



OPEN

DATA DESCRIPTOR

A chromosome-level genome assembly of the pig-nosed turtle (*Carettochelys insculpta*)

Ye Li^{1,3}, Yuxuan Liu^{1,3}, Jiangmin Zheng^{1,3}, Baosheng Wu^{1,2,3}, Xinxin Cui^{1,3}, Wenjie Xu¹, Chenglong Zhu¹, Qiang Qiu¹✉ & Kun Wang¹✉

The pig-nosed turtle (*Carettochelys insculpta*) represents the only extant species within the Carettochelyidae family, is a unique Trionychia member fully adapted to aquatic life and currently facing endangerment. To enhance our understanding of this species and contribute to its conservation efforts, we employed high-fidelity (HiFi) and Hi-C sequencing technology to generate its genome assembly at the chromosome level. The assembly result spans 2.18 Gb, with a contig N50 of 126 Mb, encompassing 34 chromosomes that account for 99.6% of the genome. The assembly has a BUSCO score above 95% with different databases and strong collinearity with Yangtze giant softshell turtles (*Rafetus swinhoei*), indicating its completeness and continuity. A total of 19,175 genes and 46.86% repetitive sequences were annotated. The availability of this chromosome-scale genome represents a valuable resource for the pig-nosed turtle, providing insights into its aquatic adaptation and serving as a foundation for future turtle research.

Background & Summary

The pig-nosed turtle is a remarkable and unique organism within the world of Chelonians, standing as the only extant species within the genus *Carettochelys*¹. Known for its distinct pig-like nose, paddle-shaped fore flippers adorned with claws, and a shell that can reach up to 50 cm in length in females, this species has evolved a fascinating suite of phenotype that set it apart from other turtles¹. The closest living relative of the softshell turtle, pig-nosed turtle thrives in habitats such as large rivers, swamps, lagoons, and freshwater environments found in southern Lrian Jaya (Indonesia), southern Papua New Guinea, and the major river systems of the northwestern Northern Territory in Australia^{2,3}.

Unfortunately, despite its evolutionary and ecological significance, the pig-nosed turtle is facing a high risk of extinction due to a combination of habitat loss, overexploitation for food and the pet trade, and other anthropogenic pressures⁴. These threats have caused a precipitous decline in its wild population, leading to its classification as “Insufficiently Known” in the 1982 Red Data Book⁵ and as “Vulnerable” in the 1996 Red List⁶. These conservation challenges highlight the urgent need for comprehensive molecular data to inform effective protection strategies.

Despite extensive research has been conducted on turtle sex determination and other traits^{7,8}, the research for pig-nosed turtle, as a unique Trionychia member fully adapted to aquatic life, has not yet been systematically analyzed due to the lack of a high-quality genome data. In this study, we present a chromosome-level draft genome of the pig-nosed turtle (~2.18 Gb) with a contig N50 of 126 Mb achieved through advanced PacBio high-fidelity (HiFi) sequencing and chromosome conformation capture (Hi-C) sequencing techniques. Conserved core genes (BUSCO score) and genome synteny confirmed the continuity and accuracy of the assembly, representing the highest quality Testudines genome assembled to date. Overall, the high-quality genome serves as a valuable resource for future research on chelonian evolution and conservation.

¹Shaanxi Key Laboratory of Qinling Ecological Intelligent Monitoring and Protection, School of Ecology and Environment, Northwestern Polytechnical University, Xi'an, 710072, China. ²Guangdong Key Laboratory of Animal Conservation and Resource Utilization, Institute of Zoology, Guangdong Academy of Sciences, Guangzhou, 510260, China. ³These authors contributed equally: Ye Li, Yuxuan Liu, Jiangmin Zheng, Baosheng Wu, Xinxin Cui. ✉e-mail: qiuqiang@lzu.edu.cn; wk8910@gmail.com

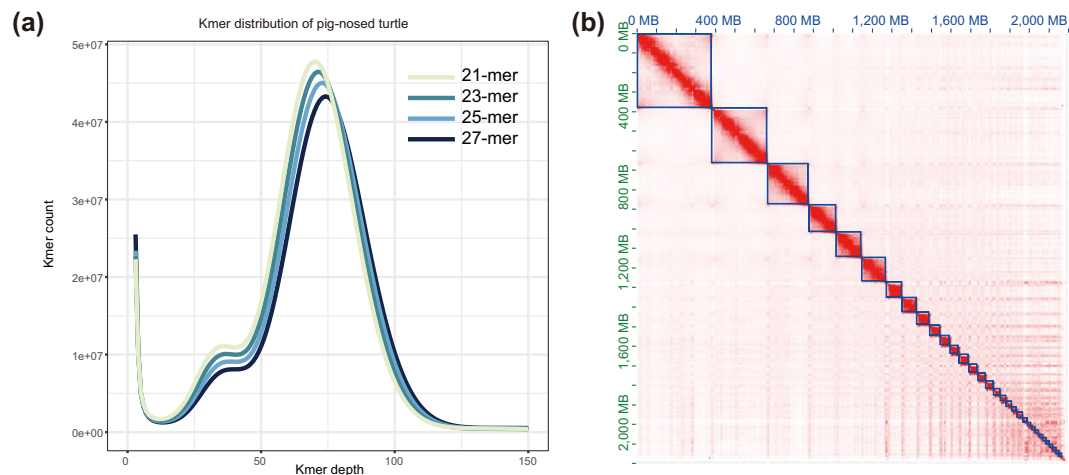


Fig. 1 Genome assembly of pig-nosed turtle. **(a)** Different k -mer ($k = 21, 23, 25, 27$) distribution of the pig-nosed turtle genome. Genome size inferred by k num divided by k depth. The range of estimated genome size from 2.15 to 2.17 Gb for pig-nosed turtle. **(b)** Hi-C linkage density heat map of the pig-nosed turtle. The x-axis and y-axis represent genomic positions. Red dots indicate regions with a high density of paired reads, suggesting that they are more likely to be on the same chromosome.

Methods

Sample collection and sequencing. An artificially bred pig-nosed turtle was obtained from an aquarium in Xiong County, Hebei Province, China. The turtle was anesthetized, and its abdominal cavity was exposed to collect tissue samples, including liver, muscle, kidney, spleen, trachea, and lungs. These samples were promptly frozen in liquid nitrogen and stored at -80°C . For genomic DNA extraction, muscle tissues were used for both short-reads sequencing and HiFi sequencing, while liver was used for Hi-C sequencing. Additionally, various tissues including muscle, liver, kidney, spleen, lung, and trachea were used for RNA-Seq to obtain a comprehensive annotation of protein-coding genes. Animal care and experimental protocols were approved by the Northwestern Polytechnic University Ethics Committee Institutional Review Board (approval number 202301024).

The Illumina sequencing library was generated using the NEB Next[®] Ultra[™] DNA Library Prep Kit (NEB, USA) and the sequencing was performed on the HiSeq 2000 platform. Raw sequencing reads were filtered for adapter sequences, low-quality reads, and trimmed using Fastp v0.20⁹ with default parameters and yielding a total of 189.18 Gb of clean short reads. To generate HiFi reads, we followed PacBio's standard protocol (Pacific Biosciences, CA, USA), which yielded 74.23 Gb cleaned long reads with the PacBio Sequel II platform. For Hi-C library construction, we followed the standard protocol described in a published study¹⁰. The HiSeq X Ten platform was utilized for sequencing, producing a total of 234 Gb of cleaned reads with Fastp v0.20⁹.

To obtain RNA sequences from multiple tissues, the TRIzol kit (TIANGEN, Cat # DP424, China) was used to extract RNA. After constructing libraries for each sample, the Illumina NovaSeq 6000 platform was used for sequencing, Fastp v0.20⁹ was used to clean the raw data and each sample obtains more than 6 G of data.

Chromosome level genome assembly of pig-nosed turtle. To estimate the genome size, we performed k -mer analysis based on Illumina sequencing reads. The optimal k -mer value was calculated by Jellyfish v2.2.10¹¹ and GenomeScope v2.0¹² was used to estimate the genome size for corresponding different k values. We utilized four distinct k -mer values—21, 23, 25, and 27 and the corresponding genome size estimates are 2.16, 2.15, 2.17, and 2.16 Gb, respectively. (Fig. 1a). The HiFi reads were assembled into contigs using Hifiasm v0.16.1¹³ with default parameters. The resulting contig-level assembly spanning 2.18 Gb comprised 155 contigs with an N50 of 126.6 Mb (Table 1). The size of the genome assembly is very close to the results predicted by k -mer analysis. Then, Hi-C sequencing reads were aligned to the contig assembly utilizing BWA v0.7.12¹⁴. Following this, we used YaHS v1.1a-r3¹⁵ to generate the scaffolded genome assembly. JuiceBox v1.11.08¹⁶ was used to visually correct the assembly based on the strength of chromosomal interactions. Contigs with no obvious interaction relationship were processed as individual scaffolds. Contigs assigned to one chromosome were joined with 200 'N' to build the final chromosome-level genome assembly. Reflecting the $2N = 68$ karyotype reported in study¹⁷, our final assembly comprised 34 chromosomes with an anchoring rate of 99.6% (Fig. 1b).

Assessment of the genome assembly. The assembled genome was evaluated for completeness using BUSCO v5.5.0¹⁸. The analysis revealed a genome completeness score of 97.6% including 97.2% single-copy orthologous genes using the tetrapoda_odb10 lineage database. The BUSCO scores based on the sauropsida_odb10 (95.5%) and vertebrata_odb10 (98.3%) lineage databases were also all above 95%. We compared the genome completeness by Busco score and contig N50 length with 13 published turtle genomes, including *Pelodiscus sinensis* (GCF_000230535)¹⁹, *Chelonia mydas* (GCF_015237465)¹⁹, *Rafetus swinhoei* (GCA_019425775.1)²⁰, *Caretta caretta* (GCF_023653815)²¹, *Pelochelys cantorii* (GCA_032595735.1)²², *Dermochelys coriacea* (GCF_009764565.3)²³, *Chelydra serpentina* (GCA_018859375.1)²⁴, *Dermatemys mawii* (GCA_007922305.1),

| Term | Contig assembly | | Hi-C assembly | |
|-----------------|-----------------|--------|---------------|--------|
| | Size (bp) | Number | Size (bp) | Number |
| N90 | 20,314,305 | 24 | 26,117,499 | 21 |
| N80 | 40,037,570 | 17 | 40,688,678 | 15 |
| N70 | 50,809,117 | 12 | 54,081,386 | 10 |
| N60 | 70,227,638 | 8 | 76,607,989 | 7 |
| N50 | 126,464,372 | 5 | 130,115,908 | 5 |
| Max length (bp) | 383,128,740 | — | 383,128,740 | — |
| Total size (bp) | 2,180,686,862 | — | 2,180,689,262 | — |
| Total number | — | 152 | — | 143 |

Table 1. Statistics of the genome assembly.

| Library | tetrapoda_odb10 | sauropsida_odb10 | vertebrata_odb10 |
|-------------------------------------|-----------------|------------------|------------------|
| Complete BUSCOs (C) | 5,180 | 7,139 | 3,297 |
| Complete and single-copy BUSCOs (S) | 5,159 | 7,075 | 3,270 |
| Complete and duplicated BUSCOs (D) | 21 | 64 | 27 |
| Fragmented BUSCOs (F) | 23 | 52 | 18 |
| Missing BUSCOs (M) | 107 | 289 | 39 |
| Total BUSCO groups searched | 5,310 | 7,480 | 3,354 |
| Summarize | 97.60% | 95.5% | 98.3% |

Table 2. Statistics of the genome assembly on the presence of conserved BUSCO orthologs.

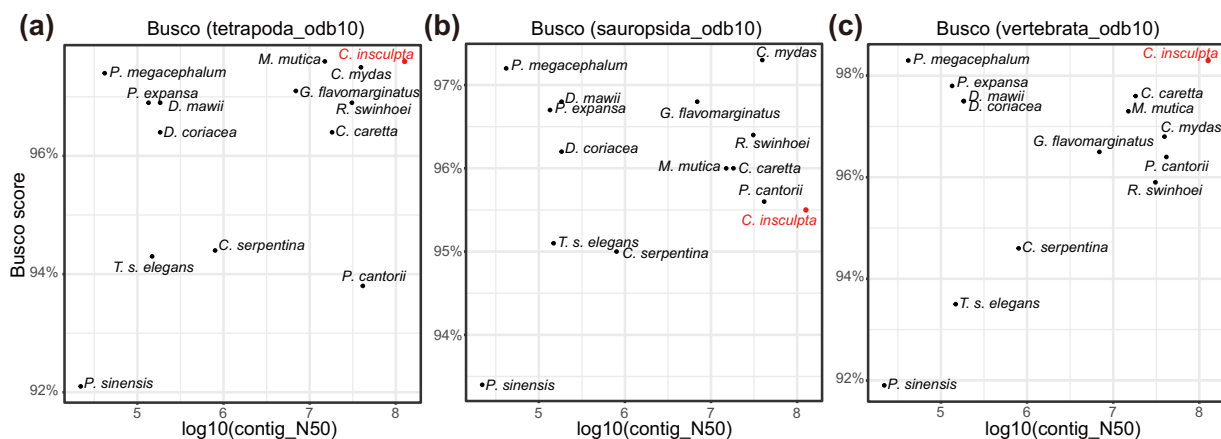


Fig. 2 Quality statistics of genomes of the representative turtles. (a–c) The x-axis indicates the contig N50 of the genome assembly, while the y-axis indicates the genome BUSCO scores based on the tetrapoda_odb10, sauropsida_odb10 and vertebrata_odb10 lineage databases, respectively.

Gopherus flavomarginatus (GCF_025201925.1), *Mauremys mutica* (GCF_020497125)²⁵, *Platysternon megacephalum* (GCA_003942145.1)²⁶, *Trachemys scripta elegans* (GCF_013100865.1)²⁷ and *Podocnemis expansa* (GCA_007922195.1)²⁸ from NCBI. Our assembly is comparable to the achievements observed in other genomes (Fig. 2, Table 2). Furthermore, to assess the continuity of the assembly, we performed a genome alignment between the pig-nosed turtle and the closely related Yangtze giant softshell turtle using LAST v1282²⁹ with parameters “-P 20 -i 2 G -m 10” to identify syntenic regions. The genome alignment revealed intact synteny between the two genomes, further validating the quality of the pig-nosed turtle genome assembly (Fig. 3a).

Genome annotation. A comprehensive genome annotation was performed for pig-nosed turtle, focusing on repetitive sequences, protein-coding genes, and functional predictions. For repetitive sequence annotation, approximately 1 Gb of repetitive regions in the assembled sequence were detected by *de novo* annotation and homology annotation. For *de novo* annotation, Tandem Repeats Finder (TRF) v4.0.9³⁰ was employed to annotated simple tandem repeats. The parameters used were “2 5 7 80 10 50 2000 -d -h -ngs”. Additionally, we utilized RepeatModeler v1.0.11 (<http://www.repeatmasker.org/RepeatModeler/>) to *de novo* annotate transposable elements. To perform homology-based annotation of transposable elements, we employed with RepeatMasker v4.0.6³¹ and RepeatProteinMask v1.0.8³¹ with parameters “-engine ncbi -noLowSimple -pvalue 1e-04”. The annotated repetitive sequence were then converted to lowercase letters in the genome using BEDtools v2.29.2³² to

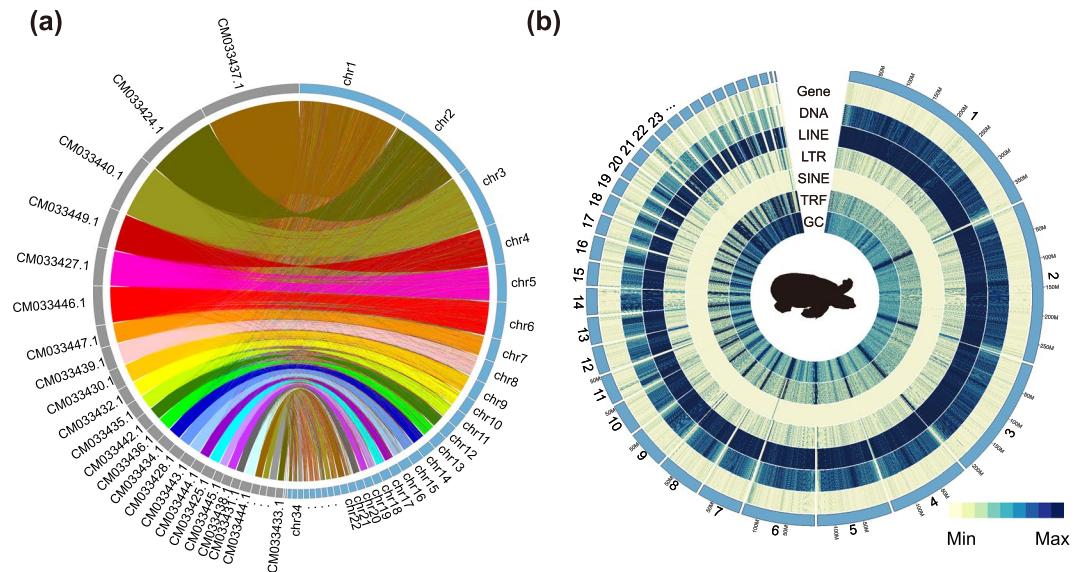


Fig. 3 Overview of the pig-nosed turtle genome. **(a)** Synteny alignment in pig-nosed turtles and *R. swinhoei*. The blue circle represents chromosomes of pig-nosed turtle, while the grey circle represents chromosomes of *R. swinhoei*. **(b)** The densities of protein coding genes and different types of repeat sequences density and GC content are shown in the inner rims with a window size of 10 Mb. The numbers 1 to 34 correspond to the chromosomes of the pig-nosed turtle.

| Term | Number | Percent (%) |
|-----------------|--------|-------------|
| InterPro | 17,917 | 93.44 |
| GO | 12,412 | 64.73 |
| KEGG | 15,036 | 78.41 |
| Swissprot | 18,589 | 96.94 |
| TrEMBL | 18,954 | 98.85 |
| COG | 6,977 | 36.39 |
| NR | 19,074 | 99.47 |
| Annotated genes | 19,117 | 99.70 |
| Missing genes | 58 | 0.30 |
| Total genes | 19,175 | — |

Table 3. Statistics of the functional annotation of protein-coding genes.

generate soft-masked sequences. The repetitive sequences account for 46.86% of the total genome length and include 310 Mb of DNA elements (14.23% of total length), 346.89 Mb of long interspersed nuclear elements (LINEs) (15.91% of total length), 25.38 Mb of short interspersed nuclear elements (SINEs) (1.16% of total length), and 121.3 Mb of long terminal repeats (LTRs) (5.56% of total length) (Fig. 3b and Table S1).

For protein-coding gene prediction, we employed a combination of three methods: *de novo* prediction, homology-based prediction, and transcript-based prediction. For *de novo* prediction, we used Augustus v2.5.5³³. For homology-based prediction, we downloaded protein sets of closely related species mentioned above. Gene structures were predicted based on these homology proteins applied by miniport v0.12-r237³⁴. Additionally, we obtained protein sets of the pig-nosed turtle using TransDecoder v5.5.0 based on RNA-Seq. The detailed steps include that SPAdes v3.1.1³⁵ was performed transcriptome assembly and then TransDecoder v5.5.0 (<https://github.com/TransDecoder/>) used to predict protein structure. We aligned it to pig-nosed turtle genome using BLAST v2.6.0³⁶. Gene structures were predicted using GeneWise v2.2.0³⁷. The results of the three prediction methods were integrated into a final gene set using EvidenceModeler v1.1.1³⁸. A total of 19,175 protein-coding genes were annotated in the pig-nosed turtle genome assembly, with a BUSCO completeness of 97.6% using tetrapoda_odb10 lineage database.

Functional annotation of proteins was applied by comparing protein sequences with public databases including Gene Ontology (GO) annotations (<http://geneontology.org/>), Cluster of Protein Orthologous Groups (COG: <https://www.ncbi.nlm.nih.gov/COG/>), Swiss-Prot (www.uniprot.org), TrEMBL (www.uniprot.org) and non-redundant proteins (NR: <https://ftp.ncbi.nlm.nih.gov/blast/db>) and Kyoto Encyclopedia of Genes and Genomes (KEGG: <https://www.kegg.jp/>) and InterPro. 99.70% proteins and were annotated in these databases (Table 3).

Data Records

All sequencing data and genome assembly have been deposited in the National Center for Biotechnology Information (NCBI) database (<https://www.ncbi.nlm.nih.gov/bioproject/PRJNA1037723>), which include SRR26796054³⁹, SRR26796055⁴⁰, SRR26796056⁴¹, SRR26796057⁴², SRR26796058⁴³, SRR26796059⁴⁴, SRR26796045, SRR26796061⁴⁶ in SRA and JAWWUY000000000⁴⁷ in GenBank. The genome annotation deposited in the Figshare database (<https://doi.org/10.6084/m9.figshare.24630915>)⁴⁸.

Technical Validation

The integrity of the extracted DNA was checked by agarose gel electrophoresis, and the concentration of DNA was determined using Qubit fluorometer using the 1 × dsDNA HS kit. Contig N50 (126.6 Mb) and BUSCO score (97.60%) are much higher quality than genomes of other turtles (Fig. 2). Higher anchoring rate (99.6%) and a strong linear relationship between the genome assembly of pig-nosed turtle and Yangtze giant softshell turtle validating that we obtained a contiguous pig-nosed turtle genome.

Code availability

The data analyses were performed all software and parameters were mentioned in Methods. The core code is available at <https://github.com/YuXuanLiua/pig-nosed-turtle/>.

Received: 28 November 2023; Accepted: 15 March 2024;

Published online: 23 March 2024

References

- Ramsay, E. P. On a new genus and species of fresh water tortoise from the Fly River, New Guinea. *Proceedings of the Linnaean Society of New South Wales* **1**, 158–162 (1887).
- Heaphy, L. J. The ecology of the pig nosed turtle, *Carettochelys insculpta*, in Northern Australia. <https://doi.org/10.26190/UNSWORKS/7247> (UNSW Sydney, 1990).
- Davenport, J., Davenport, J., Diong, C.-H. & Low, K. H. Swimming and bipedal bottom-running in the pig-nosed turtle *Carettochelys insculpta* Ramsay, 1886. *J. Nat. Hist.* **50**, 2097–2105 (2016).
- Eisemberg, C., van Dijk, P. P., Georges, A. & Amepou, Y. *Carettochelys insculpta*. (The IUCN Red List of Threatened Species, 2018).
- Wright, L. *The IUCN Amphibia-Reptilia Red Data Book*. vol. 1 (IUCN, 1982).
- Baillie, J., Gärdenfors, U., Groombridge, B., Rabb, G. & Stattersfield, A. J. 1996 IUCN Red List of threatened animals. (1996).
- Hui, H.-B. *et al.* *Sox9* is indispensable for testis differentiation in the red-eared slider turtle, a reptile with temperature-dependent sex determination. *Zool. Res.* **42**, 721–725 (2021).
- Wu, P.-F. *et al.* Role of *Cyp19a1* in the female pathway of a freshwater turtle species (*Mauremys reevesii*) with temperature-dependent sex determination. *Zool. Res.* **43**, 81–84 (2022).
- Chen, S., Zhou, Y., Chen, Y. & Gu, J. fastp: an ultra-fast all-in-one FASTQ preprocessor. *Bioinformatics* **34**, i884–i890 (2018).
- Belton, J.-M. *et al.* Hi-C: a comprehensive technique to capture the conformation of genomes. *Methods* **58**, 268–276 (2012).
- Marçais, G. & Kingsford, C. A fast, lock-free approach for efficient parallel counting of occurrences of *k*-mers. *Bioinformatics* **27**, 764–770 (2011).
- Vurtture, G. W. *et al.* GenomeScope: Fast reference-free genome profiling from short reads. *Bioinformatics* **33**, 2202–2204 (2017).
- Cheng, H. *et al.* Haplotype-resolved assembly of diploid genomes without parental data. *Nature Biotechnology* **40**, 1332–1335 (2022).
- Li, H. & Durbin, R. Fast and accurate short read alignment with Burrows–Wheeler transform. *Bioinformatics* **25**, 1754–1760 (2009).
- Zhou, C., McCarthy, S. A. & Durbin, R. YaHS: yet another Hi-C scaffolding tool. *Bioinformatics* **39**, btac808 (2023).
- Durand, N. C. *et al.* Juicer provides a one-click system for analyzing loop-resolution Hi-C experiments. *Cell Syst.* **3**, 95–98 (2016).
- Bickham, J. W. & Legler, J. M. Karyotypes and evolutionary relationships of trionychoid turtles. *Cytologia* **48**, 177–183 (1983).
- Manni, M., Berkeley, M. R., Seppely, 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).
- Wang, Z. *et al.* The draft genomes of soft-shell turtle and green sea turtle yield insights into the development and evolution of the turtle-specific body plan. *Nat. Genet.* **45**, 701–706 (2013).
- Ren, Y. *et al.* Genomic insights into the evolution of the critically endangered soft-shelled turtle *Rafetus swinhoei*. *Mol. Ecol. Resour.* **22**, 1972–1985 (2022).
- Chang, G. *et al.* The genome sequence of the Loggerhead sea turtle, *Caretta caretta* Linnaeus 1758. *F1000Res* **12**, 336 (2023).
- Hong, X. *et al.* A chromosome-level genome assembly of the Asian giant softshell turtle *Pelochelys cantorii*. *Sci. Data* **10**, 754 (2023).
- Blair P. *et al.* Divergent sensory and immune gene evolution in sea turtles with contrasting demographic and life histories. *Proc Natl Acad Sci USA* **14**, 120(7), e2201076120 (2023).
- Das, D. *et al.* Draft Genome of the Common Snapping Turtle, *Chelydra serpentina*, a Model for Phenotypic Plasticity in Reptiles. *G3 (Bethesda)* **10**, 4299–4314 (2020).
- Liu, X. *et al.* Chromosome-level genome assembly of Asian yellow pond turtle (*Mauremys mutica*) with temperature-dependent sex determination system. *Sci. Rep.* **12**, 7905 (2022).
- Cao, D., Wang, M., Ge, Y. & Gong, S. Draft genome of the big-headed turtle *Platysternon megacephalum*. *Sci. Data* **6**, 60 (2019).
- Brian Simison, W., Parham, J. F., Papenfuss, T. J., Lam, A. W. & Henderson, J. B. An Annotated Chromosome-Level Reference Genome of the Red-Eared Slider Turtle (*Trachemys scripta elegans*). *Genome Biol. Evol.* **12**, 456–462 (2020).
- Noronha, R. C. R. *et al.* New insights of karyoevolution in the Amazonian turtles *Podocnemis expansa* and *Podocnemis unifilis* (Testudines, Podocnemidae). *Mol. Cytogenet.* **9**, 73 (2016).
- Kielbasa, S. M., Wan, R., Sato, K., Horton, P. & Frith, M. C. Adaptive seeds tame genomic sequence comparison. *Genome Res.* **21**, 487–493 (2011).
- Benson, G. Tandem repeats finder: a program to analyze DNA sequences. *Nucleic Acids Res.* **27**, 573–580 (1999).
- Bedell, J. A., Korf, I. & Gish, W. MaskerAid: a performance enhancement to RepeatMasker. *Bioinformatics* **16**, 1040–1041 (2000).
- Quinlan, A. R. & Hall, I. M. BEDTools: a flexible suite of utilities for comparing genomic features. *Bioinformatics* **26**, 841–842 (2010).
- Stanke, M. & Waack, S. Gene prediction with a hidden Markov model and a new intron submodel. *Bioinformatics* **19**(Suppl 2), ii215–225 (2003).
- Li, H. Protein-to-genome alignment with miniprot. *Bioinformatics* **39**, btad014 (2023).
- Prijbelski, A., Antipov, D., Meleshko, D., Lapidus, A. & Korobeynikov, A. Using SPAdes De Novo Assembler. *Curr Protoc Bioinformatics* **70**, e102 (2020).
- Zhang, Z., Schwartz, S., Wagner, L. & Miller, W. A greedy algorithm for aligning DNA sequences. *J. Comput. Biol.* **7**, 203–214 (2000).
- Birney, E., Clamp, M. & Durbin, R. GeneWise and Genomewise. *Genome Res.* **14**, 988–995 (2004).

38. Haas, B. J. *et al.* Automated eukaryotic gene structure annotation using EVIDENCEModeler and the Program to Assemble Spliced Alignments. *Genome Biol.* **9**, R7 (2008).
39. NCBI Sequence Read Archive <https://identifiers.org/ncbi/insdc.sra:SRR26796054> (2023).
40. NCBI Sequence Read Archive <https://identifiers.org/ncbi/insdc.sra:SRR26796055> (2023).
41. NCBI Sequence Read Archive <https://identifiers.org/ncbi/insdc.sra:SRR26796056> (2023).
42. NCBI Sequence Read Archive <https://identifiers.org/ncbi/insdc.sra:SRR26796057> (2023).
43. NCBI Sequence Read Archive <https://identifiers.org/ncbi/insdc.sra:SRR26796058> (2023).
44. NCBI Sequence Read Archive <https://identifiers.org/ncbi/insdc.sra:SRR26796059> (2023).
45. NCBI Sequence Read Archive <https://identifiers.org/ncbi/insdc.sra:SRR26796060> (2023).
46. NCBI Sequence Read Archive <https://identifiers.org/ncbi/insdc.sra:SRR26796061> (2023).
47. NCBI GenBank <https://identifiers.org/ncbi/insdc:JAWWUY000000000> (2023).
48. Li, Y. Genome annotations of *Carettochelys insculpta*. Figshare <https://doi.org/10.6084/m9.figshare.24630915> (2023).

Acknowledgements

The project was supported by the National Key R&D Program of China (2022YFC3400300) and National Natural Science Foundation of China (32370452, 32122021 and 32225009).

Author contributions

K.W. and Q.Q. designed this project and research aspects. Y.L., J.Z., B.W. and X.C. performed sample collection. J.Z. performed genome assembly. Y.X.L. performed genome annotation and analysis of phylogeny relationship and innovations of pig-nosed turtles. Y.L. conducted the remaining data analysis, figure design and the manuscript. W.X. and C.Z. provided valuable suggestions for the study. K.W. amended the manuscript.

Competing interests

The authors declare no competing interests.

Additional information

Supplementary information The online version contains supplementary material available at <https://doi.org/10.1038/s41597-024-03157-8>.

Correspondence and requests for materials should be addressed to Q.Q. or K.W.

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 licence, and indicate if changes were made. The images or other third party material in this article are included in the article's Creative Commons licence, unless indicated otherwise in a credit line to the material. If material is not included in the article's Creative Commons licence 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 licence, visit <http://creativecommons.org/licenses/by/4.0/>.

© The Author(s) 2024