scientific data

Check for updates

OPEN

DATA DESCRIPTOR

# Chromosome-level genome assembly of the estuarine *Corbicula flumnalis* from the Yangtze River estuary in China

Yangxin Tang[1,2], Jie Pi[1,2,3], Siting Yang[1,2], Xinhua Liu[1,2], Jianguo Xiang[1,2], Chao Bian[4] & Deliang Li[1,2] ✉

***Corbicula*** **clams are of important economic values.** ***Corbicula*** **clams have diverse reproductive characteristics, inhabiting both freshwater and estuary. Here, we constructed a chromosome-level genome of the estuarine** ***C. fluminalis*** **from China by combining HiFi sequencing and Hi-C technology. The assembled genome is 1,140.30 Mb with a scaffold N50 length of 58.80 Mb, and is anchored onto 19 pseudo-chromosomes. A total of 25,363 protein-coding genes were identified and 94.03% of the identified genes were annotated. Approximately 63.50% of the genome is consisted of repetitive elements. The genome synteny between** ***C. fluminalis*** **and** ***Mercenaria mercenaria*** **is highly conserved. The chromosome genome assembly of** ***C. fluminalis*** **is a valuable resource for active ingredient research, phylogenomic studies and comparative genomics of the** ***Corbicula*** **clams and related species.**

## Background & Summary

*Corbicula* is a benthic shellfish group native to Asia, the Middle East, Australia and Africa and has now extended to an almost global distribution[1,2]. Within its native Asia range, *Corbicula* clams have a wide distribution and are important economic bivalves served as food or supplement for its nutritional and medicinal values[3–5]. Recent studies have revealed that the meat or extract of *Corbicula* clams is rich in proteins and polysaccharides, showing protective effects against acute liver injury and effects of lowering liver fat deposition[3,4]. The function investigation and supplement development of active ingredient of *Corbicula* clams are worth for further study.

*Corbicula* clams inhabit from estuarine to freshwater environments and harbor highly diverse sexual strategies (male, female and hermaphrodite)[6–8], reproductive modes (sexual dioecy and asexual hermaphrodite)[7,9] and ploidy status (diploid, triploid and tetraploid)[10,11]. The sexual dioecious *Corbicula* clams are restricted to be diploid in estuarine and freshwater while the asexual hermaphrodites are diploid, triploid and tetraploid in freshwater[12]. These characteristics render *Corbicula* as a potential model for investigation of the bivalve radiations from the seawater to the freshwater environments[7,13]. Since many freshwater asexual *Corbicula* populations were reported within the Yangtze River basin in China[6,7,11,14], *Corbicula* from the estuary of the Yangtze River (designated as *C. fluminalis*[15]) have provided an important opportunity to study the radiation process. Decoding the genome of *C. fluminalis* inhabiting the estuary of the Yangtze River may provide new insights into the evolution and adaption of *Corbicula*. In addition, the phylogenetic relationships among distinct *Corbicula* populations or lineages remain unresolved[16,17]. Previous studies attempting to clarify the relationships among distinct *Corbicula* lineages rely on limited number of molecular markers[12,17,18], therefore further phylogenomic studies are crucial to comprehensively resolve the genetic relationships within this species complex[7,17,19].

Here, the chromosome-level genome of *C. fluminalis* from the estury of the Yangtze River has been successfully assembled by using PacBio HiFi sequencing and Hi-C technology. The genome assembly of *C. fluminalis* was 1140.30 Mb with a scaffold N50 length of 58.80 Mb. This assembly was anchored onto 19

[1]College of Fisheries, Hunan Agricultural University, Changsha, 410128, China. [2]Hunan Engineering Technology Research Center of Featured Aquatic Resources Utilization, Hunan Agricultural University, Changsha, 410128, China. [3]Hunan Applied Technology University, Changde, 415100, China. [4]Laboratory of Aquatic Genomics, College of Life Sciences and Oceanography, Shenzhen University, Shenzhen, 518060, China. ✉e-mail: lidl@hunau.edu.cn

| | Contig | Scaffold |
|---|---|---|
| Total Number | 195 | 155 |
| Total Length (bp) | 1,244,015,206 | 1,149,555,071 |
| N50 (bp) | 31,933,900 | 58,803,969 |
| N90 (bp) | 11,337,362 | 51,136,350 |
| Max Length (bp) | 75,405,073 | 81,131,101 |
| Average Length (bp) | 6,379,565.16 | 7,416,484.33 |
| GC Content (%) | 35.00 | 34.99 |

**Table 1.** The statistics of the de novo assembled *C. fluminalis* genome.

pseudo-chromosomes. The assembled genome provides important resource for further adaptive evolution and phylogenomic research on *Corbicula* clams.

## Methods

**Sampling and sequencing.** Live samples of *C. fluminalis* were collected from the Yangtze River estuary at Chongming island, Shanghai, China (121.83°E, 31.44°N). One sample was dissected. After removing the shells, all soft parts, including mantle, adductor muscle, foot, gill and visceral mass, were immediately frozen in liquid nitrogen and stored at −80 °C. Genomic DNA was isolated from the foot using a modified cetyltrimethylammonium bromide (CTAB) method. The quality and concentration of the isolated genomic DNA were evaluated using a NanoDrop One spectrophotometer (Thermo Fisher Scientific, USA) and 0.75% agarose gel electrophoresis.

The isolated genomic DNA was concentrated and purified to construct a long-read library using a SMRTbell Template Prep Kit according to the manufacture's protocols (Pacific Biosciences, USA). The HiFi library was subsequently sequenced on the PacBio Sequel II platform (Pacific Biosciences, USA) following the manufacturer's instructions. A total of 55.61 Gb of PacBio HiFi clean reads were obtained after this process. A short-read library was constructed using a Nextera DNA Flex Library Prep Kit (Illumina, USA), and then paired-end sequencing was conducted on the DNBSEQ-T7 platform (MGI, China) following the manufacturer's recommendation, generating 201.48 Gb clean data in total after filtration. To capture the chromatin conformation, the foot and adductor muscle were used to prepare a Hi-C library according to the manufacturer's instructions. The Hi-C library was sequenced on the Illumina Novaseq. 6000 platform (Illumina, USA), yielding a total of 171.89 Gb clean data.

To polish genome structure annotation, total RNA was extracted from the soft parts using the TRIzol reagent to generate transcriptome libraries. A conventional transcriptome library was constructed following the protocol and sequenced on the DNBSEQ-T7 platform using this qualified RNA, resulting in 24.7 Gb data. A full-length transcriptome library was prepared and was sequenced on the Nanopore PromethION platform (Oxford Nanopore Technologies, UK), which yielded 13.92 Gb data in total.

**Genome assembly and assessment.** To estimate the genome feature of *C. fluminalis*, the k-mer method was used in Jellyfish (v 2.2.10)[20] and Genomescope (v 2.0)[21] software with the 201.48 Gb paired-end data in 19-mer frequencies. The genome size was estimated to be 1,039.15 Mb with 1.65% heterozygosity and 52.02% sequence repeats.

PacBio HiFi clean reads were used for initial genome assembly with default parameters in Hifiasm (v 0.16.1)[22]. Overlaps among clean reads were subjected to three rounds of error correction and then recomputed from the corrected reads. Haploid duplications were removed by haplotype-aware read and phased string graph construction using Hifiasm. The final assembly was 1,244.02 Mb, consisting of 195 contigs with a contig N50 size of 31.93 Mb (Table 1). Subsequently, the 201.48 Gb short-read data were mapped onto the genome assembly to assess its completeness using BWA software (v 0.7.17)[23] with default parameters, which suggested that 93.60% of short-read data could be mapped and covered 99.87% of the assembled genome.

To obtain a chromosome-level assembly, the Hi-C clean data were mapped to the contig-level genome using BWA. Contigs were clustered into chromosome groups and contigs within the same chromosome group were ordered and oriented using ALLHIC (v 0.9.8)[24]. Subsequently, Juicebox (v 1.11.08)[25] was employed to correct the processed contigs and remove redundant contigs. Finally, a final assembly in 1,140.30 Mb size was obtained and 91.66% of the assembled sequences were anchored onto 19 pseudo-chromosomes with a scaffold N50 length of 58.80 Mb (Table 1, Figs. 1 and 2).

**Repeat annotation and gene prediction.** Both de novo and homology comparison methods were used to annotate repetitive elements in the genome of *C. fluminalis*. RepeatModeler (v 1.0.11)[26] and LTR FINDER (v 2.9)[27] were used to identify repetitive elements and build a de novo repetitive element database with default parameters. An integrated database was then constructed by combining the de novo database with Repbase database (v 19.06)[28] and RepeatMasker (v 4.0.9)[29] was employed to identify and classify the repetitive elements in the integrated database. A total of 730.01 Mb of repetitive elements were identified, constituting 63.50% of the *C. fluminalis* genome (Table 2).

Gene structure prediction was realized using integrative methods of transcriptome prediction, homology annotation and de novo prediction. For transcriptome prediction, the clean data of short-read transcriptome and full-length transcriptome were mapped onto the *C. fluminalis* genome by using Hisat2 (v 2.1.0)[30] and minimap2 (v 2.17)[31], respectively. The transcripts were subsequently assembled using StringTie2 (v 2.1.4)[32] and the

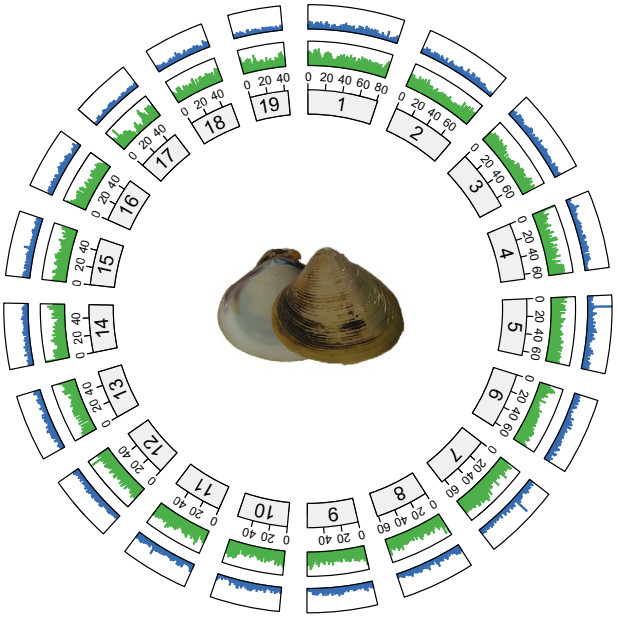

**Fig. 1** Circos plot of the *C. fluminalis* genome assembly. The circos plot showing the 19 pseudo-chromosomes in *C. fluminalis* genome. From outer to inner layers: gene density and GC content in 1Mb windows.

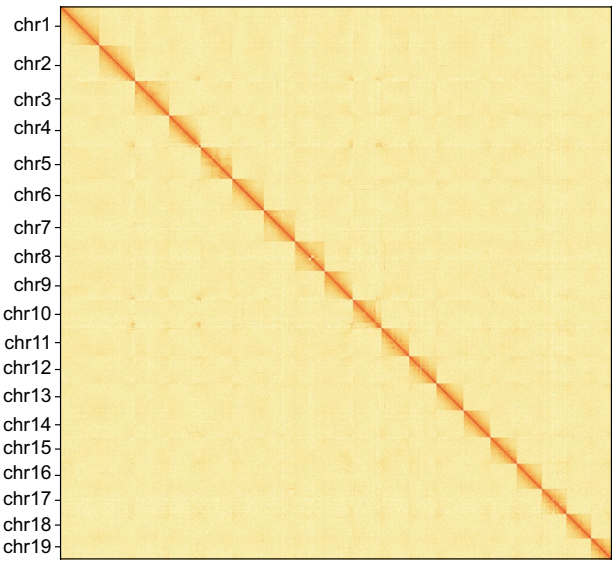


**Fig. 2** Hi-C contact heatmap of the *C. fluminalis* genome assembly.

| Type | RepeatMasker TEs | | De novo + Repbase TEs | | Combined TEs | |
|---|---|---|---|---|---|---|
| | Length (bp) | Rate (%) | Length (bp) | Rate (%) | Length (bp) | Rate (%) |
| DNA | 7,050,529 | 0.61 | 45,504,568 | 3.96 | 49,088,410 | 4.27 |
| LINE | 22,753,943 | 1.98 | 50,792,619 | 4.42 | 56,062,579 | 4.88 |
| SINE | 0 | 0 | 3,090,996 | 0.27 | 3,090,996 | 0.27 |
| LTR | 21,447,433 | 1.87 | 213,615,036 | 18.58 | 216,311,354 | 18.82 |
| Other | 126 | 0 | 10,242 | 0 | 10,368 | 0 |
| Unknown | 4,875 | 0 | 436,539,487 | 37.97 | 436,544,292 | 37.98 |
| Total | 42,640,068 | 3.71 | 678,803,102 | 59.05 | 730,014,121 | 63.50 |

**Table 2.** The statistics of repetitive elements in the *C. fluminalis* genome.

| Item | Number | Average length (bp) |
|---|---|---|
| Gene | 25,633 | 17,821.47 |
| Exon | 201,287 | 382.34 |
| Intron | 175,654 | 2,155.73 |
| **Database** | **Number** | **Rate (%)** |
| GO | 10,636 | 41.49 |
| KEGG | 6,373 | 24.86 |
| NR | 23,091 | 90.08 |
| UniProt | 21,018 | 82.00 |
| InterPro | 22,903 | 89.35 |
| Total | 24,103 | 94.03 |

**Table 3.** The statistics of gene annotation of the *C. fluminalis* genome.

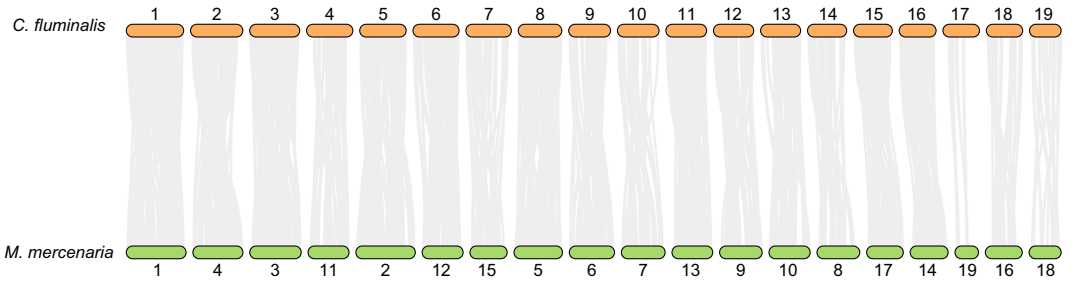


**Fig. 3** Synteny analysis of chromosomes between *C. fluminalis* and *M. mercenaria*.

coding sequences were identified in TransDecoder (v 5.1.0, https://github.com/TransDecoder/TransDecoder). For homology search, protein sequences of *Mizuhopecten yessoensis* (GCA_002113885.2), *Mercenaria mercenaria* (GCA_021730395.1), *Crassostrea gigas* (GCA_902806645.1), *Crassostrea virginica* (GCA_002022765.4) and *Dreissena polymorpha* (GCA_020536995.1) were downloaded from NCBI. BLAST + (v 2.7.1)[33] was used to align these protein sequences onto the *C. fluminalis* genome. And then Exonerate (v 2.4.0)[34] was employed to search gene structures of protein regions. For de novo prediction, the assembled transcripts were used as training set and the coding regions on the repeat-masked *C. fluminalis* genome were predicted in Augustus (v 3.3.2)[35]. Finally, the predicted gene sets from transcriptome prediction, homology annotation and de novo prediction were integrated in MAKER (v 2.31.10)[36] and a final non-redundant gene set was generated. A total of 25,633 genes with a mean length of 17,821.47 bp were predicted (Table 3). The final gene set was annotated by searching against the GO, KEGG, NR, UniProt and InterPro databases, 24,103 genes (94.03% of the total predicted genes) were ultimately annotated in at least one database (Table 3).

**Genome synteny analysis.** The genome syntenic relationships between estuarine *C. fluminalis* (this study) and marine *M. mercenaria* (GCA_021730395.1) were explored using MCScanX[37] implemented in the TBtools-II[38] with default parameters. The results suggested highly conserved genome synteny between *C. fluminalis* and *M. mercenaria* with generally one-to-one correspondence between their 19 chromosomes (Fig. 3).

## Data Records

All raw sequencing data have been deposited in the NCBI database under the accession numbers SRR31869663-68[39–44] within the BioProject PRJNA1205514. The chromosome-level genome assembly was deposited at GenBank under the accession number JBLWFO000000000[45]. The genome assembly and annotations can also be accessed via Figshare[46] (https://doi.org/10.6084/m9.figshare.28093580).

## Technical Validation

**Assessment of the genome assembly.** To assess the quality of the genome assembly, the completeness of the chromosome-level genome assembly was estimated using BUSCO (v 5.8.2)[47] against the "mollusca_odb10" database. The BUSCO results showed an overall completeness of 90.01% of the genome assembly of *C. fluminalis* (Table 4). In detail, 88.01% complete single-copy genes, 2.00% complete duplicated genes and 0.94% fragmented genes were identified in the genome assembly. In addition, 9.05% missing genes were found in the genome assembly. Generally, these findings suggest the high quality of the *C. fluminalis* assembly.

**Evaluation of the gene annotation.** BUSCO analysis was also conducted against the "mollusca_odb10" database to evaluate the completeness of the gene annotation of the chromosome-level genome assembly. The BUSCO results revealed an overall completeness of 90.84% of the gene annotations (Table 5). Specifically, 88.31% complete single-copy genes, 2.53% complete duplicated genes and 0.98% fragmented genes were presented in the annotated gene set with 8.18% missing genes. These results support the quality of the *C. fluminalis* genome annotation.

| Type | Number | Rate (%) |
|---|---|---|
| Complete BUSCOs | 4766 | 90.01 |
| Complete Single-Copy BUSCOs | 4660 | 88.01 |
| Complete Duplicated BUSCOs | 106 | 2.00 |
| Fragmented BUSCOs | 50 | 0.94 |
| Missing BUSCOs | 479 | 9.05 |
| Total BUSCO groups searched | 5295 | 100 |

**Table 4.** BUSCO assessment of the *C. fluminalis* genome assembly.

| Type | Number | Rate (%) |
|---|---|---|
| Complete BUSCOs | 4810 | 90.84 |
| Complete Single-Copy BUSCOs | 4676 | 88.31 |
| Complete Duplicated BUSCOs | 134 | 2.53 |
| Fragmented BUSCOs | 52 | 0.98 |
| Missing BUSCOs | 433 | 8.18 |
| Total BUSCO groups searched | 5295 | 100 |

**Table 5.** BUSCO assessment of the *C. fluminalis* genome annotation.

## Code availability

No custom codes or scripts were used in this work. Data analyses were performed following the manuals or protocols of the corresponding software as described in Methods.

Received: 9 January 2025; Accepted: 19 May 2025;
Published online: 24 May 2025

## References

1. Crespo, D., Dolbet, H. M., Leston, S., Sousa, R. & Pardal, M. Â. Distribution of *Corbicula fluminea* (Müller, 1774) in the invaded range: a geographic approach with notes on species traits variability. *Biol Invasions.* **17**, 2087–2101 (2015).
2. Modesto, V. *et al*. What we know and do not know about the invasive Asian clam *Corbicula fluminea*. *Hydrobiologia.* **852**, 1183–1214 (2025).
3. Xie, K. *et al*. Effects of *Corbicula fluminea* meal in high-fat diet on growth, lipid metabolism and intestinal microbiota of juvenile rice field eel Monopterus albus. *Aquaculture.* **590**, 741064 (2024).
4. Yan, J. K. *et al*. Structural characteristics and ameliorative effect of a polysaccharide from *Corbicula fluminea* industrial distillate against acute liver injury induced by CCl4 in mice. *Int J Biol Macromol.* **227**, 391–404 (2023).
5. Wang, G. P., Zhang, T., Zhang, J., Li, D. L. & Xiao, T. Y. Morphological and molecular differentiation of genus *Corbicula* suggests that two species are sympatrically distributed in Datong Lake in the Central Yangtze River Basin. *Zool Stud.* **53**, 64 (2014).
6. Pi, J. *et al*. Temperature drives reproductive activity in a rare trioecy population of *Corbicula* clams. *Hydrobiologia.* **852**, 1215–1229 (2025).
7. Tang, Y. X. *et al*. Niche similarity accompanying COI genetic differentiations in asexual and sexual *Corbicula* lineages within East and Southeast Asia. *Hydrobiologia.* **851**, 4521–4534 (2024).
8. Tejima, K., Yamada, M., Houki, S. & Komaru, A. Coexistence of hermaphrodites and males in androgenetic clam *Corbicula fluminea* Müller in Shirakawa River, Kyoto, Japan. *J Shellfish Res.* **39**, 337–344 (2020).
9. Komaru, A., Kawagishi, T. & Konishi, K. Cytological evidence of spontaneous androgenesis in the freshwater clam *Corbicula leana* Prime. *Dev Genes Evol.* **208**, 46–50 (1998).
10. Okamoto, A. & Arimoto, B. Chromosomes of *Corbicula japonica*, *C. sandai* and C.(*Corbiculina*) *leana* (Bivalvia: Corbiculidae). *Venus.* **45**, 194–202 (1986).
11. Qiu, A. D., Shi, A. J. & Komaru, A. Yellow and brown shell color morphs of *Corbicula fluminea* (Bivalvia: Corbiculidae) from Sichuan Province, China, are triploids and tetraploids. *J Shellfish Res.* **20**, 323–328 (2001).
12. Pigneur, L.-M. *et al*. Genetic uniformity and long-distance clonal dispersal in the invasive androgenetic *Corbicula* clams. *Mol Ecol.* **23**, 5102–5116 (2014).
13. Zeng, C. *et al*. Salinity appears to be the main factor shaping spatial COI diversity of *Corbicula* lineages within the Chinese Yangtze River Basin. *Divers Distrib.* **29**, 364–378 (2023).
14. Pi, J. *et al*. Reproductive activity of an androdioecious population of clams of the genus *Corbicula*. *J Mollus Stud.* **90**, eyae017 (2024).
15. Park, J.-K. & Kim, W. Two *Corbicula* (Corbiculidae: Bivalvia) mitochondrial lineages are widely distributed in Asian freshwater environment. *Mol Phylogenet Evol.* **29**, 529–539 (2003).
16. Pigneur, L.-M. *et al*. Impact of invasive Asian clams, *Corbicula* spp., on a large river ecosystem. *Freshw Biol.* **59**, 573–583 (2014).
17. Vastrade, M. *et al*. Substantial genetic mixing among sexual and androgenetic lineages within the clam genus *Corbicula*. *Peer Community J.* **2**, e73 (2022).
18. Pigneur, L.-M. *et al*. Phylogeny and androgenesis in the invasive *Corbicula* clams (Bivalvia, Corbiculidae) in Western Europe. *BMC Evol Biol.* **11**, 147 (2011).
19. Bespalaya, Y. V. *et al*. Who inhabits the world's deepest crater lake? A taxonomic review of *Corbicula* (Bivalvia: Cyrenidae) clams from Lake Toba, North Sumatra, Indonesia. *J Zool Syst Evol Res.* **59**, 400–410 (2021).
20. Marçais, G. & Kingsford, C. A fast, lock-free approach for efficient parallel counting of occurrences of k-mers. *Bioinformatics* **27**, 764–770 (2011).
21. Ranallo-Benavidez, T. R., Jaron, K. S. & Schatz, M. C. GenomeScope 2.0 and Smudgeplot for reference-free profiling of polyploid genomes. *Nat Commun.* **11**, 1432 (2020).
22. Cheng, H., Concepcion, G. T., Feng, X., Zhang, H. & Li, H. Haplotype-resolved de novo assembly using phased assembly graphs with hifiasm. *Nat Methods.* **18**, 170–175 (2021).

23. Li, H. & Durbin, R. Fast and accurate short read alignment with Burrows–Wheeler transform. *Bioinformatics* **25**, 1754–1760 (2009).
24. Zhang, X., Zhang, S., Zhao, Q., Ming, R. & Tang, H. Assembly of allele-aware, chromosomal-scale autopolyploid genomes based on Hi-C data. *Nat Plants.* **5**, 833–845 (2019).
25. Durand, N. C. *et al*. Juicebox provides a visualization system for Hi-C contact maps with unlimited zoom. *Cell Syst.* **3**, 99–101 (2016).
26. Flynn, J. M. *et al*. RepeatModeler2 for automated genomic discovery of transposable element families. *P Natl Acad Sci USA*. **117**, 9451–9457 (2020).
27. Xu, Z. & Wang, H. LTR_FINDER: an efficient tool for the prediction of full-length LTR retrotransposons. *Nucleic Acids Res.* **35**, W265–W268 (2007).
28. Bao, W., Kojima, K. K. & Kohany, O. Repbase Update, a database of repetitive elements in eukaryotic genomes. *Mob DNA.* **6**, 11 (2015).
29. Tarailo-Graovac, M. & Chen, N. Using RepeatMasker to identify repetitive elements in genomic sequences. *Curr Protoc Bioinformatics.* **25**, 4.10.11–14.10.14 (2009).
30. Kim, D., Paggi, J. M., Park, C., Bennett, C. & Salzberg, S. L. Graph-based genome alignment and genotyping with HISAT2 and HISAT-genotype. *Nat Biotechnol.* **37**, 907–915 (2019).
31. Li, H. Minimap2: pairwise alignment for nucleotide sequences. *Bioinformatics* **34**, 3094–3100 (2018).
32. Kovaka, S. *et al*. Transcriptome assembly from long-read RNA-seq alignments with StringTie2. *Genome Biol.* **20**, 278 (2019).
33. Gertz, E. M., Yu, Y.-K., Agarwala, R., Schäffer, A. A. & Altschul, S. F. Composition-based statistics and translated nucleotide searches: Improving the TBLASTN module of BLAST. *BMC Biol.* **4**, 41 (2006).
34. Slater, G. S. C. & Birney, E. Automated generation of heuristics for biological sequence comparison. *BMC Bioinformatics.* **6**, 31 (2005).
35. Stanke, M. & Morgenstern, B. AUGUSTUS: a web server for gene prediction in eukaryotes that allows user-defined constraints. *Nucleic Acids Res.* **33**, W465–W467 (2005).
36. Holt, C. & Yandell, M. MAKER2: an annotation pipeline and genome-database management tool for second-generation genome projects. *BMC Bioinformatics.* **12**, 491 (2011).
37. Wang, Y. *et al*. MCScanX: a toolkit for detection and evolutionary analysis of gene synteny and collinearity. *Nucleic Acids Res.* **40**, e49–e49 (2012).
38. Chen, C. *et al*. TBtools-II: A "one for all, all for one" bioinformatics platform for biological big-data mining. *Mol Plant.* **16**, 1733–1742 (2023).
39. *NCBI Sequence Read Archive* https://identifiers.org/ncbi/insdc.sra:SRR31869663 (2025).
40. *NCBI Sequence Read Archive* https://identifiers.org/ncbi/insdc.sra:SRR31869664 (2025).
41. *NCBI Sequence Read Archive* https://identifiers.org/ncbi/insdc.sra:SRR31869665 (2025).
42. *NCBI Sequence Read Archive* https://identifiers.org/ncbi/insdc.sra:SRR31869666 (2025).
43. *NCBI Sequence Read Archive* https://identifiers.org/ncbi/insdc.sra:SRR31869667 (2025).
44. *NCBI Sequence Read Archive* https://identifiers.org/ncbi/insdc.sra:SRR31869668 (2025).
45. Tang, Y. X. & Li, D. L. *GenBank* https://identifiers.org/ncbi/insdc:JBLWFO000000000 (2025).
46. Tang, Y. X. & Li, D. L. Chromosome-level genome assembly of the estuarine *Corbicula flumnalis* from the Yangtze River estuary in China. *Figshare* https://doi.org/10.6084/m9.figshare.28093580 (2025).
47. Manni, M., Berkeley, M. R., Seppey, M., Simão, F. A. & Zdobnov, E. M. BUSCO update: novel and streamlined workflows along with broader and deeper phylogenetic coverage for scoring of eukaryotic, prokaryotic, and viral genomes. *Mol Biol Evol.* **38**, 4647–4654 (2021).


## Acknowledgements

This study was funded by the National Natural Science Foundation of China (31772832), the National Key Research and Development Program of China (No. 2023YFD2400902) and the Hunan Agriculture Research Systems (HARS-07).

## Author contributions

D.L.L. and Y.X.T. conceived the study. Y.X.T., J.P. and S.T.Y. carried out the lab work. Y.X.T. and J.P. performed the bioinformatic analyses. Y.X.T. wrote the initial manuscript. J.P., S.T.Y., X.H.L. and J.G.X. revised the manuscript. C.B. evaluated the results and revised the manuscript. All authors have read and approved the manuscript.

## Competing interests

The authors declare no competing interests.

## Additional information

**Correspondence** and requests for materials should be addressed to D.L.