*Journal of Heredity*, 2026, **117**, 141–150
https://doi.org/10.1093/jhered/esaf055
Advance access publication 16 August 2025
**Genome Resources**

American Genetic Association
OXFORD

Genome Resources

# A chromosome-level genome assembly and annotation for the beauty snake *Elaphe taeniura*

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

[1]Institute for Advanced Study, Chengdu University, Chengdu, China, [2]Faculty of Agriculture, Forestry and Food Engineering, Yibin University, Yibin, China, [3]School of Ecology and Environment, Xizang University, Lhasa, China

*Corresponding authors: Faculty of Agriculture, Forestry and Food Engineering, Yibin University, No. 8 Jiushen Road, East Section of Wuliangye Avenue, Yibin City, Sichuan Province 644000, China. Email: ybguop@163.com; Institute for Advanced Study, Chengdu University, No. 2025, Chengluo Avenue, Chengdu City, Sichuan Province 610106, China. Email: dulianming@cdu.edu.cn

## Abstract

The genus *Elaphe* Fitzinger, a large species clade within Colubridae, comprises 18 non-venomous snake species. Among them, *Elaphe taeniura* serves as a representative species due to its wide distribution and strong adaptability. However, genomic studies on this group remain limited. Here, we present a chromosome-level, high-quality reference genome for this species. The genome is assembled by integrating high-accuracy PacBio HiFi long-read sequencing and Hi-C chromatin conformation capture technologies and comprehensively annotated with the aid of transcriptomic data. The genome of *E. taeniura* spans 1.62 Gb, with a scaffold N50 of 206.4 Mb and the longest scaffold reaching 344.4 Mb. The genome exhibits a BUSCO completeness score of 98.2% and contains 22,246 protein-coding genes. Comparative analysis between the assembled genome and three other colubrid species revealed high synteny. In addition, several chromosomal fusion and fission events were observed. This reference genome provides a valuable resource for studying the taxonomy, conservation, and evolutionary history of the widely distributed *Elaphe* species.

*Key words:* *Elaphe taeniura*, genome assembly, Colubridae, snake, comparative genomics

## Introduction

Snakes are distributed across all continents except the polar regions and have undergone extensive adaptive evolution, resulting in over 30 families and more than 4,000 species (Zug et al. 2001; Feldman et al. 2015). In the long evolutionary process, various branches have occupied unique ecological positions through significant adaptive differentiation: the Elapidae and Viperidae have independently evolved efficient venom delivery systems (Kerkkamp et al. 2017); the Scolecophidia have evolved regressive vision and specialized body scales adapted for burrowing (Avila et al. 2006); Hydrophiinae have evolved salt-excreting glands and paddle-like tails among other marine adaptations (Shine 2005). This kind of cross-ecological phenotypic differentiation makes snakes a typical species for parsing the adaptive radiation of vertebrates (Pyron et al. 2013). In recent years, some high-quality chromosome-level genomes of snakes have been published, providing valuable insights into the evolution of snakes and even vertebrates (Peng et al. 2023). Despite their wide distribution and extensive species differentiation, snakes still lack high-quality reference genomes, which hinders the study of conservation and molecular biology of this group. High-quality reference genomes are crucial for studying genetic variation in a species, especially when combining re-sequencing data with high-quality annotation, which can accurately characterize structural variation (Ejigu and Jung 2020).

*Elaphe taeniura* (Cope, 1861), commonly known as the beauty snake, belongs to the genus *Elaphe* within the Colubridae family and is widely distributed across Southeast and East Asia. This nonvenomous snakes typically prey on rodents and birds and their lengths can reach up to 2 meters (Asato et al. 2022). They inhabit a variety of environments, including plains, hills, and mountainous regions, at elevations ranging from sea level to 3,000 m. Due to dramatic population decline, this species has been classified as Vulnerable (VU) on the IUCN Red List in 2021 (Li et al. 2021). *E. taeniura* once belonged to the genus *Orthriophis* and was named *O. taeniurus* (Wang et al. 2020). With the advance of molecular taxonomy, the intricate phylogenetic relationships among numerous species have been rigorously reexamined, resulting

**Fig. 1.** Photograph of *Elaphe taeniura* taken by Songwen Tan in Yibin, September 2023.

Received on 16 June 2025; revised on 14 August 2025; accepted on 14 August 2025


**Fig. 2.** Circos plot showing the distribution of genomic features. a) Chromosomes, with numbers indicating their lengths in Mb; b) repetitive sequences; c) GC ratio; d) gene density; the innermost track is synteny among chromosomes.

in the reclassification of New World and most Old World taxa into newly established or resurrected genera (Helfenberger 2001; Lenk et al. 2001; Chen et al. 2017). Therefore, *Orthrioph* was synonymized with the genus *Elaphe*, and *O. taeniurus* has accordingly been named as *E. taeniura*. According to the most recent entries in the Reptile Database (http://www.reptile-database.org/), the genus *Elaphe* consists of 18 species, *E. taeniura* includes nine recognized subspecies, most of which are distributed in Asia (Utiger et al. 2002). However, this widely distributed and highly adaptable species lacks a reference genome, posing challenges for genomic research on this group.

Here, we present a chromosome-level and high-quality genome assembly of *E. taeniura*, derived from *Elaphe taeniura* in mainland China. We employed PacBio Revio System for long-read sequencing and integrated Hi-C data to achieve a chromosome-level assembly. Genome annotation was conducted using RNA-seq data from multiple tissues, published genomes of closely related species, and de novo prediction methods. We assessed the quality of the genome assembly using BUSCO. Overall, this study ultimately produced a high-quality reference genome for the genus *Elaphe*.

## Methods

### Biological materials

A single adult male *E. taeniura* (Fig. 1) was captured in Yibin City, Sichuan Province, in September 2023. After several months of captivity, we extracted muscle, blood, heart, pancreas, spleen, kidney, liver, Duvernoy's gland, and lung samples, which were rapidly frozen at −80 °C. Muscle tissue was used for PacBio HiFi long-read sequencing and Hi-C sequencing, while the remaining eight tissues were used for RNA-seq.

### HiFi library preparation and sequencing

High-quality genomic DNA (gDNA) was extracted using the QIAGEN Genomic Kit. The quantity and quality of the extracted DNA were assessed using a NanoDrop 2000 spectrophotometer (Thermo Fisher Scientific) and a Qubit fluorometer (Invitrogen). DNA integrity was evaluated by 1% agarose gel electrophoresis. After the detection, DNA purification was performed using AMPure PB beads (Pacbio Cat #100-265-900). The library of PacBio SMRTbell was constructed using SMRTbell Express Template Prep Kit 3.0. The size and concentration of library fragments were detected with Agilent 2100 Bioanalyzer (Agilent Technologies, USA). Qualified libraries were evenly loaded on SMRT Cell and sequenced for 30 h using Sequel II/IIe system (Pacific Biosciences, CA, USA).

### Hi-C library preparation and sequencing

Muscle tissue was fixed at room temperature under vacuum in MS buffer (10 mM potassium phosphate, pH 7.0; 50 mM sodium chloride; 0.1 M sucrose) containing 1% formaldehyde

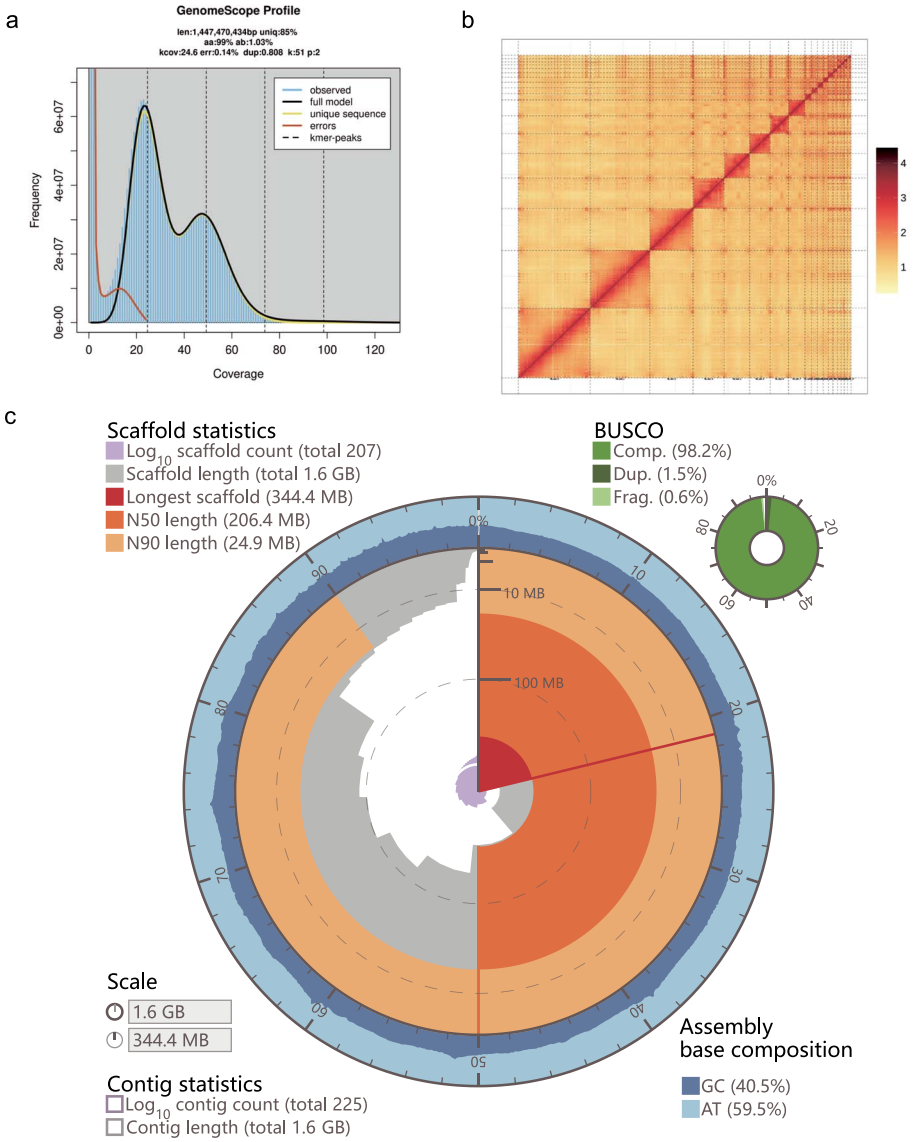


**Fig. 3.** Visual overview of genome assembly metrics. a) *K*-mer ($K = 51$) spectrum generated using GenomeScope. The bimodal distribution observed is characteristic of a diploid genome. b) The Hi-C contact map is based on a genome assembly partitioned into 500 kb bins. The number of Hi-C read pairs connecting any two bins was used to represent interaction intensity. Eighteen chromosome groups can be clearly distinguished. Interaction signals are stronger along the diagonal, reflecting high contact frequency between adjacent genomic regions. c) The BlobToolKit Snail plot provides a graphical summary of the *E. taeniura* assembly quality metrics presented in Table 2. The plot circle represents the full size of the assembly. From the inside out, the central plot covers length-related metrics in percentage of the assembly. The line indicating the longest scaffold is highlighted within the central plot. All other scaffolds are arranged clockwise in descending size and are shown in contrast, starting from the outer edge of the central plot.

to cross-link DNA and proteins. The crosslinked DNA was subsequently digested with the Hind III restriction enzyme. Sticky ends were repaired, and biotin was simultaneously added to label the oligonucleotide ends. T4 DNA ligase was added to the end-repaired DNA, followed by incubation at 16 °C for 6 h to facilitate ligation. Proteinase K was added to uncross-link the DNA, followed by overnight incubation at 65 °C. The unligated ends were then removed, and the purified DNA was fragmented to a size of 300 to 500 bp, with subsequent end repair. The labeled DNA fragments of biotin were finally isolated on Dynabeads M-280 Streptavidin (Life Technologies). The obtained DNA fragments were used to construct a second-generation sequencing library, followed by PE150 sequencing on the Illumina NovaSeq 6000 platform.

## RNA sequencing

Total RNA was extracted from multiple tissues of the same individual. mRNA was purified using poly-T oligo-attached magnetic beads. After cDNA synthesis and purification, library quality control was performed, with the 5′ end of each library undergoing phosphorylation and circularization. Circular amplification was then carried out to generate DNA nanoballs, which were loaded onto a flow cell for sequencing on the DNBSEQ-T7 platform.

## Predicting genome size and heterozygosity

Prior to genome assembly, the genome size and heterozygosity of *E. taeniura* were estimated using Jellyfish

**Table 1.** Assembly pipeline and software used.

| Procedures | Software and options | Version |
|---|---|---|
| **Initial assembly** | | |
| Quality control of the raw polymerase reads | PacBio SMRT-Analysis | |
| *K*-mer counting | Jellyfish ($k = 51$) | 2.3.0 |
| | Genomescope | 2.0 |
| ccs reads generate | SMRTLink (–min-passes = 3 –min-rq = 0.99) | 9.0 |
| De novo assembly | Hifiasm | 0.19.7 |
| **HI-C** | | |
| quality controlled | fastp (-q 20) | 0.21.0 |
| SAM/BAM processing | samtools | 1.15.1 |
| Map to contigs | hicup | 0.8.0 |
| contigs clustering, sorting and orienting | ALLHiC | 0.9.8 |
| fine-tune clustering results | JuiceBox | 1.11.08 |
| **Genome quality assessment** | | |
| assembly quality assess | BUSCO (-m geno -l vertebrata_odb10) | 5.5.0 |
| **Repeat annotation** | | |
| Repetitive element identification | RepeatMasker | 4.1.2 |
| Predict LTR | LTR_FINDER | 1.07 |
| | PILER | |
| | RepeatScout | 1.0.5 |
| | RepeatModeler | 2.0.5 |
| | TRF | 4.09.1 |
| **Structural annotation** | | |
| Homology-based gene prediction | GenblastA | 1.0.4 |
| Transcriptome-based gene prediction | Tophat | 2.1.1 |
| | cufflinks | |
| De novo gene prediction | PASA | 2.5.1 |
| | Augustus | 3.3.1 |
| Integration of gene prediction | EVidenceModeler | 1.1.1 |
| Update of gene annotation | PASA | 2.5.1 |

**Table 2.** Sequencing and assembly statistics.

| | | |
|---|---|---|
| BioProjects | NGDC BioProject | PRJCA037373 |
| | NGDC Biosample | SAMC5075381 |
| | | SAMC5075382 |
| | | SAMC5075383 |
| | | SAMC5075384 |
| | | SAMC5075385 |
| | | SAMC5075386 |
| | | SAMC5075387 |
| | | SAMC5075388 |
| | | SAMC5091416 |
| | | SAMC5091417 |
| | NGDC GSA | CRA025507 |
| | | CRA025653 |
| Genome Sequence | PacBio HiFi reads | PacBio_SMRT (Revio) cell: 6.6 million reads, 113.95Gbases |
| | Hi-C reads | DNBSEQ-T7: 2.06 billion reads, 309.40 Gbases |
| | RNA-seq reads | ILLUMINA (Illumina NovaSeq 6000):190.57 Gbases from eight organizations |
| Genome Assembly Quality Metrics | Number of contigs | 225 |
| | Contig N50 | 202,669,265 |
| | Contig N50 number | 3 |
| | Longest contigs | 344,398,495 |
| | Number of Scaffolds | 207 |
| | Scaffold N50 | 206,420,752 |
| | Scaffold N50 number | 3 |
| | Longest Scaffolds | 344,398,495 |
| | Size of final assembly | 1,624,145,665 |
| | BUSCO completeness (tetrapoda_odb10) | C:98.2% [S:96.7%, D:1.5%], F:0.6%, M:1.2%, n:3354, E:3.7% |

**Table 3.** Statistical outcomes regarding repetitive sequences in the *E. taeniura* genome.

| | Number of elements | Length occupied | Percentage of sequence (%) |
|---|---|---|---|
| **Retroelements** | 1,697,456 | 439,096,101 | 27.04 |
| **SINEs** | 399,563 | 76,710,039 | 4.72 |
| **LINEs** | 874,204 | 240,920,889 | 14.83 |
| L2/CR1/Rex | 617,430 | 158,486,427 | 9.76 |
| R2/R4/NeSL | 30,844 | 8,008,372 | 0.49 |
| RTE/Bov-B | 89,297 | 24,099,611 | 1.48 |
| L1/CIN4 | 59,224 | 38,024,486 | 2.34 |
| **LTR elements** | 423,689 | 121,465,173 | 7.48 |
| BEL/Pao | 227 | 114,525 | 0.01 |
| Ty1/Copia | 95,590 | 25,086,546 | 1.54 |
| Gypsy/DIRS1 | 168,229 | 59,529,893 | 3.67 |
| Retroviral | 4,954 | 3,409,141 | 0.21 |
| **DNA transposons** | 39,730 | 11,501,646 | 0.71 |
| hobo-Activator | 7,601 | 1,735,049 | 0.11 |
| Tc1-IS630-Pogo | 30,643 | 9,225,237 | 0.57 |
| PiggyBac | 238 | 102,634 | 0.01 |
| **Rolling-circles** | 1,461 | 357,801 | 0.02 |
| **Unclassified** | 1,376,183 | 236,186,661 | 14.54 |
| **Total interspersed repeats** | | 686,784,408 | 42.29 |
| **Small RNA** | 34,986 | 4,062,456 | 0.25 |
| **Satellites** | 14 | 3,655 | |
| **Simple repeats** | 555,220 | 33,743,591 | 2.08 |
| **Low complexity** | 67,423 | 7,543,176 | 0.46 |

(Marçais and Kingsford 2011) and GenomeScope2 (Ranallo et al. 2020). Jellyfish was employed to generate *k*-mer ($k = 51$) (Peng et al. 2023) frequency distribution tables from clean short reads, which were then analyzed by GenomeScope2 to derive genome prediction information.

## De novo genome assembly

Raw polymerase reads were quality controlled using the PacBio SMRT Analysis software suite (https://www.pacb.com). HiFi reads were generated using SMRT Link v9.0 and subsequently assembled into contigs using Hifiasm (Cheng et al. 2021). HI-C data were quality controlled using fastp (Chen et al. 2018). The clean reads were aligned to the contigs using HiCUP software (Wingett et al. 2015). Contigs were then clustered, ordered, and oriented using ALLHiC software (Zhang et al. 2019). Finally, the orientation results were fine-tuned with JuiceBox (Durand et al. 2016) to generate a chromosome-level genome. To assess the assembly quality, BUSCO (Simão et al. 2015) was used with the vertebrata_odb10 database for evaluation. Snail plot (Challis et al. 2020) was generated using assembly-stats to visualize the assembly quality. To assess assembly accuracy, short reads were mapped to the assembled genome using BWA (Li and Durbin 2009).

## Genome annotation

The known repeat sequences were identified using RepeatMasker (Chen 2004). The de novo prediction was performed by constructing a repeat library with tools LTR_FINDER (Xu and Wang 2007), PILER (Edgar and Myers 2005), RepeatScout (Price et al. 2005), RepeatModeler (Flynn et al. 2020), and TRF (Benson 1999). Then, RepeatMasker was adopted to annotate repeat sequences based on the constructed repeat library.

In genomic structural annotation of *E. taeniura*, three methods were used to predict protein-coding genes, including homology prediction, transcriptome prediction, and de novo prediction. The results were then integrated to obtain the final comprehensive annotation. In homologous prediction, the genomes of five related species (*Pantherophis guttatus*, *Ahaetulla prasina*, *Thamnophis elegans*, *T. sirtalis*, *Eryx carinatus*) were compared using the GenBlastA tool (She et al. 2009). Nonredundant gene structures predicted by PASA (Haas et al. 2008) were used to train gene model parameters, and then Augustus (Stanke et al. 2008) was used for de novo gene prediction. Transcriptome prediction was conducted by mapping clean RNA-seq reads to the assembly sequences using Tophat (Trapnell et al. 2009), followed by transcript assembly with Cufflinks (Trapnell et al. 2012). Finally, we used EVidenceModeler (Haas et al. 2008) to merge the prediction results. Genomic structural information was visualized with Circos (Fig. 2) (Krzywinski et al. 2009).

Functional annotation of protein-coding genes was performed by aligning the predicted protein sequences to the NR (Non-Redundant protein Database) and Swiss-Prot databases (Bairoch and Apweiler 1999) using Diamond (Buchfink et al. 2021). eggNOGmapper (Cantalapiedra et al. 2021) was used to predict gene function by aligning protein sequences against the eggNOG database (Huerta-Cepas et al. 2019). InterProscan (Hunter et al. 2009) was used to classify proteins and annotate functional domains and functional sites. Both InterProscan and eggNOGmapper performed GO (Gene Ontology) (Ashburner et al. 2000) annotation of the sequences. In addition, eggNOGmapper also produces the KEGG (Kyoto Encyclopedia of Genes and Genomes) (Kanehisa and Goto 2000) pathway annotation.

For noncoding RNAs, tRNAs were identified using tRNAscan-SE (Chan et al. 2021), rRNAs, miRNAs, and snRNAs were predicted using the Infernal (Nawrocki and

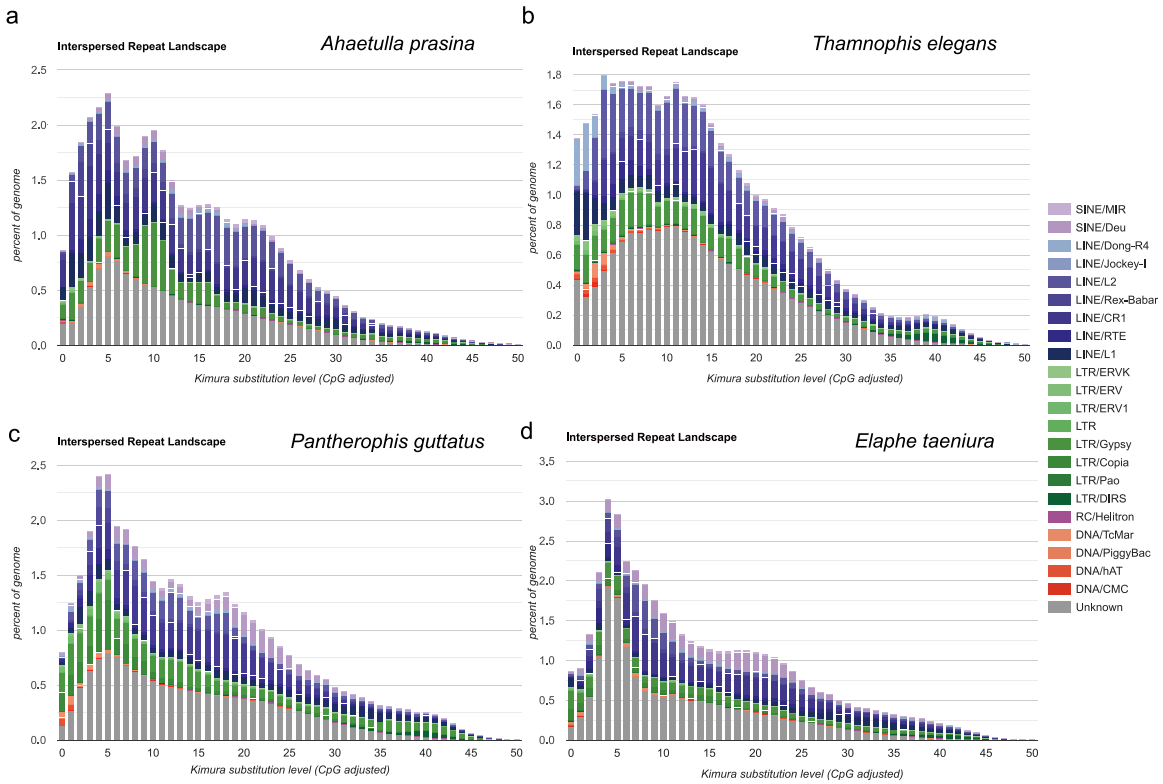


**Fig. 4.** Interspersed repeat sequence landscape. The x-axis represents the mutation rate of transposon sequences relative to their consensus sequences; higher values indicate earlier insertion times. a) *Ahaetulla prasina*. b) *Thamnophis elegans*. c) *Pantherophis guttatus*. d) *Elaphe taeniura*.

Eddy 2013) software with Rfam database (Ontiveros-Palacios et al. 2025).

### Genome synteny and gene family analysis

Peptide sequences of *E. taeniura* were compared with the chromosome-level genomes of other Colubridae species, including *A. prasina* (NCBI accession GCF_028640845.1), *T. elegans* (NCBI accession GCA_009769535.1), and *P. guttatus* (NGDC accession GWHBOZR00000000). Collinearity analysis was conducted using MCScanX (Wang et al. 2012). Protein family identification was carried out using OrthoVenn3 (Sun et al. 2023), and the species divergence time was inferred using the Mcmctree (Reis and Yang 2011). Two additional species from different families, *Naja naja* (NCBI accession GCA_009733165.1) and *Argyrophis diardii* (NGDC accession GWHBWDT00000000), were also included in the analysis. The species tree was constructed using OrthoFinder (Emms and Kelly 2019).

## Results

### Genome assembly and annotation

For genome survey analysis, a total of 76.24 Gb high-quality data were obtained from Illumina sequencing. *K*-mer distribution analysis estimated the genome size of 1,447.47 Mb, heterozygous ratio of 1.03% (Fig. 3a). The software and parameters used in the assembly are listed in Table 1. PacBio HiFi sequencing generated 113.95 Gb of data, with approximately 6.6 million reads and N50 length of 17 244 bp. The Hi-C sequencing library yielded 309.49 Gb clean data with a Q30 rate of 96.64% and an average GC content of 41.38%. RAN-seq produced 190.57 Gb clean data from eight tissues (Table 2).

The Hi-C data was aligned to contigs to facilitate chromosome-level scaffolding, successfully anchoring sequences to 18 chromosomes (Fig. 3b), with a chromosome anchoring rate of 98.27%. The final assembly of *E. taeniura* consisted of 207 scaffolds, with a total genome size of approximately 1.62 Gb, with scaffold N50 of 206.4 Mb, N90 of 24.9 Mb, and the longest scaffold of 344.4 Mb (Fig. 3c). BUSCO evaluation identified 98.2% complete single-copy genes, indicating high assembly completeness (Table 2).

For repeat annotation, we identified 6,574,206 repeat elements with a total length of 732.5 Mb accounting for 45.1% of the assembled genome. Among these, long interspersed nuclear elements (LINEs, 14.83%) and long terminal repeat sequences (LTRs, 7.48%) have the largest number of family members (Table 3). In the interspersed repeat landscape, *E. taeniura* exhibits a substitution level peak at 3–5, indicating recent transposable element activity (Fig. 4a). *P. guttatus* and *A. prasina* show similar peak patterns to *E. taeniura*, but with slightly more gradual activity (Fig. 4b and c). *T. elegans* displays a broader peak between 5 and 15, dominated by LINEs (Fig. 4d). The results indicate that *E. taeniura* has recently experienced a transposon expansion event, with a more pronounced expansion trend compared with other species in the same family.

A total of 22,246 protein-coding genes were predicted (Table 4), functional annotation against major protein databases resulted in successful annotation of 85.88% of these genes (Fig. 5). For noncoding RNA annotations, we have detected 255 miRNAs, 552 tRNAs, 451 rRNAs, and 525 snRNAs from the assembled genome (Table 5).

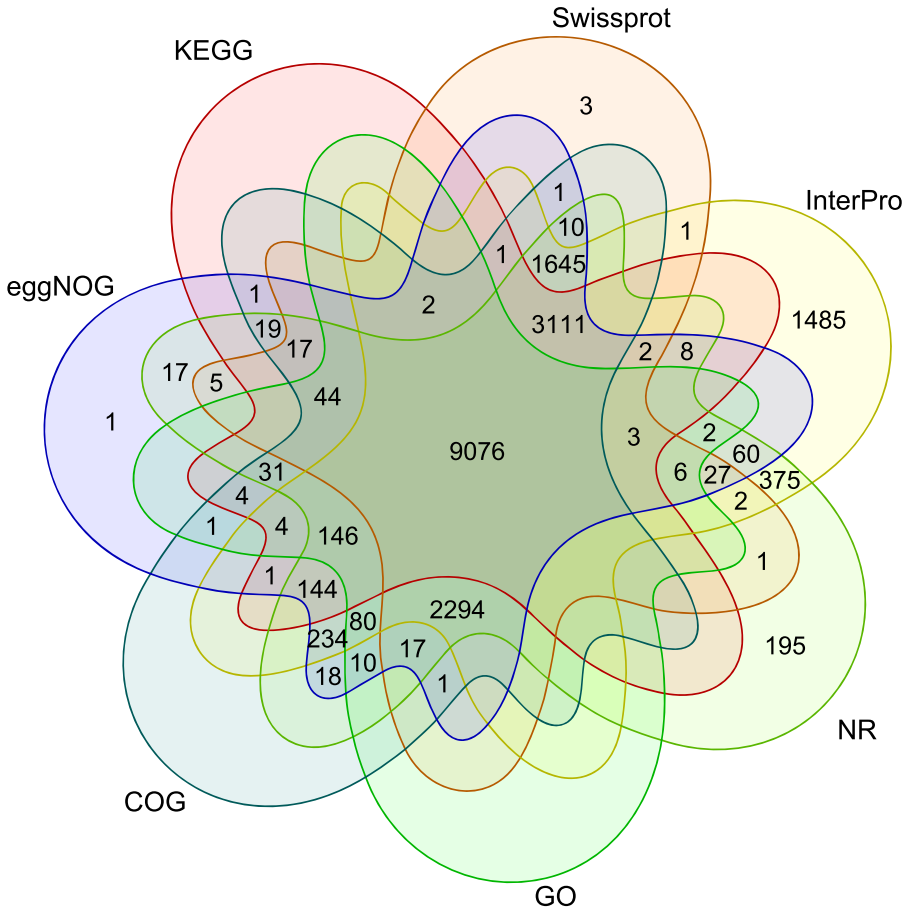


**Fig. 5.** Functional annotation of protein-coding genes across multiple databases.

**Table 4.** Summary of the functionally annotated protein-coding genes.

| | Database | Number | Percent (%) |
|---|---|---|---|
| Annotation | Swissprot | 16,269 | 73.13 |
| | NR | 17,599 | 79.11 |
| | KEGG | 12,614 | 56.70 |
| | GO | 11,721 | 52.69 |
| | eggNOG | 17,043 | 76.61 |
| | COG | 16,912 | 76.02 |
| | Interpro | 18,709 | 84.10 |
| Total | Annotated | 19,105 | 85.88 |
| | All | 22,246 | 100 |

**Table 5.** Statistical results of noncoding RNA annotation of the *E. taeniura* genome.

| Type | Number | Total length (bp) |
|---|---|---|
| tRNAs | 552 | 57,581 |
| miRNAs | 255 | 20,546 |
| rRNAs | 451 | 64,6790 |
| snRNAs | 525 | 66,044 |
| Total | 1,783 | 79,0961 |

## Orthologous cluster identification

Among the 22,246 protein-coding sequences of *E. taeniura*, 15,064 were assigned to orthologous gene families. A total of 10,466 orthologous clusters were shared among all six species. Notably, *E. taeniura* exhibited the highest number of singleton genes (5,986) compared with the other five species (Fig. 6).

## Chromosomal synteny analysis

Among species of the same genus, *E. taeniura* shows strong chromosomal collinearity with other species, although some chromosomal rearrangements are also present. The syntenic correspondence between chromosome 4 of *E. taeniura* and chromosome 12, 16 of *P. guttatus* suggests the occurrence of chromosomal fusion. In contrast, *T. elegans* exhibited more extensive chromosomal rearrangements. The longer chromosomes (>99 Mb) in *E. taeniura*, including chr1, chr2, chr3, chr5, and chr6, appear to have formed through the fusion of 2 to 3 chromosomes from *T. elegans*, while shorter chromosomes (<30 Mb) (chr9-16) resulted from chromosomal fission events (Fig. 7a). To further characterize these structural changes, synteny dot plots were generated and confirmed the identified rearrangement patterns (Fig. 7b, c, and d). These extensive chromosomal rearrangements may have contributed to the genetic diversification within the family Colubridae.

## Discussion

In the Genebank database (accessed on 26 April 2025), 26 snake genome assemblies at the chromosome level are available, with genome sizes ranging from 1.3 Gb in *Crotalus viridis viridis* (NCBI accession GCA_003400415.2) (Pasquesi et al. 2018) to 2.1 Gb in *Hydrophis major* (NCBI accession

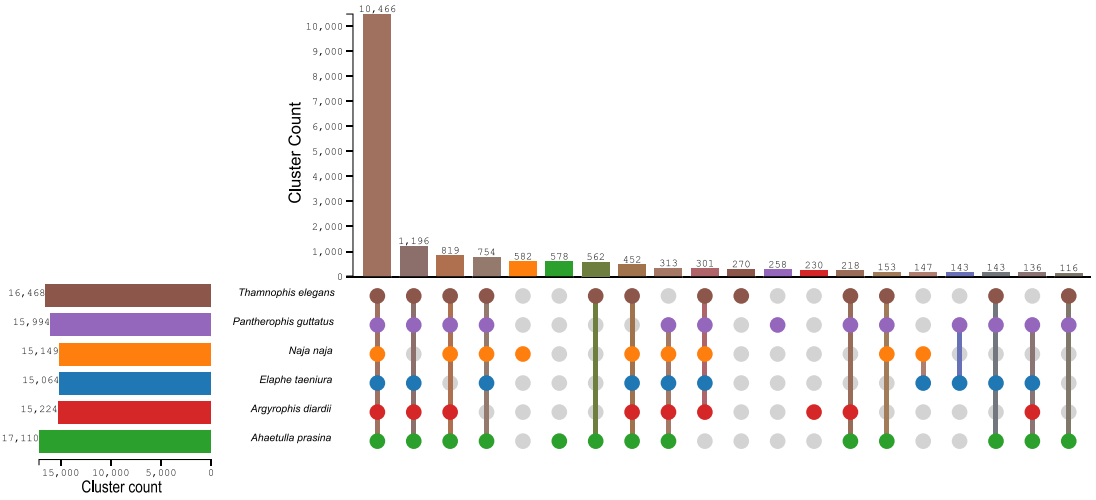


**Fig. 6.** Orthologous cluster identification. Homology clustering of protein sequences across selected species. Bars represent the number of gene families per species; dots indicate the presence of each gene family.

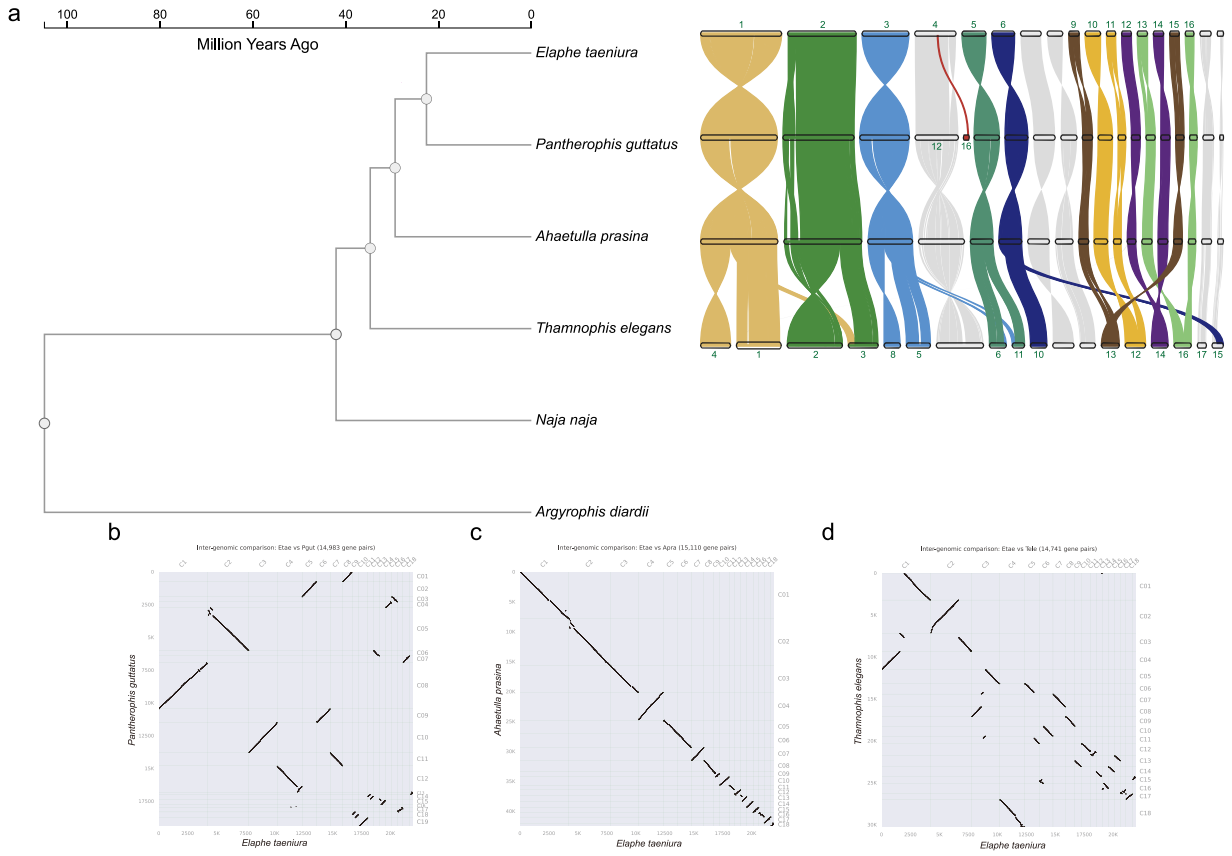


**Fig. 7.** Species tree and chromosomal synteny. Divergence times (in millions of years) are shown above the species tree. a) Pairwise alignment of four Colubridae genomes (*E. taeniura, P. guttatus, A. prasina, T. elegans*). b) Collinearity plot for *E. taeniura* and *P. gluttatus*. c) Collinearity plot for *E. taeniura* and *A. prasina*. d) Collinearity plot for *E. taeniura* and *T. elegans*.

GCA_033807585.1) (Ludington et al. 2023) and scaffold N50 values varying from 100.9 Mb in *T. elegans* to 268.3 Mb in *H. major*. The *E. taeniura* genome spans 1.62 Gb with a scaffold N50 of 206.4 Mb, placing it well within the range of existing chromosome-level snake genomes.

This study presents a chromosome-level, high-quality genome assembly for the genus *Elaphe*. Although a genome for the *Elaphe carinata* was previously published (Fan et al. 2023), it was not assembled at the chromosome level. Its genome size of 1.56 Gb is comparable to that of *E. taeniura* 1.62 Gb. Compared with reference genomes of species in the family Colubridae in GeneBank, which range from 1.7 to 1.9 Gb, the genomes of the *Elaphe* genus appear to be relatively smaller.

*Elaphe taeniura* comprises nine subspecies, and its high level of differentiation makes it a promising model for investigating adaptive evolution within the genus *Elaphe*. Moreover, the availability of a high-quality reference genome facilitates

further studies on adaptive morphology, extreme sex determination, and the inheritance of traits relevant to conservation (Brandies et al. 2019; Paez et al. 2022). The genome presented here provides a valuable resource for future research on the evolutionary history and conservation of this genus.

## Author contributions

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

## Funding

This work was supported by the National Natural Science Foundation of China (32200525, 32370486) and the second Tibetan Plateau Scientific Expedition and Research Program (2024QZKK0200).

## Data availability

Data generated for this study is available under NGDC (National Genomics Data Center) BioProject PRJCA037373. The dataset includes the final assembled genome, PacBio HiFi sequencing data, Hi-C sequencing data, and RNA-seq data. No custom scripts were employed in this study. All analyses were conducted using standard bioinformatics tools following their official manuals and protocols.

## References

Asato H, Sueyoshi K, Nakagawa M, Sugawara R, Kadota Y, Kawauchi N, Kobayashi S, Izawa M, Toda M. Food habits of the Taiwan beauty snake, *Elaphe taeniura friesi*, as an introduced species on Okinawajima Island. *Curr Herpetol*. 2022;**41**:205–214

Ashburner M, Ball CA, Blake JA, Botstein D, Butler H, Cherry JM, Davis AP, Dolinski K, Dwight SS, Eppig JT, et al. Gene ontology: tool for the unification of biology. *Nat Genet*. 2000;**25**:25–29. https://doi.org/10.1038/75556

Avila RW, Ferreira VL, Souza VB. Biology of the blindsnake *Typhlops brongersmianus* (Typhlopidae) in a semideciduous forest from Central Brazil. *Herpetol J*. 2006;**16**:403–405

Bairoch A, Apweiler R. The SWISS-PROT protein sequence data bank and its supplement TrEMBL in 1999. *Nucleic Acids Res*. 1999;**27**:49–54. https://doi.org/10.1093/nar/27.1.49

Benson G. Tandem repeats finder: a program to analyze DNA sequences. *Nucleic Acids Res*. 1999;**27**:573–580. https://doi.org/10.1093/nar/27.2.573

Brandies P, Peel E, Hogg CJ, Belov K. The value of reference genomes in the conservation of threatened species. *Genes*. 2019;**10**:846. https://doi.org/10.3390/genes10110846

Buchfink B, Reuter K, Drost H-G. 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

Cantalapiedra CP, Hernández-Plaza A, Letunic I, Bork P, Huerta–Cepas J. eggNOG-mapper v2: functional annotation, orthology assignments, and domain prediction at the metagenomic scale. *Mol Biol Evol*. 2021;**38**:5825–5829. https://doi.org/10.1093/molbev/msab293

Challis R, Richards E, Rajan J, Cochrane G, Blaxter M. BlobToolKit–interactive quality assessment of genome assemblies. G3: genes, genomes. *Genetics*. 2020;**10**:1361–1374

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 N. Using repeat masker to identify repetitive elements in genomic sequences. *Curr Protoc Bioinformatics*. 2004;**5**:11–14.10.14. https://doi.org/10.1002/0471250953.bi0410s05

Chen X, Lemmon AR, Lemmon EM, Pyron RA, Burbrink FT. Using phylogenomics to understand the link between biogeographic origins and regional diversification in ratsnakes. *Mol Phylogenet Evol*. 2017;**111**:206–218. https://doi.org/10.1016/j.ympev.2017.03.017

Chen S, Zhou Y, Chen Y, Gu J. fastp: an ultra-fast all-in-one FASTQ preprocessor. *Bioinformatics*. 2018;**34**:i884–i890. https://doi.org/10.1093/bioinformatics/bty560

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

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. https://doi.org/10.1093/bioinformatics/bti1003

Ejigu GF, Jung J. Review on the computational genome annotation of sequences obtained by next-generation sequencing. *Biology*. 2020;**9**:295. https://doi.org/10.3390/biology9090295

Emms DM, Kelly S. OrthoFinder: phylogenetic orthology inference for comparative genomics. *Genome Biol*. 2019;**20**:1–14. https://doi.org/10.1186/s13059-019-1832-y

Fan J, Huang R, Yang D, Gong Y, Cui Z, Wang X, Su Z, Yu J, Zhang Y, Zhang T, et al. Genome assembly and annotation of the king ratsnake, *Elaphe carinata*. *GigaByte*. 2023;**2023**:1–11. https://doi.org/10.46471/gigabyte.101

Feldman A, Bauer AM, Castro-Herrera F, Chirio L, Das I, Doan TM, Maza E, Meirte D, de Campos NC, Nagy ZT. The geography of snake reproductive mode: a global analysis of the evolution of snake viviparity. *Glob Ecol Biogeogr*. 2015;**24**:1433–1442. https://doi.org/10.1111/geb.12374

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*. 2020;**117**:9451–9457. https://doi.org/10.1073/pnas.1921046117

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**:1–22. https://doi.org/10.1186/gb-2008-9-1-r7

Helfenberger N. Phylogenetic relationships of Old World ratsnakes based on visceral organ topography, osteology, and allozyme variation. *Russ J Herpetol*. 2001;**8**:1–62

Huerta-Cepas J, Szklarczyk D, Heller D, Hernández-Plaza A, Forslund SK, Cook H, Mende DR, Letunic I, Rattei T, Jensen LJ, et al. eggNOG 5.0: a hierarchical, functionally and phylogenetically annotated orthology resource based on 5090 organisms and 2502 viruses. *Nucleic Acids Res*. 2019;**47**:D309–D314. https://doi.org/10.1093/nar/gky1085

Hunter S, Apweiler R, Attwood TK, Bairoch A, Bateman A, Binns D, Bork P, Das U, Daugherty L, Duquenne L. InterPro: the integrative protein signature database. *Nucleic Acids Res*. 2009;**37**:D211–D215. https://doi.org/10.1093/nar/gkn785

Kanehisa M, Goto S. KEGG: Kyoto encyclopedia of genes and genomes. *Nucleic Acids Res*. 2000;**28**:27–30. https://doi.org/10.1093/nar/28.1.27

Kerkkamp HM, Casewell NR, Vonk FJ. Evolution of the snake venom delivery system. In: Gopalakrishnakone, P., Malhotra, A, eds. *Evolution of venomous animals and their toxins*. Springer; 2017. p. 1–11. https://doi.org/10.1007/978-94-007-6727-0_11-1

Krzywinski M, Schein J, Birol I, Connors J, Gascoyne R, Horsman D, Jones SJ, Marra MA. Circos: an information aesthetic for comparative genomics. *Genome Res*. 2009;**19**:1639–1645. https://doi.org/10.1101/gr.092759.109

Lenk P, Joger U, Wink M. Phylogenetic relationships among European ratsnakes of the genus Elaphe Fitzinger based on mitochondrial DNA sequence comparisons. *Amphibia-Reptilia*. 2001;**22**:329–339. https://doi.org/10.1163/156853801317050124

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

Li P, Zhou Z, Ghosh A. *Elaphe taeniura. The IUCN red list of threatened species*. 2021: e.T192204A2055202; https://doi.org/10.2305/IUCN.UK.2021-3.RLTS.T192204A2055202.en.

Ludington AJ, Hammond JM, Breen J, Deveson IW, Sanders KL. New chromosome-scale genomes provide insights into marine adaptations of sea snakes (Hydrophis: Elapidae). *BMC Biol*. 2023;**21**:284. https://doi.org/10.1186/s12915-023-01772-2

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

Ontiveros-Palacios N, Cooke E, Nawrocki EP, Triebel S, Marz M, Rivas E, Griffiths-Jones S, Petrov AI, Bateman A, Sweeney B. Rfam 15: RNA families database in 2025. *Nucleic Acids Res*. 2025;**53**:D258–D267. https://doi.org/10.1093/nar/gkae1023

Paez S, Kraus RH, Shapiro B, MTP G, Jarvis ED, Group VGPC, Al-Ajli FO, Ceballos G, Crawford AJ, Fedrigo O. Reference genomes for conservation. *Science*. 2022;**377**:364–366. https://doi.org/10.1126/science.abm8127

Pasquesi GI, Adams RH, Card DC, Schield DR, Corbin AB, Perry BW, Reyes-Velasco J, Ruggiero RP, Vandewege MW, Shortt JA, et al. Squamate reptiles challenge paradigms of genomic repeat element evolution set by birds and mammals. *Nat Commun*. 2018;**9**:2774. https://doi.org/10.1038/s41467-018-05279-1

Peng C, Wu D-D, Ren J-L, Peng Z-L, 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 e2922. https://doi.org/10.1016/j.cell.2023.05.030

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

Pyron RA, Burbrink FT, Wiens JJ. A phylogeny and revised classification of Squamata, including 4161 species of lizards and snakes. *BMC Evol Biol*. 2013;**13**:1–54

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

Reis M, Yang Z. Approximate likelihood calculation on a phylogeny for Bayesian estimation of divergence times. *Mol Biol Evol*. 2011;**28**:2161–2172. https://doi.org/10.1093/molbev/msr045

She R, Chu JS-C, 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

Shine R. All at sea: aquatic life modifies mate-recognition modalities in sea snakes (*Emydocephalus annulatus*, Hydrophiidae). *Behav Ecol Sociobiol*. 2005;**57**:591–598. https://doi.org/10.1007/s00265-004-0897-z

Simão FA, Waterhouse RM, Ioannidis P, Kriventseva EV, Zdobnov EM. BUSCO: assessing genome assembly and annotation completeness with single-copy orthologs. *Bioinformatics*. 2015;**31**:3210–3212. https://doi.org/10.1093/bioinformatics/btv351

Stanke M, Diekhans M, Baertsch R, Haussler D. Using native and syntenically mapped cDNA alignments to improve de novo gene finding. *Bioinformatics*. 2008;**24**:637–644. https://doi.org/10.1093/bioinformatics/btn013

Sun J, Lu F, Luo Y, Bie L, Xu L, Wang Y. OrthoVenn3: an integrated platform for exploring and visualizing orthologous data across genomes. *Nucleic Acids Res*. 2023;**51**:W397–W403. https://doi.org/10.1093/nar/gkad313

Trapnell C, Pachter L, Salzberg SL. TopHat: discovering splice junctions with RNA-Seq. *Bioinformatics*. 2009;**25**:1105–1111. https://doi.org/10.1093/bioinformatics/btp120

Trapnell C, Roberts A, Goff L, Pertea G, Kim D, Kelley DR, Pimentel H, Salzberg SL, Rinn JL, Pachter L. Differential gene and transcript expression analysis of RNA-seq experiments with TopHat and Cufflinks. *Nat Protoc*. 2012;**7**:562–578. https://doi.org/10.1038/nprot.2012.016

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

Wang Y, Tang H, DeBarry JD, Tan X, Li J, Wang X, Lee T-H, Jin H, Marler B, Guo H, et al. MCScanX: a toolkit for detection and evolutionary analysis of gene synteny and collinearity. *Nucleic Acids Res*. 2012;**40**:e49–e49. https://doi.org/10.1093/nar/gkr1293

Wang K, Ren J, Chen H, Lyu Z, Guo X, Jiang K, Chen J, Li J, Guo P, Wang Y. The updated checklists of amphibians and reptiles of China. *Biodivers Sci*. 2020;**28**:189

Wingett S, Ewels P, Furlan-Magaril M, Nagano T, Schoenfelder S, Fraser P, Andrews S. HiCUP: pipeline for mapping and processing Hi-C data. *F1000Res*. 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

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

Zug GR, Vitt L, Caldwell JP. *Herpetology: an introductory biology of amphibians and reptiles*. San Diego, CA, USA: Academic Press; 2001.




*Journal of Heredity*, 2026, 117, 141–150
https://doi.org/10.1093/jhered/esaf055
Genome Resources