

# Chromosome-level genome assembly of the common chiton, *Liolophura japonica* (Lischke, 1873)

Hong Kong Biodiversity Genomics Consortium<sup>\*,†</sup>

## ABSTRACT

Chitons (Polyplacophora) are marine molluscs that can be found worldwide from cold waters to the tropics, and play important ecological roles in the environment. However, only two chiton genomes have been sequenced to date. The chiton *Liolophura japonica* (Lischke, 1873) is one of the most abundant polyplacophorans found throughout East Asia. Our PacBio HiFi reads and Omni-C sequencing data resulted in a high-quality near chromosome-level genome assembly of ~609 Mb with a scaffold N50 length of 37.34 Mb (96.1% BUSCO). A total of 28,233 genes were predicted, including 28,010 protein-coding ones. The repeat content (27.89%) was similar to that of other Chitonidae species and approximately three times lower than that of the Hanleyidae chiton genome. The genomic resources provided by this work will help to expand our understanding of the evolution of molluscs and the ecological adaptation of chitons.

**Subjects** Genetics and Genomics, Animal Genetics, Marine Biology

**Submitted:** 11 January 2024  
**Accepted:** 15 April 2024  
**Published:** 25 April 2024

\* Correspondence on behalf of the consortium. E-mail: [jeromehui@cuhk.edu.hk](mailto:jeromehui@cuhk.edu.hk)

† Collaborative Authors: Entomological experts who validated the dataset and their affiliations appears at the end of the document

Published by GigaScience Press.

Preprint submitted at <https://doi.org/10.1101/2024.01.15.575488>

Included in the series: **Hong Kong Biodiversity Genomics** ([https://doi.org/10.46471/GIGABYTE\\_SERIES\\_0006](https://doi.org/10.46471/GIGABYTE_SERIES_0006))

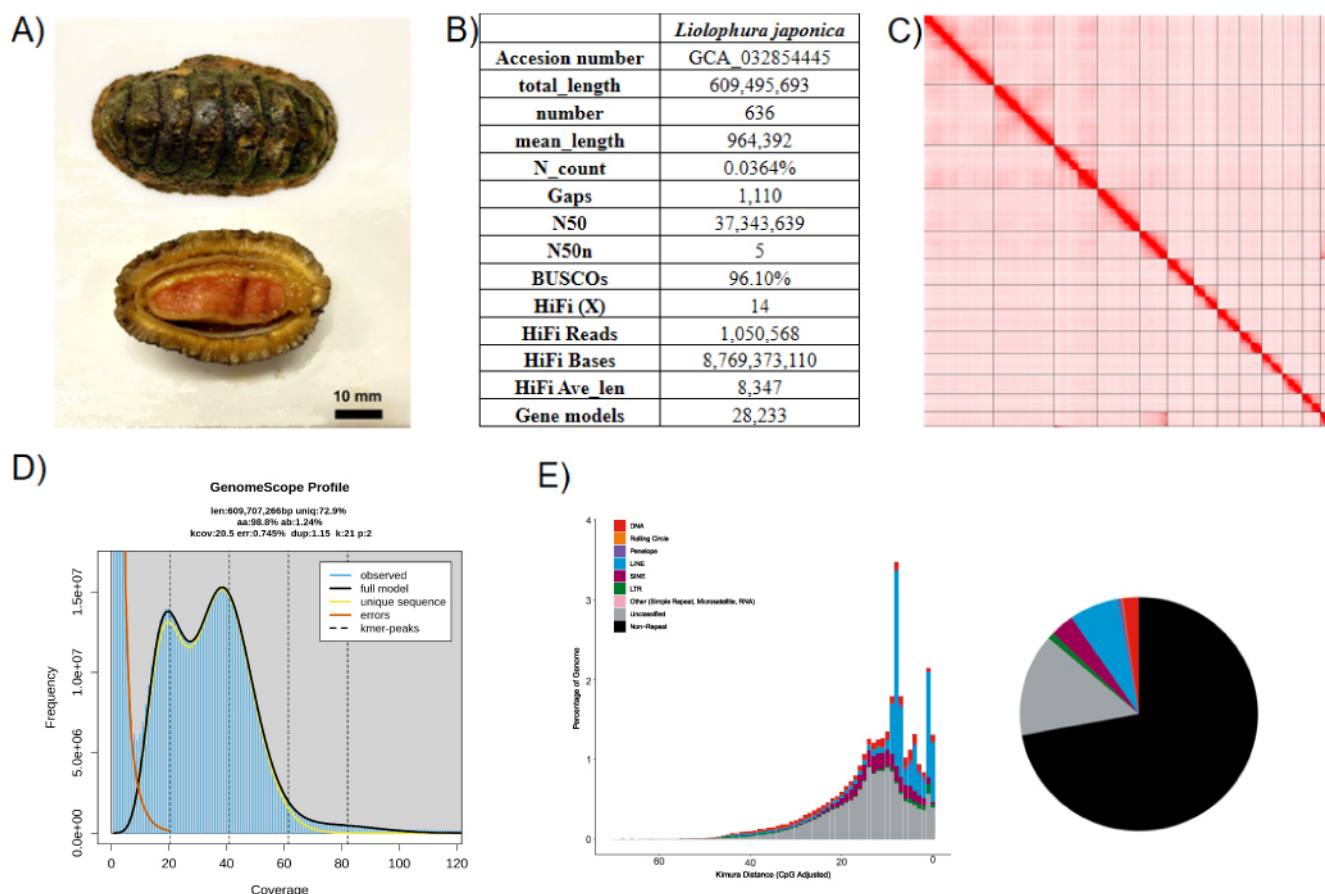
This is an Open Access article distributed under the terms of the Creative Commons Attribution License (<https://creativecommons.org/licenses/by/4.0/>), which permits unrestricted reuse, distribution, and reproduction in any medium, provided the original work is properly cited.

*Gigabyte*, 2024, 1–14

## INTRODUCTION

Mollusca is the second largest animal phylum after Arthropoda and is divided into two subphyla, including the Conchifera (Monoplacophora, Bivalvia, Gastropoda, Scaphopoda, and Cephalopoda) and the Aculifera (Polyplacophora, Caudofoveata, and Solenogastres) [1–3]. Within the latter, chitons (Polyplacophora) are thought to be a relatively early diverging group of living molluscs [4]. They play a crucial role in shaping marine communities in both intertidal and subtidal systems worldwide [5, 6]. These “living fossils” are characterised by a highly evolutionary-conserved and unique type of shell formed by eight articulating aragonite plates that protect from environmental threats [7–9]. This biomineralized armour incorporates an unpigmented sensory system known as aesthetes, which is found in all chiton species [10] as a light-sensing adaptation [11]. Some members of the families Schizochitonidae and Chitonidae developed shell eyes with aragonite-based lenses, allowing the light to focus onto the pigmented photoreceptive retina [10, 12]. However, contrary to other molluscs, our understanding of the Polyplacophora is constrained to only two available genomes (*Acanthopleura granulata* and *Hanleya hanleyi*) [13, 14].

*Liolophura japonica* (Polyplacophora, Chitonidae) (Lischke, 1873) (Figure 1A) is one of the most abundant polyplacophorans found on the intertidal rocky reefs of the Asian continent, including China, Korea, and Japan. This species diverged from the last common ancestor of *Liolophura* ~184 million years ago during the early Pleistocene period [15]. On the shore, they are distributed over a wide vertical range from the mid-littoral zone to the



**Figure 1.** (A) Picture of *Liolophura japonica*; (B) Statistics of the genome assembly generated in this study; (C) Hi-C contact map of the assembly visualised using Juicebox (v1.11.08); (D) Genomescope report with k-mer = 21; (E) Repetitive elements distribution.

low-subtidal zone, where the animals experience periodic fluctuations in environmental conditions with the ebb and flow of the tide [16, 17]. Unlike many other mobile species inhabiting rocky shores, which migrate towards the low shore areas during summer [18, 19], *L. japonica* does not show any significant seasonal migration behaviour and can survive stressful low tide periods presumably by fitting itself into small refuges via its eight flexible, interlocking plates [17, 20, 21]. In terms of its feeding biology, it is a generalist consumer feeding on a wide range of microalgae and macroalgae [16, 22] and, owing to their high density, it is an important grazer in the intertidal zone of East Asia, contributing to the control of on-shore primary productivity [22, 23]. Given the importance of this species, the high-quality genome presented in this study will help to expand our understanding of the evolution of molluscs and the ecological adaptation of chitons.

## CONTEXT

Here, we report the assembled genome of the chiton *L. japonica* (Polyplacophora, Chitonidae) (Lischke 1873) (Figure 1A). *L. japonica* was selected to be one of the species sequenced by the Hong Kong Biodiversity Genomics Consortium (also known as

EarthBioGenome Project Hong Kong), which is organised by researchers from eight publicly funded universities in Hong Kong. The *L. japonica* genome presented in this study is of high quality and near chromosomal level, providing a valuable resource for the understanding of the evolutionary biology of polyplacophorans and the adaptation of its resilience under oscillating environmental changes in the intertidal zones.

## METHODS

### Collection and storage of samples, isolation of high molecular weight genomic DNA, quantification, and qualification

The chitons *L. japonica* were collected at the rocky shore in Kau Sai Chau, Hong Kong (22.380 °N, 114.310 °E) during the summer of 2022. They were kept in 35 ppt artificial seawater at room temperature until DNA isolation. High molecular weight (HMW) genomic DNA was isolated from a single individual. The foot muscle was first frozen in liquid nitrogen and ground to powder. DNA extraction was carried out with the sample powder using a Qiagen MagAttract HMW kit (Qiagen Cat. No. 67563) following the manufacturer's protocol with some modifications.

Around 1 g of sample was first put in 200 µl 1× PBS and mixed with RNase A, Proteinase K, and Buffer AL provided in the kit. The mixture was allowed to sit at room temperature (~22 °C) for 2 h. Next, the mixture was gently flicked every 30 min to allow thorough mixing of samples and digestion solution. The DNA was then isolated from the lysate with the magnetic beads provided in the kit and a magnet rack. Finally, the sample was eluted with 120 µl of elution buffer (PacBio Ref. No. 101-633-500). To prevent unintended DNA shearing during the extraction process, wide-bore tips were consistently employed whenever DNA transfer occurred. Next, the sample was quantified by the Qubit® Fluorometer, Qubit™ dsDNA HS, and BR Assay Kits (Invitrogen™ Cat. No. Q32851). Overnight pulse-field gel electrophoresis was used to examine the molecular weight of the isolated DNA, together with three DNA markers ( $\lambda$ -Hind III digest; Takara Cat. No. 3403, DL15,000 DNA Marker; Takara Cat. No. 3582A and CHEF DNA Size Standard-8-48 kb Ladder; Cat. No. 170-3707). The purity of the sample was examined by the NanoDrop™ One/OneC Microvolume UV-Vis Spectrophotometer, with A260/A280: ~1.8 and A260/A230: >2.0 as a standard.

### DNA shearing, library preparation, and sequencing

A total of 120 µl of DNA sample with 6.2 µg DNA was transferred to a g-tube (Covaris Part No. 520079). The sample was then subjected to six centrifugation steps for 2 min each at 2,000 × g. The resultant DNA was collected and stored in a 2 mL DNA LoBind® Tube (Eppendorf Cat. No. 022431048) at 4 °C until library preparation. Overnight pulse-field gel electrophoresis was used to examine the molecular weight of the isolated DNA, as described in the previous section. The electrophoresis profile was set as follows: 5 K as the lower end and 100 K as the higher end for the designated molecular weight; Gradient = 6.0 V/cm; Run time = 15 h:16 min; included angle = 120°; Int. Sw. Tm = 22 s; Fin. Sw. Tm = 0.53 s; Ramping factor: a = Linear. The gel was run in 1.0% PFC agarose in 0.5× TBE buffer at 14 °C. A SMRTbell library was constructed using the SMRTbell® prep kit 3.0 (PacBio Ref. No. 102-141-700), following the manufacturer's protocol.

The genomic DNA was first subjected to DNA repair to remove single-stranded overhangs and repair damages from the shearing step on the DNA backbone. After repair, both ends of

Table 1. Details of genome and transcriptome sequencing data.				
Genome sequencing data				
Liabrary	No. of reads	No. of bases	Accession	Coverage (X)
PacBio HiFi	1,050,568	8,769,373,110	SRX20411988	14
Omic	264,637,334	39,695,600,100	SRX21911526	65
Transcriptome sequencing data				
Sample name	No. of reads	No. of bases	Accession	Tissue type
Lj2HS_Dg_T	35,899,054	5,384,856,396	SAMN35319765	Digestive gland
Lj2HS_Ft_T	35,289,414	5,293,410,229	SAMN35319766	Foot
Lj2HS_Gl_T	34,240,482	5,136,070,054	SAMN35319767	Gill
Lj2HS_Gn_T	31,128,962	4,669,342,254	SAMN35319768	Gonad
Lj2HS_Ht_T	37,837,458	5,675,616,543	SAMN35319769	Heart

the DNA were polished and tailed with an A-overhang. The ligation of T-overhang SMRTbell adapters was performed at 20 °C for 30 min. Next, the SMRTbell library was purified with SMRTbell® cleanup beads (PacBio Ref. No. 102158-300). The concentration and size of the library were examined using the Qubit® Fluorometer, Qubit™ dsDNA HS, and BR Assay Kits (Invitrogen™ Cat. No. Q32851), and the pulse-field gel electrophoresis, respectively. A subsequent nuclease treatment step was performed to remove any non-SMRTbell structure in the library mixture. A final size-selection step was carried out to remove the small DNA fragments in the library with 35% AMPure PB beads [24]. The Sequel® II binding kit 3.2 (PacBio Ref. No. 102-194-100) was used for the final preparation for sequencing. Sequel II primer 3.2 and Sequel II DNA polymerase 2.2 were annealed and bound to the SMRTbell library, respectively. The library was loaded at an on-plate concentration of 50–90 pM using the diffusion loading mode. The sequencing was conducted on the Sequel IIE System with an internal control provided in the binding kit. The sequencing was set up and performed in 30-hour movies (with 120 min pre-extension) with the software SMRT Link v11.0 (PacBio). HiFi reads were generated and collected for further analysis. One SMRT cell was used for this sequencing. Detailed sequencing data can be found in Table 1.

### Omni-C library preparation and sequencing

An Omni-C library was constructed using the Dovetail® Omni-C® Library Preparation Kit (Dovetail Cat. No. 21005) according to the manufacturer’s instructions. Around 20 mg of flash-freezing powered tissue sample was added into 1 mL 1× PBS, where the genomic DNA was crosslinked with formaldehyde, and the fixed DNA was digested with endonuclease DNase I. Next, the concentration and fragment size of the digested sample were checked by the Qubit® Fluorometer, Qubit™ dsDNA HS, and BR Assay Kits (Invitrogen™ Cat. No. Q32851), and the TapeStation D5000 HS ScreenTape, respectively. Following the quality examination, both ends of the DNA were polished. Ligation of a biotinylated bridge adaptor was conducted at 22 °C for 30 min, and the subsequent proximity ligation between crosslinked DNA fragments was performed at 22 °C for 1 h. After the ligation events, the DNA was reverse crosslinked and then purified with SPRIselect™ Beads (Beckman Coulter Product No. B23317) to remove the biotin that was not internal to the ligated fragments. The Dovetail™ Library Module for Illumina (Dovetail Cat. No. 21004) was used for the end repair and adapter ligation. During this process, the DNA was tailed with an A-overhang, which allowed Illumina-compatible adapters to ligate to the DNA fragments at 20 °C for 15 min. The Omni-C library was then sheared into small fragments with USER Enzyme Mix and purified with SPRIselect™ Beads. Next, Streptavidin Beads were added to isolate the DNA

fragments with internal biotin. Universal and Index PCR Primers from the Dovetail™ Primer Set for Illumina (Dovetail Cat. No. 25005) were used to amplify the library. The final size selection step was carried out with SPRIselect™ Beads to pick only the DNA fragments ranging between 350 bp and 1,000 bp. At last, the concentration and fragment size of the sequencing library were examined by the Qubit® Fluorometer, Qubit™ dsDNA HS, and BR Assay Kits, and the TapeStation D5000 HS ScreenTape, respectively. After the quality check, the library was sequenced on an Illumina HiSeq-PE150 platform. Detailed sequencing data information can be found in Table 1.

### Transcriptome sequencing

Total RNA and small RNA (<200 nt) from different tissues (i.e., digestive gland, foot, gill, gonad, and heart) of the other individual were isolated using the TRIzol reagent (Invitrogen) and the mirVana™ miRNA Isolation Kit (Ambion) following the manufacturer's protocol respectively. The quality of the extracted RNA was checked using NanoDrop™ One/OneC Microvolume UV-Vis Spectrophotometer (Thermo Scientific™ Cat. No. ND-ONE-W) and gel electrophoresis. The qualified transcriptome samples were sent to Novogene Co. Ltd (Hong Kong, China) for the library construction for polyA-selected RNA sequencing using the TruSeq RNA Sample Prep Kit v2 (Illumina Cat. No. RS-122-2001), and 150 bp paired-end sequencing. Agilent 2100 Bioanalyser (Agilent DNA 1000 Reagents) was used to measure the insert size and concentration of the final libraries. Details of the sequencing data are listed in Table 1.

### Genome assembly, gene model prediction, repeat analysis, and genome size estimation

*De novo* genome assembly was performed using Hifiasm (version 0.16.1-r375; RRID:SCR\_021069) with default parameters [25]. Haplotypic duplications were identified and removed using purge\_dups (RRID:SCR\_021173) based on the depth of HiFi reads [26]. Proximity ligation data from the Omni-C library were used to scaffold the genome assembly by YaHS (version 1.2a.2) with default parameters [27]. Transposable elements (TEs) were annotated as previously described using the automated Earl Grey TE annotation pipeline (version 1.2) [28]. The mitochondrial genome was assembled using MitoHiFi (v2.2) [29]. RNA sequencing data were first processed with Trimmomatic (version 0.39; RRID:SCR\_011848) with parameters “TruSeq3-PE.fa:2:30:10 SLIDINGWINDOW:4:5 LEADING:5 TRAILING:5 MINLEN:25” [30]. Gene models were then trained and predicted by funannotate (version 1.8.15) [31] using the parameters “--repeats2evm --protein\_evidence uniprot\_sprot.fasta --genemark\_mode ET --optimize\_augustus --organism other --max\_intronlen 350000”. Briefly, the genome was soft-masked by redmask (v0.0.2) [32]. In the funannotate-train step, the transcripts assembled by Trinity (v2.8.5; RRID:SCR\_013048) with parameters “--stranded RF --max\_intronlen 350000” [33] were used to map to the repeat soft-masked genome by minimap2 (v2.2.1; RRID:SCR\_018550) with default parameters [34]. Next, the Trinity transcript alignments were converted to the GFF3 format and used as input for the PASA alignment (v2.5.3; RRID:SCR\_014656) with default parameters [35] in the Launch\_PASA\_pipeline.pl process to get the PASA models trained by TransDecoder (v5.7.1; RRID:SCR\_017647) with default parameters [36]. Finally, Kallisto TPM data (v 0.46.1; RRID:SCR\_016582) [37] was used to select the PASA gene models. The PASA gene models were used to train Augustus (RRID:SCR\_008417) in the funannotate-predict step. The gene



models from several prediction sources, including GeneMark (v3.68\_lic; RRID:SCR\_011930) [38], high-quality Augustus predictions, PASA (v2.5.3) [35], Augustus (v3.5.0) [39], GlimmerHM (v 3.0.4) [40], and SNAP (v2006-07-28; RRID:SCR\_007936) [41], were passed to Evidence Modeler (v1.1.1; RRID:SCR\_014659) to generate the gene model annotation files. Untranslated regions (UTRs) were captured in the funannotate-update step using PASA. Briefly, PASA was run twice in the funannotate-update step, and the transcripts generated in the previous funannotate-train step were mapped to the genome. These data were automatically parsed and used to update the UTR data using the PASA comparison method according to the PASA built-in process. In total, 10,327 UTRs were updated in this study. The protein-coding genes were searched with BLASTP (RRID:SCR\_001010) against the NR and swissprot databases by DIAMOND (v0.9.24; RRID:SCR\_016071) [42] with parameters “--more-sensitive --evalue  $1 \times 10^{-3}$ ” and mapped by HISAT2 (version 2.1.0; RRID:SCR\_015530) [43] with transcriptome reads. A total of 73.9% and 55.8% of the 28,032 protein-coding genes were mapped to the NR and swissprot databases, respectively. The BUSCO (RRID:SCR\_015008) of the protein-coding genes was 90.9%, including 83.5% complete and single-copy genes, 7.4% complete and duplicated genes, 5.0% fragmented genes, while 4.1% of the BUSCOs genes were missed (metazoa\_odb10 with 954 total BUSCOs genes) (v 5.5.0) [44].

TEs were annotated as previously described [28] using the automated Earl Grey TE annotation pipeline (version 1.2) with “-r eukarya” to search the initial mask of known elements and other default parameters. Briefly, this pipeline first identified known TEs from Dfam (RRID:SCR\_021168) with RBRM (release 3.2) and Repbase (v20181026; RRID:SCR\_021169). *De novo* TEs were then identified, and the consensus boundaries were extended using an automated BLAST (RRID:SCR\_004870), Extract, and Extend process with five iterations and 1,000 flanking bases added at each round. Redundant sequences were removed from the consensus library before the genome assembly was annotated with the combined known and *de novo* TE libraries. Annotations were processed to remove overlap and defragment annotations prior to the final TE quantification.

The 21 k-mer count and histogram were generated using Omni-C reads and Jellyfish (v2.3.0; RRID:SCR\_005491) with the parameters “count -C -m 21 -s 1000000000 -t 10” [45]. Next, the reads.histo were uploaded to GenomeScope (v2.0; RRID:SCR\_017014) to estimate the genome heterozygosity, repeat content, and size using default parameters [46].

Hi-C contact maps were generated using the Juicer tools (version 1.22.01; RRID:SCR\_017226), following the Omni-C manual [47]. Briefly, Omni-C reads were mapped and aligned by BWA (RRID:SCR\_010910) with parameters “mem -5SP -T0”. Next, the parsing module of the pairtools pipeline (v1.0.2) [48] was used to find ligation junctions with parameters “--min-mapq 40 --walks-policy 5unique --max-inter-align-gap 30 --nproc-in 8 --nproc-out 8”. The parsed pairs were then sorted using pairtools sort with default parameters, and PCR duplicate pairs were removed using pairtools dedup with parameters “--nproc-in 8 --nproc-out 8 --mark-dups”. The pairs’ file was generated using pairtools split with default parameters and used to generate the contact matrix using juicertools and Juicebox (v1.11.08; RRID:SCR\_021172) [49].

## RESULTS AND DISCUSSION

A total of 8.77 Gb of HiFi bases of the common chiton *L. japonica* were generated with an average HiFi read length of 8,347 bp with 14X data coverage. After scaffolding with ~397 Gb

Table 2. Summary of the genome statistics of <i>Liolophura japonica</i> , <i>Acanthopleura granulata</i> [13], and <i>Hanleya hanleyi</i> [14].			
	<i>Liolophura japonica</i>	<i>Acanthopleura granulata</i>	<i>Hanleya hanleyi</i>
Total length (bp)	609,495,693	606,536,932	2,516,608,230
Number	632	87	57,495
Mean length (bp)	964,392	6,971,689	43,771
Longest	96,433,062	50,906,754	801,547
Shortest	1,000	43,755	1,532
N count	0.0364%	10.1930%	0.0005%
Gaps	1,110	5,849	216
N50	37,343,639	23,921,462	65,037
N50n	5	9	10,427
BUSCO (genome, metazoa_odb10)	C:96.1% [S:95.2%, D:0.9%], F:1.9%, M:2.0%, n:954	C:96.0% [S:95.3%, D:0.7%], F:2.1%, M:1.9%, n:954	C:80.0% [S:74.9%, D:5.1%], F:12.1%, M:7.9%, n:954
Total length of protein-coding genes (AA)	12,529,260	17,148,789	22,920,296
Number of protein-coding genes	28,010	37,872	69,284
Mean length of protein-coding genes (AA)	447	453	331
BUSCO (Proteome, metazoa_odb10)	C:90.9% [S:83.5%, D:7.4%], F:5.0%, M:4.1%, n:954	C:93.4% [S:71.4%, D:22.0%], F:2.8%, M:3.8%, n:954	C:81.8% [S:75.6%, D:6.2%], F:11.1%, M:7.1%, n:954
Reference	This study	[13]	[14]

Table 3. Scaffold information of 13 pseudomolecules.			
Chr_number	scaffold id	scaffold length	Cumulative % of the whole genome
1	scaffold_1	96,433,062	15.82%
2	scaffold_2	82,493,220	29.35%
3	scaffold_3	60,359,117	39.25%
4	scaffold_4	58,388,612	48.83%
5	scaffold_5	37,343,639	54.96%
6	scaffold_6	36,638,032	60.97%
7	scaffold_7	32,819,725	66.35%
8	scaffold_8	31,077,156	71.45%
9	scaffold_9	30,102,928	76.39%
10	scaffold_10	28,395,274	81.05%
11	scaffold_11	27,489,247	85.56%
12	scaffold_12	25,057,479	89.67%
13	scaffold_13	19,479,325	92.86%

Omni-C data, the assembled genome size was 609.5 Mb, with 632 scaffolds, a scaffold N50 of 37.34 Mb, and the complete BUSCO estimation of 96.1% (metazoa\_odb10) (Figure 1B; Table 2). A total of 13 pseudomolecules of chromosomal length were anchored from the Omni-C data (Figure 1C; Table 3). This result is close to the karyotype of *L. japonica* ( $2n = 24$ ), indicating the assembly is near chromosome-level. The assembled *L. japonica* genome has a genome size close to the estimation performed by GenomeScope, which was 609.7 Mb with a heterozygosity rate of 1.24% (Figure 1D; Table 4), and similar to the *Acanthopleura granulata* chiton genome (*A. granulata*: 606 Mb) [13] (Table 2). Telomeres can also be found in 7 out of 13 pseudomolecules (Table 5).

Total RNA sequencing data from different tissues, including the digestive gland, foot, gill, gonad, and heart, was used to assemble the transcriptome of *L. japonica*. The final transcriptome assembly contained 294,118,260 transcripts, with 192,010 Trinity-annotated genes (average length of 1,100 bp and N50 length of 2,373 bp). The resultant transcriptomes

Table 4. GenomeScope statistics report at K-mer = 21.		
Property	min	max
Homozygous (aa)	98.74%	98.78%
Heterozygous (ab)	1.22%	1.26%
Genome haploid length (bp)	608,187,534	609,707,265
Genome repeat length (bp)	165,056,812	165,469,254
Genome unique length (bp)	443,130,722	444,238,011
Model fit	78.27%	99.09%
Read error rate	0.74%	0.74%

Table 5. List of telomeric repeats found in nine scaffolds.			
Scaffold id	Strand	Position	Sequence
scaffold_1	Forward	start	CTAACCCCTAACCCCTAACCCCTAACCCCTAACCCCTAACCCCTAACCCCT
scaffold_2	Reverse	end	TAGGGTTAGGGTTAGGGTTAGGGTTAGGGTTAGGGTTAGGGTTAGGGTTA
scaffold_5	Forward	start	CTAAACCCTAACCCCTAACCCCTAACCCCTAACCCCTAACCCCTAACCC
scaffold_6	Forward	start	CTAACCCCTAACCCCTAACCCCTAACCCCTAACCCCTAACCCCTAACCTAAC
scaffold_6	Reverse	end	TAGGGTTAGGGTTAGGGTTAGGGTTAGGGTTAGGGTTAGGGTTAGGGTTA
scaffold_7	Forward	start	TATCCCTAACCCCTAACCCCTAACCCCTAACCCCTAACCCCTAACCCCTATACC
scaffold_8	Forward	start	CCTAACCCCTAACCCCTAACCCCTAACCCCTAACCCCTAACCCCTAACCCCTAACCC
scaffold_9	Forward	start	AACCCTAACCCCTAACCCCTAACCCCTAACCCCTAACCCCTAACCCCTAACCCCTAA
scaffold_9	Reverse	end	GGGTTAGGGTTAGGGTTAGGGTTAGGGTTAGGGTTAGGGTTAGGGTTAGGGTTAGG

Table 6. Catalogue of repeat elements in the <i>Liolophura japonica</i> genome.				
Classification	Total length (bp)	Count	Proportion (%)	No. of distinct classifications
DNA	13,496,516	29,294	2.21	5,097
LINE	40,857,525	12,510	6.70	4,445
LTR	5,801,101	8,948	0.95	2,645
Other (simple repeat, microsatellite, RNA)	471,212	641	0.08	408
Penelope	3,180,531	6,938	0.52	2,484
Rolling circle	896,188	1,714	0.15	968
SINE	19,737,403	85,584	3.24	1,124
Unclassified	85,585,554	221,825	14.04	6,188
SUM:	170,026,030	367,454	27.89	23,359

were used to predict the gene models, and 28,233 gene models were generated with 28,010 predicted protein-coding genes, a mean coding sequence length of 447 amino acids (AA), and the proteome complete BUSCO estimation of 90.9% (metazoa\_odb10) (Table 2).

For the repeat elements, a total repeat content of 27.89% was found in the genome assembly, including 14.04% of unclassified elements. This result is comparable to the estimated genome repeat length with kmer 21 (27.14%) and the Chitonidae species *A. granulata* (23.56%) [13] (Figure 1E; Tables 6, 4). Among the remaining repeats, long interspersed nuclear elements (LINEs) were the most abundant (6.70%), followed by short interspersed nuclear elements (SINEs) (3.24%) and DNA (2.21%), whereas long terminal repeats (LTRs), Penelope, rolling circle, and others were only present in low proportions (LTR: 0.95%, Penelope: 0.52%, rolling circle: 0.15%, other: 0.08%).



## DATA VALIDATION AND QUALITY CONTROL

To assess the quality of samples in DNA extraction and PacBio library preparation, NanoDrop™ One/OneC Microvolume UV–Vis Spectrophotometer, Qubit® Fluorometer, and overnight pulse-field gel electrophoresis were performed. Furthermore, the Omni-C library quality was assessed by Qubit® Fluorometer and TapeStation D5000 HS ScreenTape.

During genome assembly, the Hifiasm output was mapped with PacBio HiFi reads with minimap2 (v2.15-r905) with parameters “-ax asm20” (Li, H. 2018) and compared to the NT database via BLAST with parameters “-task megablast -outfmt ‘6 qseqid staxids bitscore std’ -max\_target\_seqs 1 -max\_hsps 1 -evalue  $1 \times 10^{-25}$ ”, which was used as the input for Blobtools (v1.1.1; RRID:SCR\_017618) [50]. Scaffolds recognised as possible contaminations were removed from the assembly (Figure 2). The GC content ranged from 27.73% to 54.34%, and the coverage ranged from 0.9576 to 216.3843 for the remaining contigs. Moreover, a kmer-based statistical approach was employed to estimate the genome heterozygosity, repeat content, and genome size from sequencing Omni-C reads using Jellyfish (v2.3.0) [45] and GenomeScope (v2.0) [46] (Figure 1D, Table 4). In addition, telomeric repeats were screened by using FindTelomeres [51]. BUSCO (v5.5.0) [44] was used to assess the completeness of the genome assembly and gene annotation with a metazoan dataset (metazoa\_odb10).

## CONCLUSION AND FUTURE PERSPECTIVE

The high-quality, near chromosome-level *L. japonica* genome presented in this study is a valuable resource for gaining new insights into the environmental adaptations of *L. japonica* in residing the intertidal zones and for future investigations on the evolutionary biology in polyplacophorans and other molluscs.

## DATA AVAILABILITY

The final assemblies, Omni-C data, and PacBio HiFi reads were submitted to NCBI under the accession number GCA\_032854445.1. The raw reads generated in this study were deposited to the NCBI database under the BioProject accession PRJNA973839, with PacBio HiFi reads SRX20411988, Omni-c reads SRX21911526 and transcriptome data (SAMN35319765, SAMN35319766, SAMN35319767, SAMN35319768, SAMN35319769). The genome annotation files were deposited and are publicly available in figshare [52].

## ABBREVIATIONS

AA, amino acids; HMW, high molecular weight; LINE, long interspersed nuclear element; LTR, long terminal repeat; SINE, short interspersed nuclear element; TE, transposable elements; UTR, untranslated region.

## DECLARATIONS

### Ethics approval and consent to participate

The authors declare that ethical approval was not required for this type of research.

### Competing interests

The authors declare that they have no competing interests.

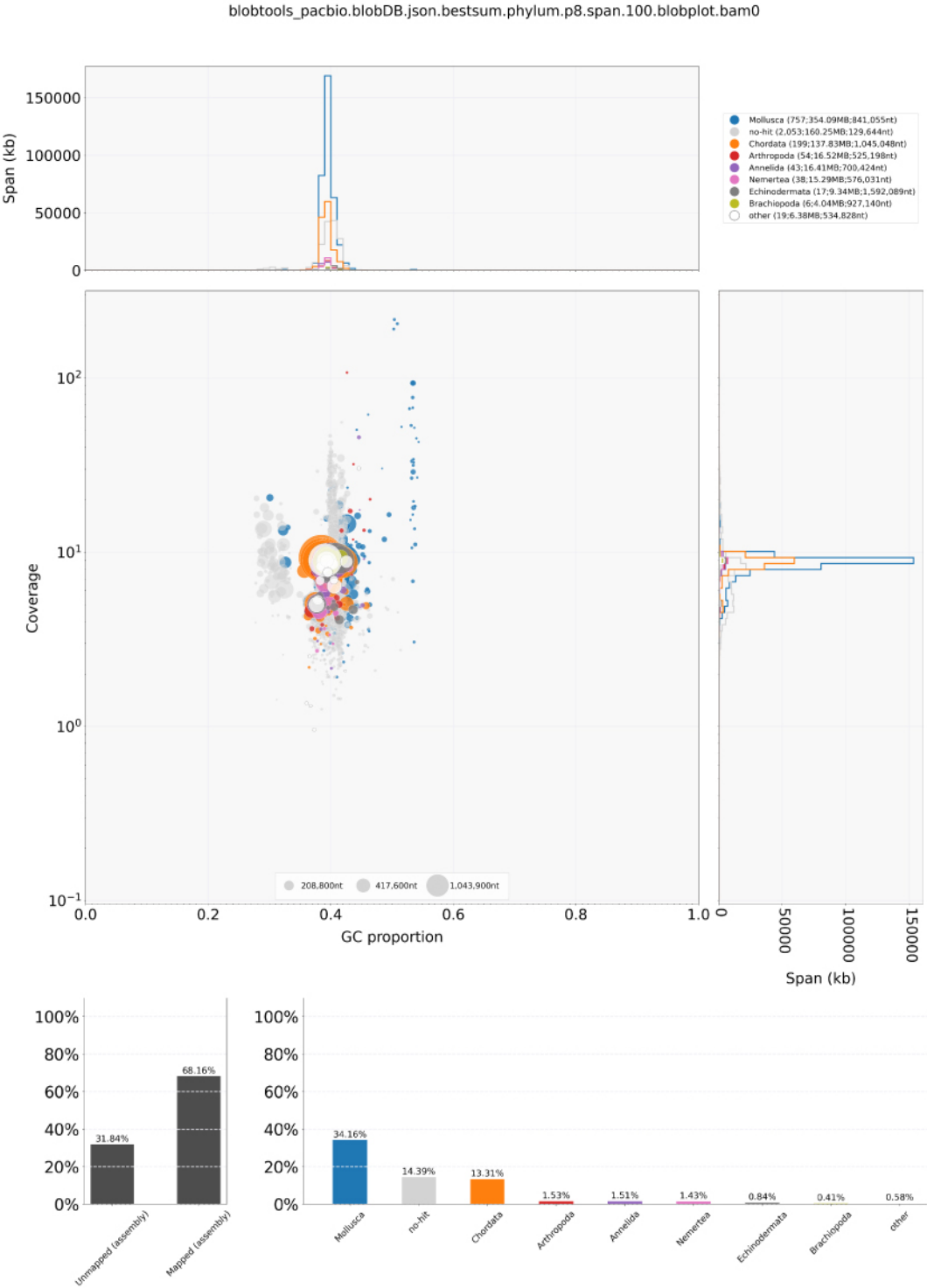


Figure 2. Genome assembly quality control and contaminant/cobiont detection for the *Liolophura japonica*.

## Authors' contributions

JHLH, TFC, LLC, SGC, CCC, JKHF, JDG, SCKL, YHS, CKCW, KYLY, and YW conceived and supervised the study. MFFA and WLS carried out the DNA extraction, library preparation, and sequencing. WN performed the genome assembly and gene model prediction. MFFA conducted the homeobox gene annotation. TYH, BKHL, and GAW collected the chiton samples.

## Funding

This work was funded and supported by the Hong Kong Research Grant Council Collaborative Research Fund C4015-20EF, CUHK Strategic Seed Funding for Collaborative Research Scheme (3133356), and CUHK Group Research Scheme (3110154).

## REFERENCES

- 1 Kocot KM, Poustka AJ, Stöger I et al. New data from Monoplacophora and a carefully-curated dataset resolve molluscan relationships. *Sci. Rep.*, 2020; **10**(1): 101. doi:10.1038/s41598-019-56728-w.
- 2 Kocot KM, Cannon JT, Todd C et al. Phylogenomics reveals deep molluscan relationships. *Nature*, 2011; **477**(7365): 452–456. doi:10.1038/nature10382.
- 3 Smith SA, Wilson NG, Goetz FE et al. Resolving the evolutionary relationships of molluscs with phylogenomic tools. *Nature*, 2011; **480**(7377): 364–367. doi:10.1038/nature10526.
- 4 Sigwart JD, Sutton MD. Deep molluscan phylogeny: synthesis of palaeontological and neontological data. *Proc. R. Soc. B*, 2007; **274**(1624): 2413–2419. doi:10.1098/rspb.2007.0701.
- 5 Paine RT. Trophic control of production in a Rocky Intertidal Community. *Science*, 2002; **296**(5568): 736–739. doi:10.1126/science.1069811.
- 6 Tejada-Martinez D, López DN, Bonta CC et al. Positive and negative effects of mesograzers on early-colonizing species in an intertidal rocky-shore community. *Ecol. Evol.*, 2016; **6**(16): 5761–5770. doi:10.1002/ece3.2323.
- 7 Eernisse DJ, Reynolds PD. Polyplacophora. In: Harrison FW, Kohn AJ (eds), *Microscopic Anatomy of Invertebrates*. vol. 5, 1994; pp. 56–110, ISBN: 9780471561125.
- 8 Scherholz M, Redl E, Wollesen T et al. Aplacophoran mollusks evolved from ancestors with polyplacophoran-like features. *Curr. Biol.*, 2013; **23**(21): 2130–2134. doi:10.1016/j.cub.2013.08.056.
- 9 Connors MJ, Ehrlich H, Hog M et al. Three-dimensional structure of the shell plate assembly of the chiton *Tonicella marmorea* and its biomechanical consequences. *J. Struct. Biol.*, 2012; **177**(2): 314–328. doi:10.1016/j.jsb.2011.12.019.
- 10 Liu X, Sigwart JD, Sun J. Phylogenomic analyses shed light on the relationships of chiton superfamilies and shell-eye evolution. *Mar. Life Sci. Technol.*, 2023; **5**(4): 525–537. doi:10.1007/s42995-023-00207-9.
- 11 Chappell DR, Speiser DI. Polarization sensitivity and decentralized visual processing in an animal with a distributed visual system. *J. Exp. Biol.*, 2023; **226**(4): jeb.244710. doi:10.1242/jeb.244710.
- 12 Li L, Connors MJ, Kolle M et al. Multifunctionality of chiton biomineralized armor with an integrated visual system. *Science*, 2015; **350**(6263): 952–956. doi:10.1126/science.aad1246.
- 13 Varney RM, Speiser DI, McDougall C et al. The iron-responsive genome of the chiton *Acanthopleura granulata*. *Genome Biol. Evol.*, 2021; **13**(1): evaa263. doi:10.1093/gbe/evaa263.
- 14 Varney RM, Yap-Chiongco MK, Mikkelsen NT et al. Genome of the lepidopleurid chiton *Hanleya hanleyi* (Mollusca, Polyplacophora). *F1000Research*, 2022; **11**: 555. doi:10.12688/f1000research.121706.1.
- 15 Choi EH, Yeo MY, Kim G et al. Liolophura species discrimination with geographical distribution patterns and their divergence and expansion history on the northwestern Pacific coast. *Sci. Rep.*, 2021; **11**(1): 17602. doi:10.1038/s41598-021-96823-5.
- 16 Harper KD. The ecology of *Acanthopleura japonica* (Mollusca: Polyplacophora): a dominant herbivore on Hong Kong rocky shores. Thesis, University of Hong Kong, 1996; <http://hdl.handle.net/10722/34725>.
- 17 Harper KD, Williams GA. Variation in abundance and distribution of the chiton *Acanthopleura japonica* and associated molluscs on a seasonal, tropical, rocky shore. *J. Zool.*, 2001; **253**(3): 293–300. doi:10.1017/S0952836901000279.

- 18 Lewis J. Observations on a high-level population of limpets. *J. Animal Ecol.*, 1954; 23(1): 85–100. doi:10.2307/1661.
- 19 Williams G, Morritt D. Habitat partitioning and thermal tolerance in a tropical limpet, *Cellana grata*. *Mar. Ecol. Prog. Ser.*, 1995; 124(1–3): 89–103. doi:10.3354/meps124089.
- 20 Ng JSS, Williams GA. Intraspecific variation in foraging behaviour: influence of shore height on temporal organization of activity in the chiton *Acanthopleura japonica*. *Mar. Ecol. Prog. Ser.*, 2006; 321: 183–192. doi:10.3354/meps321183.
- 21 McMahon RF. Aerial respiratory behavior, desiccation resistance, thermal tolerance and respiratory response to hypoxia of the intertidal chiton, *Liolophura japonica*. *Am. Zool.*, 1986; A123–A123.
- 22 Nishihama S. Distribution, diet and activity of the chiton *Liolophura japonica*. *Publ. Amakusa Mar. Biol. Lab., Kyushu Univ.*, 1986; 8: 113–123.
- 23 Dethier MN, Duggins DO. An “indirect commensalism” between marine herbivores and the importance of competitive hierarchies. *Am. Natural.*, 1984; 124: 205–219. doi:10.1086/284264.
- 24 Pacific Biosciences of California, Inc: AMPure® PB Beads. <https://www.pacb.com/wp-content/uploads/Procedure-Checklist-%E2%80%93-Using-AMPure-PB-Beads-for-Size-Selection.pdf>.
- 25 Cheng H, Concepcion GT, Feng X et al. Haplotype-resolved de novo assembly using phased assembly graphs with hifiasm. *Nat. Methods*, 2021; 18(2): 170–175. doi:10.1038/s41592-020-01056-5.
- 26 Guan D, Guan D, McCarthy SA et al. Identifying and removing haplotypic duplication in primary genome assemblies. *Bioinformatics*, 2020; 36(9): 2896–2898. doi:10.1093/bioinformatics/btaa025.
- 27 Zhou C, McCarthy SA, Durbin R. YaHS: yet another Hi-C scaffolding tool. *Bioinformatics*, 2023; 39(1): btac808. doi:10.1093/bioinformatics/btac808.
- 28 Baril T, Galbraith J, Hayward A. Earl Grey: a fully automated user-friendly transposable element annotation and analysis pipeline. *Mol. Biol. Evol.*, 2024; 41(4): msae068. doi:10.1093/molbev/msae068.
- 29 Allio R, Schomaker-Bastos A, Romiguier J et al. MitoFinder: efficient automated large-scale extraction of mitogenomic data in target enrichment phylogenomics. *Mol. Ecol. Resour.*, 2020; 20(4): 892–905. doi:10.1111/1755-0998.13160.
- 30 Bolger AM, Lohse M, Usadel B. Trimmomatic: a flexible trimmer for Illumina sequence data. *Bioinformatics*, 2014; 30(15): 2114–2120. doi:10.1093/bioinformatics/btu170.
- 31 Palmer JM, Stajich J. Funannotate v1.8.1: eukaryotic genome annotation. Zenodo, 2020; <https://doi.org/10.5281/ZENODO.4054262>.
- 32 Girgis HZ. Red: an intelligent, rapid, accurate tool for detecting repeats de-novo on the genomic scale. *BMC Bioinform.*, 2015; 16: 227. doi:10.1186/s12859-015-0654-5.
- 33 Grabherr MG, Haas BJ, Yassour M et al. Trinity: reconstructing a full-length transcriptome without a genome from RNA-Seq data. *Nat. Biotechnol.*, 2011; 29(7): 644–652. doi:10.1038/nbt.1883.
- 34 Li H. Minimap2: pairwise alignment for nucleotide sequences. *Bioinformatics*, 2018; 34(18): 3094–3100. doi:10.1093/bioinformatics/bty191.
- 35 Haas BJ, Salzberg SL, Zhu W et al. Automated eukaryotic gene structure annotation using Evidencemodeler and the program to assemble spliced alignments. *Genome Biol.*, 2008; 9(1): R7. doi:10.1186/gb-2008-9-1-r7.
- 36 Haas BJ. TransDecoder. GitHub. <https://github.com/TransDecoder/TransDecoder>.
- 37 Bray NL, Pimentel H, Melsted P et al. Near-optimal probabilistic RNA-seq quantification. *Nat. Biotechnol.*, 2016; 34: 525–527. doi:10.1038/nbt.3519.
- 38 Lomsadze A, Burns PD, Borodovsky M. Integration of mapped RNA-Seq reads into automatic training of eukaryotic gene finding algorithm. *Nucleic Acids Res.*, 2014; 42(15): e119. doi:10.1093/nar/gku557.
- 39 Stanke M, Keller O, Gunduz I et al. AUGUSTUS: ab initio prediction of alternative transcripts. *Nucleic Acids Res.*, 2006; 34(Suppl\_2): W435–W439. doi:10.1093/nar/gkl200.
- 40 Majoros WH, Pertea M, Salzberg SL. TigrScan and GlimmerHMM: two open source ab initio eukaryotic gene-finders. *Bioinformatics*, 2004; 20(16): 2878–2879. doi:10.1093/bioinformatics/bth315.
- 41 Korf I. Gene finding in novel genomes. *BMC Bioinform.*, 2004; 5: 59. doi:10.1186/1471-2105-5-59.
- 42 Buchfink B, Xie C, Huson DH. Fast and sensitive protein alignment using DIAMOND. *Nat. Methods*, 2015; 12: 59–60. doi:10.1038/nmeth.3176.

- 43 Kim D, Paggi JM, Park C et al. Graph-based genome alignment and genotyping with HISAT2 and HISAT-genotype. *Nat. Biotechnol.*, 2019; 37: 907–915. doi:10.1038/s41587-019-0201-4.
- 44 Manni M, Berkeley MR, Seppey M et al. BUSCO Update: novel and streamlined workflows along with broader and deeper phylogenetic coverage for scoring of eukaryotic, prokaryotic, and viral genomes. *Mol. Biol. Evol.*, 2021; 38(10): 4647–4654. doi:10.1093/molbev/msab199.
- 45 Marçais G, Kingsford C. A fast, lock-free approach for efficient parallel counting of occurrences of k-mers. *Bioinformatics*, 2011; 27: 764–770. doi:10.1093/bioinformatics/btr011.
- 46 Ranallo-Benavidez TR, Jaron KS, Schatz MC. GenomeScope 2.0 and Smudgeplot for reference-free profiling of polyploid genomes. *Nat. Commun.*, 2020; 11: 1432. doi:10.1038/s41467-020-14998-3.
- 47 Dovetail Genomics. Omni-C. Generating Contact Matrix. 2021; [https://omni-c.readthedocs.io/en/latest/contact\\_map.html](https://omni-c.readthedocs.io/en/latest/contact_map.html).
- 48 Abdennur N, Fudenberg G, Flyamer IM et al. Pairtools: from sequencing data to chromosome contacts. bioRxiv. 2023; <https://doi.org/10.1101/2023.02.13.528389>.
- 49 Durand NC, Robinson JT, Shamim MS et al. Juicebox provides a visualization system for Hi-C contact maps with unlimited zoom. *Cell Syst.*, 2016; 3(1): 99–101. doi:10.1016/j.cels.2015.07.012.
- 50 Laetsch DR, Blaxter ML. BlobTools: interrogation of genome assemblies. *F1000Research*, 2017; 6: 1287. doi:10.12688/f1000research.12232.1.
- 51 Jana Sperschneider. FindTelomeres. GitHub. <https://github.com/JanaSperschneider/FindTelomeres>.
- 52 Hong Kong Biodiversity Genomics Consortium. Chromosome-level genome assembly of the common chiton, *Liolophura japonica* (Lischke, 1873). figshare. [Dataset]. 2024; <https://doi.org/10.6084/m9.figshare.24208719.v5>.

## DETAILS OF COLLABORATIVE AUTHORS

### • List of authors in Hong Kong Biodiversity Genomics Consortium

Jerome H. L. Hui,<sup>1</sup> Ting Fung Chan,<sup>2</sup> Leo Lai Chan,<sup>3</sup> Siu Gin Cheung,<sup>4</sup> Chi Chiu Cheang,<sup>5</sup> James Kar-Hei Fang,<sup>6</sup> Juan Diego Gaitan-Espitia,<sup>7</sup> Stanley Chun Kwan Lau,<sup>8</sup> Yik Hei Sung,<sup>9,10</sup> Chris Kong Chu Wong,<sup>11</sup> Kevin Yuk-Lap Yip,<sup>12,13</sup> Yingying Wei,<sup>14</sup> Ming Fung Franco Au,<sup>1</sup> Wai Lok So,<sup>1</sup> Wenyan Nong,<sup>1</sup> Tin Yan Hui,<sup>9,15</sup> Brian Kai Hin Leung,<sup>15</sup> Gray A. Williams<sup>15</sup>

<sup>1</sup>School of Life Sciences, Simon F.S. Li Marine Science Laboratory, State Key Laboratory of Agrobiotechnology, Institute of Environment, Energy and Sustainability, The Chinese University of Hong Kong, Hong Kong, China

<sup>2</sup>School of Life Sciences, State Key Laboratory of Agrobiotechnology, The Chinese University of Hong Kong, Hong Kong SAR, China

<sup>3</sup>State Key Laboratory of Marine Pollution and Department of Biomedical Sciences, City University of Hong Kong, Hong Kong SAR, China

<sup>4</sup>State Key Laboratory of Marine Pollution and Department of Chemistry, City University of Hong Kong, Hong Kong SAR, China

<sup>5</sup>Department of Science and Environmental Studies, The Education University of Hong Kong, Hong Kong SAR, China

<sup>6</sup>State Key Laboratory of Marine Pollution, Research Institute for Future Food, and Department of Food Science and Nutrition, The Hong Kong Polytechnic University, Hong Kong SAR, China

<sup>7</sup>The Swire Institute of Marine Science and School of Biological Sciences, The University of Hong Kong, Hong Kong SAR, China

<sup>8</sup>Department of Ocean Science, The Hong Kong University of Science and Technology, Hong Kong SAR, China

<sup>9</sup>Science Unit, Lingnan University, Hong Kong SAR, China



<sup>10</sup>School of Allied Health Sciences, University of Suffolk, Ipswich, IP4 1QJ, UK

<sup>11</sup>Croucher Institute for Environmental Sciences, and Department of Biology, Hong Kong Baptist University, Hong Kong SAR, China

<sup>12</sup>Department of Computer Science and Engineering, The Chinese University of Hong Kong, Hong Kong SAR, China

<sup>13</sup>Sanford Burnham Prebys Medical Discovery Institute, La Jolla, CA, USA

<sup>14</sup>Department of Statistics, The Chinese University of Hong Kong, Hong Kong SAR, China

<sup>15</sup>The Swire Institute of Marine Science and School of Biological Sciences, The University of Hong Kong, Hong Kong SAR, China