scientific data

Check for updates

OPEN

DATA DESCRIPTOR

A high-quality chromosome-level genome assembly of the common carp (*Cyprinus carpio* var. *Xiangxi*)

Zhe Li[1,2], Zi'ao Wang[3], Yinghai Peng[1], Yazhou Hu[2] ✉ & Xianwen Zhou[1] ✉

***Cyprinus carpio* var. *Xiangxi* is a regionally farmed rice-field carp endemic to the Xiangxi region of Hunan Province, China. To adapt to the rice-fish farming environment, this variety has developed a range of genetic and physiological adaptations. However, genomic research on this strain remains limited. Here, we assembled a high-quality chromosome-level genome of *C. carpio* var. *Xiangxi* to fill this knowledge gap. Using PacBio HiFi, DNBSEQ, and Hi-C technology, We generated a 1.54 Gb genome with a contig N50 of 28.12 Mb. Using Hi-C data, 99.38% of the assembly (1,535.34 Mb) was anchored to 50 pseudochromosomes, yielding a scaffold N50 of 30.26 Mb. The quality value and Benchmarking Universal Single-Copy Ortholog score were 62.32 and 97.4%, respectively, indicating that our genome sequence is of high quality and completeness. A total of 46,362 protein-coding genes were predicted, of which 45,409 (97.94%) were functionally annotated. This study provides a valuable genomic resource for the conservation of *C. carpio* var. *Xiangxi* and offers new insights into the evolutionary history of common carp.**

Background & Summary

The common carp (*Cyprinus carpio*) is an important freshwater species native to Asia and is now widely distributed worldwide, including China, Japan, Europe, and North America. It plays a vital role in aquaculture, culture, and ecology[1,2]. As one of the earliest domesticated aquatic species, common carp has been shaped by both natural and artificial selection over its evolutionary history, leading to the development of several selectively bred varieties adapted to different environmental conditions and human needs. These include *C. carpio* var. *vuyuanensis*, *C. carpio* var. *specularis*, *C. carpio* var. *hongguoensis*, *C. carpio* var. *wananensis*, and *C. carpio haematopterus*, all of which have contributed valuable genetic resources for breeding research.

Advancements in third-generation sequencing technologies, such as Single-Molecule Real-Time (SMRT) sequencing and Nanopore sequencing[3], have revolutionized whole-genome sequencing, providing a robust basis for studying species evolution and adaptive mechanisms. Genomic data for several *C. carpio* varieties, including *C. carpio* var. *vuyuanensis*[4], *C. carpio* var. *specularis*[2], and *C. carpio haematopterus*[2], have been sequenced and are publicly available, revealing patterns in carp ploidy evolution and mechanisms of environmental adaptation.

Despite the extensive genomic studies of various common carp species, information on the genome of *C. carpio* var. *Xiangxi*, a locally bred strain from the mountainous regions of Xiangxi, Hunan Province, China, remains lacking. This strain, developed through long-term rice-farming practices[5], is renowned for its high adaptability, disease resistance, docile behavior, and low escape rate, making it particularly well-suited for rice-fish farming systems, and it plays a crucial role in the regional rice farming industry[5]. However, the genetic variation and adaptive mechanisms underlying its domestication remain poorly characterized, largely due to the absence of high-quality, chromosome-level genomic resources.

In this study, we assembled a high-quality chromosome-level genome of *Cyprinus carpio* var. *Xiangxi* through a combination of PacBio HiFi sequencing, DNBSEQ, and Hi-C technology. A total of 1,535.34 Mb (99.38%) of sequences were successfully anchored to 50 pseudochromosomes, with a scaffold N50 of 30.26 Mb and a BUSCO completeness score of 97.40%. A total of 45,409 protein-coding genes were identified through genome annotation. This high-quality genome provides valuable resources for the conservation of *C. carpio* var. *Xiangxi* and offers important insights into the evolutionary studies of common carp.

[1]Affair center of animal husbandry and aquaculture in XiangXi Autonomous Prefecture, Jishou, 416000, Hunan, China. [2]Hunan Agricultural University, Changsha, 410128, Hunan, China. [3]Hunan Biological and Electromechanical Polytechnic, Changsha, 410127, Hunan, China. ✉e-mail: huyazhou@hunau.edu.cn; zhouxianwen2008@163.com

Type	Total base(Gb)	Reads Numbers	Max length (bp)	Mean length (bp)	N50(bp)	Depth(X)
HiFi	91.01	5,496,785	63,852	16,557	16,550	58.9
Hi-C	345.84	1,152,793,314	150	150	150	222.7
RNA	26.78	178,558,442	150	150	150	—

Table 1. Statistical information of PacBio HiFi and DNBSEQ sequencing data.

Methods

Sample collection and sequencing. The experimental fish were collected from the Xiangxi Tujia and Miao Autonomous Prefecture, Hunan Province, China. After being identified as female, the fish were anesthetized with MS-222. Muscle, liver, pancreas and swim bladder samples were then rapidly collected and immediately frozen in liquid nitrogen. A multi-platform sequencing approach involving PacBio, DNBSEQ, and Hi-C was employed to assemble a chromosome-level genome.

For HiFi sequencing, High-quality genomic DNA was extracted using a modified CTAB method. The quality and quantity of the extracted DNA were examined using a NanoDrop 2000 spectrophotometer (NanoDrop Technologies, Wilmington, DE, USA), Qubit dsDNA HS Assay Kit on a Qubit 3.0 Fluorometer (Life Technologies, Carlsbad, CA, USA) and electrophoresis on a 0.8% agarose gel, respectively. The SMRT Bell library was prepared using the SMRTbell Express Template Prep Kit 2.0. Large fragments (>15 kb) were enriched using the BluePippin size selection system (Sage Science). The quality of the selected libraries was analyzed using Femto Pulse, and their concentrations were measured with a Qubit fluorometer (Life Technologies). Sequencing was performed at Frasergen Bioinformatics (Wuhan, China) using a PacBio Sequel II instrument on PacBio SMRT cells 8 M. This resulted in the generation of 91.01 Gb of HiFi long pass reads (Table 1).

For Hi-C sequencing, a high-throughput chromatin conformation capture (Hi-C) library was prepared by formaldehyde crosslinking, MboI enzyme digestion, biotin labeling, ligation, and purification. The qualified libraries (with insert sizes ranging from 200 to 600 bp) were subsequently subjected to sequencing on a DNBSEQ-T7 platform, yielding a total of 345.84 Gb of 150-bp paired-end reads, which covered approximately 222.7 × of the genome (Table 1).

For genome-assisted annotation, Total RNA from a pooled sample of muscle, liver, pancreas, and lung tissues was extracted using Trizol reagent (Magen, GuangDong, China). RNA purity and integrity was monitored by NanoDrop 2000 spectrophotometer (NanoDrop Technologies, Wilmington, DE, USA) and a Agilent 2100 Bioanalyzer (Agilent Technologies, CA, USA). RNA contamination was assessed by 1.5% agarose gel. Sequencing was performed on a DNBSEQ-T7 platform using paired-end 150 bp reads, generating a total of 26.78 Gb of clean data (Table 1).

Genome assembly. HiFi reads were used to construct highly accurate contigs through de novo assembly, while Hi-C data were subsequently applied to organize these contigs into chromosome-scale scaffolds. The draft genome assembly was performed using HiFiasm (v0.16.1)[6] with default parameters, and the sequence graph in GFA format was converted to FASTA format using gfatools (https://github.com/lh3/gfatools). Trimmomatic (v0.40)[7] was used to trim high-quality paired-end reads to remove low-quality bases and adapter sequences from the reads. The draft genome assembly spans a total length of 1,544.93 Mb, with a contig N50 of 28.12 Mb across 216 contigs. And all filtered reads were aligned to contigs using a juicer (https://github.com/aidenlab/juicer, v3)[8] to calculate the contact frequency. Then, 3D-DNA (v180922)[9] was used with two iterative rounds for misjoin correction ($-r^2$) using default parameters. The oriented scaffolds were utilized to generate the interaction matrices with a juicer to inspect and manually correct with Juicebox assembly tools (v1.11.08)[8]. Subgenome identification was performed using Subphaser software[10] based on 15 k-mer clustering, with chromosome numbering determined based on Xu's results[2](Fig. 1). A total of 1,535.34 Mb (99.38%) of contig data were anchored to 50 chromosomes, with a scaffold N50 of 30.26 Mb. (Fig. 2 and Table 2). Benchmarking Universal Single-Copy Orthologue (BUSCO, v3.0.2)[11] analyses were performed to assess the completeness of the genome, and 97.4% of the core conserved genes were found in the genome. Long reads were mapped using Minimap2 (v2.24)[12] with the parameters "-ax map-ont/map-hifi". we obtained a mapping rate of 99.98%, a genome coverage of 99.99%, demonstrating the high quality of the assembled genome. Additionally, k-mer-based quality estimates (k = 19 bp) for the genome were assessed using the Merqury pipeline (v1.3)[13] with HiFi reads resulting in a quality value (QV) score of 62.32 (Table 2).

Genome repeats annotation. Repetitive sequences, including tandem repeats and transposable elements (TEs), were identified using a combination of de novo and homology-based approaches. Tandem repeats were annotated using Tandem Repeats Finder (TRF, v4.09.1)[14] with the parameters 2 7 7 80 10 50 2000. To identify TEs at the DNA level, LTR_FINDER (v1.0.7)[15] was used to detect LTR retrotransposons (LTR-RTs), and RepeatModeler (v2.0.1)[16] was employed to construct a de novo repeat library, which included a repeat consensus database with classification information. Additionally, RepeatMasker (v4.1.2)[17] was used to search for TEs in both the known Repbase TE library[18,19] and the de novo repeat library. At the protein level, RepeatProteinMask within the RepeatMasker package was utilized to search against the TE protein database using the WU-BLASTX engine. This analysis revealed that 46.46% (717.75 Mb) of the genome comprises transposable element sequences, among the TE repeats, short interspersed nuclear elements (SINEs), long interspersed nuclear elements (LINEs), long terminal repeats (LTRs), DNA transposons, and other unclassified repeats accounted for 0.85%, 9.50%, 10.11%, 29.36%, and 3.23%, respectively (Table 3).

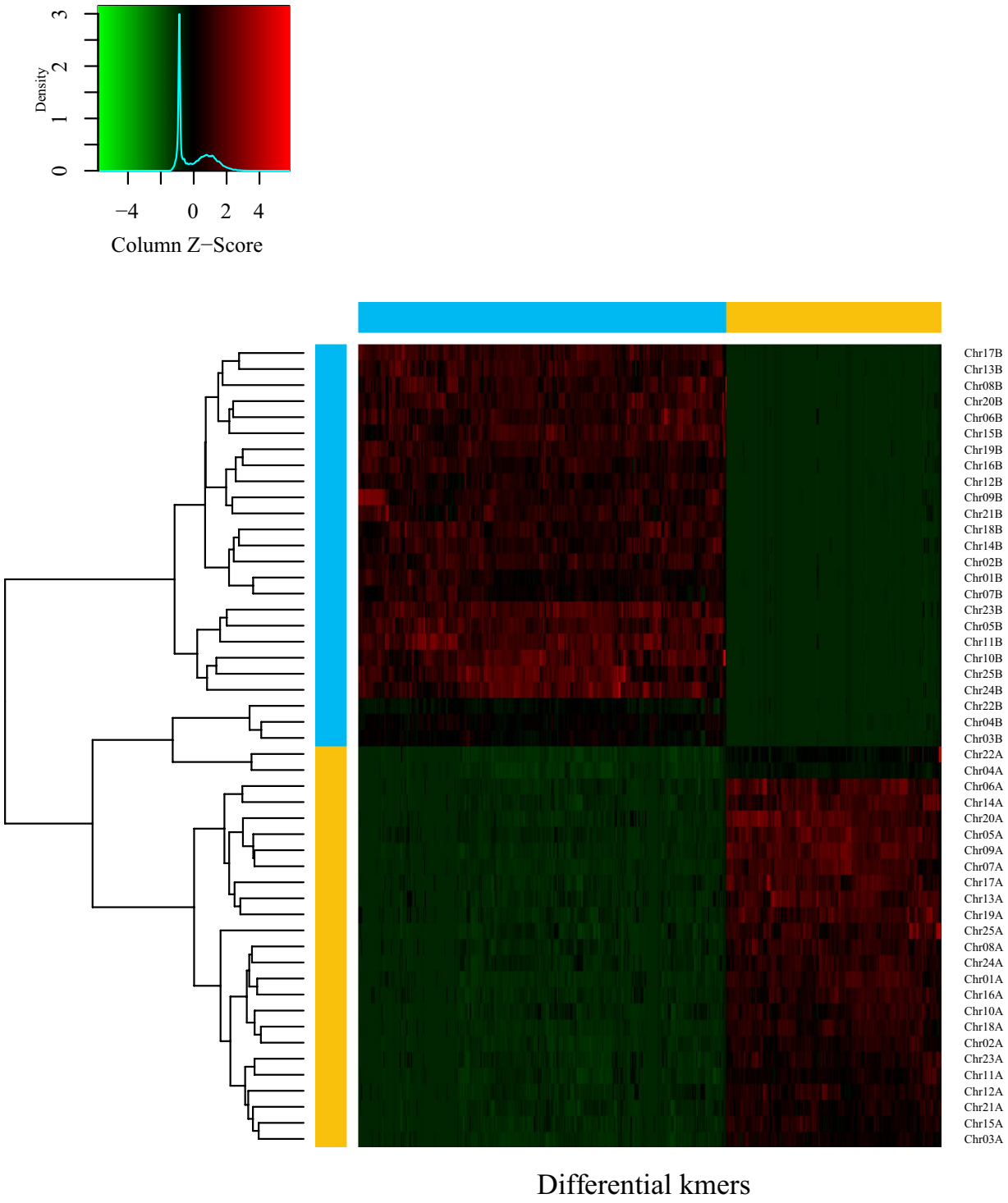

Fig. 1 *C. carpio* var. *Xiangxi* subgenome clustering map based on 15 k-mer.

Gene annotation. Gene annotation was performed using homologous, ab initio, and transcriptome-assisted approaches. For homologous annotation, TBLASTN (v2.11.0 +)[20] was employed to align sequences from 5 species, including *Cyprinus carpio* (European Carp (GCF_018340385.1))[21], *Cyprinus carpio* (Yellow river carp (GWHAATB00000000)), *Cyprinus carpio* (German mirror carp (GWHAATC00000000))[22], *Cyprinus carpio* (Hebao red carp (GWHAATD00000000))[23] and *Carassius auratus* (GCF_003368295.1)[24], to the reference genome, and the aligned sequences and their corresponding proteins were filtered and further refined using Exonerate (v2.4.0)[25]. For de novo gene prediction, Augustus (v3.4.0)[26–28] and GlimmerHMM (v3.0.4)[29] were used. In addition, transcriptome data were incorporated into the annotation process using transcriptome assemblies. RNA-seq reads were aligned using HiSat2 (v.2.2.1)[30], followed by further assembly of RNA-seq alignments into

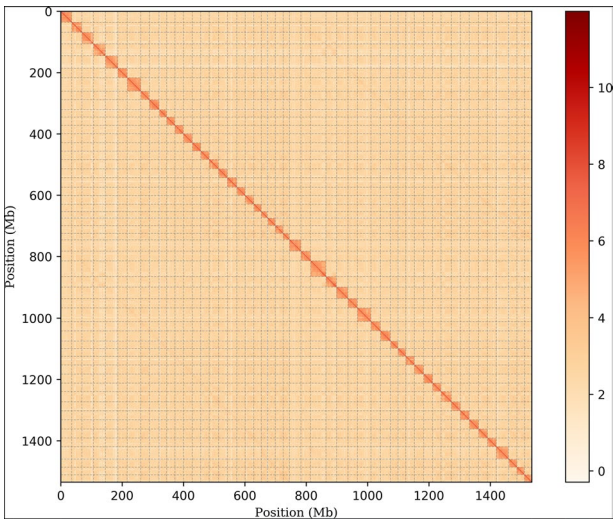

Fig. 2 The heat map of Hi-C contact matrix of the *C. carpio* var. *Xiangxi* genome assembly. The horizontal axis (left to right) corresponds to chromosomes Chr1A–Chr25A and Chr1B–Chr25B.

Assembly	*C. carpio* var. *Xiangxi*
Number of contigs	216
Assembly length(bp)	1,544,934,851
Contig N50(bp)	28,128,084
Scaffold N50(bp)	30,265,916
anchor ratio(%)	99.38
HiFi reads mapping rate(%)	99.98
HiFi reads coverage(%)	99.99
BUSCO(%)	97.40
QV	62.32

Table 2. Genome assembly results and evaluation.

Type	RepeatMasker TEs Length (bp)	RepeatProteinMask TEs Length (bp)	De novo Length (bp)	Combined TEs Length (bp)	Combined TEs % in genome
DNA	255987590	2303552	305626771	453560812	29.36
LINE	66278192	39414203	117227068	146830039	9.5
SINE	6285136	0	7815603	13156535	0.85
LTR	59474559	27063836	134655927	156268031	10.11
Other	53749	0	0	53749	0
Unknown	4152522	0	46407869	49884144	3.23
Total TE	365761998	68768622	562244186	717750237	46.46

Table 3. Classification and summary of transposable elements.

transcripts with the genome-guided assembler Stringtie (v.2.1.7)[31]. The transcriptome was also assembled de novo using Trinity (v2.8.5)[32]. For gene set integration, Maker (v3.01.03)[33] was used to merge the predicted gene sets into a non-redundant, more complete, and reliable gene set. The PASA pipeline (v2.4.1)[34] was employed to update Maker's consensus predictions, adding UTR annotations and models for alternatively spliced isoforms. A total of 46,362 protein-coding genes were successfully predicted, with BUSCO results indicating 98.2% complete BUSCOs, 0.5% fragmented BUSCOs, and 1.3% missing BUSCOs. Gene functions were inferred according to the best match of the alignments to the National Center for Biotechnology Information (NCBI) Non-Redundant (NR), Kyoto Encyclopedia of Genes and Genomes (KEGG) database[35], Gene Ontology (GO)[36], TrEMBL[37] and Swiss-Prot[37] protein databases using Diamond BLASTP (v2.0.7)[38] with an E-value threshold of 1E-5. The protein domains were annotated using InterProScan (v5.50–84.0)[39] based on InterPro[40] protein databases. In summary, the function of 45,409 protein-coding genes (97.94%) was successfully annotated (Table 4). We used tRNAscan-SE (v2.0.9) algorithms[41] with default parameters to identify the genes associated with tRNA. RNAmmer (v1.2)[42] was

	Number	Percent(%)
Total	46,362	
InterPro	38,518	83.08
GO	32,954	71.08
KEGG_ALL	44,803	96.64
KEGG_KO	30,348	65.46
Swissprot	40,082	86.45
TrEMBL	44,179	95.29
NR	45,383	97.89
Annotated	45,409	97.94
Unannotated	953	2.06

Table 4. Summary of functional annotation statistics.

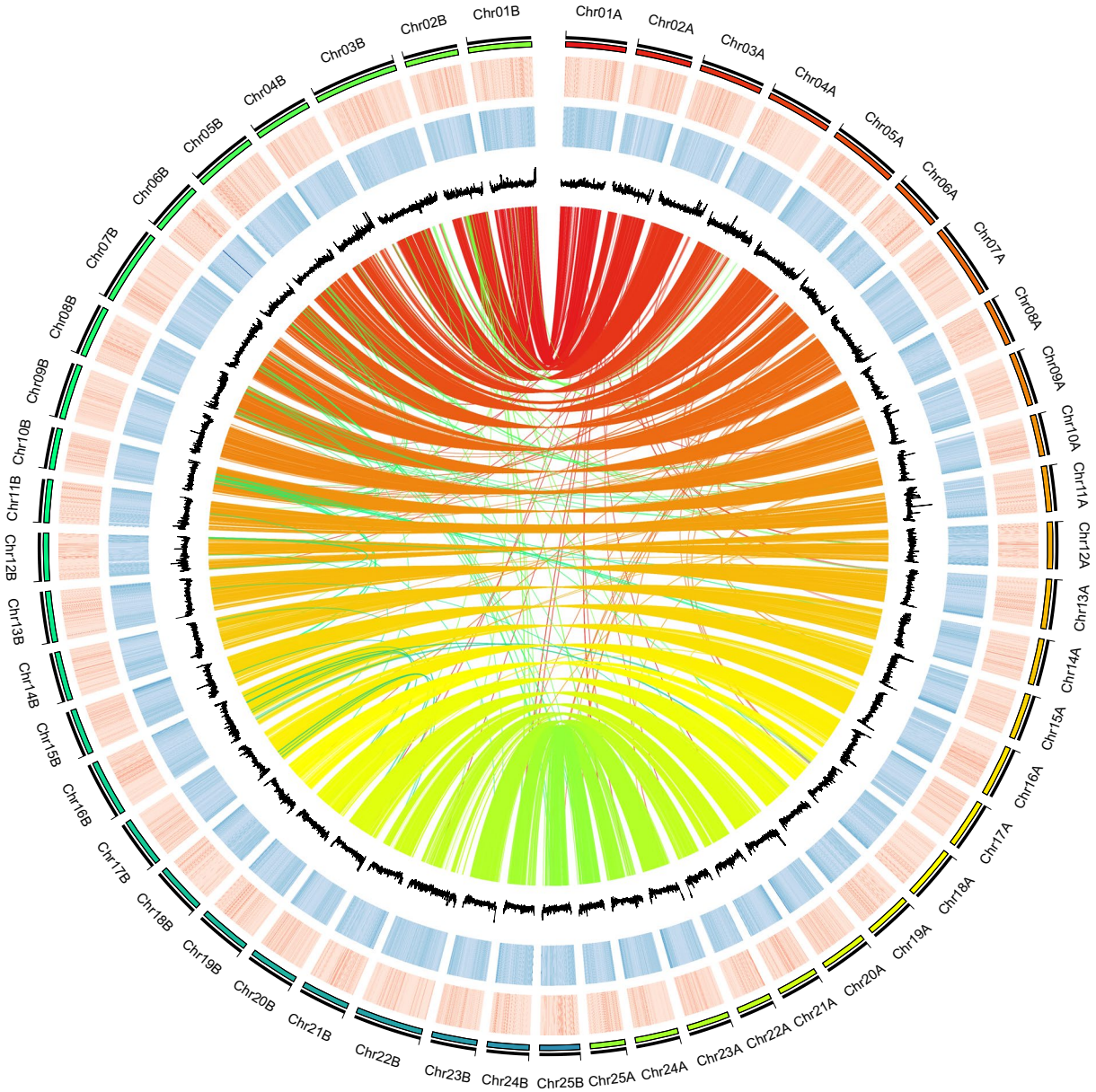

Fig. 3 The collinearity of the *C. carpio* var. *Xiangxi* assembled genome. From the outermost to the innermost circles, the tracks represent gene density, repeat sequence density, and GC content, respectively, calculated in 50 kb windows.

used to predict rRNA sequences. snoRNAs are a class of small RNA molecules that guide chemical modifications of other RNAs, mainly ribosomal RNAs, transfer RNAs and small nuclear RNAs. MiRNAs and snRNAs were identified by Infernal (v1.1.2)[43] software against the Rfam (v14.6) database[44] with default parameters. In total, 2,604 miRNAs, 28,459 tRNAs, 10,009 rRNAs, and 3,984 snRNAs were identified. In addition, the overview of this genome including the syntenic block links among the chromosomes, GC content, and gene density were analyzed and visualized using MCScanX[45] and Advanced Circos[46] in TBtools-II[47] (Fig. 3).

Ethics declarations. This work was approved by the Bioethical Committee of affair center of animal husbandry and aquaculture in XiangXi Autonomous Prefecture. All the methods used in this study were carried out following approved guidelines.

Data Records

The genome assembly of *C. carpio* var. *Xiangxi* is available at GenBank (JBNVZQ000000000.1)[48] and GWH (GWHFQUJ00000000.1)[49]. The raw data of Cyprinus carpio var. Xiangxi were deposited at the CNCB (China National Center for Bioinformation), and the accession number of RNA-seq sequencing data, Hi-C sequencing data and PacBio HiFi sequencing data is CRA023823[50].

Technical Validation

High-quality genomic DNA was extracted for sequencing library construction. The DNA samples exhibited a concentration of 166.8 ng/μl, OD260/280 ratio of 1.8, OD260/230 ratio of 2.58, and a fragment size ≥ 23 kb. The ratio of Nanodrop to Qubit concentration (N/Q) was 0.96, indicating reliable quantification and high purity. Total RNA samples used for transcriptome sequencing met stringent quality requirements, with a total yield of 25.374 μg (sufficient for three library preparations), a concentration of 845.8 ng/μl, volume ≥ 30 μl, OD260/280 ratio of 1.95, OD260/230 ratio of 1.83, a normal absorption peak at 260 nm, and an RNA integrity number (RIN) ≥ 6.3. Comprehensive quality assessments were conducted to evaluate the assembled genome of *Cyprinus carpio* var. *Xiangxi*. These assessments included genome completeness analysis using BUSCO, sequence accuracy (QV score), and structural integrity via Hi-C interaction heatmaps. The results demonstrated high completeness, accuracy, and chromosomal collinearity, supporting the robustness of the final genome assembly.

Data availability

The genome assembly of *Cyprinus carpio* var. *Xiangxi* has been deposited in GenBank under accession number JBNVZQ000000000.1 and in the Genome Warehouse (GWH) under accession number GWHFQUJ00000000.1. The raw sequencing data, including RNA-seq, Hi-C, and PacBio HiFi reads, are available at the China National Center for Bioinformation under project accession CRA023823.

Code availability

This study did not generate or use any custom code.

Received: 27 March 2025; Accepted: 2 September 2025;
Published online: 13 October 2025

References

1. Food and Agriculture Organization of the United Nations. *Yearbook of Fishery and Aquaculture Statistics*. https://www.fao.org/family-farming/detail/en/c/1145050/ (2018).
2. Xu, P. *et al*. The allotetraploid origin and asymmetrical genome evolution of the common carp Cyprinus carpio. *Nat Commun* **10**, 4625 (2019).
3. Burian, A. N., Zhao, W., Lo, T.-W. & Thurtle-Schmidt, D. M. Genome sequencing guide: An introductory toolbox to whole-genome analysis methods. *Biochem Mol Biol Educ* **49**, 815–825 (2021).
4. Li, B.-J. *et al*. Intercross population study reveals that co-mutation of mitfa genes in two subgenomes induces red skin color in common carp (*Cyprinus carpio* wuyuanensis). *zr* **44**, 276–286 (2023).
5. Qin, Q. & Zhou, X. The advantages of Cyprinus carpio var. Xiangxi. *Hunan Agricuture* 18 (2004).
6. Cheng, H. *et al*. Haplotype-resolved assembly of diploid genomes without parental data. *Nat Biotechnol* **40**, 1332–1335 (2022).
7. Bolger, A. M., Lohse, M. & Usadel, B. Trimmomatic: a flexible trimmer for Illumina sequence data. *Bioinformatics* **30**, 2114–2120 (2014).
8. Durand, N. C. *et al*. Juicer Provides a One-Click System for Analyzing Loop-Resolution Hi-C Experiments. *Cell Syst* **3**, 95–98 (2016).
9. Dudchenko, O. *et al*. De novo assembly of the Aedes aegypti genome using Hi-C yields chromosome-length scaffolds. *Science* **356**, 92–95 (2017).
10. Jia, K.-H. *et al*. SubPhaser: a robust allopolyploid subgenome phasing method based on subgenome-specific k-mers. *New Phytol* **235**, 801–809 (2022).
11. 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**, 3210–3212 (2015).
12. Li, H. Minimap2: pairwise alignment for nucleotide sequences. *Bioinformatics* **34**, 3094–3100 (2018).
13. Rhie, A., Walenz, B. P., Koren, S. & Phillippy, A. M. Merqury: reference-free quality, completeness, and phasing assessment for genome assemblies. *Genome Biol* **21**, 245 (2020).
14. Benson, G. Tandem repeats finder: a program to analyze DNA sequences. *Nucleic Acids Res* **27**, 573–580 (1999).
15. Xu, Z. & Wang, H. LTR_FINDER: an efficient tool for the prediction of full-length LTR retrotransposons. *Nucleic Acids Res* **35**, W265–268 (2007).
16. Flynn, J. M. *et al*. RepeatModeler2 for automated genomic discovery of transposable element families. *Proc Natl Acad Sci USA* **117**, 9451–9457 (2020).
17. Tarailo-Graovac, M. & Chen, N. Using RepeatMasker to identify repetitive elements in genomic sequences. *Curr Protoc Bioinformatics* Chapter 4, 4.10.1–4.10.14 (2009).
18. Jurka, J. Repbase update: a database and an electronic journal of repetitive elements. *Trends Genet* **16**, 418–420 (2000).
19. Jurka, J. *et al*. Repbase Update, a database of eukaryotic repetitive elements. *Cytogenet Genome Res* **110**, 462–467 (2005).

20. 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).
21. Li, J. T. *Cyprinus carpio* isolate SPL01, whole genome shotgun sequencing project. *GenBank* https://identifiers.org/ncbi/insdc:JAE000000000.1 (2021).
22. Xu, P. *Cyprinus carpio* haematopterus(Yellow River carp) genome assembly. Genome Warehouse, China National Center for Bioinformation https://ngdc.cncb.ac.cn/gwh/Assembly/GWHAATB00000000 (2019).
23. Xu, P. *Cyprinus carpio* 'wuyuanensis' (Wuyuan red carp) genome assembly. Genome Warehouse, China National Center for Bioinformation https://ngdc.cncb.ac.cn/gwh/Assembly/GWHAATD00000000 (2019).
24. Chen, Z. *Carassius auratus* strain Wakin, whole genome shotgun sequencing project. *GenBank* https://identifiers.org/ncbi/insdc:QPKE00000000.1 (2018).
25. Slater, G. S. C. & Birney, E. Automated generation of heuristics for biological sequence comparison. *BMC Bioinformatics* **6**, 31 (2005).
26. Stanke, M. *et al*. AUGUSTUS: ab initio prediction of alternative transcripts. *Nucleic Acids Res* **34**, W435–439 (2006).
27. Stanke, M., Steinkamp, R., Waack, S. & Morgenstern, B. AUGUSTUS: a web server for gene finding in eukaryotes. *Nucleic Acids Res* **32**, W309–312 (2004).
28. Stanke, M. & Morgenstern, B. AUGUSTUS: a web server for gene prediction in eukaryotes that allows user-defined constraints. *Nucleic Acids Res* **33**, W465–467 (2005).
29. Majoros, W. H., Pertea, M. & Salzberg, S. L. TigrScan and GlimmerHMM: two open source ab initio eukaryotic gene-finders. *Bioinformatics* **20**, 2878–2879 (2004).
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. Kovaka, S. *et al*. Transcriptome assembly from long-read RNA-seq alignments with StringTie2. *Genome Biol* **20**, 278 (2019).
32. Grabherr, M. G. *et al*. Full-length transcriptome assembly from RNA-Seq data without a reference genome. *Nat Biotechnol* **29**, 644–652 (2011).
33. Cantarel, B. L. *et al*. MAKER: an easy-to-use annotation pipeline designed for emerging model organism genomes. *Genome Res* **18**, 188–196 (2008).
34. Haas, B. J. *et al*. Improving the Arabidopsis genome annotation using maximal transcript alignment assemblies. *Nucleic Acids Res* **31**, 5654–5666 (2003).
35. Kanehisa, M., Goto, S., Sato, Y., Furumichi, M. & Tanabe, M. KEGG for integration and interpretation of large-scale molecular data sets. *Nucleic Acids Res* **40**, D109–114 (2012).
36. Ashburner, M. *et al*. Gene ontology: tool for the unification of biology. The Gene Ontology Consortium. *Nat Genet* **25**, 25–29 (2000).
37. Boeckmann, B. *et al*. The SWISS-PROT protein knowledgebase and its supplement TrEMBL in 2003. *Nucleic Acids Res* **31**, 365–370 (2003).
38. Buchfink, B., Reuter, K. & Drost, H.-G. Sensitive protein alignments at tree-of-life scale using DIAMOND. *Nat Methods* **18**, 366–368 (2021).
39. Jones, P. *et al*. InterProScan 5: genome-scale protein function classification. *Bioinformatics* **30**, 1236–1240 (2014).
40. Mitchell, A. *et al*. The InterPro protein families database: the classification resource after 15 years. *Nucleic Acids Res* **43**, D213–221 (2015).
41. Lowe, T. M. & Eddy, S. R. tRNAscan-SE: a program for improved detection of transfer RNA genes in genomic sequence. *Nucleic Acids Res* **25**, 955–964 (1997).
42. Lagesen, K. *et al*. RNAmmer: consistent and rapid annotation of ribosomal RNA genes. *Nucleic Acids Res* **35**, 3100–3108 (2007).
43. Nawrocki, E. P., Kolbe, D. L. & Eddy, S. R. Infernal 1.0: inference of RNA alignments. *Bioinformatics* **25**, 1335–1337 (2009).
44. Kalvari, I. *et al*. Rfam 14: expanded coverage of metagenomic, viral and microRNA families. *Nucleic Acids Res* **49**, D192–D200 (2021).
45. Wang, Y. *et al*. MCScanX: a toolkit for detection and evolutionary analysis of gene synteny and collinearity. *Nucleic Acids Res* **40**, e49 (2012).
46. Krzywinski, M. *et al*. Circos: an information aesthetic for comparative genomics. *Genome Res* **19**, 1639–1645 (2009).
47. Chen, C. *et al*. TBtools-II: A "one for all, all for one" bioinformatics platform for biological big-data mining. *Molecular Plant* **16**, 1733–1742 (2023).
48. Li, Z. A high-quality chromosome-level genome assembly of the *Cyprinus carpio* var. *Xiangxi*. *GenBank* https://identifiers.org/ncbi/insdc:JBNVZQ000000000.1 (2025).
49. Li, Z. A high-quality chromosome-level genome assembly of the *Cyprinus carpio* var. *Xiangxi*. Genome Warehouse, China National Center for Bioinformation https://ngdc.cncb.ac.cn/gwh/Assembly/92499/show (2025).
50. Li, Z. Raw data for *Cyprinus carpio* var. *Xiangxi*. SRA, China National Center for Bioinformation https://ngdc.cncb.ac.cn/gsa/search?searchTerm=CRA023823 (2025).

Acknowledgements

This study was funded by Hunan Provincial Natural Science Foundation under Grant No. 2024JJ7550 and the earmarked fund for Hunan modern agricultural industry technology system (HARS-07).

Author contributions

Zhe Li conducted the writing and data analysis. Ziao Wang and Yinghai Peng participated in data analysis. Yazhou Hu and Xianwen Zhou provided guidance on experimental design and manuscript revision. Xianwen Zhou also provided project support.

Competing interests

The authors declare no competing interests.

Additional information

Correspondence and requests for materials should be addressed to Y.H. or X.Z.