



OPEN

The chromosome-level genome of *Cherax quadricarinatus*

DATA DESCRIPTOR

Honglin Chen^{1,5}, Rui Zhang^{2,5}, Feng Liu¹, Changwei Shao³, Fangfang Liu¹, Weidong Li⁴, Jindong Ren¹, Baolong Niu¹, Haipeng Liu⁴✉ & Bao Lou¹✉

Red claw crayfish (*Cherax quadricarinatus*) is an aquatic crustacean with considerable potential for the commercial culture and an ideal model for studying the mechanism of sex determination. To provide better genomic resources, we assembled a chromosome-level genome with a size of 5.26 Gb and contig N50 of 144.33 kb. Nearly 90% of sequences were anchored to 100 chromosomes, which represents the high-quality crustacean genome with the largest number of chromosomes ever reported. The genome contained 78.69% repeat sequences and 20,460 protein-coding genes, of which 82.40% were functionally annotated. This chromosome-scale genome would be a valuable reference for assemblies of other complex genomes and studies of evolution in crustaceans.

Background & Summary

Crustaceans are a diverse and ancient group of arthropods¹, and are not only essential components of the marine and freshwater environments, but also an interesting model for the study of evolutionary biology and developmental biology. However, due to the high complexity, assembly of complete and exact crustacean genomes is difficult, let alone genomes at the chromosome level².

Cherax quadricarinatus, also known as the red claw crayfish, is a large tropical freshwater crustacean with significant commercial interest for global aquaculture³. Intersexuality appears relatively widespread throughout gonochoristic crustaceans and has been reported in several crayfish species⁴. In red claw crayfish, the intersex individuals undergo a dramatic morphological and physiological sex shift, which makes it a fascinate model to study the mechanisms underlying sex determination and differentiation of crustacean. Although a genome of this species has been reported previously, with uncomplete and fragmental genome assembly (assembled genome size, 3.24 Gb and Contig N50, 33 kb), it still prevents many studies from going deep⁵. Here, we *de novo* assembled a chromosome-level genome of red claw crayfish with the assembled genome size of 5.26 Gb and contig N50 of 144,316 bp. This high-quality genome would enrich the genomic resources of crustaceans and provides basic data for further genome-wide selective breeding.

Methods

Sample collection and genomic sequencing. All samples used in this study were from a healthy male adult red claw crayfish farmed in Honghai Co., LTD., Zhejiang, China. Fresh muscle and haemolymph were used for whole genomic sequencing and Hi-C sequencing, respectively. Seven tissues including muscle, intestine, eyestalk, hepatopancreas, gills, stomach, and antennal gland were used for transcriptomic sequencing. Isolation of DNA/RNA, construction of libraries and genomic sequencing were carried out according to protocols from <https://www.protocols.io/widgets/doi?uri=dx.doi.org/10.17504/protocols.io.bs8inhue>.

For whole genomic sequencing (WGS), the genomic DNA was sonicated into ~250 bp fragments that used to build the 100 bp paired-end (PE100) sequencing library. The library was then sequenced on the BGISEQ-500 platform and generated 280.51 Gb raw data, which covered ~58X of the estimated genome (Table 1).

¹State Key Laboratory for Managing Biotic and Chemical Threats to the Quality and Safety of Agro-products, Institute of Hydrobiology, Zhejiang Academy of Agricultural Sciences, Hangzhou, China. ²BGI-Qingdao, Qingdao, China. ³Key Lab of Sustainable Development of Marine Fisheries, Ministry of Agriculture and Rural Affairs, Yellow Sea Fisheries Research Institute, Chinese Academy of Fishery Sciences, 266072, Qingdao, Shandong, China. ⁴State Key Laboratory of Marine Environmental Science; State-Province Joint Engineering Laboratory of Marine Bioproducts and Technology; College of Ocean and Earth Sciences, Xiamen University; Xiamen, 361102, Fujian, China. ⁵These authors contributed equally: Honglin Chen, Rui Zhang. ✉e-mail: haipengliu@xmu.edu.cn; loubao6577@163.com

Types	Sample	Raw reads (Gb)	Clean data (Gb)
PE100 DNA	Muscle	280.51	238.09
20 kb PacBio CLR	Muscle	568.55	
PE100 Hi-C	Hemolymph	542.71	
PE100 RNA	Intestines	15.79	6.91
	Antennal Gland	25.41	11.06
	Hepatopancreas	18.50	10.51
	Muscle	20.17	8.33
	Gill	16.30	7.05
	Stomach	14.81	7.12
	Eyestalk	25.98	13.69

Table 1. Statistics of sequencing data.

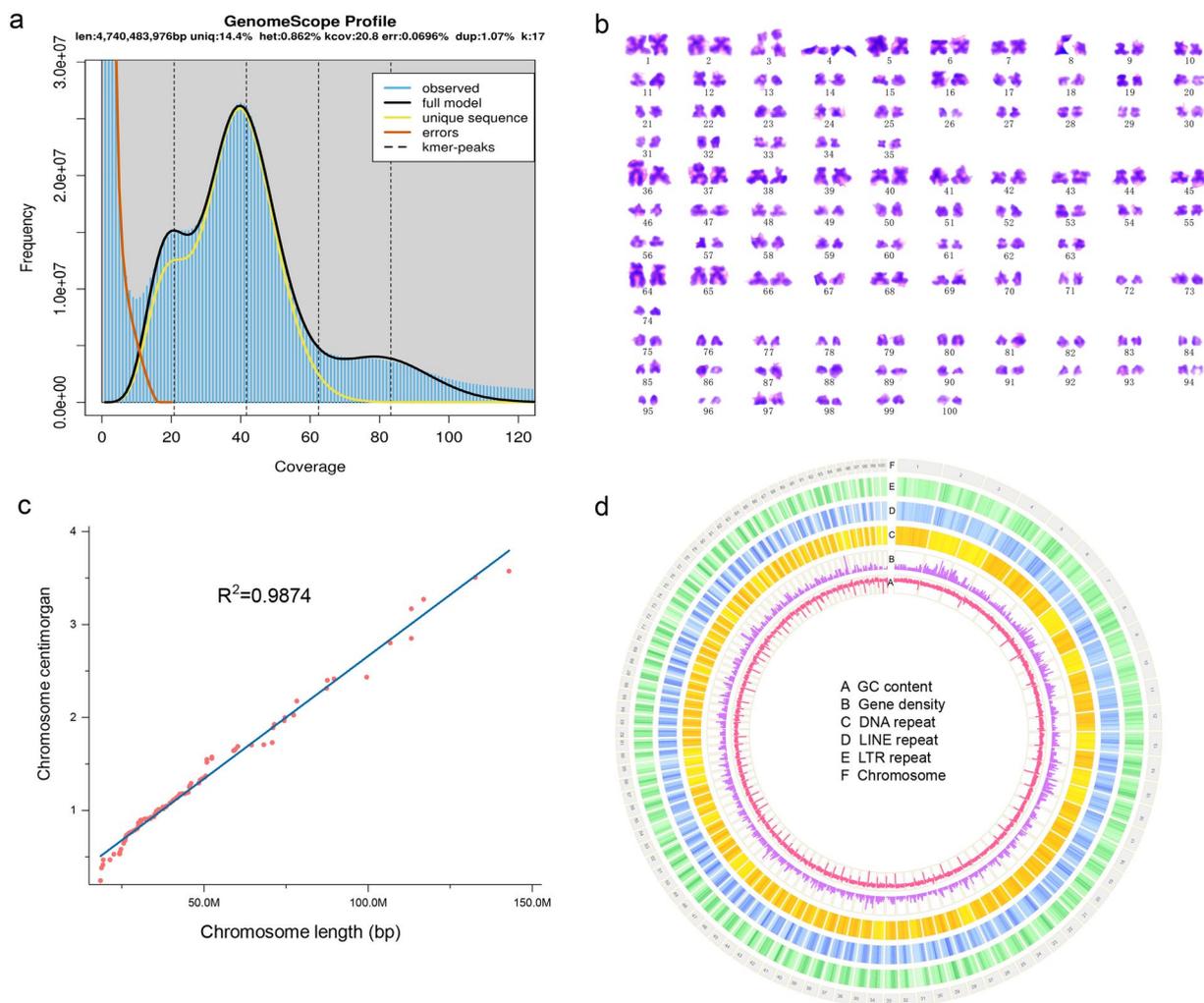


Fig. 1 Genome assembly of the red claw crayfish. (a) The 17-mer analysis of the genome. (b) The karyotypic analysis. The karyotype formula of the male is $n = 100 = 36m + 33sm + 14st + 17t$. (c) The linear regression analysis between sequence length and physical length of chromosomes. (d) Genomic features.

For PacBio Continuous Long Reads (CLR) sequencing, seven sequencing libraries were constructed using ~20Kb high-quality molecular DNA fragments. All libraries were sequenced on the PacBio Sequel II platform, which generated 568.55 Gb raw data with an N50 of 17,393 bp (Table 1).

For the construction of Hi-C library, DNA was fixed with formaldehyde solution and isolated from nuclei, and digested with MboI, the digested fragments were labeled with biotinylated nucleotides. Eight libraries were sequenced on the BGISEQ-500 platform and produced a total of 542.71 Gb raw data, which covered ~105X of the estimated genome (Table 1).

	PacBio		Hi-C	
	Scaffold	Contig	Scaffold	Contig
Total number	99,922	100,361	46,864	100,373
Total length of (bp)	5,229,209,719	5,229,209,280	5,255,744,719	5,229,209,280
Gap number (bp)	439	0	26,535,439	0
Average length (bp)	52,333	52,104,00	112,149	52,097,77
N50 Length (bp)	145,977	144,333	45,061,517	144,316
N90 Length (bp)	19,668	19,628	56,367	19,628
Maximum length (bp)	2,570,330	2,570,330	142,949,047	2,570,330
Minimum length (bp)	638	76	638	76
GC content (%)	42.21%	42.21%	42.21%	42.21%

Table 2. Summary of the genome assembly of red claw crayfish.

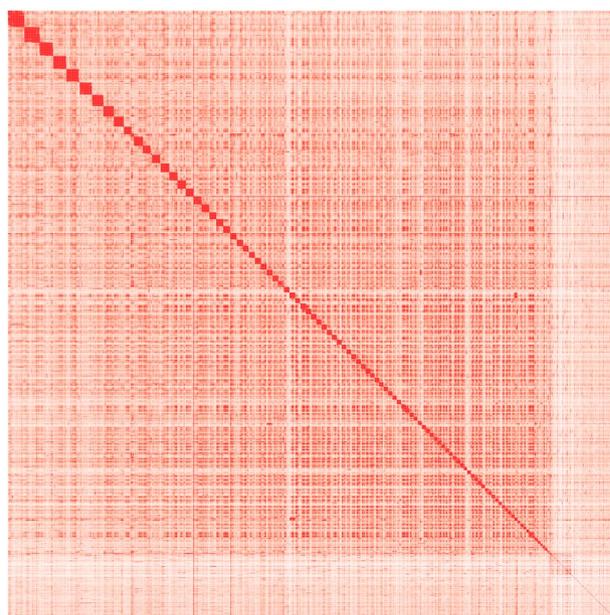


Fig. 2 The chromosome matrix heatmap.

Seven RNA libraries were constructed according to the protocols and sequenced on the BGISEQ-500 platform, generating a total of 136.96 Gb raw data (Table 1).

Genome survey. Raw PE100 reads were firstly filtered by SOAPnuke (v1.6.5)⁶ with parameters of “-M 1 -d -A 0.4 -n 0.05 -l 10 -q 0.4 -Q 2 -G -5 0”, and 240 Gb clean data were retained (Table 1). Then Jellyfish (v2.2.6)⁷ was used to count k-17mers and GenomeScope⁸ was used to estimate the size, heterozygosity, and repetitive sequences of the genome at 4.74 Gb, 0.86% and 85.6%, respectively (Fig. 1a).

Chromosome karyotyping. The number and length of chromosomes in red claw crayfish were obtained by karyotyping experiment using 15 male adults, according to the published pipeline⁹. Chromosomes were measured using Adobe Photoshop CS6 measurement tools under a magnification of 600 \times . The chromosome pairs were classified following the nomenclature of Levan (1964)¹⁰ into m = metacentric (long arm/short arm (r) = 1–1.7), sm = submetacentric (r = 1.7–3), st = subtelocentric (r = 3–7), and a = acrocentric (r > 7). The karyotype formula of the male red claw crayfish is n = 100 = 36 m + 33 sm + 14 st + 17 t (Fig. 1b), and the arm lengths data were listed in Supplementary Table 1.

Genome assembly. Reads longer than 5 kb were kept from raw Pacbio CLR reads and corrected by Canu (v1.5)¹¹, based on which the draft genome was assembled by Wtdbg2¹² with parameters of “-p 21 -E 2 -S 4 -s 0.05 -L 5000 -X 40”. The draft genome was further polished by Pilon¹³ using clean PE100 reads with default parameters, giving an assembly with the size of 5.26 Gb and the contig N50 of 144.33 kb (Table 2).

Based on the polished genome, 84.34 Gb Hi-C data were validated through quality control by Hi-C-Pro (v. 2.8.0)¹⁴, which were then applied for chromosomal reconstruction by Juicer (v1.5)¹⁵ and 3D-DNA (3D-de novo assembly)¹⁶. To get more precise chromosomes, we manually made some adjustments according to the

Type	Repeat Size(bp)	% of genome
Tandem Repeat Finder	1,188,877,157	22.621
RepeatMasker	445,491,828	8.476
RepeatProteinMask	626,288,466	11.916
<i>De novo</i>	3,479,664,132	66.207
Total	4,135,818,061	78.691

Table 3. Summary of repetitive sequences.

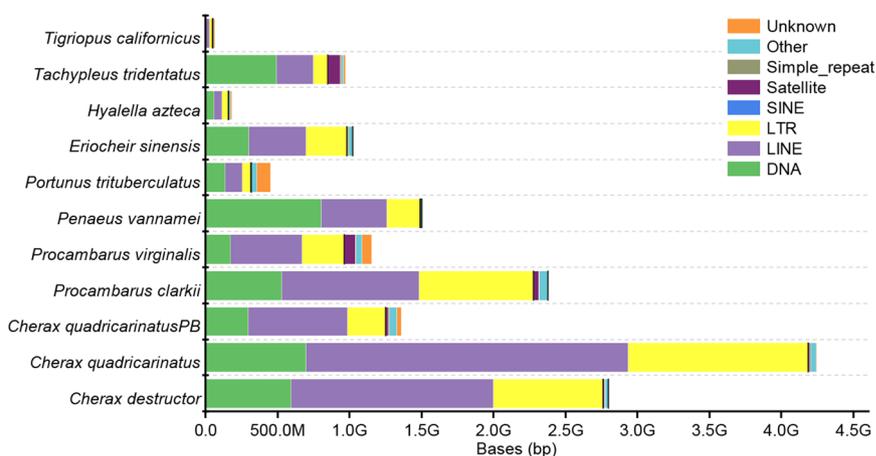


Fig. 3 Composition of the major TEs among 11 crustacean species.

chromosomal interaction heatmap by Juicebox¹⁷ (Fig. 2). Finally, a total of 4.70 Gb sequences were anchored to 100 chromosomes, of which the longest is 142.95 Mb and the shortest is 18.54 Mb (Supplementary Table 2). The linear regression analysis of karyotyping and assembly showed a high correlation ($R^2 = 0.9874$) between the physical length and sequence length of 100 chromosomes (Fig. 1c), indicating the high-quality crustacean genome with the largest number of chromosomes ever reported.

Repeat annotation. Based on aligning the genome to the Repbase library by TRF (v.4.09)¹⁸, repetitive sequences were predicted by RepeatMasker (v. 3.3.0) and RepeatProteinMask (v. 3.3.0)¹⁹. In addition, transposable elements (TEs) were constructed and RepeatModeler (v1.0.8)²⁰ (Table 3). All the above results together showed that red claw crayfish contains 78.69% repetitive sequences, among which TEs were most abundant (3,482 Mb) (Fig. 3, Table 4). Compared with other decapod crustaceans, the proportion of TES in crayfish was generally much higher.

Gene prediction. For homology-based gene prediction, the encoded protein sequences of six crustacean species include *Cherax quadricarinatus* (previous version), *Eriocheir sinensis*, *Hyalella azteca*, *Macrobrachium nipponense*, *Penaeus vannamei*, and *Procambarus virginialis* were aligned with the genomic sequence of red claw crayfish using BLAST²⁰ and Genewise²¹ with default parameters. Augustus (v3.2.3)²² and Genscan²³ were used for *de novo* gene prediction²⁴. RNA reads were mapped to the genome by HISAT2 (v2.1.0)²⁵ and gene structure were predicted by Stringtie (v1.2.2)²⁶. Meanwhile, transcriptome was *de novo* assembled by Trinity (v2.1.1)²⁷ and splicing variations were identified by PASApipeline (v2.4.1)²⁸. EVIDENCEModeler (v1.1)²⁹ was applied to integrate the above evidence and a total of 20,460 protein-coding genes were predicted, with average gene length and exon number per gene of 40,182.55 bp and 6.5, respectively (Tables 5, 6).

These genes were then functionally annotated through BLAST against NCBI non-redundant proteins (NR), TrEMBL, Gene Ontology (GO), SwissProt, and Kyoto Encyclopedia of Genes and Genomes (KEGG) protein databases. Finally, 16,859 genes accounting for 82.40% of the total were successfully annotated with at least one public functional database (Table 7).

The tRNAscan-SE³⁰ was used to annotate the tRNAs based on annotated features such as isotype, anticodon, and tRNAscan-SE bit score. The rRNA sequences were annotated from homologous references in close species. MiRNAs and snRNAs were predicted by the INFERNAL³¹ based on the covariance model of the Rfam database. Totally 6,954 non-coding RNAs were predicted, including 25 miRNA, 1,448 rRNA, 5,023 tRNA and 458 snRNA genes (Table 8).

Data Records

The genomic WGS sequencing data were deposited in the SRA at NCBI SRR22412649³², SRR22412641³³.

The genomic PacBio sequencing data were deposited in the SRA at NCBI SRR22412654³⁴.

The transcriptomic sequencing data were deposited in the SRA at NCBI SRR22412651³⁵, SRR22412652³⁶, SRR22412653³⁷, SRR22412637³⁸, SRR22412638³⁹, SRR22412639⁴⁰, SRR22412640⁴¹.

Type	Rebase TEs		Protein TEs		Denovo TEs		Combined TEs	
	Length(bp)	% of genome	Length(bp)	% of genome	Length(bp)	% of genome	Length(bp)	% of genome
DNA	246,065,669	4.68	21,028,400	0.40	519,239,284	9.88	697,212,505	13.27
LINE	197,906,764	3.77	552,300,634	10.51	2,046,408,761	38.94	2,238,906,979	42.60
SINE	6,813,035	0.13	0	0	2,033,568	0.04	8,812,197	0.17
LTR	54,064,213	1.03	53,065,444	1.01	1,213,361,102	23.09	1,251,832,766	23.82
Other	253,460	0.01	0	0	0	0	253,460	0.01
Unknown	0	0	0	0	8,011,372	0.15	8,011,372	0.15
Total	445,491,828	8.48	626,288,466	11.92	3,436,709,566	65.39	3,481,829,970	66.25

Table 4. Summary of different TE repeat sequences. Note: TEs, transposable elements; LINE, long interspersed nuclear elements; SINE, short interspersed nuclear elements; LTR, long terminal repeats.

	Gene set	Gene number	Gene length(bp)	CDS number	Intron length(bp)	Exon length(bp)	Exon per gene	BUSCO
Homolog	<i>C. quadricarinatus</i>	37,558	25,503.02	665.20	11,214.81	206.92	3.21	C:76.50%
	<i>E. sinensis</i>	10,669	29,723.88	913.70	8,792.97	213.66	4.28	C:51.60%
	<i>H. azteca</i>	6,232	47,229.48	1,022.85	11,887.37	209.30	4.89	C:45.50%
	<i>M. nipponense</i>	15,693	56,965.30	913.50	20,215.13	242.13	3.77	C:37.30%
	<i>P. vannamei</i>	14,642	128,560.17	1,217.39	31,515.52	241.52	5.04	C:67.30%
	<i>P. virginalis</i>	22,924	42,488.16	667.11	22,422.71	232.84	2.87	C:61.70%
Denovo	Augustus	197,737	13,764.73	1,220.68	7,260.74	447.52	2.73	C:65.30%
	Genscan	296,975	9,888.34	1,255.71	2,914.62	316.95	3.96	C:66.30%
Transcriptome	Hisat + Stringtie	78,739	36,776.39	841.27	10,097.69	269.31	3.12	C:82.60%
EVM		234,118	12,363.55	1,186.88	4,680.45	350.32	3.39	C:80.40%
Final		20,460	40,182.55	1,753.24	6,992.06	269.89	6.50	C:88.30%

Table 5. Statistical results of gene structure prediction.

Type	Percentage
Complete BUSCOs	88.3% (941)
Complete Single-Copy BUSCOs	86.2% (919)
Complete Duplicated BUSCOs	2.1% (22)
Fragmented BUSCOs	5.7% (61)
Missing BUSCOs	6.0% (64)
Total	100% (1066)

Table 6. BUSCO evaluation of gene annotation in red claw crayfish.

Database	Number	Percentage
Total	20,460	100%
NR	15,959	78.00%
Swissprot-Annotated	12,318	60.21%
KEGG-Annotated	13,570	66.32%
TrEMBL-Annotated	16,303	79.68%
Interpro-Annotated	12,536	61.27%
GO-Annotated	8,861	43.31%
Overall	16,859	82.40%

Table 7. Summary of gene annotation in red claw crayfish.

The Hi-C sequencing data were deposited in the SRA at NCBI SRR22412642⁴², SRR22412643⁴³, SRR22412644⁴⁴, SRR22412645⁴⁵, SRR22412646⁴⁶, SRR22412647⁴⁷, SRR22412648⁴⁸, SRR22412650⁴⁹.

The final chromosome assembly was deposited in GenBank at NCBI JAPQEV000000000⁵⁰.

The genome annotation file is available in figshare⁵¹.

Type		Copy	Average length(bp)	Total length(bp)	% of genome
miRNA		25	77.64	1,941	0.0000
tRNA		5023	71.63	359,796	0.0068
rRNA	rRNA	724	138.72	100,431	0.0019
	18 S	225	162.96	36,665	0.0007
	28 S	374	158.15	59,147	0.0011
	5.8 S	17	52.06	885	0.0000
	5 S	108	34.57	3,734	0.0001
	8 S	0	0	0	0
snRNA	snRNA	229	137.39	31,462	0.0006
	CD-box	1	199.00	199	0.0000
	HACA-box	0	0	0	0
	splicing	228	137.12	31,263	0.0006
	scaRNA	0	0	0	0

Table 8. Statistics of annotated non-coding RNAs.

Technical Validation

The quality and quantity of total DNA was checked using agarose gel electrophoresis, and the concentration was determined using a NanoDrop 2000 spectrophotometer. RNA integrity was evaluated using an Agilent 2100 Bioanalyzer (Agilent Technologies, CA, USA). The sample used in our study had an RNA integrity number (RIN) larger than 8. To further assess the quality of the genome, clean PE100 reads were aligned back to the genome by BWA⁵², showing the mapping rate as high as 99.03%. The depth and GC content were also statistically analyzed within a 10Kb sliding window. Moreover, 85.7% completed and 6.2% fragmented BUSCOs⁵³ (Benchmarking Universal Single-Copy Orthologs, v4.0) in arthropoda_odb9 database were identified, which showed a noticeable improvement than the previous version (81.3%).

Code availability

No specific code was developed in this work. The parameters of all commands and pipelines used for data processing are described in the Methods section. If no detailed parameters are mentioned for a software, the default parameters were used, as suggested by the developer.

Received: 16 December 2022; Accepted: 30 March 2023;

Published: 17 April 2023

References

- Stillman, J. H. *et al.* Recent advances in crustacean genomics. *Integr. Comp. Biol.* **48**(6), 852–868 (2008).
- Meehan, D., Xu, Z., Zuniga, G. & Alcivar-Warren, A. High frequency and large number of polymorphic microsatellites in cultured shrimp, *Penaeus (Litopenaeus) vannamei* [Crustacea: Decapoda]. *Mar. Biotechnol.* **5**(4), 311–330 (2003).
- Saoud, I. P., Ghanawi, J., Thompson, K. R. & Webster, C. D. A review of the culture and diseases of redclaw crayfish *Cherax quadricarinatus* (von Martens 1868). *J WORLD AQUACULT SOC.* **44**(1), 1–29 (2013).
- Ford, A. T., Fernandes, T. F., Read, P. A., Robinson, C. D. & Davies, I. M. The costs of intersexuality: a crustacean perspective. *Mar. Biol.* **145**(5), 951–957 (2004).
- Tan, M. H. *et al.* A giant genome for a giant crayfish (*Cherax quadricarinatus*) with insights into cox1 pseudogenes in decapod genomes. *Front. Genet.* **11**, 201 (2020).
- Chen, Y. *et al.* SOAPnuke: a MapReduce acceleration-supported software for integrated quality control and preprocessing of high-throughput sequencing data. *GigaScience.* **7**(1), gix120 (2018).
- Li, R. *et al.* The sequence and de novo assembly of the giant panda genome. *Nature.* **463**(7279), 311–317 (2010).
- Vurture, G. W. *et al.* GenomeScope: fast reference-free genome profiling from short reads. *Bioinformatics.* **33**(14), 2202–2204 (2017).
- Shi, L. L., Xu, X. H., Zhang, L. & Li, Y. H. Comparative analysis of karyotype in female and male *Procambarus clarkii*. *Journal of Anhui Agricultural University.* **46**(2), 234–241 (in Chinese) (2019).
- Levan, A., Fredga, K. & Sandberg, A. A. Nomenclature for centromeric position on chromosomes. *Hereditas.* **52**(2), 201–220 (1964).
- Koren, S. *et al.* Canu: scalable and accurate long-read assembly via adaptive k-mer weighting and repeat separation. *Genome Res.* **27**(5), 722–736 (2017).
- Ruan, J. & Li, H. Fast and accurate long-read assembly with wtdbg2. *Nat. Methods.* **17**(2), 155–158 (2020).
- Walker, B. J. *et al.* Pilon: an integrated tool for comprehensive microbial variant detection and genome assembly improvement. *PLoS one.* **9**(11), e112963 (2014).
- Servant, N. *et al.* HiC-Pro: an optimized and flexible pipeline for Hi-C data processing. *Genome Biol.* **16**(1), 1–11 (2015).
- Durand, N. C. *et al.* Juicer provides a one-click system for analyzing loop-resolution Hi-C experiments. *Cell Syst.* **3**(1), 95–98 (2016).
- Dudchenko, O. *et al.* De novo assembly of the *Aedes aegypti* genome using Hi-C yields chromosome-length scaffolds. *Science.* **356**(6333), 92–95 (2017).
- Durand, N. C. *et al.* Juicebox provides a visualization system for Hi-C contact maps with unlimited zoom. *Cell Syst.* **3**(1), 99–101 (2016).
- Benson, G. Tandem repeats finder: a program to analyze DNA sequences. *Nucleic Acids Res.* **27**(2), 573–580 (1999).
- Tarailo-Graovac, M. & Chen, N. Using RepeatMasker to identify repetitive elements in genomic sequences. *Curr. Protoc. Bioinformatics.* **4**, 1–4 (2009).
- McGinnis, S. & Madden, T. L. BLAST: at the core of a powerful and diverse set of sequence analysis tools. *Nucleic Acids Res.* **32** (suppl_2), W20–W25 (2004).
- Birney, E., Clamp, M. & Durbin, R. GeneWise and genomewise. *Genome Res.* **14**(5), 988–995 (2004).
- Stanke, M. *et al.* AUGUSTUS: ab initio prediction of alternative transcripts. *Nucleic Acids Res.* **34**, W435–W439 (2006).
- Burge, C. & Karlin, S. Prediction of complete gene structures in human genomic DNA. *Journal of Mol. Biol.* **268**(1), 78–94 (1997).

24. Majoros, W. H., Pertea, M. & Salzberg, S. L. TigrScan and GlimmerHMM: Two open source ab initio eukaryotic gene-finders. *Bioinformatics*. **20**(16), 2878–2879 (2004).
25. Kim, D., Langmead, B. & Salzberg, S. L. HISAT: a fast spliced aligner with low memory requirements. *Nat. Methods*. **12**(4), 357–360 (2015).
26. Pertea, M., Kim, D., Pertea, G. M., Leek, J. T. & Salzberg, S. L. Transcript-level expression analysis of RNA-seq experiments with HISAT, StringTie and Ballgown. *Nat. Protoc.* **11**(9), 1650–1667 (2016).
27. Grabherr, M. G. *et al.* Trinity: reconstructing a full-length transcriptome without a genome from RNA-Seq data. *Nat. Biotechnol.* **29**(7), 644 (2011).
28. Roberts, A., Pimentel, H., Trapnell, C. & Pachter, L. Identification of novel transcripts in annotated genomes using RNA-Seq. *Bioinformatics*. **27**(17), 2325–2329 (2011).
29. Haas, B. J. *et al.* Automated eukaryotic gene structure annotation using EVidenceModeler and the Program to Assemble Spliced Alignments. *GENOME BIOL EVOL.* **9**(1), 1–22 (2008).
30. Lowe, T. M. & Eddy, S. R. tRNAscan-SE: a program for improved detection of transfer RNA genes in genomic sequence. *Nucleic Acids Res.* **25**(5), 955–964 (1997).
31. Nawrocki, E. P., Kolbe, D. L. & Eddy, S. R. Infernal 1.0: inference of RNA alignments. *Bioinformatics*. **25**(10), 1335–1337 (2009).
32. NCBI Sequence Read Archive <https://identifiers.org/insdc.sra:SRR22412649> (2022).
33. NCBI Sequence Read Archive <https://identifiers.org/insdc.sra:SRR22412641> (2022).
34. NCBI Sequence Read Archive <https://identifiers.org/insdc.sra:SRR22412654> (2022).
35. NCBI Sequence Read Archive <https://identifiers.org/insdc.sra:SRR22412651> (2022).
36. NCBI Sequence Read Archive <https://identifiers.org/insdc.sra:SRR22412652> (2022).
37. NCBI Sequence Read Archive <https://identifiers.org/insdc.sra:SRR22412653> (2022).
38. NCBI Sequence Read Archive <https://identifiers.org/insdc.sra:SRR22412637> (2022).
39. NCBI Sequence Read Archive <https://identifiers.org/insdc.sra:SRR22412638> (2022).
40. NCBI Sequence Read Archive <https://identifiers.org/insdc.sra:SRR22412639> (2022).
41. NCBI Sequence Read Archive <https://identifiers.org/insdc.sra:SRR22412640> (2022).
42. NCBI Sequence Read Archive <https://identifiers.org/insdc.sra:SRR22412642> (2022).
43. NCBI Sequence Read Archive <https://identifiers.org/insdc.sra:SRR22412643> (2022).
44. NCBI Sequence Read Archive <https://identifiers.org/insdc.sra:SRR22412644> (2022).
45. NCBI Sequence Read Archive <https://identifiers.org/insdc.sra:SRR22412645> (2022).
46. NCBI Sequence Read Archive <https://identifiers.org/insdc.sra:SRR22412646> (2022).
47. NCBI Sequence Read Archive <https://identifiers.org/insdc.sra:SRR22412647> (2022).
48. NCBI Sequence Read Archive <https://identifiers.org/insdc.sra:SRR22412648> (2022).
49. NCBI Sequence Read Archive <https://identifiers.org/insdc.sra:SRR22412650> (2022).
50. Chen, H. L. Genbank https://identifiers.org/ncbi/insdc.gca:GCA_026875155.1 (2022).
51. Chen, H. L. Cqu.final.last.gff. [figshare https://doi.org/10.6084/m9.figshare.21599397](https://doi.org/10.6084/m9.figshare.21599397) (2022).
52. Li, H. & Durbin, R. Fast and accurate short read alignment with Burrows–Wheeler transform. *Bioinformatics*. **25**(14), 1754–1760 (2009).
53. Simão, 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**(19), 3210–3212 (2015).

Acknowledgements

The work was supported by the Zhejiang Science and Technology Major Program (2021C02069-4). We thank Dr. Jun Wang for his kindly help in manuscript polishing.

Author contributions

Haipeng Liu, Bao Lou and Changwei Shao conceived the study and supervised the project. Honglin Chen collected the samples and wrote the manuscript. Rui Zhang performed the data analysis, figures drawing and manuscript writing. Feng Liu, Fangfang Liu, Jindong Ren and Baolong Niu supported in chromosome karyotyping and data uploading. Weidong Li assisted in data analysis.

Competing interests

Here we declare together, all data of this project were obtained from our independent research. There's no conflict of interest in this manuscript.

Additional information

Supplementary information The online version contains supplementary material available at <https://doi.org/10.1038/s41597-023-02124-z>.

Correspondence and requests for materials should be addressed to H.L. or B.L.

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.



Open Access This article is licensed under a Creative Commons Attribution 4.0 International License, which permits use, sharing, adaptation, distribution and reproduction in any medium or format, as long as you give appropriate credit to the original author(s) and the source, provide a link to the Creative Commons license, and indicate if changes were made. The images or other third party material in this article are included in the article's Creative Commons license, unless indicated otherwise in a credit line to the material. If material is not included in the article's Creative Commons license and your intended use is not permitted by statutory regulation or exceeds the permitted use, you will need to obtain permission directly from the copyright holder. To view a copy of this license, visit <http://creativecommons.org/licenses/by/4.0/>.

© The Author(s) 2023, corrected publication 2023