, ecology, and invasion and environmental sciences[6]. Their transparent, rapidly developing embryos, compact genomes, and tractable laboratory culture facilitate molecular and developmental studies, while their sessile, filter-feeding adults allow detailed ecological and environmental investigations[7,8]. Moreover, many ascidians have rapidly colonized diverse habitats and adapted to novel conditions, making them valuable for studying invasion biology, adaptive evolution, and stress responses[9–11]. Human-mediated dispersal further exposes them to fluctuating environments, creating quasi-natural experiments to explore adaptation, phenotypic plasticity, and host-microbiome interactions[8,12]. Collectively, tunicates, particularly ascidians, serve as crucial model systems that integrate cross-disciplinary perspectives and provide unique biological insights.

The two most extensively studied ascidians, *Ciona robusta* (formerly *C. intestinalis* type A) and *C. savignyi*, are globally distributed due to their high invasiveness and adaptability to diverse environmental conditions. For example, *C. robusta* has been documented in salinities ranging from 12‰ to 40‰ and temperatures from 3 °C to 30 °C[13,14], including an invasion of the high-salinity, high-temperature Red Sea in 2015[15,16]. Meanwhile, *C. savignyi* has been found to tolerate temperatures from 1.7 °C to 26.7 °C and salinities between 18‰ and 40‰[17,18]. These broad tolerances and high invasiveness underpin their success as model systems for adaptive evolution and environmental stress response.

The draft genome of *C. robusta* was first sequenced and assembled in 2002 using a whole-genome shotgun (WGS) approach[19], representing the first genome sequence of an invertebrate chordate. The initially fragmented

[1]Research Center for Eco-Environmental Sciences, Chinese Academy of Sciences, 18 Shuangqing Road, Haidian District, Beijing, 100085, China. [2]University of Chinese Academy of Sciences, Chinese Academy of Sciences, 19A Yuquan Road, Shijingshan District, Beijing, 100049, China. ✉e-mail: zhanaibin@hotmail.com; azhan@rcees.ac.cn

Library type	Purpose	Clean data (Gb)	Clean data number	Average read length (bp)
WGS Illumina	Genome survey	54.68	369,408,624	150
Nanopore	Draft genome assembly	22.33	4,288,156	5,207
Hi-C	Chromosome-scale scaffolding	24.06	160,403,204	150
ONT RNA-Seq	Transcriptome evidence for annotation	28.37	40,709,166	719

Table 1. Statistics of sequencing data for the *Ciona savignyi* genome assembly and annotation.

assembly was substantially improved in 2008 through the integration of paired-end expressed sequence tags (ESTs), bacterial artificial chromosome (BAC) sequences, and chromosomal *in situ* hybridization data[20]. A further major advancement was achieved in 2019 by incorporating long reads generated from PacBio RSII sequencing and Hi-C chromatin interaction data, yielding a nearly complete genome with 95% of sequences assigned to chromosomes[21]. Subsequently, the gene models in this high-quality assembly were manually curated[22], producing the most complete and accurate *C. robusta* genome to date. This refined genome provides a valuable resource for evolutionary studies, facilitating multi-disciplinary studies[23,24]. In contrast, the first draft genome of *C. savignyi* was sequenced and assembled using a WGS approach later than that of *C. robusta*[25]. The genome, with a size of 157 Mb, exhibited an exceptionally high heterozygosity rate of 4.6%. This initial assembly was highly fragmented due to the extreme polymorphism, but it was subsequently improved using a multistep assembly strategy, producing a 174 Mb nonredundant reference genome comprising 374 reftigs with an N50 of 1.8 Mb[26]. The chromosomes of *C. savignyi* were later reconstructed by integrating a comprehensive genetic map with the reference genome sequence[27]. In contrast to the frequent updates and major improvements made to the *C. robusta* genome over the past two decades, the *C. savignyi* genome has not been updated since 2008, and the CSAV2.0 version (https://asia.ensembl.org/index.html) remains the most widely used reference in recent studies[28–30]. However, its highly fragmented nature has substantially limited its utility for investigating the genetic architecture of complex traits, including stress tolerance, thermal adaptation, and acclimation to novel environments. Furthermore, the high heterozygosity of *C. savignyi*, a potential source of adaptive potential, has historically posed significant challenges for achieving a high-quality genome assembly.

Here, we present a substantially improved, chromosome-scale genome assembly by integrating long reads generated using Oxford Nanopore Technology (ONT), short reads from the Illumina platform, and high-throughput chromosomal conformation capture (Hi-C) data. The assembled genome spans 200.63 Mb, with a contig N50 of 1.66 Mb and a scaffold N50 of 14.37 Mb. Notably, 96% (192.60 Mb) of the total sequence was anchored onto 13 chromosomes, and 89.62 Mb (44.67%) was identified as repetitive elements. A total of 15,327 protein-coding genes were predicted, of which 96.2% are functionally annotated. This high-quality, chromosome-level genome assembly of *C. savignyi* provides a valuable resource for multi-disciplinary studies.

Methods

Sample collection and species identification. Adult *C. savignyi* individuals were collected from the coast of Dalian (38°49′13″N, 121°24′20″E), Liaoning Province, China, in September 2022. All specimens were acclimated for five days under laboratory conditions without feeding. Due to the high morphological similarity between *C. robusta* and *C. savignyi*, molecular identification was performed on all samples following Smith *et al.*[31] to ensure accurate species discrimination. Body wall muscle from a single individual was dissected, immediately frozen in liquid nitrogen, and stored at −80 °C until DNA extraction for genome sequencing and Hi-C analysis. Additionally, five individuals were collected for ONT RNA sequencing (RNA-seq) to assist gene annotation. For these RNA-seq samples, body wall muscle tissue was similarly dissected, flash-frozen in liquid nitrogen, and stored at −80 °C until RNA extraction.

Genome sequencing. Genomic DNA was extracted using the DNeasy Blood & Tissue Kit (Qiagen, Germany), and DNA quality and concentration were assessed by 1.0% agarose gel electrophoresis, NanoDrop 2000 spectrophotometer (Thermo Fisher Scientific, USA), and Qubit 3.0 fluorometer (Life Technologies, USA). Sequencing libraries were then constructed from the extracted DNA. For genome survey, a paired-end library with a 300 bp insert size was prepared using the Nextera DNA Flex Library Prep Kit (Illumina, San Diego, CA, USA) following the manufacturer's instructions and sequenced on the Illumina NovaSeq 6000 platform, yielding 54.68 Gb of clean reads. For ONT long-read sequencing, libraries were prepared *via* DNA end repair and ligation using the ONT Template Prep Kit (SQK-LSK109, UK) according to standard protocols and subsequently sequenced on the ONT PromethION P48 platform (Oxford Nanopore Technologies, Oxford, UK). Raw reads were base-called with Guppy v4 in high-accuracy mode, resulting in 22.33 Gb of long-read data for genome assembly. For Hi-C sequencing, libraries were constructed following the protocol[32] and sequenced on the Illumina NovaSeq 6000 platform with 150 bp paired-end reads, producing 24.06 Gb of high-throughput Hi-C data (Table 1).

RNA extraction and sequencing. For long-read RNA sequencing, total RNA from each ascidian individual was extracted using TRIzol (Thermo Fisher Scientific, USA) following the manufacturer's instructions. RNA quality and quantity were evaluated using an Agilent 2100 Bioanalyzer (Agilent Technologies). Five RNA libraries were then constructed according to the standard protocol provided by Oxford Nanopore Technologies and sequenced on the PromethION platform (Oxford Nanopore Technologies, Oxford, UK). After adapter removal

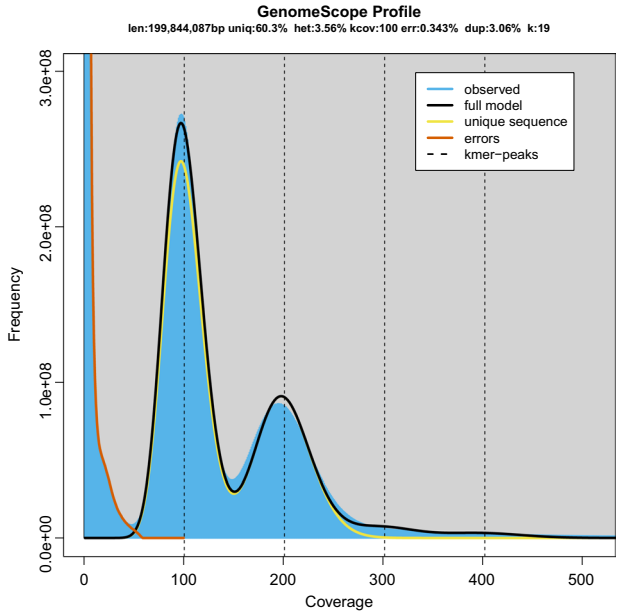

Fig. 1 Genome size and content estimation using GenomeScope. The X-axis is the k-mer depth (k = 19), and Y-axis represents the frequency of the k-mer for a given depth.

K-mer	K-mer Number	K-mer Depth	Genome Size (Mb)	Heterozygous Ratio (%)	Duplication Ratio (%)
19	40,087,585,781	200	199.84	3.56	39.72

Table 2. Key metrics derived from k-mer analysis of the genome.

and filtering of low-quality reads, a total of 28.37 Gb of long-read transcriptome data was obtained, with an average read length of 719 bp (Table 1).

Genome size estimation. Adapters and low-quality reads were removed from the raw Illumina paired-end data using TRIMMOMATIC v0.36[33]. The k-mer method was employed to survey the genomic features of *C. savignyi* using 54.68 Gb of clean data. Genome size, heterozygosity, and repetitive sequence content (Fig. 1) were estimated using Jellyfish v2.2.10 and GenomeScope v1.0 (http://genomescope.org)[34,35]. Based on the formula Genome Size = K-mer number/peak depth, the genome size of *C. savignyi* was estimated at 199.84 Mb, with a repetitive sequence ratio of 39.72% and a heterozygosity rate of 3.56% (Table 2), indicating a highly heterozygous genome.

Chromosome assembly. Nanopore long reads (22.33 Gb) were used for de novo assembly of the *C. savignyi* draft genome with NextDenovo v2.5.0[36]. To enhance assembly accuracy, the draft genome was polished in two rounds with ONT long reads using Racon v1.4.11[37], followed by two cycles of polishing with Illumina short reads using Pilon v1.23 with default parameters[38]. Finally, the two-step polished genome assembly was processed with Purge_Haplotigs v1.0.4[39] to remove heterozygous redundancies, yielding a contig-level draft genome of 200.63 Mb. This assembly consists of 223 contigs, with a contig N50 of 1.66 Mb and a GC content of 37% (Table 3).

Genome scaffolding by Hi-C sequencing. To further anchor the genome assembly at the chromosome level, HiCUP v0.8.0 was first used to remove erroneous mappings and duplicate contigs, generating cleaned mapped data[40]. Approximately 15.44 million valid paired-end reads (~82.15%) were subsequently used to cluster, order, and orient scaffolds onto chromosomes using the 3D-DNA pipeline with default settings[41]. Scaffolds were subsequently refined with Juicebox v1.11.08 to correct instances of chromosome translocations and inversions[42]. Following Hi-C scaffolding, 96% of the 223 contigs were anchored to 13 pseudo-chromosomes (Table 4), which were clearly distinguishable in the chromatin contact matrix heatmap, showing strong interaction signals along the diagonal (Fig. 2). The final chromosome-scale assembly has a total size of 200.63 Mb and a scaffold N50 of 14.37 Mb (Table 3), with genomic features (Fig. 3) visualized using Circos[43]. Hi-C scaffolding substantially improved the assembly N50 from 1.66 Mb to 14.37 Mb (Table 3).

Repeat sequence and non-coding RNA annotation. Repetitive elements in the *C. savignyi* genome were annotated using both ab initio prediction and homology-based approaches. First, a de novo species-specific repeat library was generated using RepeatModeler v2.0.5[44] and LTR_FINDER v1.1[45] with default parameters, and RepeatMasker v4.0.9[46] was employed to annotate novel repeat elements. Second, known repeat elements were identified by querying the genome against the Repbase database using RepeatMasker v4.0.9 in combination with

Labels	Statistics
Draft genome assembly	
Number of contigs	223
Contig N50 (bp)	1,661,210
Contig N90 (bp)	384,193
Maximum contig size (bp)	6,667,610
Total size (bp)	200,629,497
Chromosome-level genome assembly	
GC_content (%)	37
Number of scaffolds	53
Scaffold N50 (bp)	14,366,484
Scaffold N90 (bp)	10,471,772
Number of chromosomes	13
Total size (bp)	200,629,497

Table 3. Summary of the *Ciona savignyi* genome assembly statistics.

Chromosome	Length (bp)	contig number	% of genome
chr1	31,887,471	17	15.89
chr2	16,267,953	16	8.11
chr3	15,348,036	15	7.65
chr4	14,634,314	21	7.29
chr5	14,607,172	23	7.28
chr6	14,366,484	20	7.16
chr7	14,158,382	14	7.06
chr8	13,713,942	14	6.84
chr9	13,704,960	19	6.83
chr10	13,577,752	21	6.77
chr11	11,075,334	14	5.52
chr12	10,471,772	11	5.22
chr13	8,806,540	14	4.39
unplaced	8,029,985	40	4.00

Table 4. Chromosome length statistics for the 13 chromosomes of the *Ciona savignyi* genome.

RepeatProteinMask v4.1.0[47]. The integration of these two approaches yielded a comprehensive repeat annotation, revealing 89.62 Mb of repetitive sequences, accounting for 44.67% of the genome (Table 5). Among these, DNA transposons, LINEs, SINEs, and LTRs represent 4.65%, 2.41%, 4.29%, and 5.94% of the genome, respectively (Table 5).

Non-coding RNAs (ncRNAs) were annotated as follows: tRNA genes were predicted using tRNAscan-SE v1.23 with default parameters[48], and rRNAs were predicted with INFERNAL v1.1.2[49] based on the Rfam database[50]. In total, 3,392 tRNAs, 380 rRNAs, 16 miRNAs, and 434 snRNAs were identified in the *C. savignyi* genome (Table 6).

Gene prediction and annotation. Using the repeat-masked genome, we integrated three methods for the prediction of protein-coding genes within the *C. savignyi* genome, including ab initio predictions, homology-based prediction, and mRNA-based prediction. For ab initio prediction, two gene prediction programs, Augustus v3.3.2[51] with parameters of "–uniqueGeneId = true –noInFrameStop = true –gff3 = on –strand = both" and Genscan v1.0[52] tools with parameter of HumanIso.smat were used to de novo predict genes. For homology-based prediction, we firstly retrieved the protein sequences from five well-annotated species, comprising *Ciona_robusta* (KY21 verssion, https://ghost.zool.kyoto-u.ac.jp/download_ht.html), *Oikopleura_dioica* (GenBank assembly accession: GCA_907165135.1[53]), *Styela_clava* (GenBank assembly accession: GCA_013122585.2[54]), *Branchiostoma_floridae* (GenBank assembly accession: GCA_000003815.2[55]), and *C. savignyi* itself (GenBank assembly accession: GCA_000149265.1[56]). These protein sequences were downloaded and then aligned to the repeat-masked genome assembly in our present study using tBLASTn v2.7.1[57] with an E-value ≤ 1e-5 and Exonerate v2.4.0[58] with the parameters of -model protein2genome –showtargetgff 1. For mRNA-based prediction, a total of 28.37 Gb of long read RNA clean data generated from five individuals was aligned to the assembled genome utilizing minimap2 v2.17[59] with parameters of -ax map-ont -xsplice -G 1000000, and the putative transcript structures were assembled using StringTie v2.1.4[60], subsequently TransDecoder v5.1.0

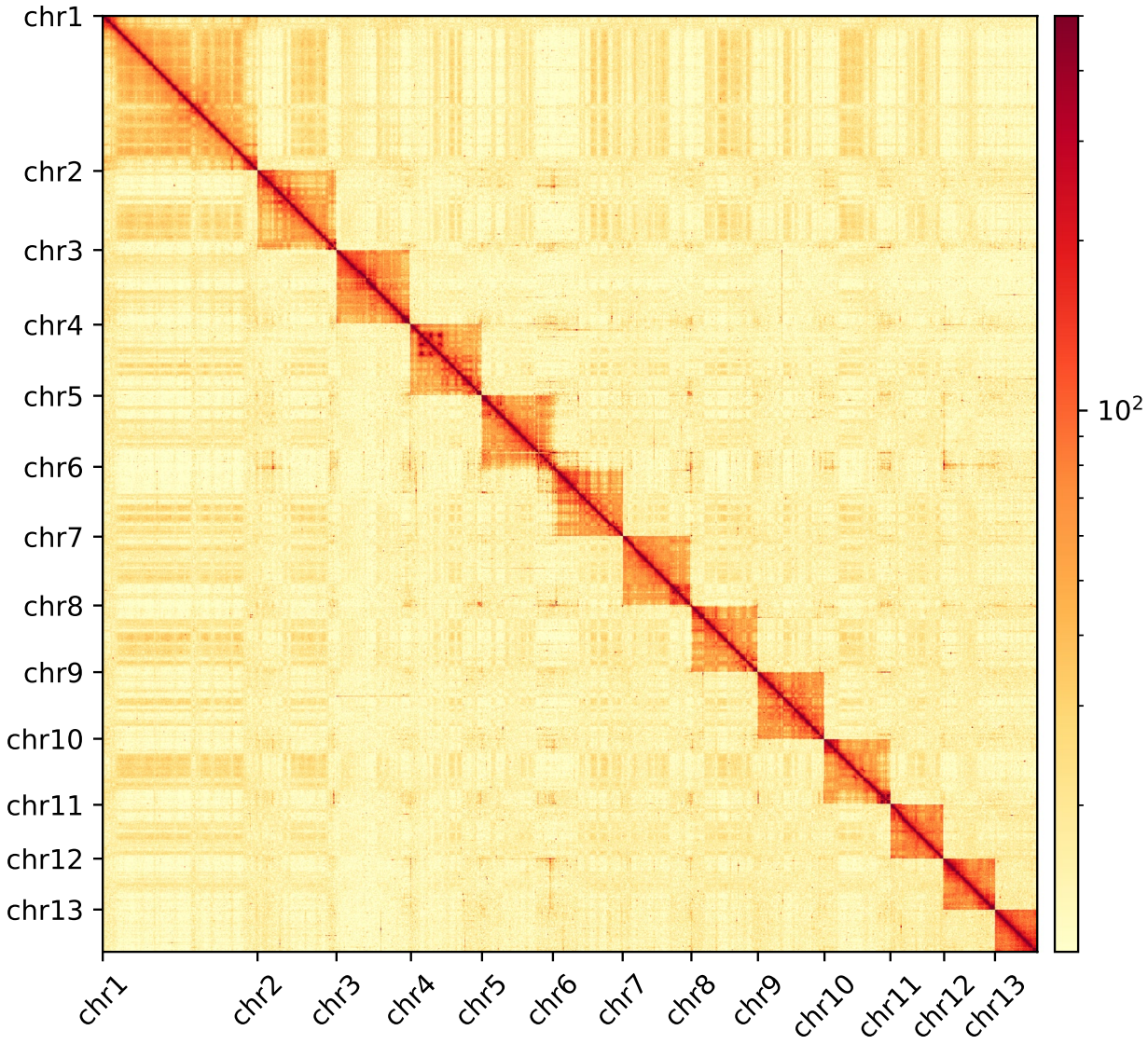

Fig. 2 Hi-C intra-chromosomal contact map of *Ciona savignyi* genome assembly.

(https://github.com/TransDecoder/TransDecoder) was used to predict genes based on the assembled transcripts with default parameters. Finally, MAKER v2.31.10[61] was used to integrate the above three gene annotations to eliminate redundancy with parameters of maker_exe.ctl maker_opts.ctl maker_bopts.ctl –ignore_nfs_tmp -fix_nucleotides, resulting in a weighted and nonredundant final gene set. A total of 15,327 protein-coding genes were predicted with an average gene length of 7,034 bp and an average exon number of 7.82 within a gene (Table 7).

Gene functional annotations. The predicted protein-coding gene sets were functionally annotated using BLASTp v2.0.11[62] with parameter of E-value $\leq$ 1e-5 by searching public biological functional databases, including NCBI nonredundant protein (NR), Swiss-Prot (http://www.gpmaw.com/html/swiss-prot.html), TrEMBL, Gene Ontology (GO) and Kyoto Encyclopedia of Genes and Genomes (KEGG) databases. InterProScan v5.36[63] was used to identify conserved protein domains by integrating multiple databases such as InterPro and Pfam databases. Ultimately, 14,744 genes (96.20% of total predicted genes) were functionally annotated by at least one of these databases (Table 8).

Data Records

The raw Illumina sequencing data for genome survey, Oxford Nanopore sequencing raw data for genome assembly, Hi-C sequencing data and Nanopore long-read RNA-seq data have been deposited on the NCBI Sequence Read Archive database under the accession numbers: SRR35976859[64], SRR35984044[65], SRR35984038[66] and SRR36000873-77[67–71], respectively. The final genome assembly has been deposited in European Nucleotide Archive (ENA) under the accession number GCA_978045305[72]. Both the assembly and its annotation are also publicly available via the Figshare repository (https://doi.org/10.6084/m9.figshare.30596336[73]).

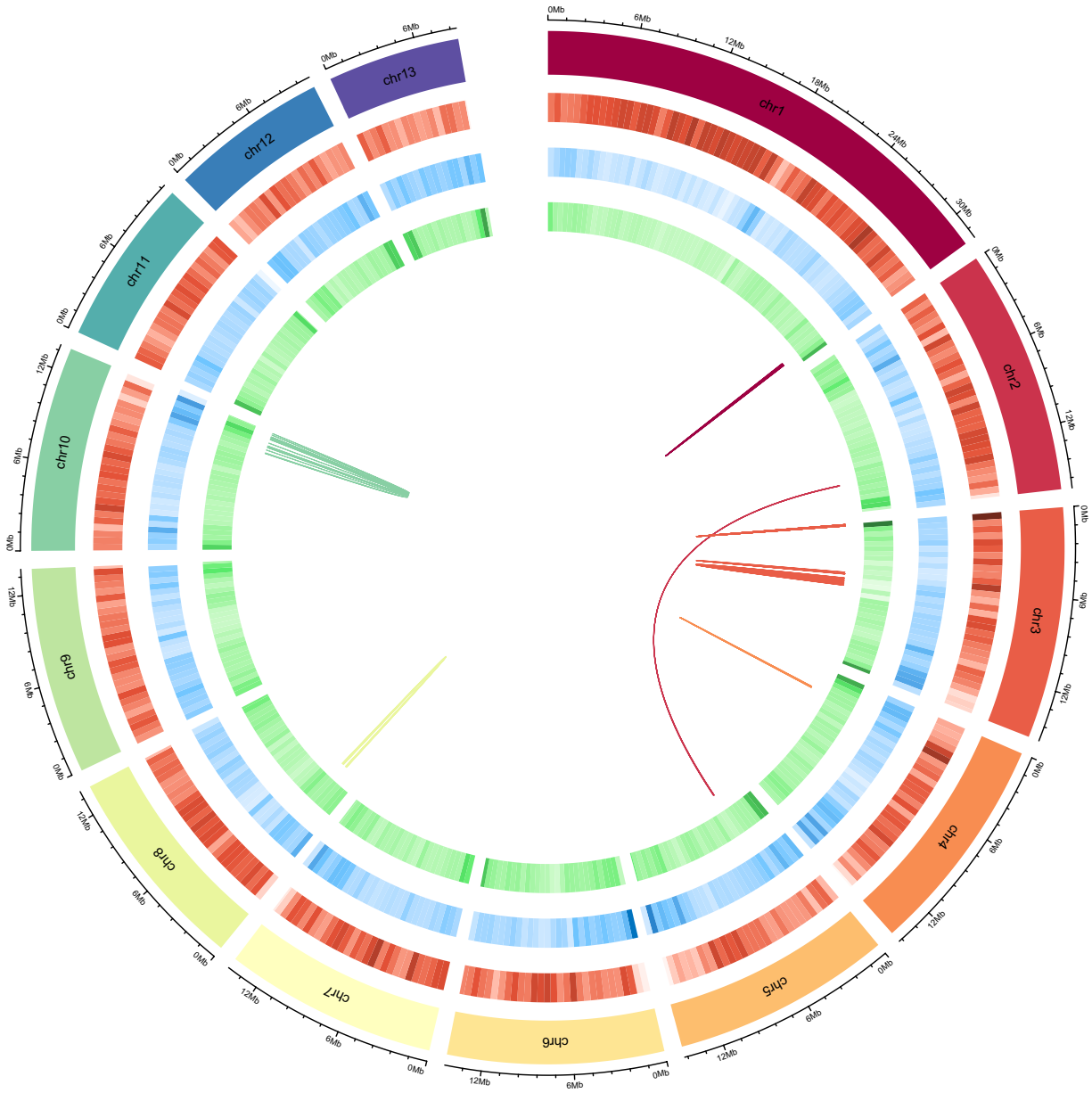

Fig. 3 Chromosome-level genome architecture of *Ciona savignyi*. The Circos plot illustrates, from the outer to the inner layers: (**a**) gene density, (**b**) repeat density, (**c**) GC content, and (**d**) intra-genomic synteny. Panels (**a**–**d**) were calculated using 50-kb sliding windows.

Type	TE protiens		De novo + repbase		Combined TEs	
	Length (bp)	% in genome	Length (bp)	% in genome	Length (bp)	% in genome
DNA	1,390,628	0.69	8,958,294	4.46	9,330,827	4.65
LINE	2,436,823	1.21	4,622,433	2.3	4,825,668	2.41
SINE	0	0	8,605,166	4.29	8,605,166	4.29
LTR	4,359,368	2.17	11,086,576	5.53	11,915,304	5.94
Satellite	0	0	1,090,412	0.54	1,090,412	0.54
Simple repeat	0	0	5,815	0	5,815	0
Other	270	0	0	0	270	0
Unknown	11,427	0.01	49,836,956	24.84	49,848,359	24.84
Total	8,195,461	4.08	79,180,465	39.46	89,624,369	44.67

Table 5. Statistics on transposable elements in the *Ciona savignyi* genome.

Type	Number	Average length (bp)	Total length (bp)	% of genome
miRNA	16	81	1,293	0.000644
tRNA	3,392	75	254,752	0.126963
rRNA	380	224	85,006	0.042365
snRNA	434	150	64,978	0.032384

Table 6. Statistics of non-coding RNAs in the *Ciona savignyi* genome.

Method	Software	Species	Gene number	Average gene length (bp)	Average CDS length (bp)	Average exon per gene	Average exon length (bp)	Average intron length (bp)
Ab initio	Genscan		13,334	10,751.87	1,739.35	8.37	207.69	1,222.08
	AUGUSTUS		21,608	5,745.80	1,394.43	7.28	191.62	693.22
Homology-based	Exonerate	*Oikopleura dioica*	11,220	7,274.71	607.92	3.32	183.25	2,876.76
	Exonerate	*Ciona savignyi*	27,073	10,826.95	842.72	5.03	167.6	2,478.68
	Exonerate	*Styela clava*	22,894	11,144.53	835.21	4.2	198.77	3,219.81
	Exonerate	*Branchiostoma floridae*	19,618	10,141.82	741.16	3.97	186.54	3,161.87
	Exonerate	*Ciona robusta*	28,885	13,452.94	1,025.43	5.27	194.57	2,910.26
RNAseq	TransDecoder		6,409	6,466.80	837.33	6.5	263.42	864.68
Integration	Maker		14,966	7,859.16	1,561.44	8.82	207.11	772.03
Final set	Anno-self		15,327	7,034.22	1,346.86	7.82	214.59	777.72

Table 7. Statistics of gene predictions in the *Ciona savignyi* genome.

Type	Number	Percentage %
NR	13,666	89.16
Uniprot	13,923	90.84
Pfam	11,957	78.01
GO	11,361	74.12
KEGG	8,245	53.79
Pathway	4,632	30.22
Interproscan	13,863	90.45
Annotated	14,744	96.20
Unannotated	583	3.8
All	15,327	100

Table 8. Functional annotation statistics of the *Ciona savignyi* genome.

Technical Validation

The quality and completeness of the *C. savignyi* genome assembly were evaluated across multiple aspects. (i) Sufficient sequencing depth: Illumina genome survey data (54.68 Gb), ONT long reads for genome assembly (22.33 Gb), Hi-C sequencing data (24.06 Gb), and ONT RNA-seq data for gene annotation (28.37 Gb) provided coverage of 272 ×, 111 ×, 120 ×, and 141 ×, respectively, ensuring high-quality assembly, chromosome-scale scaffolding, and comprehensive gene annotation. (ii) High mapping rate: Re-mapping of Illumina reads to the assembly showed that 99.17% of reads aligned, covering 98.78% of the genome, indicating excellent completeness and base-level accuracy. (iii) Improved contiguity: The scaffold N50 reached 14.37 Mb, eight times higher than the previous assembly[26] (contig N50 = 1.8 Mb), and 96% of the genome was contained within the 13 scaffolded chromosomes confirmed by Hi-C, achieving a chromosome-level assembly of exceptional continuity. (iv) BUSCO assessment: Using the metazoa_odb10 database and BUSCO v2.0[74], 88.90% of universal single-copy orthologs were complete (87.50% single-copy, 1.40% duplicated; Table 9), supporting high annotation integrity. Taken together, these assessments demonstrate that the *C. savignyi* genome assembly represents a substantial improvement in quality and completeness compared with previous reference genomes.

Data availability

All sequencing data (Illumina, Nanopore, Hi-C, RNA-seq) and the final genome assembly are available in public repositories: raw reads are in the SRA (SRR35976859, SRR35984044, SRR35984038, SRR36000873-77), the assembly is in the ENA (GCA_97804530572). Moreover, both the assembly and its annotation are also accessible via Figshare (https://doi.org/10.6084/m9.figshare.30596336).

BUSCO parameters	Contig	Hi-C	Annotation
Complete BUSCOs (C)	90.8	87.9	88.9
Complete and single-copy BUSCOs (S)	88.9	87.1	87.5
Complete and duplicated BUSCOs (D)	1.9	0.8	1.4
Fragmented BUSCOs (F)	3.1	3.2	1.6
Missing BUSCOs (M)	6.1	8.9	9.5

Table 9. BUSCO evaluation (%) of genome assemblies and annotations.

Code availability

No custom code or script was used in this study. The data processing commands are executed by the manuals and protocols provided by the creators of the corresponding bioinformatics software, and the parameter settings were outlined in the methods section.

Received: 15 November 2025; Accepted: 28 February 2026;
Published online: 06 March 2026

References

1. Delsuc, F., Brinkmann, H., Chourrout, D. & Philippe, H. Tunicates and not cephalochordates are the closest living relatives of vertebrates. *Nature* **439**, 965–968, https://doi.org/10.1038/nature04336 (2006).
2. Schubert, M., Escriva, H., Xavier-Neto, J. & Laudet, V. Amphioxus and tunicates as evolutionary model systems. *Trends Ecol. Evol.* **21**, 269–277, https://doi.org/10.1016/j.tree.2006.01.009 (2006).
3. Ferrier, D. E. Tunicates push the limits of animal evo-devo. *BMC Biol.* **9**, 3, https://doi.org/10.1186/1741-7007-9-3 (2011).
4. Ambrosino, L. *et al*. Functional conserved non-coding elements among tunicates and chordates. *Dev. Biol.* **448**, 101–110, https://doi.org/10.1016/j.ydbio.2018.12.012 (2019).
5. Kumano, G. & Nishida, H. Ascidian embryonic development: an emerging model system for the study of cell fate specification in chordates. *Dev. Dyn.* **236**, 1732–1747, https://doi.org/10.1002/dvdy.21108 (2007).
6. Corbo, J. C., Gregorio, A. D. & Levine, M. The Ascidian as a Model Organism in Developmental and Evolutionary Biology. *Cell* **106**, 535–538, https://doi.org/10.1016/S0092-8674(01)00481-0 (2001).
7. Lambert, G. Invasive sea squirts: a growing global problem. *J. Exp. Mar. Biol. Ecol.* **342**, 3–4, https://doi.org/10.1016/j.jembe.2006.10.009 (2007).
8. Procaccini, G., Affinito, O., Toscano, F. & Sordino, P. A new animal model for merging ecology and evolution. In: Pontarotti P (ed) *Evolutionary biology—concepts, biodiversity, macroevolution and genome evolution*. Springer, Berlin, pp 91–106 (2011).
9. Zhan, A., Briski, E., Bock, D. G., Ghabooli, S. & MacIsaac, H. J. Ascidians as models for studying invasion success. *Mar. Biol.* **162**, 2449–2470, https://doi.org/10.1007/s00227-015-2734-5 (2015).
10. Huang, X., Platin, R., Unger, A., Shenkar, N. & Zhan, A. Interplay of Multilayered Transcriptomic Plasticity in Response to Temperature Fluctuations During Biological Invasions: A Case Study in Invasive *Styela plicata*. *Mol. Ecol.* **34**, e70013, https://doi.org/10.1111/mec.70013 (2025).
11. Huang, X. *et al*. Rapid response to changing environments during biological invasions: DNA methylation perspectives. *Mol. Ecol.* **26**, 6621–6633, https://doi.org/10.1111/mec.14382 (2017).
12. Colautti, R. I. & Lau, J. A. Contemporary evolution during invasion: evidence for differentiation, natural selection, and local adaptation. *Mol. Ecol.* **24**, 1999–2017, https://doi.org/10.1111/mec.13162 (2015).
13. Carver, C. E., Mallet, A. L. & Vercaemer, B. Biological synopsis of the solitary tunicate *Ciona intestinalis*. Dartmouth, Nova Scotia: Bedford institute of Oceanography (2006).
14. Bouchemousse, S., Bishop, J. D. D. & Viard, F. Contrasting global genetic patterns in two biologically similar, widespread and invasive *Ciona* species (Tunicata, Ascidiacea). *Sci. Rep.* **6**, 24875, https://doi.org/10.1038/srep24875 (2016).
15. Shenkar, N., Shmuel, Y. & Huchon, D. The invasive ascidian *Ciona robusta* recorded from a Red Sea marina. *Mar. Biodivers.* **48**, 2211–2214, https://doi.org/10.1007/s12526-017-0699-y (2018).
16. Chen, Y. *et al*. Rapid microevolution during recent range expansion to harsh environments. *BMC Evol. Biol.* **18**, 187, https://doi.org/10.1186/s12862-018-1311-1 (2018).
17. Lord, J. P., Calini, J. M. & Whitlatch, R. B. Influence of seawater temperature and shipping on the spread and establishment of marine fouling species. *Mar. Biol.* **162**, 2481–2492, https://doi.org/10.1007/s00227-015-2737-2 (2015).
18. Zvyagintsev, A. Y., Sanamyan, K. E. & Kashenko, S. D. On the introduction of the Ascidian *Ciona savignyi* Herdman, 1882 into Peter the Great Bay, Sea of Japan. *Russ. J. Mar. Biol.* **33**, 133–136, https://doi.org/10.1134/S1063074007020083 (2007).
19. Dehal, P. *et al*. The draft genome of *Ciona intestinalis*: insights into chordate and vertebrate origins. *Science* **298**, 2157–2167, https://doi.org/10.1126/science.1080049 (2003).
20. Satou, Y. *et al*. Improved genome assembly and evidence-based global gene model set for the chordate *Ciona intestinalis*: new insight into intron and operon populations. *Genome Biol.* **9**, R152, https://doi.org/10.1186/gb-2008-9-10-r152 (2008).
21. Satou, Y. *et al*. A Nearly Complete Genome of *Ciona intestinalis* Type A (*C. robusta*) Reveals the Contribution of Inversion to Chromosomal Evolution in the Genus *Ciona*. *Genome Biol. Evol.* **11**, 3144–3157, https://doi.org/10.1093/gbe/evz228 (2019).
22. Satou, Y. *et al*. A Manually Curated Gene Model Set for an Ascidian, *Ciona robusta* (*Ciona intestinalis* Type A). *Zoolog. Sci.* **39**, 253–260, https://doi.org/10.2108/zs210102 (2022).
23. Ishida, T. & Satou, Y. Ascidian embryonic cells with properties of neural-crest cells and neuromesodermal progenitors of vertebrates. *Nat. Ecol. Evol.* **8**, 1154–1164, https://doi.org/10.1038/s41559-024-02387-8 (2024).
24. Todorov, L. G., Oonuma, K., Kusakabe, T. G., Levine, M. S. & Lemaire, L. A. Neural crest lineage in the protovertebrate model *Ciona*. *Nature* **635**, 912–916, https://doi.org/10.1038/s41586-024-08111-7 (2024).
25. Vinson, J. P. *et al*. Assembly of polymorphic genomes: Algorithms and application to *Ciona savignyi*. *Genome Res.* **15**, 1127–1135, https://doi.org/10.1101/gr.3722605 (2005).
26. Small, K. S., Brudno, M., Hill, M. M. & Sidow, A. A haplome alignment and reference sequence of the highly polymorphic *Ciona savignyi* genome. *Genome Biol.* **8**, R41, https://doi.org/10.1186/gb-2007-8-3-r41 (2007).
27. Hill, M. M. *et al*. The *C. savignyi* genetic map and its integration with the reference sequence facilitates insights into chordate genome evolution. *Genome Res* **18**, 1369–1379, https://doi.org/10.1101/gr.078576.108 (2008).
28. Huang, X. & Zhan, A. Highly dynamic transcriptional reprogramming and shorter isoform shifts under acute stresses during biological invasions. *RNA Biol* **18**, 340–353, https://doi.org/10.1080/15476286.2020.1805904 (2021).

29. Huang, X., Li, H., Shenkar, N. & Zhan, A. Multidimensional plasticity jointly contributes to rapid acclimation to environmental challenges during biological invasions. *RNA* **29**, 675–690, https://doi.org/10.1261/rna (2023).
30. Zhang, W., Jiang, A., Yu, H. & Dong, B. Comparative Transcriptomic Analysis Reveals the Functionally Segmented Intestine in Tunicate Ascidian. *Int. J. Mol. Sci.* **24**, 6270, https://doi.org/10.3390/ijms24076270 (2023).
31. Smith, K., Thia, J., Gemmill, C., Cary, C. & Fidler, A. Barcoding of the cytochrome oxidase I (COI) indicates a recent introduction of Ciona savignyi into New Zealand and provides a rapid method for *Ciona* species discrimination. *Aquat. Invasions* **7**, 305–313, https://doi.org/10.3391/ai.2012.7.3.002 (2012).
32. Belton, J. M. *et al*. Hi-C: a comprehensive technique to capture the conformation of genomes. *Methods* **58**, 268–276, https://doi.org/10.1016/j.ymeth.2012.05.001 (2012).
33. Bolger, A. M., Lohse, M. & Usadel, B. Trimmomatic: a flexible trimmer for Illumina sequence data. *Bioinformatics* **30**, 2114–2120, https://doi.org/10.1093/bioinformatics/btu170 (2014).
34. Marcais, G. & Kingsford, C. A fast, lock-free approach for efficient parallel counting of occurrences of k-mers. *Bioinformatics* **27**, 764–770, https://doi.org/10.1093/bioinformatics/btr011 (2011).
35. Vurture, G. W. *et al*. GenomeScope: fast reference-free genome profiling from short reads. *Bioinformatics* **33**, 2202–2204, https://doi.org/10.1093/bioinformatics/btx153 (2017).
36. Hu, J. *et al*. NextDenovo: an efficient error correction and accurate assembly tool for noisy long reads. *Genome Biol.* **25**, 107, https://doi.org/10.1186/s13059-024-03252-4 (2024).
37. Vaser, R., Sovic, I., Nagarajan, N. & Sikic, M. Fast and accurate de novo genome assembly from long uncorrected reads. *Genome Res.* **27**, 737–746, https://doi.org/10.1101/gr.214270.116 (2017).
38. Walker, B. J. *et al*. Pilon: An integrated tool for comprehensive microbial variant detection and genome assembly improvement. *PLoS ONE* **9**, e112963, https://doi.org/10.1371/journal.pone.0112963.g001 (2014).
39. Roach, M. J., Schmidt, S. A. & Borneman, A. R. Purge Haplotigs: allelic contig reassignment for third-gen diploid genome assemblies. *BMC Bioinformatics* **19**, 460, https://doi.org/10.1186/s12859-018-2485-7 (2018).
40. Wingett, S. *et al*. HiCUP: pipeline for mapping and processing Hi-C data. *F1000Res*. **4**, 1310, https://doi.org/10.12688/f1000research.7334 (2015).
41. Dudchenko, O. *et al*. De novo assembly of the Aedes aegypti genome using Hi-C yields chromosome-length scaffolds. *Science* **356**, 92–95, https://doi.org/10.1126/science.aal3327 (2017).
42. Durand, N. C. *et al*. Juicebox provides a visualization system for Hi-C contact maps with unlimited zoom. *Cell Syst.* **3**, 99–101, https://doi.org/10.1016/j.cels.2015.07.012 (2016).
43. Krzywinski, M. *et al*. Circos: an information aesthetic for comparative genomics. *Genome res.* **19**, 1639–1645, https://doi.org/10.1101/gr.092759.109 (2009).
44. Flynn, J. M. *et al*. RepeatModeler2 for automated genomic discovery of transposable element families. *Proc. Natl. Acad. Sci.* **117**, 9451–9457, https://doi.org/10.1073/pnas.1921046117 (2020).
45. Xu, Z. & Wang, H. LTR_FINDER: an efficient tool for the prediction of full-length LTR retrotransposons. *Nucleic Acids Res.* **35**, W265–268, https://doi.org/10.1093/nar/gkm286 (2007).
46. Tarailo-Graovac, M. & Chen, N. Using RepeatMasker to identify repetitive elements in genomic sequences. *Curr. Protoc. Bioinformatics* **4**, 4.10.1–4.10.14, https://doi.org/10.1002/0471250953.bi0410s25 (2009).
47. Bao, W., Kojima, K. K. & Kohany, O. Repbase update, a database of repetitive elements in eukaryotic genomes. *Mobile DNA* **6**, 11, https://doi.org/10.1186/s13100-015-0041-9 (2015).
48. Chan, P. P., Lin, B. Y., Mak, A. J. & Lowe, T. M. tRNAscan-SE 2.0: improved detection and functional classification of transfer RNA genes. *Nucleic Acids Res.* **49**, 9077–9096, https://doi.org/10.1093/nar/gkab688 (2021).
49. Nawrocki, E. P. & Eddy, S. R. Infernal 1.1: 100–fold faster RNA homology searches. *Bioinformatics* **29**, 2933–2935, https://doi.org/10.1093/bioinformatics/btt509 (2013).
50. Griffiths-Jones, S. *et al*. Rfam: annotating non-coding RNAs in complete genomes. *Nucleic Acids Res.* **33**, D121–124, https://doi.org/10.1093/nar/gki081 (2005).
51. Stanke, M. *et al*. AUGUSTUS: ab initio prediction of alternative transcripts. *Nucleic Acids Res.* **34**, W435–W439, https://doi.org/10.1093/nar/gkl200 (2006).
52. Burge, C. & Karlin, S. Prediction of complete gene structures in human genomic DNA. *J. Mol. Biol.* **268**, 78–94, https://doi.org/10.1006/jmbi.1997.0951 (1997).
53. *NCBI GenBank* https://identifiers.org/ncbi/insdc.gca:GCA_907165135.1 (2021).
54. *NCBI GenBank* https://identifiers.org/ncbi/insdc.gca:GCA_013122585.2 (2021).
55. *NCBI GenBank* https://identifiers.org/ncbi/insdc.gca:GCA_000003815.2 (2020).
56. *NCBI GenBank* https://identifiers.org/ncbi/insdc.gca:GCA_000149265.1 (2004).
57. Camacho, C. *et al*. BLAST+: architecture and applications. *BMC Bioinformatics* **10**, 421, https://doi.org/10.1186/1471-2105-10-421 (2009).
58. Slater, G. S. C. & Birney, E. Automated generation of heuristics for biological sequence comparison. *BMC bioinformatics* **6**, 31, https://doi.org/10.1186/1471-2105-6-31 (2005).
59. Li, H. Minimap and miniasm: Fast mapping and de novo assembly for noisy long sequences. *Bioinformatics* **32**, 2103–2110, https://doi.org/10.1093/bioinformatics/btw152 (2016).
60. Pertea, M. *et al*. StringTie enables improved reconstruction of a transcriptome from RNA-seq reads. *Nat. Biotechnol.* **33**, 290–295, https://doi.org/10.1038/nbt.3122 (2015).
61. Holt, C. & Yandell, M. MAKER2: an annotation pipeline and genome-database management tool for second-generation genome projects. *BMC Bioinformatics* **12**, 491, https://doi.org/10.1186/1471-2105-12-491 (2011).
62. Kent, W. J. BLAT—the BLAST-like alignment tool. *Genome Res.* **12**, 656–664, https://doi.org/10.1101/gr.229202 (2002).
63. Jones, P. *et al*. InterProScan 5: genome–scale protein function classification. *Bioinformatics* **30**, 1236–1240, https://doi.org/10.1093/bioinformatics/btu031 (2014).
64. *NCBI Sequence Read Archive* https://www.ncbi.nlm.nih.gov/sra/?term=SRR35976859 (2025).
65. *NCBI Sequence Read Archive* https://www.ncbi.nlm.nih.gov/sra/?term=SRR35984044 (2025).
66. *NCBI Sequence Read Archive* https://www.ncbi.nlm.nih.gov/sra/?term=SRR35984038 (2025).
67. *NCBI Sequence Read Archive* https://www.ncbi.nlm.nih.gov/sra/?term=SRR36000873 (2025).
68. *NCBI Sequence Read Archive* https://www.ncbi.nlm.nih.gov/sra/?term=SRR36000874 (2025).
69. *NCBI Sequence Read Archive* https://www.ncbi.nlm.nih.gov/sra/?term=SRR36000875 (2025).
70. *NCBI Sequence Read Archive* https://www.ncbi.nlm.nih.gov/sra/?term=SRR36000876 (2025).
71. *NCBI Sequence Read Archive* https://www.ncbi.nlm.nih.gov/sra/?term=SRR36000877 (2025).
72. *European Nucleotide Archive* http://www.ebi.ac.uk/ena/browser/view/GCA_978045305 (2026).
73. Huang, X. Chromosome-level genome assembly of invasive ascidian *Ciona savignyi*. *figshare*. https://doi.org/10.6084/m9.figshare.30596336.v1 (2025).
74. Simao, F. A., Waterhouse, R. M., Ioannidis, P., Kriventseva, E. V. & Zdobnov, E. M. BUSCO: assessing genome assembly and annotation completeness with single-copy orthologs. *Bioinformatics* **31**, 3210–3212, https://doi.org/10.1093/bioinformatics/btv351 (2015).

Acknowledgements
This work was supported by the National Natural Science Foundation of China (grant numbers 32101352 and 32561143021).

Author contributions
X.H. and A.Z. conceived this study and designed the experiment. X.H. conducted the experiment and analyzed the data. X.H. wrote the first draft of this paper, and A.Z and X.H. revised the manuscript.

Competing interests
The authors declare no competing interests.

Additional information
Correspondence and requests for materials should be addressed to A.Z.

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.