

# The genome assembly and annotation of the white-lipped tree pit viper *Trimeresurus albolabris*

Xiaotong Niu<sup>1,2,3,†</sup>, Yakui Lv<sup>4,†</sup>, Jin Chen<sup>5,†</sup>, Yueheng Feng<sup>1</sup>, Yilin Cui<sup>1</sup>, Haorong Lu<sup>5,\*</sup> and Hui Liu<sup>1,\*</sup>

- 1 Key Laboratory of Genetics and Germplasm Innovation of Tropical Special Forest Trees and Ornamental Plants (Ministry of Education), School of Tropical Agriculture and Forestry (School of Agricultural and Rural Affairs, School of Rural Revitalization), Hainan University, Haikou, 570228, China
- 2 State Key Laboratory of Agricultural Genomics, BGI-Shenzhen, Shenzhen, 518083, China
- 3 School of Ecology, Sun Yat-sen University, Shenzhen, 510275, China
- 4 College of Ecology and Environment, Hainan University, Haikou, 570228, China
- 5 China National GeneBank, BGI-Shenzhen, Shenzhen, 518120, China

## ABSTRACT

*Trimeresurus albolabris*, also known as the white-lipped pit viper or white-lipped tree viper, is a highly venomous snake distributed across Southeast Asia and the cause of many snakebite cases. In this study, we report the first whole genome assembly of *T. albolabris* obtained with next-generation sequencing from a specimen collected in Mengzi, Yunnan, China. After genome sequencing and assembly, the genome of this male *T. albolabris* individual was 1.51 Gb in length and included 38.42% repeat-element content. Using this genome, 21,695 genes were identified, and 99.17% of genes could be annotated using gene functional databases. Our genome assembly and annotation process was validated using a phylogenetic tree, which included six species and focused on single-copy genes of nuclear genomes. This research will contribute to future studies on *Trimeresurus* biology and the genetic basis of snake venom.

**Submitted:** 07 July 2023  
**Accepted:** 19 January 2024  
**Published:** 25 January 2024

**Subjects** Genetics and Genomics, Evolutionary Biology, Zoology

\* Corresponding authors. E-mail:  
[luhaorong@genomics.cn](mailto:luhaorong@genomics.cn);  
[liuhui@hainanu.edu.cn](mailto:liuhui@hainanu.edu.cn)

† Contributed equally.

Published by GigaScience Press.

Preprint submitted at <https://doi.org/10.20944/preprints202401.1712.v1>

Included in the series: **Snake Genomes** ([https://doi.org/10.46471/GIGABYTE\\_SERIES\\_0004](https://doi.org/10.46471/GIGABYTE_SERIES_0004))

This is an Open Access article distributed under the terms of the Creative Commons Attribution License (<https://creativecommons.org/licenses/by/4.0/>), which permits unrestricted reuse, distribution, and reproduction in any medium, provided the original work is properly cited.

*Gigabyte*, 2024, 1–11

## INTRODUCTION

*Trimeresurus albolabris*, also known as the white-lipped pit viper, white-lipped tree viper, white-lipped bamboo pit viper, and green tree pit viper, is a venomous snake species belonging to the family Viperidae [1]. It is a relatively small snake, with adults typically measuring around 70–90 cm in length, and is known for its distinctive appearance, with a white stripe running down the center of its upper lip [2] (Figure 1). This species has been reported in China, Vietnam, Thailand, Laos, Cambodia, India, Bangladesh, Myanmar, and West Java and has become one of the most common venomous snakes with medical importance in Southeast Asia [3]. *T. albolabris* is a highly venomous snake. Its bite can be dangerous to humans, causing symptoms ranging from pain and swelling to more severe ones, such as shock, spontaneous bleeding, defibrination, and other complications of thrombocytopenia and leukocytosis [4, 5]. Notably, the venom of *T. albolabris* contains metalloproteinases [6, 7], a thrombin-like enzyme [8], and other venom components [5, 9].

## CONTEXT

Despite its venomous nature, *T. albolabris* is also an important research subject for its sexual dimorphism [10] and geographic variation [11]. A complete and high-quality genome



**Figure 1.** *Trimeresurus albolabris*, also known as the white-lipped pit viper, photographed by Diancheng Yang.

of this species is crucial for studying venom proteomics, particularly for drug discovery, developing antivenom therapies, and understanding the evolution of venomous species [12–14]. However, a complete genome of *T. albolabris* has not been published yet [15].

Here, we report the first whole genome with high continuity of a male *T. albolabris* individual, collected from Mengzi, Yunnan, China. The genome was generated using single-tube long fragment read (stLFR) [16] and whole genome sequencing (WGS) technologies. Our *T. albolabris* genome had a repeat element content of 38.42% and a total size of 1.51 Gb. This new genome assembly provides valuable evidence for future studies on snake venom and the genetic underpinnings of the *Trimeresurus* species.

## METHOD

The detailed stepwise protocols used in this study are gathered in a protocols.io collection, with the minor adaptations outlined below (Figure 2) [17].

### Sample collection and sequencing

A male *T. albolabris* sample was captured in Mengzi, Yunnan, China. To preserve its quality, this specimen was frozen in dry ice (at  $-80^{\circ}\text{C}$ ) immediately after collection and identification, both for storage and transportation. The protocols we used for DNA extraction, library construction, and sequencing can be found in a protocols.io protocol collection [17]. The heart, stomach, liver, and kidneys were used for RNA sequencing. Additionally, a muscle sample was used for stLFR and WGS sequencing. The genome assembly and annotation workflow is also included in the same protocols.io protocol [17].

This study, including sample collection, experimental procedures, and research design, was approved by the Institutional Review Board of Beijing Genomics Institute (BGI-IRB E22017). Throughout the research, meticulous adherence to the guidelines established by BGI-IRB was strictly followed, ensuring compliance with ethical and regulatory standards.



**Protocols for the assembly and annotation of snake genomes V.2**

DOI  
dx.doi.org/10.17504/protocols.io.5jyl8j6e9g2w/v2

Boyang Liu<sup>1</sup>, Liangyu Cui<sup>1</sup>, Zhangwen Deng<sup>2</sup>, Yue Ma<sup>1</sup>, Diancheng Yang<sup>3,4</sup>, Yanan Gong<sup>3,4</sup>, Yanchun Xu<sup>1</sup>, Shuhui Yang<sup>1</sup>, Song Huang<sup>3,4</sup>

<sup>1</sup>College of Wildlife and Protected Area, Northeast Forestry University, Harbin 150040, China;  
<sup>2</sup>Guangxi Forest Inventory and Planning Institute, Nanning 530011, China;  
<sup>3</sup>Anhui Province Key Laboratory of the Conservation and Exploitation of Biological Resource, College of Life Sciences, Anhui Normal University, Wuhu 241000, China;  
<sup>4</sup>Huangshan Noah Biodiversity Institute, Huangshan 245000, China

VERSION 2 ▾  
FEB 09, 2023

SHARE

WORKS FOR ME 1

COMMENTS 0

BOOKMARK

COPY / FORK

MORE ↓

博洋 Boyang Liu

**Figure 2.** A protocols.io collection of protocols for sequencing snake genomes [17]. <https://www.protocols.io/widgets/doi?uri=dx.doi.org/10.17504/protocols.io.4r3l27ez4g1y/v1>

## Genome assembly, annotation, and assessment

The stLFR sequencing data were subjected to assembly using Supernova [18] (v2.1.1, RRID:SCR\_016756). Subsequently, the gap-filling and redundancy removal steps were performed using GapCloser [19] (v1.12-r6, RRID:SCR\_015026) and redundans [20] (v0.14a), respectively. These processes involved utilizing the WGS data to address the gaps in the assembly and eliminate redundant sequences.

To identify known repeat elements within the genome sequences, several genomics tools were employed, including Tandem Repeats Finder [21] (v. 4.09, RRID:SCR\_022193), LTR\_Finder [22] (RRID:SCR\_015247), RepeatModeler [23] (v1.0.8), RepeatMasker [24] (v. 3.3.0, RRID:SCR\_015027), and RepeatProteinMask (v. 3.3.0) [25]. For the prediction of protein-coding genes, we used a comprehensive approach combining *de novo*, homology-based, and transcript mapping strategies. *De novo* gene prediction was performed using GlimmerHMM [26] (RRID:SCR\_002654). For RNA-seq-based predictions, RNA-seq data was first filtered by Trimmomatic [27] (v0.30, RRID:SCR\_011848). After obtaining clean RNA-seq data, transcripts were assembled using Trinity [28] (v2.13.2, RRID:SCR\_013048). Finally, PASA [29] (v2.0.2, RRID:SCR\_014656) was used to align transcripts against the white-lipped tree viper genome to obtain gene structures. Homology-based prediction was performed by mapping protein sequences of the UniProt database (release-2020\_05), *Pseudonaja textilis* (GCA\_900518735.1), *Protobothrops mucrosquamatus* (GCA\_001527695.3), *Thamnophis elegans* (GCA\_009769535.1), and *Notechis scutatus* (GCA\_900518725.1) to the white-lipped tree viper genome using Blastall (v2.2.26) [30] with an E-value cut-off of 1e-5. Next, we used GeneWise [31] (v2.4.1, RRID:SCR\_015054) to analyze the alignment results and predict gene homology. Finally, the integration of RNA-seq, homology, and *de novo* predicted genes resulted in the generation of a final gene set using the MAKER pipeline (v3.01.03, RRID:SCR\_005309) [32]. This approach, incorporating multiple genomic tools and techniques, facilitated the annotation and prediction of genes in the white-lipped tree viper genome.

Table 1. Summary statistics of <i>T. albolabris</i> of WGS paired-end (fq - fastq 1 and fastq 2) sequenced reads.						
	WGS-1		WGS-2		WGS-3	
	fq1	fq2	fq1	fq2	fq1	fq2
%Q20	96.98	97.81	97.74	94.96	95.69	97.59
%Q30	90.79	90.6	92.83	84.27	84.46	89.81
%GC	40.37	40.21	41.02	40.81	40.36	40.47
%ErrorRate	0.351809	0.233019	0.264481	0.540272	0.449364	0.257064
TotalReads	492,445,828		425,689,572		104,911,172	
TotalBases	98,489,165,600		85,137,914,400		20,982,234,400	

Table 2. Summary statistics of <i>T. albolabris</i> stLFR and RNA sequenced reads.						
	stLFR-1		stLFR-2		RNA-seq	
	fq1	fq2	fq1	fq2	fq1	fq2
%Q20	96.53	95.59	96.41	96.3	98.3	98.19
%Q30	89.83	87.26	87.98	86.37	94.3	93.71
%GC	39.34	42.16	39.28	42.15	44.11	44.07
%ErrorRate	0.403065	0.525486	0.442228	0.392415	0.194523	0.205665
TotalReads	633,976,833		161,105,172		50,828,075	
TotalBases	145,814,671,590		37,054,189,560		10,165,615,000	

Functional annotation was performed using BLAST search, comparing with several databases, including SwissProt, TrEMBL, and Kyoto Encyclopedia of Genes and Genomes (KEGG), and limiting the E-value cut-off to 1e-5. InterProScan [26] (v5.52-86.0, RRID:SCR\_005829) was used to predict motifs and domains, as well as gene ontology (GO) terms.

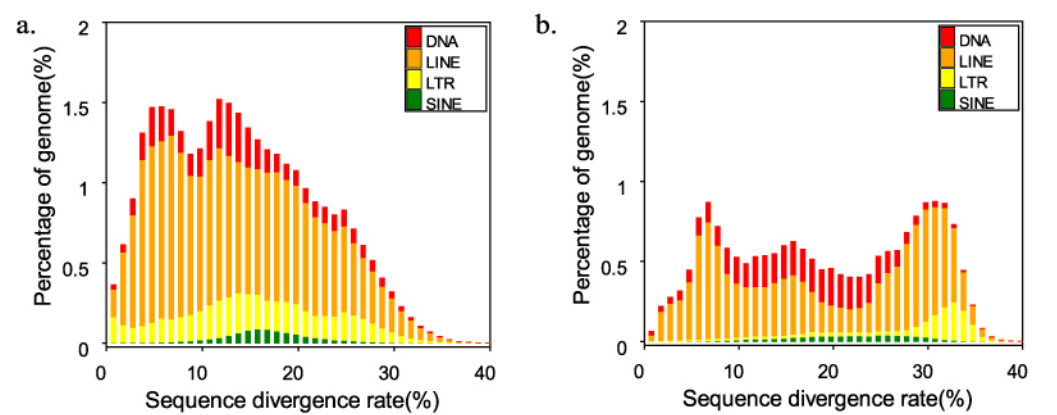
Benchmarking Universal Single-Copy Orthologs (or BUSCO, v5.2.2, RRID:SCR\_015008) with genome mode and lineage data from vertebrata\_odb10 was used to evaluate the completeness of our genome [33].

A reconstructed phylogenetic tree was generated by OrthoFinder (v2.3.7, RRID:SCR\_017118) [34], which can search for single-copy orthologs among the protein sequences of *Chelonia mydas* (GCA\_015237465.2), *Gallus gallus* (GCA\_016699485.1), *Homo sapiens* (GCA\_000001405.29), *Mus musculus* (GCA\_000001635.9), *Ophiophagus hannah* (GCA\_000516915.1), *Python bivittatus* (GCA\_000186305.2), *Xenopus tropicalis* (GCA\_000004195.4), *Alligator mississippiensis* (GCA\_000281125.4), *Danio rerio* (GCA\_000002035.4), *Anolis carolinensis* (GCA\_000090745.2), *Gopherus evgoodei* (GCA\_007399415.1), *Podarcis muralis* (GCA\_004329235.1), and *Deinagkistrodon acutus* [35].

RESULTS

This study on snake genomics resulted in a total of 387.48 Gb of paired-end (fastq 1 and fastq 2) data, which comprised 204.61 Gb of short reads data obtained through WGS sequencing and 182.87 Gb of long reads data obtained through stLFR sequencing, as shown in Tables 1 and 2.

We generated the first whole genome assembly of *T. albolabris* with high continuity, with a total genome size of 1.51 Gb, 39.97% GC content, and a scaffold N50 length of 381.55 kb (Table 3). The assembled *T. albolabris* genome consists of 10,016 contigs over 1,000 base pairs, with a total length of 1.50 Gb, accounting for 99.14% of the genome’s total length. This resource will provide valuable evidence to explore new perspectives in the study of the *Trimeresurus viper* genomics.



**Figure 3.** Distribution of transposable elements (TEs) in our *T. albolabris* genome. The TEs include DNA transposons (here indicated as DNA) and RNA transposons (i.e., DNAs, LINEs, LTRs, and SINEs). (a) Distribution of the *de novo* sequence divergence-rate. (b) Distribution of the known sequence divergence-rate.

**Table 3.** Summary of the features of the *T. albolabris* genome.

	Contigs	Contigs > (1,000 bp)	Contigs > (10,000 bp)
Total number (>)	71,131	46608	10,016
Total length of (bp)	1,513,852,334	1,501,212,553	1,355,102,082
N50 Length (bp)		381,553	
N75 Length (bp)		115,212	
GC content is (%)		39.97	

**Table 4.** Statistics of the repetitive sequences identified in our *T. albolabris* genome.

Type	Repeat size	% of genome
Trf	47,767,541	3.155363
Repeatmasker	252,985,952	16.711402
Proteinmask	185,792,360	12.272819
<i>De novo</i>	498,353,737	32.919574
Total	581,568,803	38.416482

We detected repetitive elements in the *T. albolabris* genome, accounting for 38.42% of the total genome. Among them, the highest proportion was occupied by long interspersed nuclear elements (LINEs), which accounted for 23.94% and amounted to approximately 362.35 Mb. These findings were found to be highly similar to the repetitive element content observed in previously sequenced genomes, such as those of *Thamnophis elegans* (42.02%) (accession No. [PRJNA561996](#)) and *Crotalus tigris* (42.31%) [36]. This indicates that the results we obtained are highly reliable and plausible. The remaining types of transposable elements, including DNA transposons, long terminal repeats (LTRs), and short interspersed nuclear elements (SINEs), accounted for 6.90%, 5.83%, and 1.24%, respectively (Figure 3, Tables 4, and 5).

Using homology-based, *de novo*, and RNA-sequencing annotation methods, we successfully identified 21,695 protein-coding genes in our *T. albolabris* genome assembly. We compared our assembly to those of *Notechis scutatus* ([GCA\\_900518725.1](#)), *Pseudonaja textilis* ([GCA\\_900518735.1](#)), and *Thamnophis elegans* ([GCA\\_009769535.1](#)), all of which are available from the NCBI database. Our analysis revealed no significant differences in the



Table 5. Summary of the TEs in our *T. albolabris* genome.

Type	Repbse TEs		TE protiens		De novo		Combined TEs	
	Length (bp)	% in genome	Length (bp)	% in genome	Length (bp)	% in genome	Length (bp)	% in genome
DNA	51,357,881	3.392529	2,032,636	0.134269	64,605,374	4.267614	104,513,127	6.903786
LINE	184,866,441	12.211656	157,414,659	10.398284	294,829,469	19.475444	362,351,919	23.935751
SINE	9,622,825	0.635651	0	0	13,144,889	0.868307	18,769,499	1.23985
LTR	23,685,560	1.564589	26,413,572	1.744792	74,868,256	4.945546	88,305,953	5.833195
Other	77,658	0.00513	141	0.000009	0	0	77,799	0.005139
Unknown	0	0	0	0	98,895,691	6.532717	98,895,691	6.532717
Total	252,985,952	16.711402	185,792,360	12.272819	496,342,637	32.786727	566,552,754	37.424572

Table 6. Statistics for the miRNA, tRNA, rRNA, and snRNA discerned from our *T. albolabris* genome.

Type	Copy (w)	Average length (bp)	Total length (bp)	% of genome
miRNA	250	98.992	24,748	0.001635
tRNA	179	75.70949721	13,552	0.000895
rRNA	104	137.6057692	14,311	0.000945
snRNA	301	115.1229236	34,652	0.002289

Table 7. Consequences of gene functional annotation.

Values	Total	Swissprot-annotated	KEGG-annotated	TrEMBL-annotated	Interpro-annotated	GO-annotated	Overall
Number	21,695	20,240	19,216	21,134	21,019	14,786	21,516
Percentage	100%	93.29%	88.57%	97.41%	96.88%	68.15%	99.17%

distribution of transcript mapping lengths, coding sequences (CDS) lengths, or the quantity of exons and introns. Additionally, our analysis predicted the presence of 250 miRNAs, 179 tRNAs, and 301 snRNAs within the *T. albolabris* genome (Table 6).

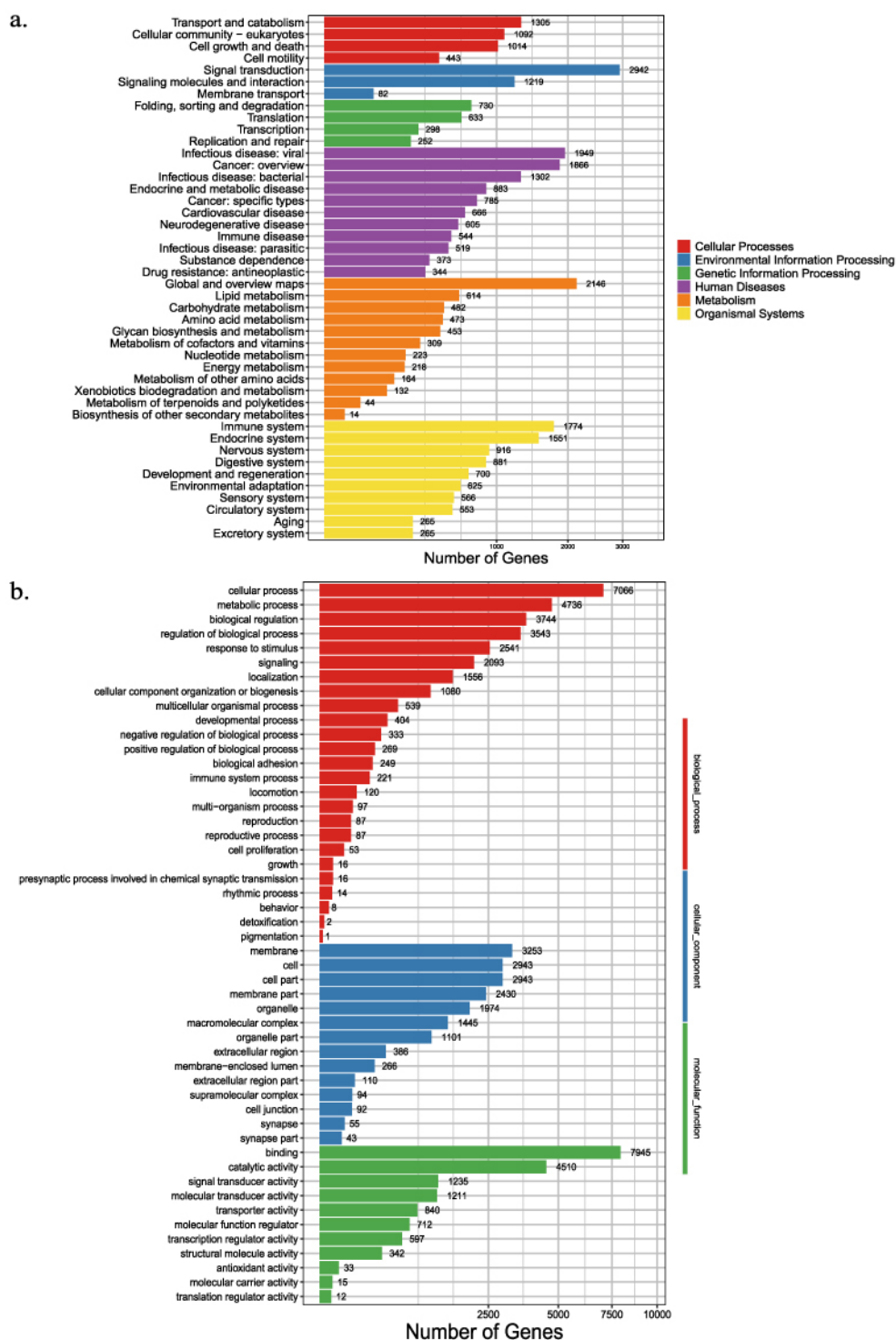
Comparing our results with various public datasets, such as InterPro [37], KEGG [38], SwissProt [39], TrEMBL [39], and GO terms, we identified 21,695 expanded gene families, including 99.17% functionally annotated genes (Table 7).

Further analyses using KEGG enrichment revealed that Environmental Information Processing, Organismal Systems, and Metabolism pathways were the most abundant, with Signal Transduction pathways being the most prominent. Among the Organismal Systems pathways, 1,774 Immune System genes and 1,551 Endocrine System genes were the most abundant (Figure 4a). In addition, based on the results of our GO analysis, we found that 7,900 genes are related to binding, while 7,740 genes are related to cellular processes (Figure 4b).

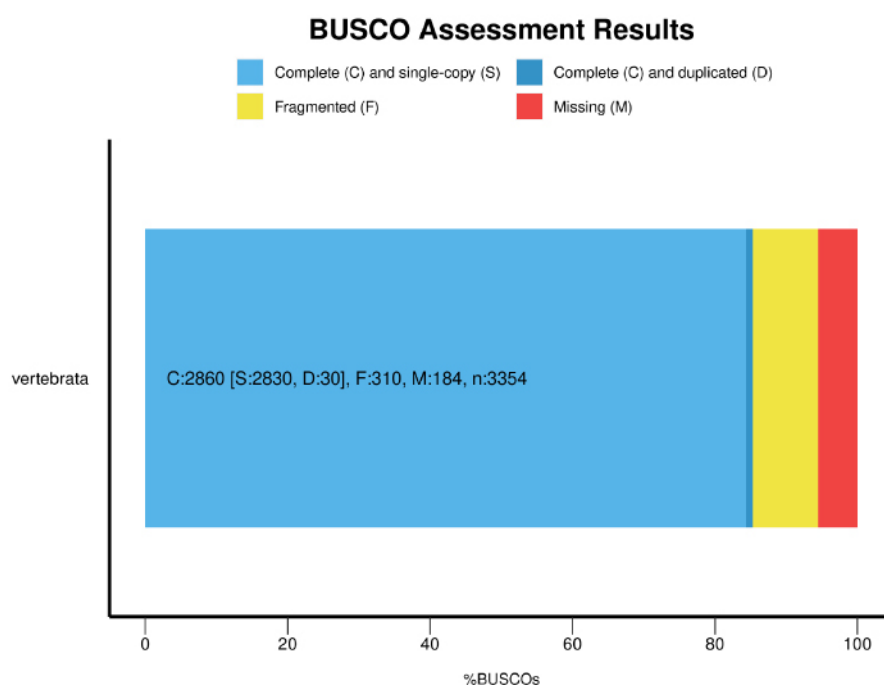
DATA VALIDATION AND QUALITY CONTROL

We employed BUSCO v5.2.2 to assess the quality and completeness of our genome assembly [40]. The results of our BUSCO analysis revealed that our assembly achieved 85.3% completeness when evaluated against the vertebrata\_odb10 database (Figure 5), indicating that our assembly is of relatively high quality and completeness.

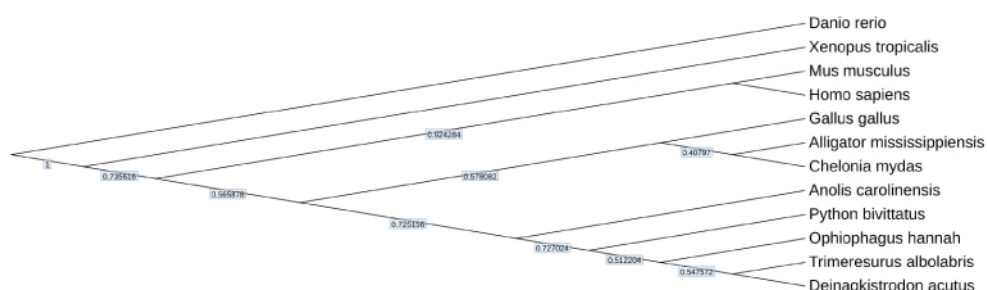
To assess the quality of our assembly, we constructed a phylogenetic tree using the protein sequences of seven different amphibian and reptile species (*Anolis carolinensis*, *Chelonia mydas*, *Deinagkistrodon acutus*, *Ophiophagus hannah*, *Python bivittatus*, *Xenopus tropicalis*, and *Alligator mississippiensis*) as well as the protein sequences of *Gallus gallus*, *Homo sapiens*, *Mus musculus*, and *Danio rerio* downloaded from NCBI. The resulting



**Figure 4.** Gene annotation information obtained from our *T. albolabris* genome. (a) KEGG enrichment. (b) GO enrichment.



**Figure 5.** BUSCO Assessment result of our *T. albolabris* genome.



**Figure 6.** Phylogenetic tree reconstructed using nuclear genome single-copy genes. The numbers on the branches of the phylogenetic tree represent the branch lengths obtained using OrthoFinder.

phylogenetic tree is consistent with previous research, indicating that our data can accurately identify related species (Figure 6).

## REUSE POTENTIAL

We presented the first genome assembly of the white-lipped tree pit viper. This data provides new resources for studying the vipers biology and evolution, as well as the genetic foundation of its venom.

## DATA AVAILABILITY

The data supporting the findings of this study have been deposited into the CNGB Sequence Archive (CNSA) [41] of the China National GeneBank DataBase (CNGBdb) [42] with accession number [CNP0004151](#). The raw data is also available in NCBI with the bioproject



number [PRJNA955401](#) (see also the machine readable [nanopublication: RAV3oIcruk](#)). Additional data is also available in GigaDB [43].

## EDITOR'S NOTE

This paper is part of a series of Data Release papers presenting the genomes of different snake species [44].

## ABBREVIATIONS

fq: fastq; GO: gene ontology; KEGG: Kyoto Encyclopedia of Genes and Genomes; LINE: long interspersed nuclear element; LTR: long terminal repeat; SINE: short interspersed nuclear element; stLFR: single-tube long fragment read; TEs: transposable elements; WGS: whole genome sequencing.

## DECLARATIONS

### Ethics approval

This study was approved by the Institutional Review Board of BGI (BGI-IRB E22017).

### Competing interests

The authors declare no competing interests.

### Author contribution

H Liu and H Lu designed and initiated the project. YC and YF performed the DNA extraction and the library construction. XN and JC performed the data analysis. XN, JC and YL wrote the manuscript. All authors read and approved the final manuscript.

## Acknowledgements

Our project was supported by the China National GeneBank (CNGB). This work was also supported by Hainan University and BGI-Shenzhen. The Anhui Normal University collected the samples. This work was supported by the Guangdong Provincial Key Laboratory of Genome Read and Write (grant no. 2017B030301011).

## REFERENCES

- 1 Gray JE. Synopsis of the species of Rattle snakes, or Family of Crotalidae. *Zoological Miscellany*, 1842; 2: 47–51. <http://repositorio.ciencias.unam.mx:8080/xmlui/handle/11154/158063>.
- 2 Leviton A, Wogan G, Koo M et al. The Dangerously venomous snakes of myanmar. Illustrated key and checklist. *Proc. Calif. Acad. Sci.*, 2003; 54: 407–462. <https://repository.si.edu/handle/10088/4542>.
- 3 Zhu F, Liu Q, Che J et al. Molecular phylogeography of white-lipped tree viper (*Trimeresurus*; Viperidae). *Zool. Scr.*, 2016; 45(3): 252–262. doi:10.1111/zsc.12156.
- 4 Cockram CS, Chan JC, Chow KY. Bites by the white-lipped pit viper (*Trimeresurus albolabris*) and other species in Hong Kong. A survey of 4 years' experience at the Prince of Wales Hospital. *J. Trop. Med. Hyg.*, 1990; 93: 79–86.
- 5 Liew JL, Tan NH, Tan CH. Proteomics and preclinical antivenom neutralization of the mangrove pit viper (*Trimeresurus purpureomaculatus*, Malaysia) and white-lipped pit viper (*Trimeresurus albolabris*, Thailand) venoms. *Acta Trop.*, 2020; 209: 105528. doi:10.1016/j.actatropica.2020.105528.
- 6 Jangprasert P, Rojnuckarin P. Molecular cloning, expression and characterization of albolamin: A type P-IIa snake venom metalloproteinase from green pit viper (*Cryptelytrops albolabris*). *Toxicon*, 2014; 79: 19–27. doi:10.1016/j.toxicon.2013.12.005.

- 7 Pinyachat A, Rojnuckarin P, Muanpasitporn C et al. Albocollagenase, a novel recombinant P-III snake venom metalloproteinase from green pit viper (*Cryptelytrops albolabris*), digests collagen and inhibits platelet aggregation. *Toxicon*, 2011; 57(5): 772–780. doi:10.1016/j.toxicon.2011.02.011.
- 8 Tan CH. Snake venomomics: Fundamentals, recent updates, and a look to the next decade. *Toxins*, 2022; 14(4): 247. doi:10.3390/toxins14040247.
- 9 Olaoba OT, Karina DSP, Selistre-de-Araujo HS et al. Snake Venom Metalloproteinases (SVMPs): A structure-function update. *Toxicon X*, 2020; 7: 100052. doi:10.1016/j.toxcx.2020.100052.
- 10 Zhu F, Chen L, Guo P et al. Sexual dimorphism and geographic variation of the white-lipped pit viper (*Trimeresurus albolabris*) in China. *Curr. Herpetol.*, 2022; 41(1): 24–34. doi:10.5358/hsj.41.24.
- 11 Malhotra A, Thorpe RS. A phylogeny of the trimeresurus group of pit vipers: New evidence from a mitochondrial gene tree. *Mol. Phylogenetics Evol.*, 2000; 16(2): 199–211. doi:10.1006/mpev.2000.0779.
- 12 Casewell NR, Wüster W, Vonk FJ et al. Complex cocktails: the evolutionary novelty of venoms. *Trends Ecol. Evol.*, 2013; 28(4): 219–229. doi:10.1016/j.tree.2012.10.020.
- 13 Calvete JJ. Venomomics: integrative venom proteomics and beyond. *Biochem. J.*, 2017; 474(5): 611–634. doi:10.1042/BCJ20160577.
- 14 Rao W, Kalogeropoulos K, Allentoft ME et al. The rise of genomics in snake venom research: recent advances and future perspectives. *GigaScience*, 2022; 11: giac024. doi:10.1093/gigascience/giac024.
- 15 Song T, Zhang C, Zhang L et al. Complete mitochondrial genome of *Trimeresurus albolabris* (Squamata: Viperidae: Crotalinae). *Mitochondrial DNA*, 2015; 26(2): 291–292. doi:10.3109/19401736.2013.825780.
- 16 Wang O, Chin R, Cheng X et al. Efficient and unique cobarcoding of second-generation sequencing reads from long DNA molecules enabling cost-effective and accurate sequencing, haplotyping, and de novo assembly. *Genome Res.*, 2019; 29: 798–808. doi:10.1101/gr.245126.118.
- 17 Liu B, Cui L, Deng Z et al. The annotation pipeline for the genome of a snake. protocols.io. 2023; <https://dx.doi.org/10.17504/protocols.io.4r3l27ez4g1y/v1>.
- 18 Weisenfeld NI, Kumar V, Shah P et al. Direct determination of diploid genome sequences. *Genome Res.*, 2017; 27: 757–767. doi:10.1101/gr.214874.116.
- 19 Luo R, Liu B, Xie Y et al. SOAPdenovo2: an empirically improved memory-efficient short-read de novo assembler. *GigaScience*, 2012; 1(1): 18. doi:10.1186/2047-217X-1-18.
- 20 Pryszcz LP, Gabaldón T. Redundans: an assembly pipeline for highly heterozygous genomes. *Nucleic Acids Res.*, 2016; 44(12): e113. doi:10.1093/nar/gkw294.
- 21 Benson G. Tandem repeats finder: a program to analyze DNA sequences. *Nucleic Acids Res.*, 1999; 27(2): 573–580. doi:10.1093/nar/27.2.573.
- 22 Xu Z, Wang H. LTR\_FINDER: an efficient tool for the prediction of full-length LTR retrotransposons. *Nucleic Acids Res.*, 2007; 35(suppl\_2): W265–W268. doi:10.1093/nar/gkm286.
- 23 Smit AFA, Hubley R, Green P. RepeatModeler Open-1.0. 2008–2015. Seattle, USA: Institute for Systems Biology, 2018. <http://www.repeatmasker.org>. Last Accessed May 1 2015.
- 24 Tarailo-Graovac M, Chen N. Using RepeatMasker to identify repetitive elements in genomic sequences. *Curr. Protoc. Bioinformatics*, 2009; 25: 4.10.1–4.10.14. doi:10.1002/0471250953.bi0410s25.
- 25 Tempel S. Using and understanding RepeatMasker. In: Mobile Genetic Elements: Protocols and Genomic Applications. Methods in Molecular Biology, vol. 859, Humana Press, 2012; pp. 29–51, doi:10.1007/978-1-61779-603-6\_2.
- 26 Stanke M, Steinkamp R, Waack S et al. AUGUSTUS: a web server for gene finding in eukaryotes. *Nucleic Acids Res.*, 2004; 32(suppl\_2): W309–W312. doi:10.1093/nar/gkh379.
- 27 Bolger AM, Lohse M, Usadel B. Trimmomatic: a flexible trimmer for Illumina sequence data. *Bioinformatics*, 2014; 30(15): 2114–2120. doi:10.1093/bioinformatics/btu170.
- 28 Haas BJ, Papanicolaou A, Yassour M et al. De novo transcript sequence reconstruction from RNA-seq using the Trinity platform for reference generation and analysis. *Nat. Protoc.*, 2013; 8: 1494–1512. doi:10.1038/nprot.2013.084.
- 29 Haas BJ, Salzberg SL, Zhu W et al. Automated eukaryotic gene structure annotation using EvidenceModeler and the Program to Assemble Spliced Alignments. *Genome Biol.*, 2008; 9: R7. doi:10.1186/gb-2008-9-1-r7.

- 30 Mount DW. Using the Basic Local Alignment Search Tool (BLAST). *Cold Spring Harb. Protoc.*, 2007; 2007: pdb.top17. doi:10.1101/pdb.top17.
- 31 Birney E, Clamp M, Durbin R. GeneWise and Genomewise. *Genome Res.*, 2004; 14: 988–995. doi:10.1101/gr.1865504.
- 32 Campbell MS, Holt C, Moore B et al. Genome annotation and curation using MAKER and MAKER-P. *Curr. Protoc. Bioinformatics*, 2014; 48(1): 4.11.1–4.11.39. doi:10.1002/0471250953.bi0411s48.
- 33 Manni M, Berkeley MR, Seppey M et al. BUSCO update: novel and streamlined workflows along with broader and deeper phylogenetic coverage for scoring of eukaryotic, prokaryotic, and viral genomes. *Mol. Biol. Evol.*, 2021; 38(10): 4647–4654. doi:10.1093/molbev/msab199.
- 34 Emms DM, Kelly S. OrthoFinder: solving fundamental biases in whole genome comparisons dramatically improves orthogroup inference accuracy. *Genome Biol.*, 2015; 16: 157. doi:10.1186/s13059-015-0721-2.
- 35 Yin W, Wang Z, Li Q et al. Supporting data for “Evolutionary trajectories of snake genes and genomes revealed by comparative analyses of five-pacer viper”. *GigaScience Database*, 2016; <http://dx.doi.org/10.5524/100196>.
- 36 Margres MJ, Rautsaw RM, Strickland JL et al. The Tiger Rattlesnake genome reveals a complex genotype underlying a simple venom phenotype. *Proc. Natl. Acad. Sci. USA*, 2021; 118(4): e2014634118. doi:10.1073/pnas.2014634118.
- 37 Jones P, Binns D, Chang H-Y et al. InterProScan 5: genome-scale protein function classification. *Bioinformatics*, 2014; 30(9): 1236–1240. doi:10.1093/bioinformatics/btu031.
- 38 Kanehisa M.. The KEGG database. In: ‘In Silico’ Simulation of Biological Processes. John Wiley & Sons, Ltd. doi:10.1002/0470857897.ch8.
- 39 Bairoch A, Apweiler R. The SWISS-PROT protein sequence database and its supplement TrEMBL in 2000. *Nucleic Acids Res.*, 2000; 28(1): 45–48. doi:10.1093/nar/28.1.45.
- 40 Simão FA, Waterhouse RM, Ioannidis P et al. BUSCO: assessing genome assembly and annotation completeness with single-copy orthologs. *Bioinformatics*, 2015; 31(19): 3210–3212. doi:10.1093/bioinformatics/btv351.
- 41 Guo X, Chen F, Gao F et al. CNSA: a data repository for archiving omics data. *Database*, 2020; 2020: baaa055. doi:10.1093/database/baaa055.
- 42 Chen FZ, You LJ, Yang F et al. CNGBdb: China National GeneBank DataBase. *Yi Chuan (Hereditas)*, 2020; 42(8): 799–809. doi:10.16288/j.ycz.20-080.
- 43 Niu X, Lv Y, Chen J et al. Supporting data for “The genome assembly and annotation of the White-Lipped Tree viper *Trimeresurus albolabris*”. *GigaScience Database*, 2024; <http://dx.doi.org/10.5524/102492>.
- 44 Snake Genomes. *GigaByte*. 2023; [https://doi.org/10.46471/GIGABYTE\\_SERIES\\_0004](https://doi.org/10.46471/GIGABYTE_SERIES_0004).