Journal of Heredity, 2026, **117**, 834–842
https://doi.org/10.1093/jhered/esag029
Published: 26 March 2026
Genome Resources

Reference genome assembly of the mandarin rat snake (*Euprepiophis mandarinus*)

Lianming Du[1], Lian Wang[2], Jiahao Chen[1,3], Songwen Tan[3], Peng Guo[3,*], Qin Liu[4,*]

[1]Institute for Advanced Study, Chengdu University, Chengdu, China
[2]Sichuan Academy of Forestry, Chengdu, China
[3]Faculty of Agriculture, Forestry and Food Engineering, Yibin University, Yibin, China
[4]College of Life Sciences, Sichuan Normal University, Chengdu, China

*Corresponding authors: Peng Guo, Faculty of Agriculture, Forestry and Food Engineering, Yibin University, Yibin 644000, China. Email: ybguop@163.com; Qin Liu, College of Life Scieneces, Sichuan Normal University, Chengdu 610101, China. Email: chimliu@sicnu.edu.cn

Abstract

The mandarin rat snake (*Euprepiophis mandarinus*), a widely distributed and ecologically important species endemic to Asia, serves as an ideal candidate for establishing a reference genome for the genus *Euprepiophis*. This is due to its wide distribution, and its overlapping distribution with and morphological similarities to the endangered *Euprepiophis perlaceus*. Here we present the first high-quality, chromosome-level reference genome of *E. mandarinus*, assembled using PacBio HiFi long-read and Hi-C sequencing. The final assembly has a length of 1.65 Gb contained in 56 scaffolds, a scaffold N50 of 210 Mb, a benchmarking universal single-copy orthologs completeness of 98.2%, and 99.12% bases anchored onto 19 chromosomes. We identified 22,332 protein-coding genes, 3,440 noncoding RNAs, and approximately 7 million repeat elements accounting for 58.38% of the assembled genome. This reference genome provides a valuable resource for taxonomic, phylogenetic, comparative, and conservation studies of the genus *Euprepiophis*.

Keywords chromosome genomics, reference genome, sequencing, snake

Introduction

Euprepiophis is a genus of nonvenomous colubrid snakes comprising three species: *Euprepiophis conspicillata*, *Euprepiophis mandarinus*, and *Euprepiophis perlaceus* (Schoch et al. 2020; Uetz et al. 2021). Among these, *E. mandarinus* and *E. perlaceus* are distributed in the continent of Asia and share numerous morphological similarities, whereas *E. conspicillata* is endemic to Japan and shows distinct morphological divergence from the other two species (Chen et al. 2014; Tang et al. 2021). *E. perlaceus* is an endangered species endemic to western China, once considered extinct (Tang et al. 2021; Song et al. 2023). *E. mandarinus*, commonly known as the mandarin trinket snake or mandarin rat snake, is a widely distributed species endemic to Asia (Ji et al. 2012; Ashaharraza et al. 2019). Although classified as Least Concern on the IUCN Red List, it has been listed as Vulnerable on the Red List of China's Vertebrates because of population declines (Ji et al. 2012; Jiang et al. 2016). Historically, these snakes were classified under several genera based on morphological traits until molecular phylogenetic evidence definitively placed them in the resurrected genus *Euprepiophis* (Utiger et al. 2002).

The mandarin rat snake is a shy and secretive species, primarily active at dawn or dusk in dense forested environments with high humidity. It predominantly feeds on small vertebrates, containing rodents, lizards, and frogs, playing a pivotal role in maintaining ecological balance by regulating populations of rodents and reptiles. Due to its mild temperament and bright colors, the mandarin rat snake has become highly sought-after pet. Despite its ecological and commercial significance, the species remains poorly studied, with a notable lack of molecular data, particularly in phylogenetic research. To date, the only genome available for the genus *Euprepiophis* is the complete mitochondrial genome of *E. perlaceus* (Wan et al. 2016). Recently, whole-genome sequencing has emerged as a revolutionary tool for producing comprehensive genetic data and has been widely used in studies of snake evolution, development, adaptation, and phylogeny (Peng et al. 2023; Yan et al. 2023; Lv et al. 2025).

The broad distribution of *E. mandarinus* not only makes it the most accessible species within the genus for sample collection but also provides a valuable opportunity to investigate population-level genetic variation. Therefore, it serves as an ideal candidate for establishing a reference genome for the genus *Euprepiophis*. In this study, we present a high-quality, chromosome-level genome assembly of *E. mandarinus* generated using PacBio, Hi-C, and Illumina sequencing data. This assembled genome not only serves as a crucial resource for investigating genetic diversity and population genetics in *E. mandarinus*, but also provides a reference

Received: 30 December 2025. **Revised:** 10 March 2026. **Accepted:** 24 March 2026

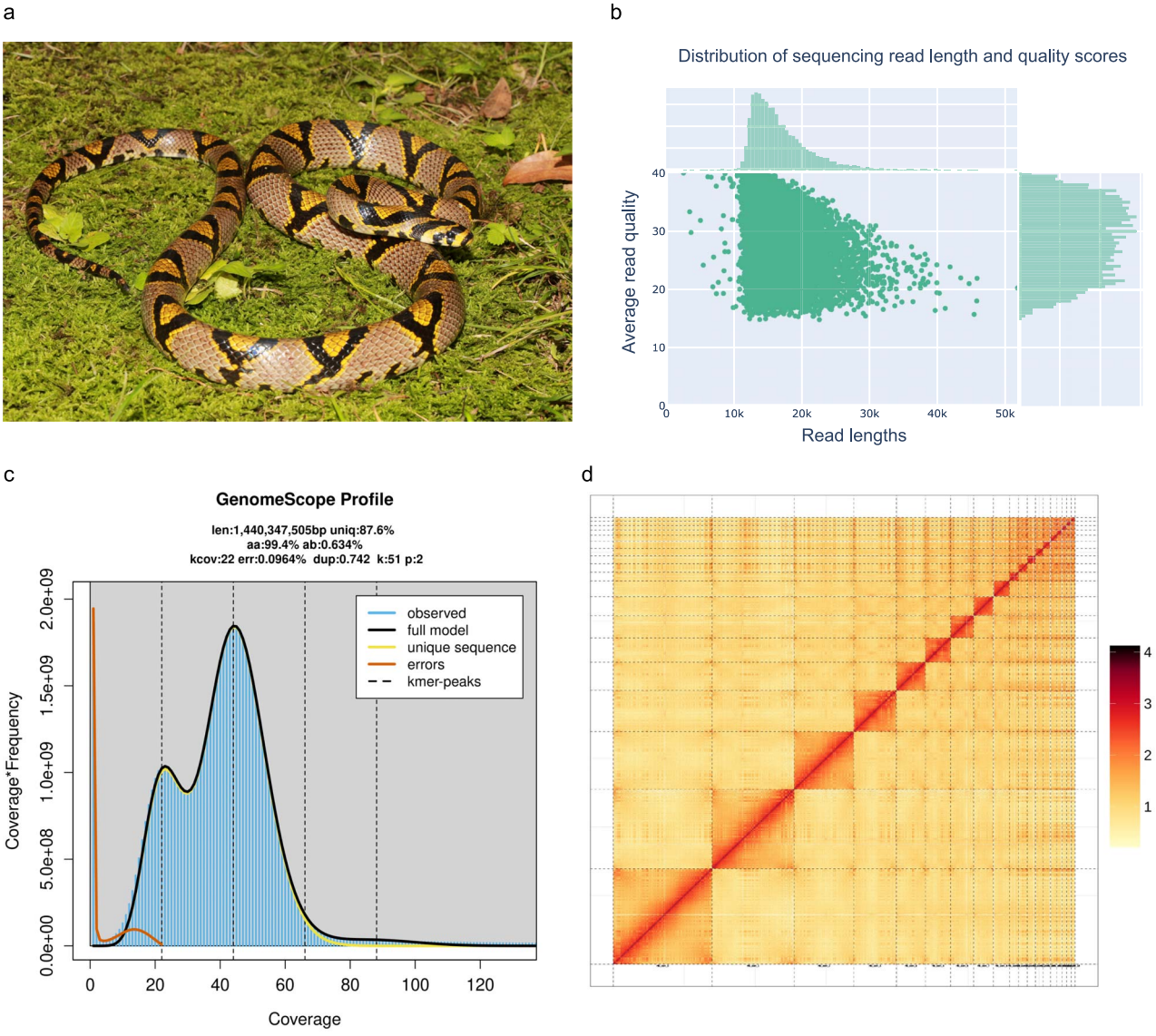

Fig. 1 Sequenced specimen and genome profile of *Euprepiophis mandarinus*. a) Photograph of the mandarin rat snake. c) The quality and length distribution of PacBio long-reads. b) The *k*-mer ($K = 51$) distribution for genome size estimation. d) Hi-C interaction heatmap for assembled genome.

framework for evolution, phylogenetics, taxonomic classification, and comparative genomics of the genus *Euprepiophis*.

Methods

Biological materials

An adult female of *E. mandarinus* was collected from Guangyuan City, Sichuan Province, China in July 2025 (Fig. 1a). For whole-genome sequencing, the genomic DNA was extracted from muscle tissue using the cetyltrimethylammonium bromide method (Doyle and Doyle 1987). The integrity of the genomic DNA was assessed using agarose gel electrophoresis. We determined the purity and quantity of the total DNA using NanoDrop 2000 Spectrophotometer (Thermo Fischer Scientific) and Qubit Fluorometer (Invitrogen). In addition, we obtained seven samples (muscle, blood, heart, kidney, liver, lung, and spleen) from the same specimen for RNA sequencing. Total RNA was isolated using RNAprep Pure Tissue Kit (TIANGEN) following the manufacturer's protocol. The voucher specimen was deposited in the Snake Collection of Yibin University under voucher number YBU250609.

Nucleic acid library preparation and sequencing

For short-read sequencing, the CWseq Universal DirectFast DNA Library Prep Kit (Illumina & MGI) was used for sample preparation to generate sequencing library according to the manufacturer's instructions. The library was analyzed using the Agilent 2100 Bioanalyzer (Agilent Technologies, USA), and its effective concentration was accurately quantified using the qPCR method. The long-read sequencing library was constructed using the SMRTbell Express Template Prep Kit 3.0. For Hi-C sequencing, muscle sample was first fixed with 1% formaldehyde to cross-link DNA and proteins. The cross-linked DNA was digested with *Hind*III restriction enzyme, labeled with biotin in end, circularized,

Table 1 Software tools and versions for genome assembly and analysis.

Procedures	Software	Version
Assembly		
***k*-mer counting**	Jellyfish	2.3.1
Estimation of genome size and heterozygosity	GenomeScope	2.0
***De novo* contig assembly**	Hifiasm	0.19.9
Scaffolding		
Hi-C data mapping	HiCUP	0.9.2
Cluster, orientate, and order the contigs	ALLHiC	0.9.14
Fine-tune the assembly	Juicebox	2.22
Genome quality assessment		
Assembly completeness	BUSCO	5.6.0
	CEGMA	2.5
	Merqury	1.3
Genome annotation		
Homology-based gene prediction	genBlastA	1.0.4
	GeneWise	2.4.1
Transcriptome-based gene prediction	HISAT	2.2.1
	StringTie	2.2.3
***De novo* gene prediction**	Augustus	3.5.0
Integration of gene prediction	EVidenceModeler	2.1.0
Update of gene annotation	PASA	2.5.3
Gene function annotation	DIAMOND	2.1.11
	InterProScan	5.59
	BlastKOALA	3.1
Repeat annotation		
Transposon identification	RepeatMasker	4.1.7
	RepeatModeler	2.0.6
Tandem repeat identification	Krait	2.0.6
Nonding-RNA annotation		
	Infernal	1.1.5
	tRNAscan-SE	2.0.12
	Barrnap	0.9

and purified. The purified DNA is fragmented to a size of 300 to 500 bp, and the biotin-labeled DNA fragment was finally separated on Dynabeads M-280 Streptavidin (Life Technologies) to construct library. For transcriptome sequencing, the cDNA libraries for each tissue sample were prepared using Hieff NGS Ultima Dual-mode RNA Library Prep Kit. The constructed long-read sequencing library was evenly loaded onto SMRT Cell chips and sequenced on the PacBio Revio platform (Pacific Biosciences, CA, USA). The short-read, Hi-C, and transcriptome sequencing libraries were sequenced on DNBSEQ-T7 platform with PE150 mode.

Genome assembly

A list of all tools and versions used for genome assembly and analysis is available in Table 1. First, we used short reads to perform genome survey analysis. The raw reads were filtered to remove low-quality reads and adaptors using fastp (Chen 2023) with default parameters. We used Jellyfish (Marçais and Kingsford 2011) to analyze the *k*-mer frequency distribution of the remaining high-quality reads with a *K* value of 51 according to a previous study (Peng et al. 2023). The *k*-mer distribution result was then imported to Genomescope (Ranallo-Benavidez et al. 2020) to predict genome size and heterozygosity.

The PacBio SMRT-Analysis software package was applied to remove low-quality polymerase reads from PacBio sequencing data. We used SMRTLink with parameters –min-passes=3 and –min-rq = 0.99 to process the remaining subreads to generate HiFi reads. The HiFi reads were assembled into contigs using Hifiasm (Cheng et al. 2021) with default parameters. The raw Hi-C sequencing data were filtered using fastp with default parameter to generate high-quality reads which were subsequently mapped against the preliminary contigs using HiCUP (Wingett et al. 2015). We employed ALLHiC (Zhang et al. 2019) to cluster, orientate, and order the contigs for scaffold-level assembly. Finally, Juicebox (Durand et al. 2016) was adopted to manually fine-tune the assembly, resulting in a chromosome-level assembly.

Genome quality assessment

First, the completeness of the genome assembly was assessed using benchmarking universal single-copy orthologs (BUSCO)

Table 2 Sequencing and assembly statistics, and accession numbers.

Accessions	NCBI BioProject	PRJNA1391878
	NCBI GenBank	JBTBFG000000000
	NGDC BioProject	PRJCA054138
	NGDC GSA	CRA035691
	NGDC GWH	GWHHOUK00000000.1
Genome sequence	Short reads	1 DNBSEQ (DNBSEQ-T7) runs: 345.32 million reads, 103.60 Gbases
	PacBio HiFi reads	1 PacBio_SMRT (Revio) cell: 6.97 million reads, 116.49 Gbases
	Hi-C reads	1 DNBSEQ (DNBSEQ-T7) runs: 664.33 million reads, 199.30 Gbases
	RNA-seq reads	7 DNBSEQ (DNBSEQ-T7) runs: 156.47 million reads, 46.94 Gbases
Genome assembly	Number of contigs	90
	Contig N50 (Mb)	147.54
	Longest contig (Mb)	313.26
	Number of scaffolds	56
	Scaffold N50 (Mb)	210.15
	Longest scaffold (Mb)	348.92
	Size of final assembly (Mb)	1,650.60
Assembly quality	BUSCO completeness	C: 98.2% [S: 97.5%, D: 0.7%], F: 0.7%, M: 1.1%, n: 3,354
	k-mer completeness	89.53%
	QV score	51.99
	CEGMA completeness	91.13%

(Manni et al. 2021) based on vertebrata_odb10 lineage dataset and core eukaryotic genes mapping approach (CEGMA) (Parra et al. 2007). Then, we mapped short clean reads to the assembled genome using BWA (Li and Durbin 2009) for accurate assessment. We further assessed the quality value (QV) and *k*-mer completeness using Merqury (Rhie et al. 2020) with 21-mers generated from short reads. We also performed chromosomal synteny analysis between *E. mandarinus* and other four snakes (*Elaphe taeniura*, *Pantherophis guttatus*, *Ahaetulla prasina*, and *Thamnophis elegans*) with well assembled genomes using JCVI (Tang et al. 2024).

Genome annotation

We have used homology-based, transcriptome-based, and *ab initio* methods to predict protein-coding gene structures. The protein sequences of five species (including *P. guttatus*, *Thamnophis sirtalis*, *T. elegans*, *A. prasina*, and *Mus musculus*) were downloaded from NCBI database for performing homology alignment using genBlastA (She et al. 2009). Gene structures were refined using GeneWise (Birney et al. 2004) based on the candidate homologous regions. After quality control, the RNA sequencing data were aligned to assembled genome with HISAT2 (Kim et al. 2019). The alignment results were analyzed using StringTie (Pertea et al. 2015) to perform genome-guided transcript assembly. We used Augustus (Keller et al. 2011), geneid (Alioto et al. 2018), and GENSCAN (Burge and Karlin 1997) to carry out *ab initio* gene prediction. The gene models from different annotation strategies were integrated using EVidenceModeler (Haas et al. 2008) into a nonredundant and complete gene set which was further corrected using PASA (Haas et al. 2003).

For functional annotation, we aligned predicted protein sequences against the NCBI non-redundant (nr) database and Swiss-Prot database using DIAMOND (Buchfink et al. 2021). The sequences were aligned with the InterPro database using InterProScan (Jones et al. 2014) to identify conserved domains, structural motifs, and functional signatures. InterProScan analysis also performed Gene Ontology assignment. We submitted the protein sequences to BlastKOALA (Kanehisa et al. 2016) server for Kyoto Encyclopedia of Genes and Genomes (KEGG) pathway identification.

We aligned the assembled genome against Rfam database for detecting non-coding RNAs (rRNAs, tRNAs, snRNAs, and miRNAs) using Infernal (Nawrocki and Eddy 2013). tRNAscan-SE (Chan et al. 2021) was used to explore tRNAs with default parameters. Barrnap was used to predict ribosomal RNAs with the –kingdom parameter set to euk.

We applied RepeatMasker and RepeatProteinMask to perform homology-based prediction of repeat elements. For *de novo* repeat prediction, LTR_Finder (Xu and Wang 2007), Piler (Edgar and Myers 2005), RepeatScout (Price et al. 2005), and RepeatModeler (Flynn et al. 2020) were used to build a library of repetitive sequences. Then, RepeatMasker was used to predict transposable elements based on the built library. Additionally, we identified tandem repeats using Krait (Du et al. 2025a) with pytrf (Du et al. 2025b) as search engine, the maximum motif size of 100 bp and minimum length of 10 bp.

Results

Sequencing data

The short-read sequencing library generated about 345.32 million paired-end reads, producing around 100.55 Gb clean bases (Fig. 1b, Supplementary Table S1). The PacBio HiFi library yielded 6,967,322 reads with a mean length of 16,719 bp and N50 length of 16,604 bp, corresponding to approximately 116.49 Gb bases (Supplementary Table S2). The Hi-C library produced 199.30

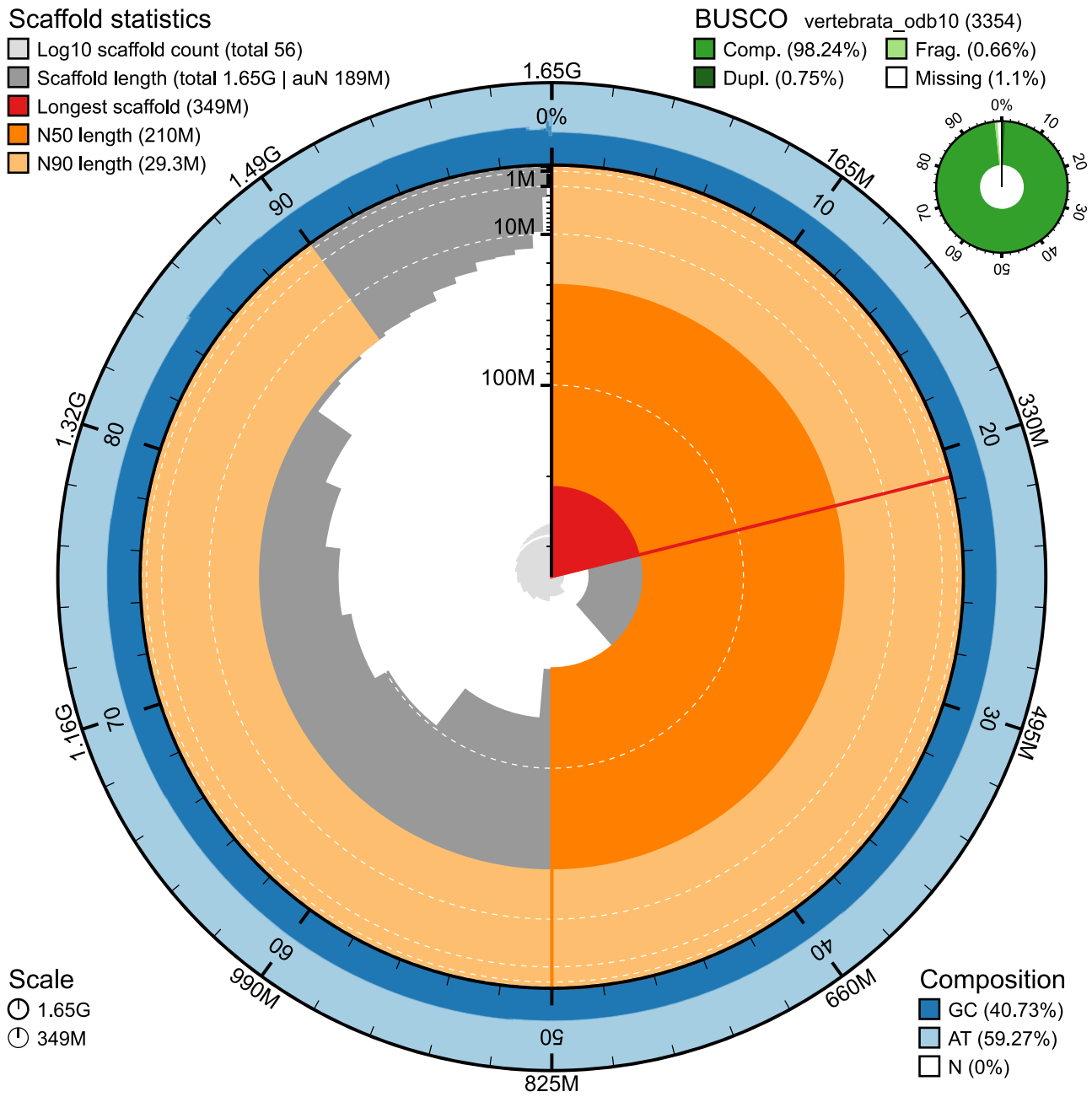

Fig. 2 Snail plot for visualization of genome assembly and assessment metrics. The plot shows N50, N90, longest contig and total genome length, GC content distribution, as well as BUSCO scores.

Gb raw data, consisting of a total of 664,327,673 read pairs (Supplementary Table S3). In total, 46.94 Gb of raw transcriptomic data were generated across seven tissues with read count ranging from 20.68 million to 25.21 million for each sample (Supplementary Table S4).

Genome assembly and evaluation

The genome size of *E. mandarinus* was estimated to be about 1440.35 Mb, with nucleotide heterozygosity rate of 0.634% and sequencing error rate of 0.0964% (Fig. 1c). The assembled contig-level genome comprises 90 contigs spanning 1,650,595,325 base pairs, with an N50 value of 147.54 Mb (Supplementary Table S5). After scaffolding, we gained a chromosome-level genome with size of 1,650,598,725 bp and 99.12% bases anchored onto 19 chromosomes (Fig. 1d, Table 2, Supplementary Table S6). The N50 length and maximum sequence length of the final assembly were 210,147,468 bp and 348,916,335 bp, respectively (Fig. 2, Table 2). The BUSCO result revealed 98.2% completeness, and 231 (93.15%) out of 248 core eukaryotic genes from CEGMA were identified in the assembled genome (Supplementary Tables S7 and S8). The mapping result indicated that 99.85% paired-end reads could be aligned to the assembled genome (Supplementary Table S9). The QV score and *k*-mer completeness were estimated as 51.99 and 89.53%, separately. A high degree of synteny was observed between our assembled genome and four other snake genomes,

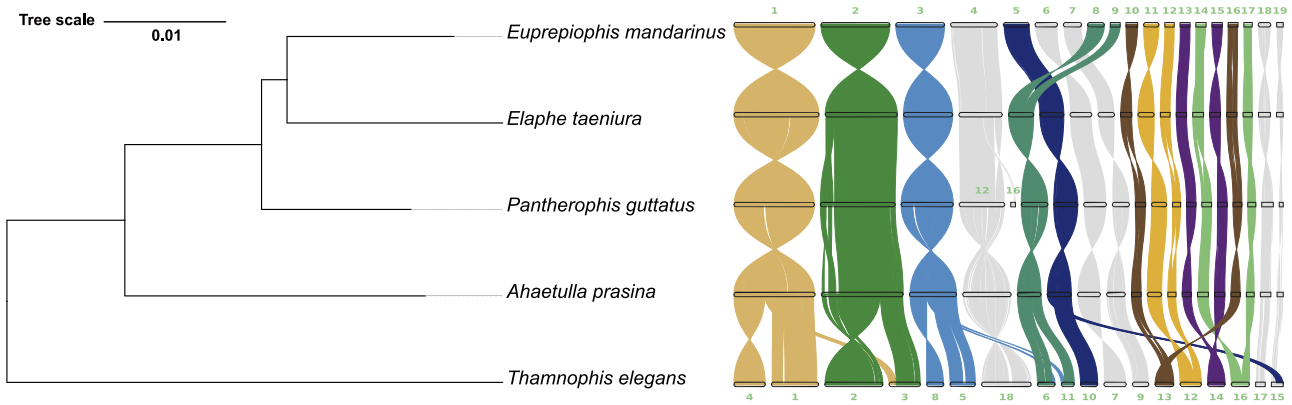

Fig. 3 Gene synteny analysis of genome chromosomes between *Euprepiophis mandarinus* and four other snakes.

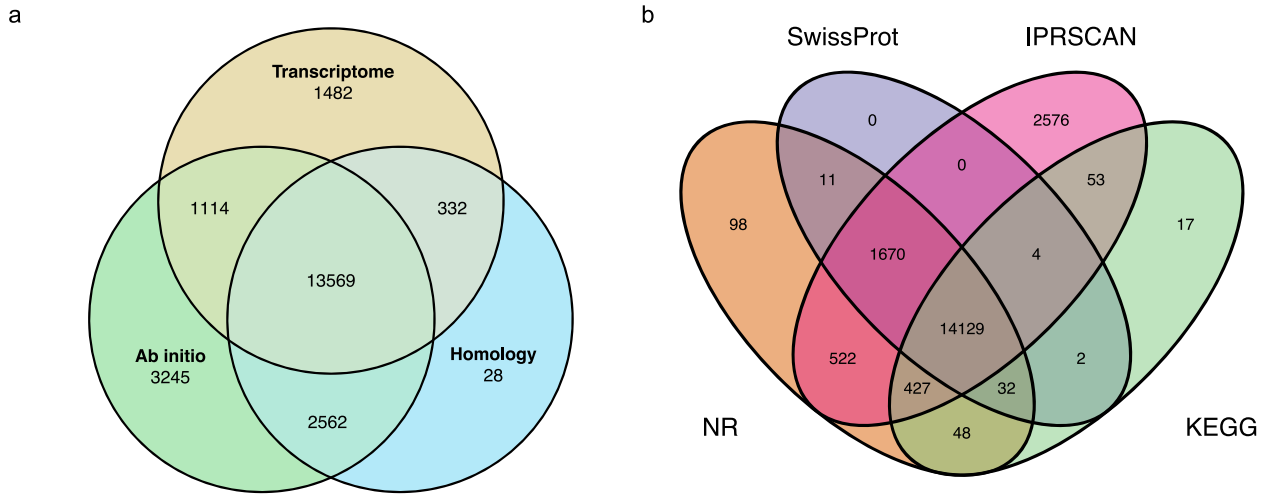

Fig. 4 Venn diagrams for protein-coding gene annotation. a) Genes annotated by different methods. b) Gene functions annotated by different databases.

further confirming the high accuracy of the *E. mandarinus* chromosomal assembly (Fig. 3).

Gene and repeat annotation

In total, we predicted 22,332 protein-coding genes, with an average gene length of 28,578.99 bp, an average CDS length of 1,406.71 bp, and an average exon number per gene of 8.21 (Fig. 4a, Supplementary Table S10). Overall, 17,531 (80.87%) predicted protein-coding genes were functionally annotated by at least one functional database (Fig. 4b, Supplementary Table S11). We also detected 3,440 non-coding RNAs including 273 miRNAs, 1,083 rRNAs, 1,549 tRNAs, and 342 snRNAs (Supplementary Table S12). Furthermore, we identified a total of 7,111,352 repeat elements with total length of 963.69 Mb accounting for 58.38% of the assembled genome (Supplementary Table S13). The distribution of these functional elements and genome features in chromosomes is shown in Fig. 5.

Discussion

Whole-genomic data has emerged as a versatile and powerful research infrastructure, enabling studies across diverse fields including comparative genomics, evolution biology, and adaptation research. For instance, a recent study employed large-scale genome analyses to decipher the evolution of limb loss in snakes (Peng et al. 2023). Despite the existence of more than 4,000 extant snake species (Title et al. 2024), genomic data remain highly limited according to the NCBI database, only 173 species of snakes have had their genomes sequenced, with merely 29 assembled to chromosome-level resolution. Our assembled genome fills a critical gap by providing a high-quality reference resource for advancing research on snake evolution and biodiversity.

Currently, phylogenetic inferences within the genus *Euprepiophis* have relied mainly on mitochondrial data (Utiger et al. 2002; Wan et al. 2016). However, many studies have reported incongruent phylogenetic signals and divergent evolutionary histories between mitochondrial and nuclear genes (Gonçalves et al. 2007; Platt et al. 2018; Laine et al. 2023). Additionally, nuclear genes have been shown to provide more robust phylogenetic resolution for closely related lineages (Near and Keck 2013; Dool et al. 2016). Whole-genomic data have recently been used to reconstruct phylogenetic relationships between 18 snake species and identify adaptive genetic variants (Roberts et al. 2024). Our assembled genome serves as a valuable reference genome for exploring phylogenetics, taxonomic classification, and conservation management of the genus *Euprepiophis*.

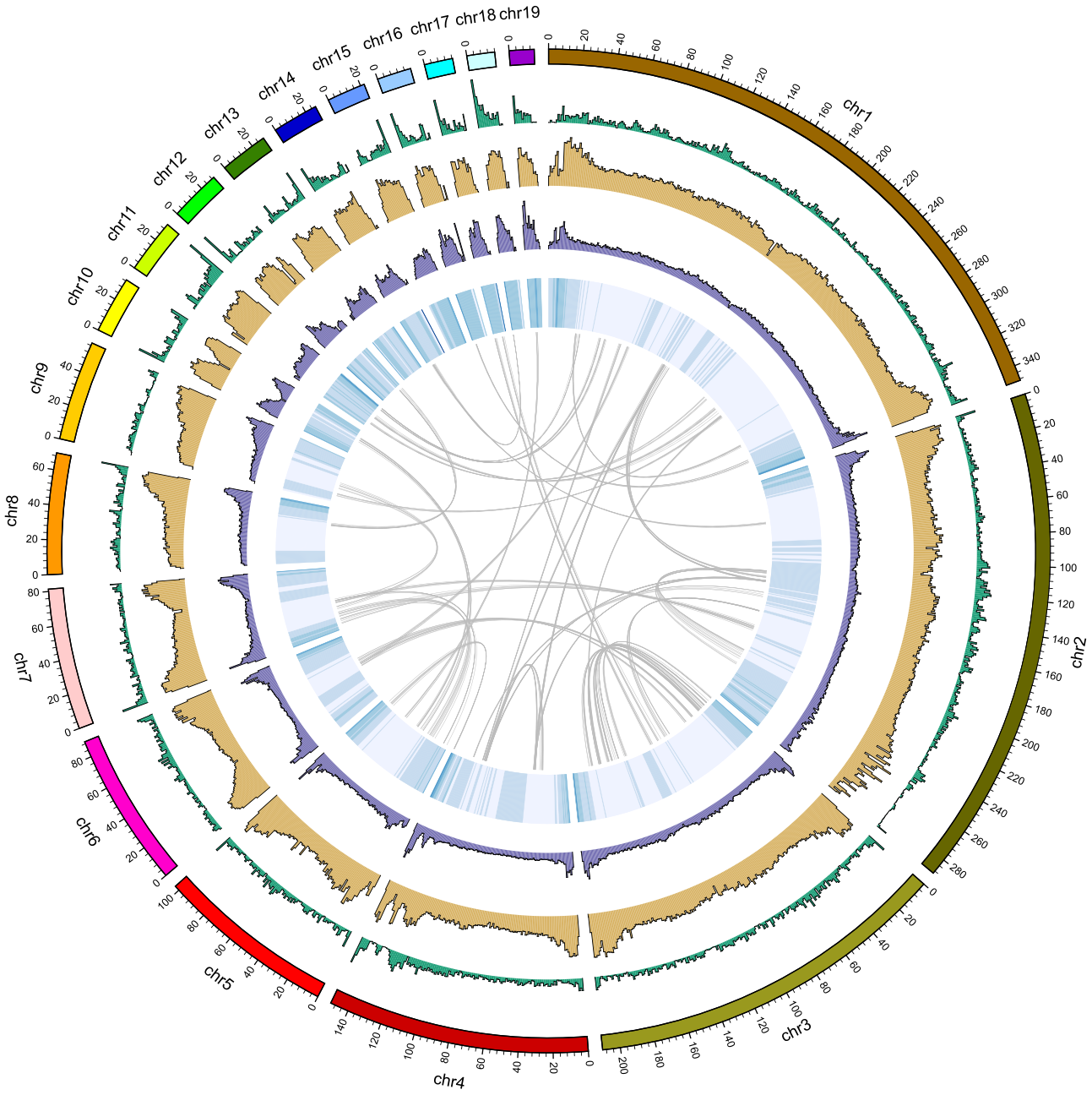

Fig. 5 Circos plot for showing distribution of genomic features. The tracks from outermost to innermost are chromosomes, gene density, transposon density, tandem repeat density, GC content, and synteny among chromosomes.

Acknowledgments

We are grateful to Haixiang Yang, Bing Lv and Cheng Zhao at Yibin University for their help in sample collection.

Author contributions

Lianming Du (Formal analysis, Visualization, Writing—original draft), Lian Wang (Data curation, Resources), Jiahao Chen (Methodology), Songwen Tan (Investigation), Peng Guo (Conceptualization, Writing—review & editing), and Qin Liu (Project administration, Supervision, Writing—review & editing)

Supplementary material

Supplementary material is available at *Journal of Heredity* online.

Funding

This work was supported by the Sichuan Science and Technology Program (2026NSFSC0139), and the second Tibetan Plateau Scientific Expedition and Research Program (2024QZKK0200).

Data availability

Data generated for this study including PacBio long-read, Hi-C, short-read, RNA-seq raw sequencing data are deposited in NGDC Genome Sequence Archive (GSA) under accession CRA035691. The chromosome-level genome assembly is available at GenBank under accession JBTBFG000000000 in BioProject PRJNA1391878. The genome assembly is also deposited at NGDC Genome Warehouse (GWH) under accession GWHHOUK00000000.1.

References

Alioto T, Blanco E, Parra G, Guigó R. Using geneid to identify genes. *Curr Protoc Bioinformatics*. 2018;64:e56. https://doi.org/10.1002/cpbi.56

Ashaharraza K, Rangasamy V, Lalremsanga HT, Biakzuala L, Sailo J, Charlton T. A new state record of the mandarin rat snake *Euprepiophis mandarinus* (Cantor, 1842) (Squamata: Colubridae: Coronellini) from Mizoram, India. *Amphib Reptile Conse*. 2019;13:230–234.

Birney E, Clamp M, Durbin R. GeneWise and genomewise. *Genome Res*. 2004;14:988–995. https://doi.org/10.1101/gr.1865504

Buchfink B, Reuter K, Drost HG. Sensitive protein alignments at tree-of-life scale using DIAMOND. *Nat Methods*. 2021;18:366–368. https://doi.org/10.1038/s41592-021-01101-x

Burge C, Karlin S. Prediction of complete gene structures in human genomic DNA. *J Mol Biol*. 1997;268:78–94. https://doi.org/10.1006/jmbi.1997.0951

Chan PP, Lin BY, Mak AJ, Lowe TM. tRNAscan-SE 2.0: improved detection and functional classification of transfer RNA genes. *Nucleic Acids Res*. 2021;49:9077–9096. https://doi.org/10.1093/nar/gkab688

Chen S. Ultrafast one-pass FASTQ data preprocessing, quality control, and deduplication using fastp. *iMeta*. 2023;2:e107. https://doi.org/10.1002/imt2.107

Chen X, Jiang K, Guo P, Huang S, Rao D, Ding L, Takeuchi H, Che J, Zhang YP, Myers EA, et al. Assessing species boundaries and the phylogenetic position of the rare Szechwan ratsnake, Euprepiophis perlaceus (Serpentes: Colubridae), using coalescent-based methods. *Mol Phylogenet Evol*. 2014;70:130–136. https://doi.org/10.1016/j.ympev.2013.09.003

Cheng H, Concepcion GT, Feng X, Zhang H, Li H. Haplotype-resolved de novo assembly using phased assembly graphs with hifiasm. *Nat Methods*. 2021;18:170–175. https://doi.org/10.1038/s41592-020-01056-5

Dool SE, Puechmaille SJ, Foley NM, Allegrini B, Bastian A, Mutumi GL, Maluleke TG, Odendaal LJ, Teeling EC, Jacobs DS. Nuclear introns outperform mitochondrial DNA in interspecific phylogenetic reconstruction: lessons from horseshoe bats (Rhinolophidae: Chiroptera). *Mol Phylogenet Evol*. 2016;97:196–212. https://doi.org/10.1016/j.ympev.2016.01.003

Doyle JJ, Doyle JL. A rapid DNA isolation procedure for small quantities of fresh leaf tissue. *Phytochem Bull*. 1987;19:11–15

Du L, Chen J, Sun D, Zhao K, Zeng Q, Yang N. Krait2: a versatile software for microsatellite investigation, visualization and marker development. *BMC Genomics*. 2025a;26:72. https://doi.org/10.1186/s12864-025-11252-2

Du L, Sun D, Chen J, Zhou X, Zhao K, Zeng Q, Yang N. Pytrf: a python package for finding tandem repeats from genomic sequences. *BMC Bioinformatics*. 2025b;26:151. https://doi.org/10.1186/s12859-025-06168-3

Durand NC, Robinson JT, Shamim MS, Machol I, Mesirov JP, Lander ES, Aiden EL. Juicebox provides a visualization system for Hi-C contact maps with unlimited zoom. *Cell Syst*. 2016;3:99–101. https://doi.org/10.1016/j.cels.2015.07.012

Edgar RC, Myers EW. PILER: identification and classification of genomic repeats. *Bioinformatics.* 2005;21:i152–i158. https://doi.org/10.1093/bioinformatics/bti1003

Flynn JM, Hubley R, Goubert C, Rosen J, Clark AG, Feschotte C, Smit AF. RepeatModeler2 for automated genomic discovery of transposable element families. *Proc Natl Acad Sci USA*. 2020;117:9451–9457. https://doi.org/10.1073/pnas.1921046117

Gonçalves H, Martínez-Solano I, Ferrand N, García-París M. Conflicting phylogenetic signal of nuclear vs mitochondrial DNA markers in midwife toads (Anura, Discoglossidae, Alytes): deep coalescence or ancestral hybridization? *Mol Phylogenet Evol*. 2007;44:494–500. https://doi.org/10.1016/j.ympev.2007.03.001

Haas BJ, Delcher AL, Mount SM, Wortman JR, Smith RK Jr, Hannick LI, Maiti R, Ronning CM, Rusch DB, Town CD, et al. Improving the Arabidopsis genome annotation using maximal transcript alignment assemblies. *Nucleic Acids Res*. 2003;31:5654–5666. https://doi.org/10.1093/nar/gkg770

Haas BJ, Salzberg SL, Zhu W, Pertea M, Allen JE, Orvis J, White O, Buell CR, Wortman JR. Automated eukaryotic gene structure annotation using EVidenceModeler and the program to assemble spliced alignments. *Genome Biol*. 2008;9:R7. https://doi.org/10.1186/gb-2008-9-1-r7

Ji X, Wang Y, Guo P. *Euprepiophis mandarinus*. The IUCN Red List of Threatened Species 2012. e.T192138A2045703. [accessed 2025 Dec 9]. https://doi.org/10.2305/IUCN.UK.2012-1.RLTS.T192138A2045703.en

Jiang Z, Jiang J, Wang Y, Zhang E, Zhang Y, Li L, Xie F, Cai B, Cao L, Zheng G, et al. Red list of China's vertebrates. *Biodivers Sci*. 2016;24:500–551.

Jones P, Binns D, Chang HY, Fraser M, Li W, McAnulla C, McWilliam H, Maslen J, Mitchell A, Nuka G, et al. InterProScan 5: genome-scale protein function classification. *Bioinformatics.* 2014;30:1236–1240. https://doi.org/10.1093/bioinformatics/btu031

Kanehisa M, Sato Y, Morishima K. BlastKOALA and GhostKOALA: KEGG Tools for Functional Characterization of Genome and Metagenome Sequences. *J Mol Bio*. 2016;428:726–731. https://doi.org/10.1016/j.jmb.2015.11.006

Keller O, Kollmar M, Stanke M, Waack S. A novel hybrid gene prediction method employing protein multiple sequence alignments. *Bioinformatics*. 2011;27:757–763. https://doi.org/10.1093/bioinformatics/btr010

Kim D, Paggi JM, Park C, Bennett C, Salzberg SL. Graph-based genome alignment and genotyping with HISAT2 and HISAT-genotype. *Nat Biotechnol*. 2019;37:907–915. https://doi.org/10.1038/s41587-019-0201-4

Laine VN, Sävilammi T, Wahlberg N, Meramo K, Ossa G, Johnson JS, Blomberg AS, Yeszhanov AB, Yung V, Paterson S, et al. Whole-genome analysis reveals contrasting relationships among nuclear and mitochondrial genomes between three sympatric bat species. *Genome Biol Evol*. 2023;15:evac17. https://doi.org/10.1093/gbe/evac175

Li H, Durbin R. Fast and accurate short read alignment with Burrows-Wheeler transform. *Bioinformatics.* 2009;25:1754–1760. https://doi.org/10.1093/bioinformatics/btp324

Lv Y, Wu W, Ren JL, Fujita MK, Song M, Wang Z, Jiang K, Jiang D, Yan C, Peng C, et al. Genomic insights into evolution of parthenogenesis and triploidy in the flowerpot snake. *Sci Adv*. 2025;11:eadt6477. https://doi.org/10.1126/sciadv.adt6477

Manni M, Berkeley MR, Seppey M, Zdobnov EM. BUSCO: assessing genomic data quality and beyond. *Curr Protoc*. 2021;1:e323. https://doi.org/10.1002/cpz1.323

Marçais G, Kingsford C. A fast, lock-free approach for efficient parallel counting of occurrences of k-mers. *Bioinformatics.* 2011;27:764–770. https://doi.org/10.1093/bioinformatics/btr011

Nawrocki EP, Eddy SR. Infernal 1.1: 100-fold faster RNA homology searches. *Bioinformatics.* 2013;29:2933–2935. https://doi.org/10.1093/bioinformatics/btt509

Near TJ, Keck BP. Free from mitochondrial DNA: nuclear genes and the inference of species trees among closely related darter lineages (Teleostei: Percidae: Etheostomatinae). *Mol Phylogenet Evol*. 2013;66:868–876. https://doi.org/10.1016/j.ympev.2012.11.009

Parra G, Bradnam K, Korf I. CEGMA: a pipeline to accurately annotate core genes in eukaryotic genomes. *Bioinformatics.* 2007;23:1061–1067. https://doi.org/10.1093/bioinformatics/btm071

Peng C, Wu DD, Ren JL, Peng ZL, Ma Z, Wu W, Lv Y, Wang Z, Deng C, Jiang K, et al. Large-scale snake genome analyses provide insights into vertebrate development. *Cell.* 2023;186:2959–2976.e22. https://doi.org/10.1016/j.cell.2023.05.030

Pertea M, Pertea GM, Antonescu CM, Chang TC, Mendell JT, Salzberg SL. StringTie enables improved reconstruction of a transcriptome from RNA-seq reads. *Nat Biotechnol*. 2015;33:290–295. https://doi.org/10.1038/nbt.3122

Platt RN, Faircloth BC, Sullivan KAM, Kieran TJ, Glenn TC, Vandewege MW, Lee TE Jr, Baker RJ, Stevens RD, Ray DA. Conflicting evolutionary histories of the mitochondrial and nuclear genomes in New World Myotis bats. *Syst Biol*. 2018;67:236–249. https://doi.org/10.1093/sysbio/syx070

Price AL, Jones NC, Pevzner PA. De novo identification of repeat families in large genomes. *Bioinformatics.* 2005;21:i351–i358. https://doi.org/10.1093/bioinformatics/bti1018

Ranallo-Benavidez TR, Jaron KS, Schatz MC. GenomeScope 2.0 and Smudgeplot for reference-free profiling of polyploid genomes. *Nat Commun*. 2020;11:1432. https://doi.org/10.1038/s41467-020-14998-3

Rhie A, Walenz BP, Koren S, Phillippy AM. Merqury: reference-free quality, completeness, and phasing assessment for genome assemblies. *Genome Biol*. 2020;21:245. https://doi.org/10.1186/s13059-020-02134-9

Roberts JR, Bernstein JM, Austin CC, Hains T, Mata J, Kieras M, Pirro S, Ruane S. Whole snake genomes from eighteen families of snakes (Serpentes: Caenophidia) and their applications to systematics. *J Hered*. 2024;115:487–497. https://doi.org/10.1093/jhered/esae026

Schoch CL, Ciufo S, Domrachev M, Hotton CL, Kannan S, Khovanskaya R, Leipe D, Mcveigh R, O'Neill K, Robbertse B, et al. NCBI taxonomy: a comprehensive update on curation, resources and tools. *Database (Oxf)*. 2020;2020:baaa062. https://doi.org/10.1093/database/baaa062

She R, Chu JS, Wang K, Pei J, Chen N. GenBlastA: enabling BLAST to identify homologous gene sequences. *Genome Res*. 2009;19:143–149. https://doi.org/10.1101/gr.082081.108

Song X, Jiang Y, Zhao L, Jin L, Yan C, Liao W. Predicting the potential distribution of the Szechwan rat Snake (*Euprepiophis perlacea*) and its response to climate change in the Yingjing Area of the Giant Panda National Park. *Animals (Basel)*. 2023;13:3828. https://doi.org/10.3390/ani13243828

Tang H, Krishnakumar V, Zeng X, Xu Z, Taranto A, Lomas JS, Zhang Y, Huang Y, Wang Y, Yim WC, et al. JCVI: a versatile toolkit for comparative genomics analysis. *Imeta.* 2024;3:e211. https://doi.org/10.1002/imt2.211

Tang SY, Li CL, Li YL, Li BG. New records of the endangered Szechwan rat Snake, *Euprepiophis perlaceus* (Stejneger, 1929) (Squamata: Colubridae: Coronellini), from Shaanxi, Chia. *Amphib Rept Conser*. 2021;15:23–30.

Title PO, Singhal S, Grundler MC, Costa GC, Pyron RA, Colston TJ, Grundler MR, Prates I, Stepanova N, Jones MEH, et al. The macroevolutionary singularity of snakes. *Science.* 2024;383:918–923. https://doi.org/10.1126/science.adh2449

Uetz P, Koo MS, Aguilar R, Brings E, Catenazzi A, Chang AT, Chaitanya R, Freed P, Gross J, Hammermann M, et al. A quarter century of reptile and amphibian databases. *Herpetol Rev*. 2021;52:246–255.

Utiger U, Helfenberger N, Schätti B, Schmidt C, Ruf M, Ziswiler V. Molecular systematics and phylogeny of old and New World Ratsnakes, Elaphe Auct., and related genera (Reptilia, Squamata, Colubridae). *Russ J Herpetol*. 2002;9:105–124.

Wan R, Liu S, Xu Q, Yue B, Zhang X. The complete mitochondrial genome of the Elaphe perlacea (Squamata: Colubridae). *Mitochondrial DNA A DNA Mapp Seq Anal*. 2016;27:12–13. https://doi.org/10.3109/19401736.2013.867436

Wingett S, Ewels P, Furlan-Magaril M, Nagano T, Schoenfelder S, Fraser P, Andrews S. HiCUP: pipeline for mapping and processing Hi-C data. *F1000Research*. 2015;4:1310. https://doi.org/10.12688/f1000research.7334.1

Xu Z, Wang H. LTR_FINDER: an efficient tool for the prediction of full-length LTR retrotransposons. *Nucleic Acids Res*. 2007;35:W265–W268. https://doi.org/10.1093/nar/gkm286

Yan C, Song MH, Jiang D, Ren JL, Lv Y, Chang J, Huang S, Zaher H, Li JT. Genomic evidence reveals intraspecific divergence of the hot-spring snake (Thermophis baileyi), an endangered reptile endemic to the Qinghai-Tibet plateau. *Mol Ecol*. 2023;32:1335–1350. https://doi.org/10.1111/mec.16687

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*. 2019;5:833–845. https://doi.org/10.1038/s41477-019-0487-8

Journal of Heredity, 2026, 117, 834–842
https://doi.org/10.1093/jhered/esag029
Genome Resources