Chromosome-Level Genome Assembly and Annotation of the Amur Rat Snake *Elaphe schrenckii*

Zexian Zhu [1,†], Xusheng Yang[2,†], Wen Kang[2], Cheng Cai[2], Qi Zhou [1,2,3,4,*]

[1]Center for Evolutionary and Organismal Biology and Women's Hospital, School of Medicine, Zhejiang University, Hangzhou, China
[2]MOE Laboratory of Biosystems Homeostasis and Protection and Zhejiang Provincial Key Laboratory for Cancer Molecular Cell Biology, Life Sciences Institute, Zhejiang University, Hangzhou, Zhejiang, China
[3]Center for Reproductive Medicine, Second Affiliated Hospital of Zhejiang University School of Medicine, Life Sciences Institute, Zhejiang University, Hangzhou, Zhejiang, China
[4]State Key Laboratory of Transvascular Implantation Devices, Zhejiang University, Hangzhou, China

[†]These authors contributed equally to this work.

*Corresponding author: E-mail: zhouqi1982@zju.edu.cn.

Accepted: April 29, 2025

Abstract

The Amur rat snake (*Elaphe schrenckii*), a widely distributed colubrid species in Northeast Asia, plays a critical role in controlling rodent and mouse populations in the wild. Despite its ecological and evolutionary significance, genomic resources for this nonvenomous species have been limited. In this study, we present a high-quality, chromosome-level genome assembly of *E. schrenckii*, generated by PacBio HiFi long-read sequencing and Hi-C chromatin interaction mapping. The assembled genome size comprises 1.69 Gb, with a scaffold N50 length of 215 Mb. Hi-C scaffolding anchored the genome into 18 chromosomes, including one that represents the conserved Z chromosome of snakes, consistent with karyotypic observations. This assembly enables further gene annotation and analysis of chromosomal synteny patterns. Repetitive elements account for 53.2% of the genome, with long interspersed nuclear element retrotransposons being the predominant class (23.2%). We identified 18,529 protein-coding genes, with 90.6% functionally annotated through homology-based methods. The genome assembly is highly complete, with a BUSCO score of 97.4% (tetrapoda_odb10). This resource provides a foundation for comparative studies of colubrid genome evolution, which also serves as a crucial reference for conservation genomics, particularly for Asian snake populations facing habitat fragmentation.

Key words: Amur rat snake, chromosome-level genome assembly, colubrid.

Significance

The Amur rat snake, *Elaphe schrenckii*, is an ecologically significant species that plays a crucial role in maintaining the balance of its native ecosystems. However, habitat loss and environmental changes pose substantial threats to its survival. Despite its ecological importance, our understanding of the genetic basis underlying its adaptation and evolution remains limited due to a lack of high-quality genomic resources. Here, we present the first chromosome-level genome assembly of *E. schrenckii*, featuring high contiguity, exceptional completeness, and comprehensive annotations of protein-coding genes and repetitive sequences. This high-quality genome provides a valuable foundation for studying evolutionary dynamics, gene family expansions, and adaptive mechanisms in snakes, and it will facilitate future research and conservation efforts dedicated to this species and other colubrids.

Introduction

Elaphe schrenckii, commonly known as the Amur rat snake or Russian rat snake, is a large, nonvenomous colubrid snake species that is widely distributed throughout Northeast Asia, including China, Russia, and the Korean Peninsula (Utiger et al. 2002). Its diet consists primarily of small mammals, birds, and frogs (Zhou and Zhou 2004), making it an important species in regulating prey populations within forest and grassland habitats. Due to its nonvenomous nature, the Amur rat snake does not pose a direct threat to humans, but disruptions in its habitat and disturbances from human activities can affect its population size. These factors, including habitat fragmentation and illegal wildlife trade, have led to population declines across its range, and it has been listed as an endangered species in regional IUCN assessments (IUCN 2024). Given the urgent situation, it is imperative to analyze the genetic diversity of the existing populations of the Amur rat snake to develop effective conservation strategies.

A comprehensive understanding of the distribution and ecological requirements of snakes is imperative for effective conservation and management strategies. For instance, the distribution of the hot springs snake (*Thermophis baileyi*) on the Tibetan Plateau was influenced by climate change during the glacial period, resulting in significant genetic differentiation due to its isolated habitat (Hofmann 2012). A similar phenomenon is observed in the yellow-bellied sea snake (*Hydrophis platurus*), which utilizes ocean currents to facilitate its extensive distribution, thereby ensuring the maintenance of a heterogeneous population within a vast marine environment (Brischoux et al. 2016).

In contrast to these naturally dispersed species, the Amur rat snake exhibits remarkable adaptation to captive environments. This species demonstrates tolerance to a wide range of climatic conditions, plasticity in habitat preference, and consistent reproductive success, making it an ideal model for investigating phenotypic plasticity and environmental adaptability (Koppel et al. 2010). Substantial research has focused on its reproductive biology, phylogenetic relationships, and morphological characteristics, yielding important insights into its evolutionary trajectory and adaptive mechanisms (Helfenberger 2001; Utiger et al. 2002; Kim et al. 2012; Lee et al. 2012).

In the broader context of studying the diversity and distribution of the *Elaphe* genus, significant disparities in morphology and genetics among species were reported. A notable example is the recent research of a new species, *Elaphe druzei*, in the Southern Levant. The study demonstrates that *E. druzei* exhibits pronounced morphological differences and genetic divergence from its close relatives, indicating that the genus may possess a greater degree of biodiversity than previously estimated (Jablonski et al. 2023).

However, the lack of a high-quality genome assembly has hindered investigations into its adaptive traits, sex determination mechanisms, and conservation genetics to monitor genetic diversity or design effective breeding strategies. A chromosome-level genome would address these gaps, enabling comparative studies on colubrid evolution and providing tools for population management. Here, we produced high-coverage PacBio long-read sequences and assembled them into contigs that were then joined into scaffolds using the Hi-C chromatin data. The assembled sequence scaffolds were ordered and oriented, yielding a total of 1.7 Gb in size. Repetitive sequences account for 53.2% of the total genome, similar to the previous reports for the other colubrids (Peng et al. 2023). Based on synteny analysis, the karyotypes of the Amur rat snake and corn snakes (*Pantherophis guttatus*) (Peng et al. 2023) are highly conserved. In contrast, multiple structural variations are evident when compared to the western terrestrial garter snake (*Thamnophis elegans*) (Rhie et al. 2021). Although all these species belong to the subfamily Colubrinae, they exhibit significant differences in ecological habits, morphology, distribution, and genomic structure (Arnold 1977; Ayres and Arnold 1983; Bronikowski and Vleck 2010). Overall, this work has advanced our understanding of the Amur rat snake genome and develops a reference genome for the *Elaphe* genus.

Results and Discussion

Genome Estimation

To have a primary assessment of genomic characteristics of *E. schrenckii*, we first generated 31.51 Gb of Illumina DNA reads (supplementary table S1, Supplementary Material online). Analysis of filtered reads using *k*-mer distribution ($k = 21$) estimated the genome size to range from 1,321 to 1,323.26 Mb. Further characterization revealed a heterozygosity rate of 0.47% and a repetitive sequence content of 26.97% (supplementary fig. S1 and table S2, Supplementary Material online). These results provide important baseline data for understanding the genomic organization of *E. schrenckii* and lay the foundation for downstream analyses, including comparative genomics and evolutionary studies. The relatively low heterozygosity rate [Compared to 0.9% in *Naja naja* and 1.2% in *Rhabdophis nuchalis* (Suryamohan et al. 2020; Duan et al. 2024)] indicates a moderate level of

genetic diversity within the sampled population, while the repetitive content is consistent with the genome size and organization typical of colubrid snakes.

Genome Assembly and Assessment

The genome of *E. schrenckii* was assembled using 66.4 Gb of PacBio long-read sequencing data, yielding approximately 53× coverage of the estimated genome size. The initial assembly resulted in a genome size of 1.7 Gb, comprising 578 contigs with a N50 length of 53.2 Mb (Table 1), exceeding the estimated genome size due to the inclusion of redundant regions. To improve assembly accuracy, one round of polishing with long reads and two rounds with Illumina short reads were conducted, followed by removals of redundant sequences.

To achieve a chromosome-level assembly, we employed 94.5 Gb of Hi-C sequencing data to anchor and orient contigs into scaffolds. This yielded a final genome assembly of 1.7 Gb, comprising 472 scaffolds with a N50 of 215.2 Mb. Among these scaffolds, the 18 largest were defined as pseudo-chromosomes, collectively accounting for 92.3% of the assembled genome (Fig. 1a and b and Table 1; supplementary fig. S2, Supplementary Material online). Chromosome lengths ranged from 10.5 to 351.7 Mb, with most chromosomes containing (TTAGGG)n repeats at one or both ends, indicating the telomeric regions (supplementary tables S3 and fig. S3, Supplementary Material online). Notably, one chromosome exhibited half the female DNA depth compared to the others, suggesting it represents the Z chromosome (Fig. 1b).

Comparative genomic analysis revealed that the chromosomes of *P. guttatus* (*Pgut*) and *E. schrenckii* (*Esch*) have undergone multiple rearrangements relative to *T. elegans* (*Tele*), a species within the same family (Colubridae). In contrast, *Esch* chromosomes exhibit high conservation with more distantly related outgroup species, such as *Ahaetulla prasina* (*Apra*) and *Cylindrophis ruffus* (*Cruf*), whose genomes show a certain degree of conservation with the ancestral snake genome (Peng et al. 2023). This pattern indicates that while some colubrid species have experienced extensive chromosomal rearrangement events, the genera *Pantherophis* and *Elaphe* have maintained a relatively stable karyotype that closely resembles that of ancestral snakes (Fig. 1c).

The quality and completeness of the genome assembly were evaluated using BUSCO analysis (against tetrapoda_odb10 database), which identified 97.4% of single-copy orthologs as complete (96.6% single-copy and 0.8% duplicated), with 0.7% fragmented and 1.9% missing (supplementary fig. S2, Supplementary Material online). Furthermore, mapping rates of 99.6%, 92.6%, and 99.9% were achieved for Illumina, RNA-seq, and PacBio data, respectively, confirming the high accuracy and integrity of the assembled genome. Interestingly, the assembled genome size was slightly larger than the estimated size, likely due to *k*-mer-based estimation methods that rely on the frequency distribution of unique *k*-mers. These methods may underestimate highly repetitive sequences, leading to a discrepancy between estimated and assembled sizes. Collectively, our high-quality chromosome-level genome assembly of *E. schrenckii* provides a robust resource for future studies on evolutionary biology and ecology.

Table 1 Genome assembly and annotation statistics of *E. schrenckii*

Elements	Value
Contig-level genome assembly statistics	
Total number of contigs	578
Contig N50 (Mb)	53.2
Scaffold-level genome assembly statistics	
Total number of scaffolds	472
Scaffold N50 (Mb)	215
Final pseudo-chromosome-level genome assembly statistics	
Total number of pseudo-chromosomes	18
Chromosome size range (Mb)	10–352
GC content of pseudo-chromosomes	50%
Total length of pseudo-chromosomes (Mb)	1556.8
BUSCO of genome (tetrapoda_odb10, $n = 5310$)	
Complete	97.4%
Single copy	96.6%
Duplicated	0.8%
Fragmented	0.7%
Missing	1.9%
Annotation	
Protein-coding genes	18529
Mean protein length (aa)	665
Mean gene length (bp)	34119
Exon/introns per gene	16.37/14.97
Exon (%)	8.1%
Mean exon length	169
Intron (%)	17.4%
Mean intron length	3963
BUSCO of annotated protein-coding sequences (tetrapoda_odb10, $n = 5310$)	
Complete	92.4%
Single copy	91.2%
Duplicated	1.2%
Fragmented	1%
Missing	6.6%

Repetitive Elements and Gene Prediction

Repetitive elements account for 53.2% of the *E. schrenckii* genome, encompassing approximately 854.8 Mb. Among these, transposable elements (TEs) constituted the majority (47.6%), with the remainder consisting of simple repeats (5.0%), low-complexity regions (0.4%), small RNAs (0.1%), and satellite sequences (0.1%) (supplementary

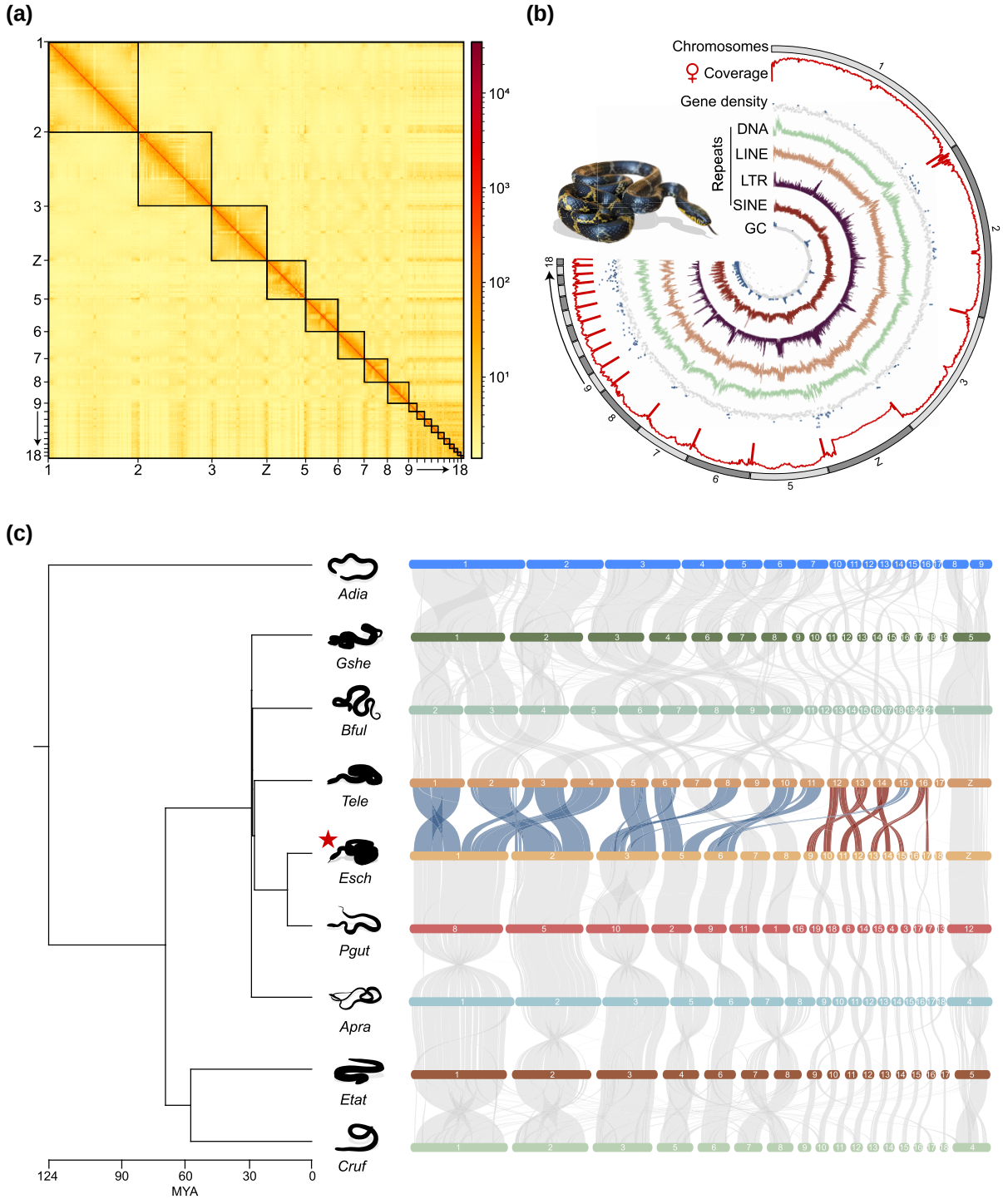

Fig. 1. Chromosome-level genome assembly of *E. schrenckii* (*Esch*). a) Hi-C interaction heatmap of *Esch*. The black squares represent the 18 chromosomes, with the color bar at the right representing the density of Hi-C interactions. b) Circos plot of the *Esch* genome assembly. The tracks from outer to inner layers were chromosomes, depth of female DNA short reads in 1 Mb sliding windows, gene densities in 1 Mb windows while blue dots indicate regions where the values greater than 25, the densities of four classes of repetitive elements in 100 kb windows, and GC content in 100 kb windows. c) The parallel linked plot shows the genome synteny between eight published snake genomes [note that the abbreviation of each species and the genome accessions can be found in supplementary table S6, Supplementary Material online (Peng et al. 2023; Tang et al. 2023; Rhie et al. 2021)] and the newly obtained *Esch* based on the genomic coding sequences. Each row represents one genome, with each chromosome displayed in separate blocks. Note that the colored blocks represent rearrangements (fusions and fissions) between *Esch* and *Tele*.

table S4, Supplementary Material online). Notably, a substantial proportion of the TEs (11.5%) remained unclassified due to the absence of homologs in known databases. Among the classified TEs, long interspersed nuclear elements (LINEs) were the most abundant (23.2%), followed by DNA transposons (7.0%) (Fig. 1b).

After masking repetitive sequences, we predicted a total of 18,529 protein-coding genes, with an average gene length of 34,118.7 bp and an average coding sequence length of 1,169.1 bp. Each gene contained an average of 16.4 exons and 15.0 introns, with mean exon and intron lengths of 169.1 bp and 3,962.7 bp, respectively (Table 1).

To assess the accuracy of gene annotation, we performed a BUSCO analysis using the longest transcript of each gene against the tetrapoda_odb10 database, which identified 92.4% of orthologs as complete, including 91.2% single-copy and 1.2% multicopy orthologs (Table 1; supplementary fig. S2, Supplementary Material online). These results further support the high quality of our gene annotation. In addition, functional annotation revealed that 90.6% (16,792 genes) had at least one transcript successfully matched to one or more databases, including EC, KEGG pathway, KEGG KO, GO, and PFAM. Among these, 19,881 transcripts were associated with Gene Ontology (GO) terms, and 10,502 transcripts were mapped to KEGG pathways. Additionally, our analysis identified 17,640 KEGG orthologous groups (KO) terms, 5,090 enzyme codes (EC), and 22,763 PFAM categories (supplementary fig. S4, Supplementary Material online). These comprehensive annotations highlight the genomic complexity of *E. schrenckii* and provide valuable insights into its gene content and functional landscape, offering a robust foundation for future evolutionary and ecological research.

Ortholog identification between *E. schrenckii* and other snake species revealed that some orthologous groups have undergone branch-specific gene family expansions involved in the immune processes and olfaction (supplementary fig. S4 and table S5, Supplementary Material online). These findings underscore the significance of the *E. schrenckii* genome as a valuable resource for advancing our understanding of snake biology and evolution.

Materials and Methods

Sample Collection and Genomic DNA Sequencing

Tissue samples were collected from captive-bred adult female offspring of *E. schrenckii*, originally obtained from the mountainous regions of Liaoning, China, and were subsequently stored at −80 °C for sequencing. The snake photo is provided by Dallin Kohler (https://www.inaturalist.org/photos/394855658) with license under http://creativecommons.org/licenses/by-nc/4.0/. To assemble the chromosome-level genome of *E. schrenckii*, we employed a combination of PacBio long-read sequencing, Hi-C chromatin interaction sequencing, and second-generation DNA sequencing. For Hi-C sequencing, an MboI digestion library was constructed.

Library Preparation and Sequencing

High-quality DNA was utilized to construct single-molecule real-time sequencing libraries with an insert size of 15 kb using the SMRTbell prep kit 3.0. Long-read sequencing was performed on the PacBio Revio platform. For short-read sequencing, paired-end libraries (150 bp insert size) were prepared using the MGIEasy Universal DNA Library Prep Kit V1.0 and sequenced on the DNBSEQ-T7RS platform. Hi-C libraries were prepared following standard proximity ligation protocols to capture chromatin conformation and sequenced on theDNBSEQ-T7RS platform. Quality control of raw Illumina reads was performed using fastp (v0.23.4) (Chen et al. 2018), ensuring clean and high-quality data. PacBio long subreads were directly generated by the SMRT Link (v13.0) sequencing system. Hi-C sequencing data underwent quality control using Juicer (v1.6) (Durand et al. 2016a, 2016b).

Genome Size Estimation

Filtered Illumina data were used for genome survey analysis to estimate genome characteristics with a *k*-mer length of 21. The characteristics were further visualized and analyzed using GenomeScope (v2.0.1) (Vurture et al. 2017). The heterozygosity ratio was calculated based on the heterozygous peak value.

Genome Assembly and Assessment

Primary contigs were constructed using Hifiasm (v0.24.0) (Cheng et al. 2021) with default parameters. Error correction was performed using Racon (v1.5.0) (Vaser et al. 2017) and Pilon (v1.24) (Walker et al. 2014), with the combination of one round of polishing with the third-generation data and two rounds with the second-generation data. The chromosome-level genome assembly was achieved using Hi-C sequencing data, where reads were aligned to the contigs with Juicer (v1.6) (Durand et al. 2016a, 2016b), followed by anchoring contigs to pseudo-chromosomes through 3D-DNA pipeline (v201008) (Dudchenko et al. 2017) and manual adjustment with Juicebox (v2.15) (Durand et al. 2016a, 2016b). Genome completeness was evaluated with BUSCO (v5.8.2) (Simão et al. 2015) against the tetrapoda_odb10 database ($n = 5{,}310$). To validate genome integrity, reads from different sources were mapped to the final assembly using Minimap2 (v2.23) (Li 2018), yielding high mapping rates for Illumina DNA reads, PacBio reads, and RNA-seq reads.

Repetitive Elements Annotation

The repeats library for *E. schrenckii* was constructed using RepeatModeler (v2.0.6) (Li 2018; Flynn et al. 2020) with default settings, combined with the consensus library from squamates. The resulting library was then applied for genome masking using RepeatMasker (v4.1.7) (Smit et al. 2013), employing rmblast (v2.14.1) as the search engine. The masked genome revealed 0.9 Gb of repetitive sequences (52.05%), including short interspersed nuclear elements at 1.8%, LINEs at 22.7%, long terminal repeats at 4.8%, DNA transposons at 6.2%, and unclassified repeats at 10.6%.

Gene Prediction and Functional Annotation

RNA-seq reads were mapped to the reference genome using STAR (v2.7.11b) (Dobin et al. 2013) with default parameters. In addition, protein sequence data from 11 snake species, annotated by NCBI RefSeq, along with the outgroup species *Gallus gallus*, were filtered to remove redundancy, forming a homolog-based data set. Gene prediction was carried out using BRAKER3 (v3.0.8) (Dobin et al. 2013; Gabriel et al. 2024).

Synteny Inferring Between Snake Genomes

Synteny blocks among nine snakes were identified using BLASTP (v2.14.1+) (Altschul et al. 1990; Gabriel et al. 2024) and MCScanX (v1.0.0) (Wang et al. 2012). Protein sequences from different species were first grouped based on phylogenetic relationships. An all-against-all BLASTP search was conducted with default parameters (*e*-value: 1e−5, maximum number of alignments: 5) after the makeblastdb step. The BLASTP output and GFF files were processed using MCScanX to identify synteny blocks. Visualization of synteny was achieved using SynVision (https://synvisio.github.io/#/) (Bandi and Gutwin 2020) after merging all collinearity and GFF files.

Gene Gain and Loss

Coding sequences from nine snakes were processed with multiple sequence alignment using OrthoFinder (v2.5.5) (Emms and Kelly 2019) followed by selecting the longest transcripts. MCMCTREE (v4.10.7) (dos Reis and Yang 2011; Reis et al. 2018; Álvarez-Carretero et al. 2019) was then used to infer species divergence time with the standard parameter settings (use data = 3, model = 4, alpha = 0.5). Next, CAFE5 (v5.1) (Mendes et al. 2021) was used to analyze changes in the size of gene families.

Supplementary Material

Supplementary material is available at *Genome Biology and Evolution* online.

Funding

Q.Z. is supported by the National Key Research and Development Program of China (2023YFA1800500, 2024YFA1802500) and National Natural Science Foundation of China (32170415).

Data Availability

The WGS, RNA-seq, and PacBio HiFi data for the *E. schrenckii* genome can be found on NCBI with the accession numbers SRR31801987 to SRR31801997, respectively, under BioProject accession number PRJNA1201621. The assembled genome is deposited in NCBI under accession number JBNFMR000000000 and is also available in the Genome Warehouse (GWH) under accession GWHFSRJ00000000.1. Annotation files are hosted on Figshare and can be accessed at https://doi.org/10.6084/m9.figshare.28595507.v2. No custom code was used in this study. The data analyses used standard bioinformatic tools specified in the methods. And all scripts used for figure performing are available on GitHub: https://github.com/zjuzexian/Rat-snake-genome/.

Literature Cited

Altschul SF, Gish W, Miller W, Myers EW, Lipman DJ. Basic local alignment search tool. J Mol Biol. 1990:215(3):403–410. https://doi.org/10.1016/S0022-2836(05)80360-2.

Álvarez-Carretero S, Goswami A, Yang Z, Dos Reis M. Bayesian estimation of species divergence times using correlated quantitative characters. Syst Biol. 2019:68(6):967–986. https://doi.org/10.1093/sysbio/syz015.

Arnold SJ. Polymorphism and geographic variation in the feeding behavior of the garter snake *Thamnophis elegans*. Science. 1977:197(4304):676–678. https://doi.org/10.1126/science.197.4304.676.

Ayres FA, Arnold SJ. Behavioural variation in natural populations. IV. Mendelian models and heritability of a feeding response in the garter snake, *Thamnophis elegans*. Heredity (Edinb). 1983:51(1):405–413. https://doi.org/10.1038/hdy.1983.45.

Bandi V, Gutwin C. Interactive exploration of genomic conservation. Graphics Interface. 2020:2020:74–83. https://doi.org/10.20380/GI2020.09.

Brischoux F, Cotté C, Lillywhite HB, Bailleul F, Lalire M, Gaspar P. Oceanic circulation models help to predict global biogeography of pelagic yellow-bellied sea snake. Biol Lett. 2016:12(8):20160436. https://doi.org/10.1098/rsbl.2016.0436.

Bronikowski A, Vleck D. Metabolism, body size and life span: a case study in evolutionarily divergent populations of the garter snake (*Thamnophis elegans*). Integr Comp Biol. 2010:50(5):880–887. https://doi.org/10.1093/icb/icq132.

Chen S, Zhou Y, Chen Y, Gu J. Fastp: an ultra-fast all-in-one FASTQ preprocessor. Bioinformatics. 2018:34(17):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(2):170–175. https://doi.org/10.1038/s41592-020-01056-5.

Dobin A, Davis CA, Schlesinger F, Drenkow J, Zaleski C, Jha S, Batut P, Chaisson M, Gingeras TR. STAR: ultrafast universal RNA-seq aligner. Bioinformatics. 2013:29(1):15–21. https://doi.org/10.1093/bioinformatics/bts635.

dos Reis M, Yang Z. Approximate likelihood calculation on a phylogeny for Bayesian estimation of divergence times. Mol Biol Evol. 2011:28(7):2161–2172. https://doi.org/10.1093/molbev/msr045.

Duan M, Yang S, Li X, Tang X, Cheng Y, Luo J, Wang J, Song H, Wang Q, Zhu GX. Chromosome-level genome assembly and annotation of the *Rhabdophis nuchalis* (Hubei keelback). Sci Data. 2024:11(1):1–11. https://doi.org/10.1038/s41597-024-03708-z.

Dudchenko O, Batra SS, Omer AD, Nyquist SK, Hoeger M, Durand NC, Shamim MS, Machol I, Lander ES, Aiden AP, et al. De novo assembly of the genome using Hi-C yields chromosome-length scaffolds. Science. 2017:356(6333):92–95. https://doi.org/10.1126/science.aal3327.

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. 2016a:3(1):99–101. https://doi.org/10.1016/j.cels.2015.07.012.

Durand NC, Shamim MS, Machol I, Rao SS, Huntley MH, Lander ES, Aiden EL. Juicer provides a one-click system for analyzing loop-resolution Hi-C experiments. Cell Syst. 2016b:3(1):95–98. https://doi.org/10.1016/j.cels.2016.07.002.

Emms DM, Kelly S. OrthoFinder: phylogenetic orthology inference for comparative genomics. Genome Biol. 2019:20(1):238. https://doi.org/10.1186/s13059-019-1832-y.

Flynn JM, Hubley R, Goubert C, Rosen J, Clark AG, Feschotte C, SmitAF. RepeatModeler2 for automated genomic discovery of transposable element families. Proc Natl Acad Sci U S A. 2020:117(17):9451–9457. https://doi.org/10.1073/pnas.1921046117.

Gabriel L, Brůna T, Hoff KJ, Ebel M, Lomsadze A, Borodovsky M, Stanke M. BRAKER3: fully automated genome annotation using RNA-seq and protein evidence with GeneMark-ETP, AUGUSTUS, and TSEBRA. Genome Res. 2024:34(5):769–777. https://doi.org/10.1101/gr.278090.123.

Helfenberger N. Phylogenetic relationships of Old World ratsnakes based on visceral organ topography, osteology, and allozyme variation. Russ J Herpetol. 2001:8:1–62. https://doi.org/10.30906/1026-2296-2001-8-0-1-62.

Hofmann S. Population genetic structure and geographic differentiation in the hot spring snake *Thermophis baileyi* (Serpentes, Colubridae): indications for glacial refuges in southern-central Tibet. Mol Phylogenet Evol. 2012:63(2):396–406. https://doi.org/10.1016/j.ympev.2012.01.014.

IUCN. The IUCN Red List of Threatened Species. Version 2024-2. 2024. https://www.iucnredlist.org.

Jablonski D, Ribeiro-Júnior MA, Simonov E, Šoltys K, Meiri S. A new, rare, small-ranged, and endangered mountain snake of the genus *Elaphe* from the Southern Levant. Sci Rep. 2023:13(1):4839. https://doi.org/10.1038/s41598-023-30878-4.

Kim D-I, Kim I-H, Kim J-K, Kim B-N, Park D-S. Movement patterns and home range of captive-bred Amur ratsnake (*Elaphe schrenckii*) juveniles in the natural habitat. J. Ecol. Environ. 2012:35(1):41–50. https://doi.org/10.5141/jefb.2012.007.

Koppel SVD, Kessel NV, Crombaghs B, Getreuer W, Lenders H. Risk analysis of the russian rat snake (*elaphe schrenckii*) in the netherlands. 2010. https://hdl.handle.net/2066/94042.

Lee J-H, Park D, Sung H-C. Large-scale habitat association modeling of the endangered Korean ratsnake (*Elaphe schrenckii*). Zoolog Sci. 2012:29(5):281–285. https://doi.org/10.2108/zsj.29.281.

Li H. Minimap2: pairwise alignment for nucleotide sequences. Bioinformatics. 2018:34(18):3094–3100. https://doi.org/10.1093/bioinformatics/bty191.

Mendes FK, Vanderpool D, Fulton B, Hahn MW. CAFE 5 models variation in evolutionary rates among gene families. Bioinformatics. 2021:36(22-23):5516–5518. https://doi.org/10.1093/bioinformatics/btaa1022.

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(16):3519. https://doi.org/10.1016/j.cell.2023.06.021.

Reis MD, Gunnell GF, Barba-Montoya J, Wilkins A, Yang Z, Yoder AD. Using phylogenomic data to explore the effects of relaxed clocks and calibration strategies on divergence time estimation: primates as a test case. Syst Biol. 2018:67(4):594–615. https://doi.org/10.1093/sysbio/syy001.

Rhie A, McCarthy SA, Fedrigo O, Damas J, Formenti G, Koren S, Uliano-Silva M, Chow W, Fungtammasan A, Kim J, et al. Towards complete and error-free genome assemblies of all vertebrate species. Nature. 2021:592(7856):737–746. https://doi.org/10.1038/s41586-021-03451-0.

Simão FA, Waterhouse RM, Ioannidis P, Kriventseva EV, Zdobnov EM. BUSCO: assessing genome assembly and annotation completeness with single-copy orthologs. Bioinformatics. 2015:31(19):3210–3212. https://doi.org/10.1093/bioinformatics/btv351.

Smit AFA, Hubley R, Green P. RepeatMasker. Open-4.0. 2013.

Suryamohan K, Krishnankutty SP, Guillory J, Jevit M, Schröder MS, Wu M, Kuriakose B, Mathew OK, Perumal RC, Koludarov I, et al. The Indian cobra reference genome and transcriptome enables comprehensive identification of venom toxins. Nat Genet. 2020:52(1):106–117. https://doi.org/10.1038/s41588-019-0559-8.

Tang CY, Zhang X, Xu X, Sun S, Peng C, Song MH, Yan C, Sun H, Liu M, Xie L, et al. Genetic mapping and molecular mechanism behind color variation in the Asian vine snake. Genome Biol. 2023:24(1):1–21. https://doi.org/10.1186/s13059-023-02887-z.

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. https://doi.org/10.30906/1026-2296-2002-9-2-105-124.

Vaser R, Sović I, Nagarajan N, Šikić M. Fast and accurate de novo genome assembly from long uncorrected reads. Genome Res. 2017:27(5):737–746. https://doi.org/10.1101/gr.214270.116.

Vurture GW, Sedlazeck FJ, Nattestad M, Underwood CJ, Fang H, Gurtowski J, Schatz MC. GenomeScope: fast reference-free genome profiling from short reads. Bioinformatics. 2017:33(14):2202–2204. https://doi.org/10.1093/bioinformatics/btx153.

Walker BJ, Abeel T, Shea T, Priest M, Abouelliel A, SakthikumarS, Cuomo CA, Zeng Q, Wortman J, Young SK, et al. Pilon: an integrated tool for comprehensive microbial variant detection and genome assembly improvement. PLoS One. 2014:9(11):e112963. https://doi.org/10.1371/journal.pone.0112963.

Wang Y, Tang H, Debarry JD, Tan X, Li J, Wang X, Lee TH, 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(7):e49. https://doi.org/10.1093/nar/gkr1293.

Zhou Z, Zhou Y. Preliminary observations on ecology of *Elaphe schrenckii* (Strauch) (Plate VII). Sichuan J Zool. 2004:23:188–190.

Associate editor: Diego Cortez