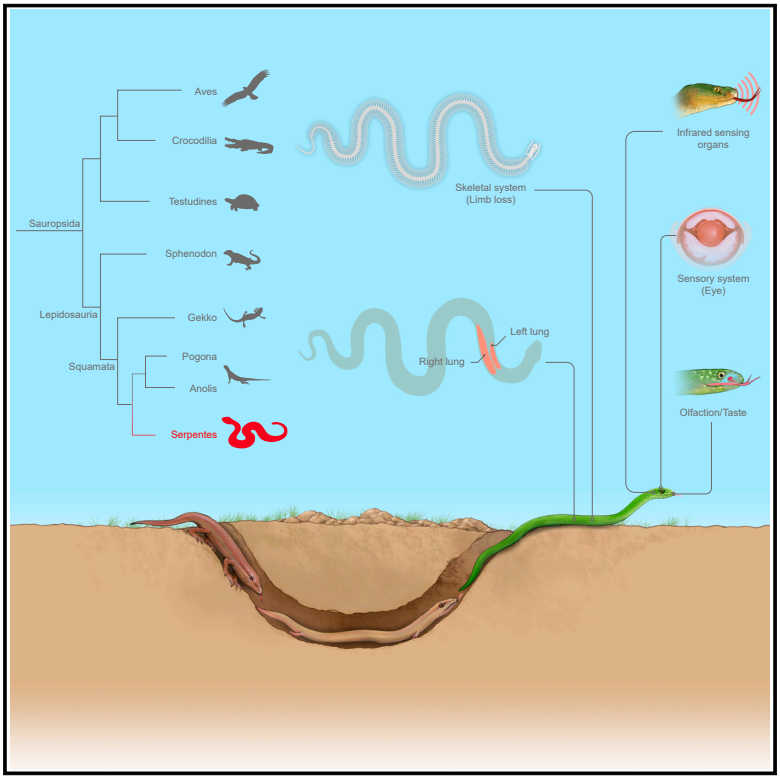
Cell

Article

Large-scale snake genome analyses provide insights into vertebrate development

Graphical abstract



Highlights

- Comprehensive chromosome-level assembly of representative snake genomes
- Reveals the genetic basis of limb loss, elongated body, and asymmetrical lungs
- Genomic variation drives blind and infrared-sensitive snake specialization
- Functional experiments validate the comparative genomic discoveries

Authors

Changjun Peng, Dong-Dong Wu, Jin-Long Ren, ..., Yin Qi, Zhi-Yi Zhang, Jia-Tang Li

Correspondence

lijt@cib.ac.cn

In brief

A vast genomic data resource sheds light on the evolution and genetic basis of unique traits of snakes.

Peng et al., 2023, Cell *186*, 2959–2976
July 6, 2023
https://doi.org/10.1016/j.cell.2023.05.030

CellPress

Cell

CellPress
OPEN ACCESS

Article

Large-scale snake genome analyses provide insights into vertebrate development

Changjun Peng,[1,2,8] Dong-Dong Wu,[3,8] Jin-Long Ren,[1,8] Zhong-Liang Peng,[1,2] Zhifei Ma,[1,2] Wei Wu,[1,2] Yunyun Lv,[1,4] Zeng Wang,[1,2] Cao Deng,[5] Ke Jiang,[1] Christopher L. Parkinson,[6] Yin Qi,[1] Zhi-Yi Zhang,[1] and Jia-Tang Li[1,2,7,9,*]

[1]CAS Key Laboratory of Mountain Ecological Restoration and Bioresource Utilization & Ecological Restoration and Biodiversity Conservation Key Laboratory of Sichuan Province, Chengdu Institute of Biology, Chinese Academy of Sciences, Chengdu 610040, China
[2]University of Chinese Academy of Sciences, Beijing 100049, China
[3]State Key Laboratory of Genetic Resources and Evolution, Kunming Institute of Zoology, Chinese Academy of Sciences, Kunming 650223, China
[4]College of Life Science, Neijiang Normal University, Neijiang, Sichuan 641100, China
[5]Departments of Bioinformatics, DNA Stories Bioinformatics Center, Chengdu 610000, China
[6]Department of Biological Sciences, Clemson University, Clemson, SC 29634, USA
[7]Southeast Asia Biodiversity Research Institute, Chinese Academy of Sciences, Yezin, Nay Pyi Taw 05282, Myanmar
[8]These authors contributed equally
[9]Lead contact
*Correspondence: lijt@cib.ac.cn
https://doi.org/10.1016/j.cell.2023.05.030

SUMMARY

Snakes are a remarkable squamate lineage with unique morphological adaptations, especially those related to the evolution of vertebrate skeletons, organs, and sensory systems. To clarify the genetic underpinnings of snake phenotypes, we assembled and analyzed 14 *de novo* genomes from 12 snake families. We also investigated the genetic basis of the morphological characteristics of snakes using functional experiments. We identified genes, regulatory elements, and structural variations that have potentially contributed to the evolution of limb loss, an elongated body plan, asymmetrical lungs, sensory systems, and digestive adaptations in snakes. We identified some of the genes and regulatory elements that might have shaped the evolution of vision, the skeletal system and diet in blind snakes, and thermoreception in infrared-sensitive snakes. Our study provides insights into the evolution and development of snakes and vertebrates.

INTRODUCTION

Snakes have successfully invaded all continents, except the polar regions,[1] and include ~4,000 extant species in more than 30 families.[2] They are efficient predators and play key roles in maintaining ecological balances. Since diverging from their burrowing ancestors, snakes have experienced various evolutionary changes, such as changes in body form (including an elongated body plan and asymmetrical lungs), limb loss, changes in chemoreception and thermoreception, and changes in feeding behavior that distinguish them from most other squamates (lizards and worm lizards). Some snakes have evolved unique lifestyles and morphological innovations. For example, blind snakes have a burrowing lifestyle and a specialized diet (mainly small arthropods, such as chitin-rich ants and termites),[3,4] and infrared-sensitive pythons, boas, and vipers have evolved specialized pit organs or receptors that enhance their thermoreception.[5,6] These unique adaptations make snakes excellent model systems for developmental biology research and studies of the evolution of complex traits.

Despite their unique biological features and ecological significance, our understanding of snakes, including their diverse phenotypes and natural history, remains poor. Furthermore, the genetic mechanisms underlying specific traits have not yet been clarified because of the limited number of high-quality snake genomes (vertebrate benchmarking universal single-copy ortholog [BUSCO] scores $\geq$ 90%, contig N50 $\geq$ 1 Mb)[7–9] and family coverage.[9–16]

To enhance our understanding of snake evolution, we *de novo* assembled the genomes of 14 snake species (Figure S1A) in 12 families in the suborder Serpentes; given that these 12 families contain approximately 84% of all snake species, our set of genomes is a highly representative sample of the snake tree of life (Table S1). We also used previously published genomes (11 snakes, six lizards, one turtle, one crocodile, and one bird), along with our newly sequenced transcriptomes (194 samples from 9 snake species, one lizard, and mouse) (Table S1), to clarify the genomic evolution of these taxa. In addition to clarifying the origin and evolution of the squamate lineage and the unique morphological adaptations of snake species, our results

Table 1. Genome assembly statistics of 14 snakes

Scientific name	English name	Abbreviation	Assembled genome size (Gb)	Scaffold N50 (bp)	Contig N50 (bp)	BUSCO (%)
Euprepiophis perlacea	Szechwan rat snake	Eper	1.64	215,947,938	54,236,136	94.1
Pantherophis guttatus	red cornsnake	Pgut	1.73	146,850,882	26,650,242	94.2
Calamaria septentrionalis	Hong Kong dwarf snake	Csep	1.95	249,103,795	13,457,423	91.5
Cylindrophis ruffus	red-tailed pipe snake	Cruf	1.55	204,348,666	47,299,786	96.1
Eryx tataricus	Tartar sand boa	Etat	1.53	210,175,343	17,592,297	95.0
Hypsiscopus plumbea	plumbeous water snake	Hplu	1.51	–	38,336,523	93.9
Psammodynastes pulverulentus	common mock viper	Ppul	1.59	216,958,053	32,196,315	93.9
Boaedon fuliginosus	African house snake	Bful	1.72	159,198,188	24,942,696	92.4
Leptotyphlops nigroterminus	black-tip worm snake	Lnig	1.97	–	3,846,211	92.5
Pareas berdmorei	keeled slug snake	Pber	1.85	144,018,271	13,457,423	93.8
Argyrophis diardii	Diard's blind snake	Adia	1.79	265,014,500	25,570,704	93.3
Gloydius shedaoensis	Shedao island pitviper	Gshe	1.52	191,459,315	6,670,067	95.6
Achalinus jinggangensis	Zong's odd-scaled snake	Ajin	1.73	226,357,597	8,252,461	92.7
Xenopeltis unicolor	sunbeam snake	Xuni	1.42	199,476,799	56,968,228	94.2

enhance our understanding of the developmental biology of vertebrates.

RESULTS

Genome sequencing, assembly, and annotation

De novo assembly using long reads (PacBio and Oxford Nanopore Technology) resulted in 14 draft snake genomes (Table S1). The assemblies were polished using BGI and Illumina short reads to obtain high-quality genomes (contig N50: 3.8–56.3 Mb, average: 26.2 Mb) (Table 1). The 12 snake genomes were further assembled at the chromosome level using Hi-C read data (Table 1), and the vast majority of karyotypes were consistent with those in previous studies (Table S1). The assembled genomes were highly complete (BUSCO scores: 91.50%–96.10%) (Table 1), indicating that all assemblies were of high quality for downstream comparative analyses.

Among the studied species, the predicted number of coding genes ranged from 18,301 to 23,299 (Table S1). The average gene, coding sequence (CDS), and exon lengths were similar to those reported in previously published snake genomes (Table S1). A user-friendly, publicly available genome browser database was established (http://herpmuseum.cib.ac.cn/snake/) to visualize the genomic and transcriptomic data presented in this study.

Phylogeny and origin of snakes

Based on the 31-taxon alignment generated by Progressive Cactus,[17] we inferred a genome-wide phylogenetic tree of 25 snakes, four lizards, one turtle, and one crocodile with high-quality genomes (contig N50 $\geq$ 180 kb or assembled to the chromosome level and with a BUSCO score $\geq$ 94%) were used as the outgroup taxa. A total of ~51 Mb of orthologous sequences were extracted, yielding a well-supported whole-genome phylogeny (Figures 1 and S1B). We also performed additional phylogenetic analyses using 1,980 sets of 1:1 orthologous protein-coding genes identified in the 31 reptile species, as well as 4-fold degenerate sites in these genes and conserved non-coding elements (CNEs). Finally, we inferred coalescent species trees using ASTRAL-III[18] with the orthologous sequences and genes. All resulting topologies were identical to the genome-wide tree (Figures 1 and S1B–S1G) and consistent with previous studies examining snake evolutionary relationships.[19,20] Serpentes was monophyletic and nested within Lacertilia, indicating that extant snakes originated from lizards. Moreover, the most recent common ancestor of snakes was sister to the (European glass lizard + Komodo dragon) clade. Within snakes, the "Scolecophidia" clade (blind snakes) comprised (Typhlopidae + Leptotyphlopidae), which was the most basal snake clade. Although the concatenated orthologous sequences and gene trees receive nearly 100% support and have the same topology as the species tree, there are a few gene tree discordances in Henophidia and Lamprophiidae, respectively (Figures S1F–S1H). The small differences in the proportions of different topologies and relatively short branch lengths suggest that the discrepancies could be attributed to incomplete lineage sorting (ILS) occurring during early species divergence. Furthermore, the estimated divergence times showed that snakes originated during the Early Cretaceous (~117.68 million years ago [mya]) and rapidly diversified after the Cretaceous-Paleogene (K-Pg, ~65 mya) mass extinction[21–23] (Figure 1).

Snake genome features and evolution

Chromosomal evolution

We used our newly assembled chromosome-level genomes and four publicly available chromosome-level genomes (Table S1) to reconstruct the hypothesized ancestral chromosomes of Serpentes. The ancestral Serpentes genome was ~1.56 Gb in length and comprised 23 chromosomes, including 8 macrochromosomes and 15 microchromosomes (Figures 2A and S2A;

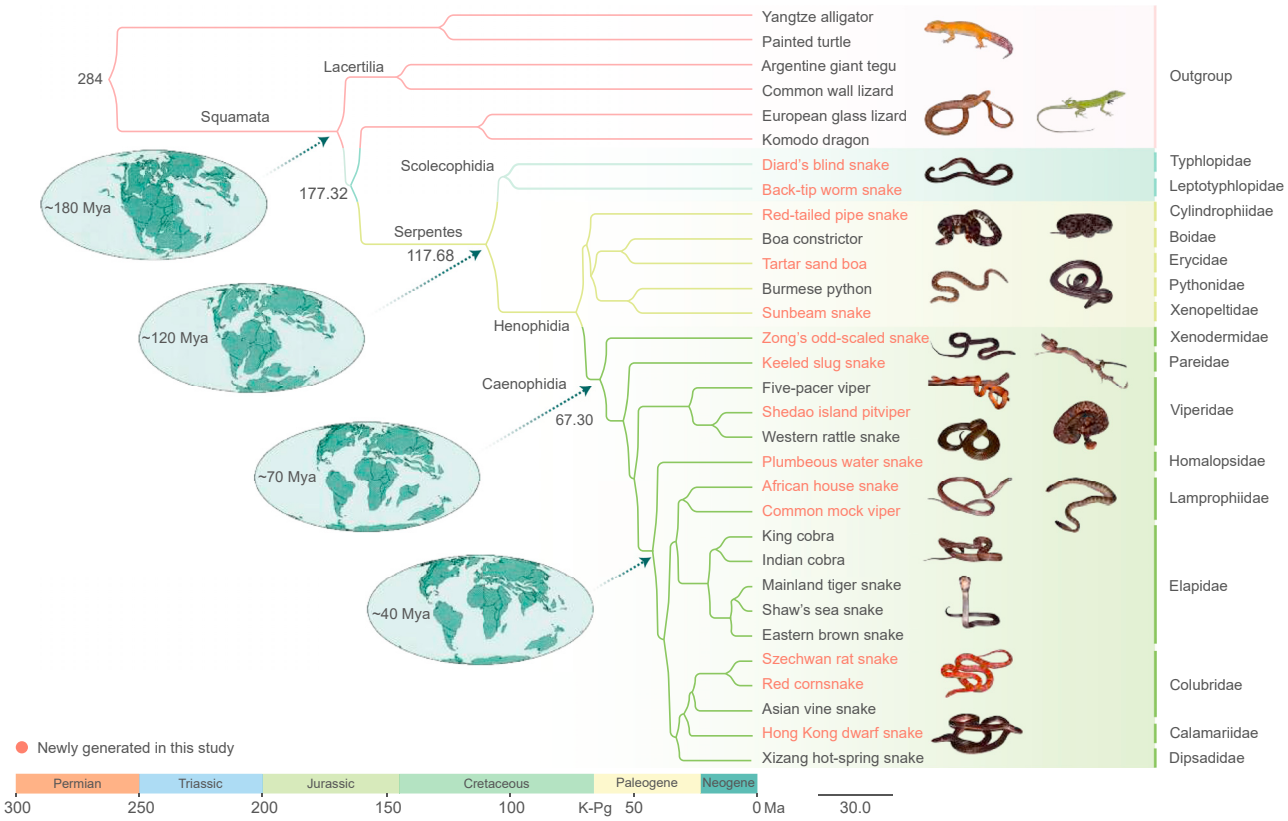


Figure 1. Phylogeny of snakes

Maximum-likelihood phylogenetic tree inferred from whole-genome sequences of 31 species. Divergence times of all nodes were estimated by r8s with whole-genome sequences using six calibration points (Figures S1I and S1J). All genomes generated in this study are in red. Maps were taken from those in a previous study.[24]

Table S2). The snake Z chromosome may have originated from lizard autosomes, and at least one break event has occurred. Although the conservation of the ancestral snake genome was evident in most snake genomes examined, the highest number of chromosome fusion and break events appears to have occurred in the Hong Kong dwarf snake (Figure S2B).

Genome size, GC content, and evolutionary rate

Our snake genome assemblies ranged from 1.42 Gb (sunbeam snake) to 1.97 Gb (red-tailed pipe snake) in size (Table 1). The average size was 1.68 Gb, which was similar to the size of the lizard genomes (average: ~1.57 Gb),[25,26] smaller than the genomes of the western painted turtle[27] and Chinese alligator (~2.27 Gb),[28] and larger than the chicken genome (~1.07 Gb)[29] (Figure 2B).

Transposable elements (TEs) comprised 35%–55% of the snake genomes. Correlation analysis confirmed that TE bursts have driven the expansion of snake, lizard, turtle, bird, and crocodile genomes, with LINE/CR1 and LINE/L2 sequences being the main drivers of genome expansion (Figure 2B). These results are in contrast to previous research showing that genome sizes in Squamata are not strongly correlated with the repeat content.[30] These discrepancies might stem from the previous use of Illumina short reads in whole-genome analyses, which might introduce bias into calculations of the repeat content.[31]

CpG island density and the GC content were lower in the genomes of all snakes than in the genomes of the outgroups and the reconstructed reptile ancestor (GC content: 38.14% vs. 40.09%; CpG island density 2.55 counts/kb vs. 3.40 counts/kb). In addition, CpG island density and the GC content were higher in basal snakes than in advanced snakes, indicating that CpG island density and the GC content have decreased over the evolutionary history of snakes (Figure 2C). These results are similar to those of Schield et al.,[13] showing that the evolutionary patterns of snake genomes (referring to the accumulation of AT) differ from those in birds[32] and mammals.[33] We found that the genome evolutionary rate estimated using r8s[34] was negatively correlated with snake body length (Figure S2C). This is consistent with the results of studies of ruminants showing that body size is significantly negatively correlated with the rate of genome evolution.[35] These results suggest that relationships between body size and rates of genome evolution are similar among vertebrate groups.

Genome structural variation

According to the genome alignment of the 31 taxa, we identified 24,352,060 bp of unique segments and 16,146,509 bp of lost segments in snake genomes compared with the outgroups (Figure S2D). Of these, 17,877 insertions and 23,808 deletions $\geq$50 bp in length were identified as snake-specific genome

Figure 2. Evolutionary features of snake genomes

(A) Chromosome evolution in snakes. A total of 23 proto-chromosomes of Serpentes were reconstructed using four lizards as outgroups (only eight species are shown here for clarity, with more species listed in Figure S2A).

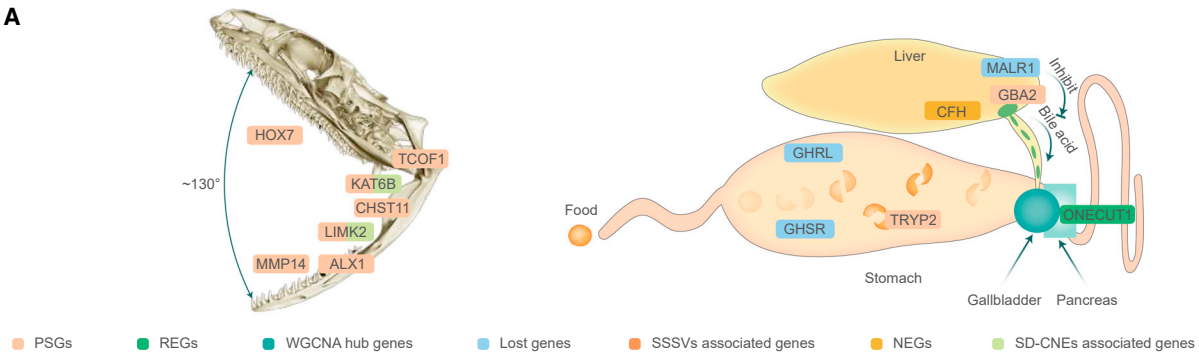
(B) Genome size, TE size, and TE type content in chicken, non-snake reptiles (four lizards, one turtle, and one crocodile), and snakes.

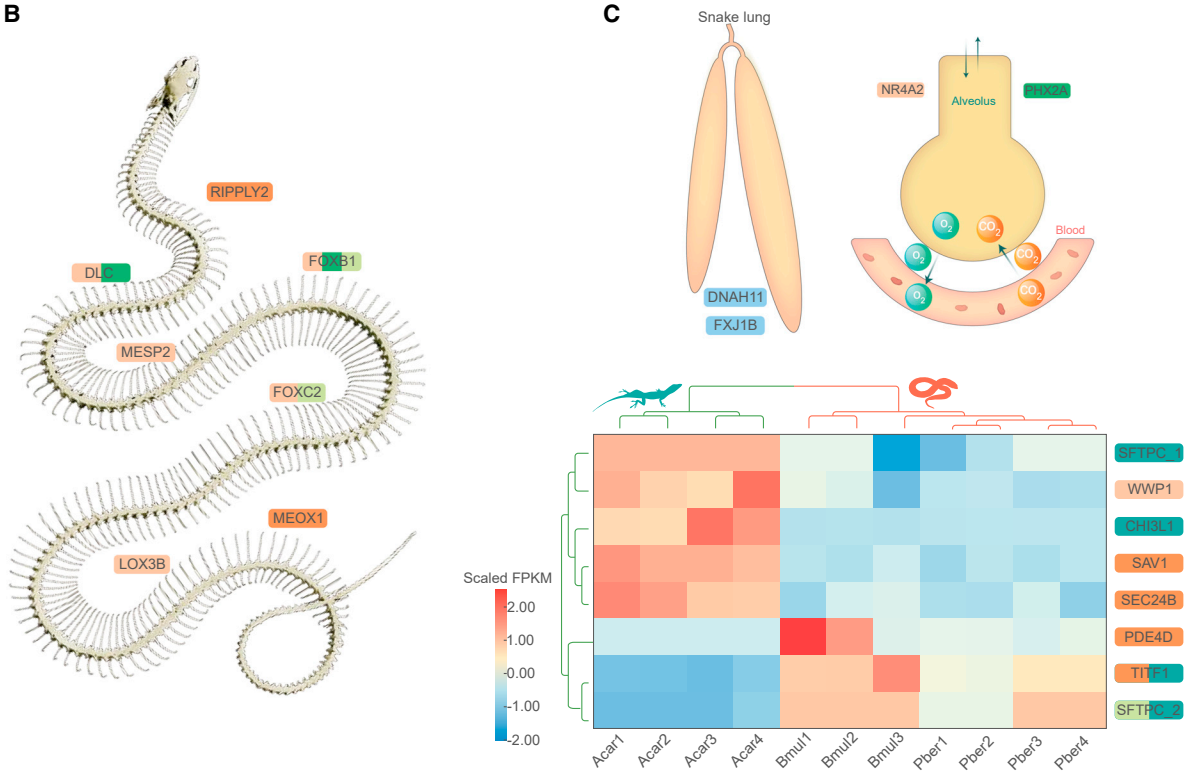
(C) GC content and CpG island density of snake genomes.

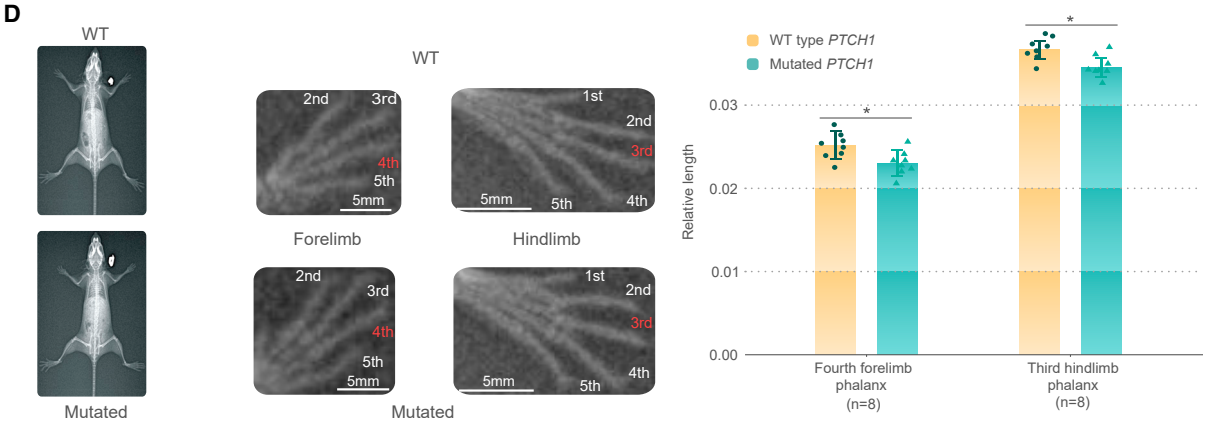
structural variations (SSSVs). Functional annotation of these SSSVs revealed that 34.7% were located either within or near (5-kb region flanking the transcriptional start and end sites) coding gene regions (Figure S3A; Table S3), including genes associated with limb, eye, eyelid, somite, and lung development, as well as respiratory function (Table S3). This suggests that SSSVs might have contributed to the evolution of snake-specific traits.

Evolution of coding and non-coding elements

Using the 31-taxon whole-genome alignment, we identified 142 newly evolved coding genes and 156 lost coding genes in snakes (Table S3). Enrichment analysis indicated that the newly







(legend on next page)

evolved and lost genes are enriched in sensory and digestive system functions. Furthermore, we identified 419 positively selected genes (PSGs) and 270 rapidly evolving genes (REGs) in snakes (Table S3). The PSGs are significantly enriched in 79 Mouse Genome Informatics (MGI) terms, including functions associated with skeletal system, lung, and eye development (Table S3). The REGs are enriched in skeletal system and lung development-related functions of MGI terms, taste transduction (ko04742), and ossification involved in bone maturation (GO:0043931) (Table S3). Thus, genes involved in skeletal and sensory systems may have played key roles in snake evolution. GERP and PhastCons scores indicated that a total of 116,707 CNEs were shared among the 31 species. We identified 1,686 snake-diverged CNEs (SD-CNEs, CNEs that were diverged in snakes but were still conserved in the outgroup), and the related genes are enriched in 48 MGI terms involved in snake traits (Figure S3D). Thus, in addition to limbs,[36] changes in other specific characteristics of snakes might also be attributed to the evolution of SD-CNEs.

Genomic elements involved in skeletal system evolution and organ adaptation

The ancestral snake phenotype was likely adapted to a burrowing and crawling lifestyle, and this was followed by the evolution of other specialized traits.[37,38] The specialized body form and skeletal system (e.g., limb loss, elongated body, and flexible mandible) of snakes are considered their signature traits. Snakes have evolved unique organ shapes, including asymmetrical lungs, to achieve an elongated body plan.[39–41] To investigate the genetic basis of these traits, we conducted comparative analyses using genomic and transcriptomic data.

Evolution of the genetic elements underlying limb loss

Extant snakes have completely lost their forelimbs, and basal snakes have retained rudimentary hindlimbs that are lacking in advanced snakes.[42,43]

To investigate the genetic mechanisms underlying limb loss, previous studies of limb development have focused on clarifying the evolution of non-coding regulators, which have diverged in snakes, suggesting that these elements may have contributed to limb loss in snakes.[44] Here, we focused on the potential contributions of coding genes and structural variation evolution to snake limb loss. Our results revealed that the forelimb development initiator RDH10[45] was lost in snakes (Table S3), and another copy of *RDH10* was surrounded by SD-CNEs (Figure S3D), implying that snake forelimb loss may be controlled by expression changes of this initiator.

In the sonic hedgehog receptor PTCH1 and cell polarity effector CPLANE1, we identified three Toxicofera-specific (i.e., snakes, anguimorphs, and iguanian lizards) and four snake-specific amino acid residue deletions, respectively (Figures S4A and S4B). PTCH1 plays an essential role in embryonic limb morphogenesis,[46,47] and CPLANE1 participates in embryonic digit morphogenesis.[48] We suggest that these deletions may alter the functions of these genes, which affects snake limb development. In addition, potential regulatory regions (5-kb region flanking the transcriptional start and end sites) of eight known limb development-related genes overlapped the SSSVs (Table S3). The 17-bp snake-specific deletion of the *SHH* gene enhancer (ZRS), which is known to contribute to abnormal limb development,[44] was observed in all newly sequenced snake genomes (Figure S4G). We also found that advanced snakes lost nearly the entire ZRS sequence (Figure S4G). Our results indicate that both coding gene mutations and structural variations may have also contributed to snake limb loss.

Although we identified an array of candidate genes that might contribute to snake limb loss, we lacked the evidence needed to directly infer the mechanism linking genes to phenotypes. Therefore, we investigated how the deletion of one coding gene (*PTCH1*) affects limb development using CRISPR-Cas9 technology. Within Toxicofera, 65% of species are limbless,[2] and we identified Toxicofera-specific amino acid residue deletions, suggesting that they potentially have limb development-related functions. These deletions were located in the extracellular topological domain (Figure S4A). The insertion of a single amino acid residue in this domain can lead to the abnormal development of the hands and feet in humans.[49] Therefore, we hypothesized that these deletions may underlie the molecular mechanism responsible for limb loss in Toxicofera.

To investigate the functional effects of these deletions *in vivo*, we edited the mouse *PTHC1* gene using CRISPR-Cas9 technology to obtain *PTCH1* mice with the 946–948 bp deleted (subsequently referred to as *PTCH1*-mutated mice with snake deletion). 1-week-old *PTCH1*-mutated mice showed no evident limb defects but had significantly lower relative body mass and shorter bodies than wild-type (WT) mice (Figure S4C). The 1-week-old *PTCH1*-mutated mice also had significantly smaller trabecular bone volume (BV) and trabecular BV per tissue volume (BV/TV) (Figure S4D). Furthermore, the fourth forelimb phalanx and third hindlimb phalanx were significantly shortened in adult *PTCH1*-mutated mice (Figures 3D and S4E). These results suggest that the deletions of this gene inhibit skeletal growth, including the growth of the limb digits. The limb buds of *PTCH1*-mutated mice contained 12 differentially expressed limb development-related genes (DELGs) in the forelimbs (including *SHH*, *ID4*, *HOXD11*, and *FGF8*) and 22 DELGs in the hindlimbs (including *SP9*, *DLX6*, *PLXNA2*, *PITX1*, and *EN1*) compared with WT mice

Figure 3. Genetic basis of skeletal system evolution and organ adaptation in snakes

PSGs, REGs, WGCNA hub genes, lost genes, SSSV-related genes, newly evolved genes, and SD-CNE-associated genes are marked in different colors and are represented by rectangles.

(A) Renewal of genomic elements contributing to snake skull development and digestion. Related genes are shown in their corresponding regions.

(B) The evolution of genomic elements has potentially facilitated the evolution of the elongated body plan of snakes.

(C) The evolution of genes related to lung development and function. The heatmap shows the expression levels of WGCNA hub genes, PSGs, and SSSV-related genes in green anole (Acar), keeled slug snake (Pber), and many-banded krait (Bmul) (The "_numbers" represents the different copies of a multi-copy gene).

(D) Whole-body X-ray images and relative lengths of the fourth forelimb phalanx and third hindlimb phalanx of wild and *PTCH1*-mutated mice (eight samples per group, the significance is indicated as *$p < 0.05$). Mean ± SD is shown by error bar. See also Figure S4 and Table S3.

(Figure S4F). These results imply that the *PTCH1* deletion can alter its function, which inhibits limb digit growth by altering the expression of genes in the limb buds. Therefore, we suggest that deletions within the Toxicofera are involved in the evolution of the skeletal system and limb loss, and this might represent an exaptation in snakes and limbless lizards.

Genetic variations related to swallowing and digestion

Given the loss of forelimbs in snakes, they have evolved other morphological traits to facilitate prey capture and feeding, such as the independently moving maxillae and mandibles and the recurved palatal teeth in the skull, which allow snakes to capture and swallow large prey[41] (Figure 3A). Here, 85 SD-CNEs were associated with 38 coding genes involved in mandibular morphogenesis (Figure S3D). In addition, several coding genes functionally related to mandible and maxilla morphogenesis showed strong signals of selection in snakes (Figure 3A; Table S3). Among them, LIM domain kinase 2 (LIMK2) and matrix metalloproteinase-14 (MMP14) are essential for normal head growth[50] and may play a role in snake head evolution. Homeobox protein HOX7 (MSX-1), a key transcriptional repressor involved in regulating the odontogenesis of dentin-containing teeth,[51,52] has evolved rapidly in snakes (Figure 3A). These results suggest that the evolution of genes and regulatory factors may have facilitated the evolution of the unique skull structure of snakes.

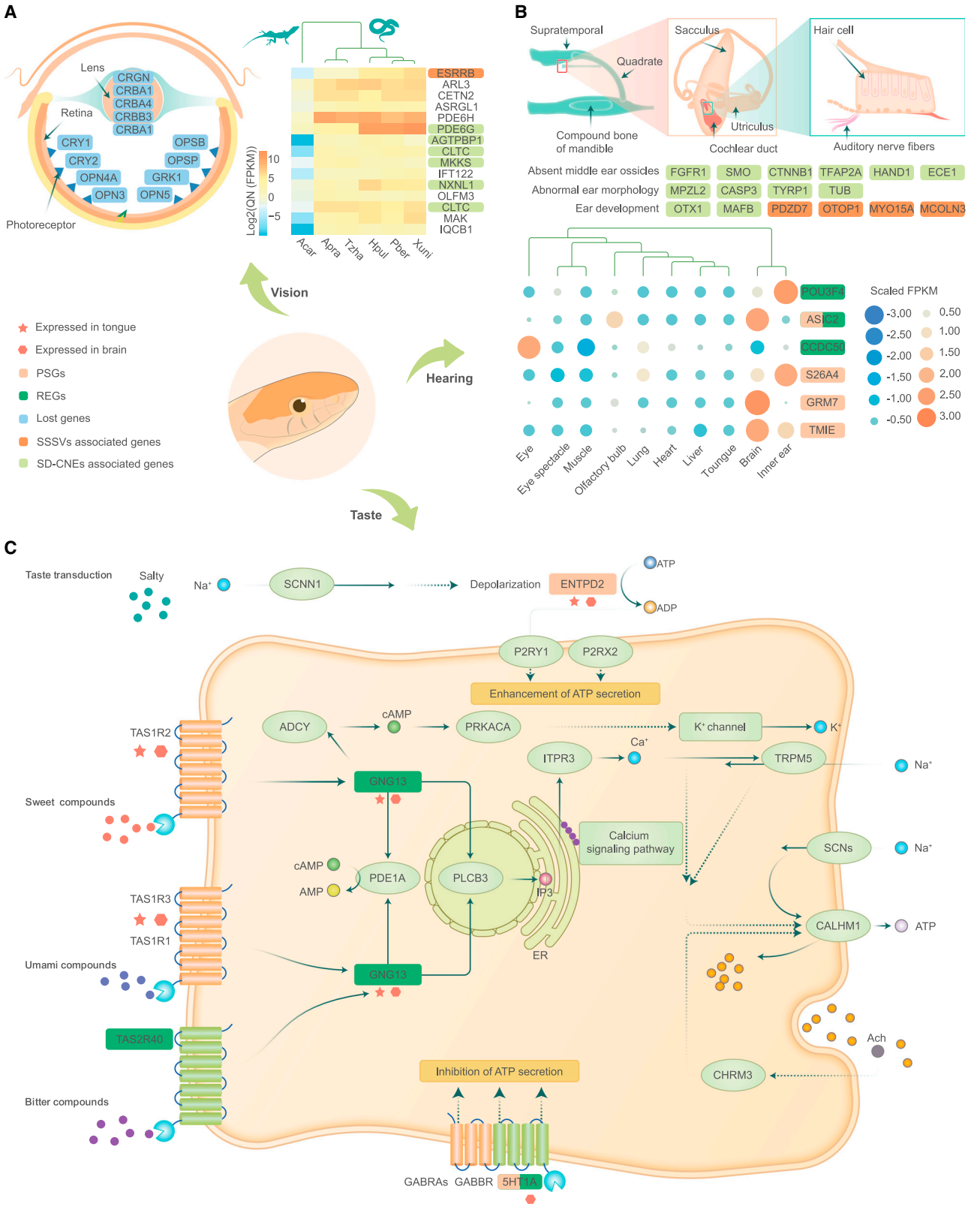
The special features of snake skulls allow them to swallow whole prey; many consume prey only a few times per year[41] and fast for many months.[53–55] Snakes have thus evolved an efficient digestive system that can shrink when not in use but quickly expand when food is consumed.[56] We found that the appetite-regulating hormone GHRL and its receptor GHSR, which are responsible for appetite stimulation in vertebrates,[57–60] have been lost in snakes (Figures 3A and S3F). The bile acid biosynthesis suppressor MALR1[61] has also been lost in snakes (Figures 3A and S3F). By contrast, two complement factor Hs (CFHs) have newly evolved in snakes (Figures 3A and S3E). Evolutionary analysis revealed that two digestive system-related genes (non-lysosomal glucosylceramidase [*GBA2*] and anionic trypsin-2 [*TRYP2*]) have experienced positive selection (Figure 3A), whereas *ONECUT1*, a pancreatic development regulator,[62] evolved rapidly. *GBA2* and *TRYP2* are responsible for digestion,[63,64] and *CFHs* respond to dietary excess.[65] The turnover and renewal of genes may explain why snakes can fast for long periods and then rapidly digest large prey.

Elongated body plan and asymmetrical lung evolution

Limb loss has a direct effect on locomotion, and snakes have evolved highly specialized movement systems. As a result, the snake skeletal system has evolved specific phenotypes, including an elongated body plan, multiple times.[66] In this study, five PSGs (*LOX3B*, *FOXB1*, *DLC*, *MESP1*, and *FOXC2*) are enriched in GO terms related to somitogenesis (Figure 3B). DLC, which evolved rapidly in snakes (Figure 3B), is a key protein responsible for somite formation, and mutation of this protein in zebrafish results in somite defects.[67,68] *FOXC2* is a well-known gene that contributes to anterior somite development in mice.[69] We also identified 16 SD-CNE-associated coding genes significantly enriched in somite size change-related MGI terms (Figure S3D). In addition, 10 *HOX* genes have experienced relaxed selection (Table S3), which is similar to the results of a previous study.[9] We found SSSV insertions in the potential regulatory regions of *RIPPLY2* and *MEOX1*, two well-known core genes involved in somite development and specification[70–73] (Table S3). These genes were also found to be expressed in the pre-ossified vertebral column of *Ptyas dhumnades* embryos (40 days post oviposition [dpo]), with *FOXB1* showing high expression compared with other tissues (Figure S4H). These genetic elements may explain the evolution of an elongated body in snakes.

The evolution of the elongated body of snakes was accompanied by changes in internal organs, including the lungs.[41] In extant snakes, the left lung is absent, residual, or developed but smaller than the right lung.[39,40] We found that one dynein-coding gene (*DNAH11*), which plays a key role in the development of left-right symmetries,[74] was absent in snakes. The forkhead box protein J1-coding gene (*FOXJ1*), which affects left-right asymmetric development[75,76] and facilitates lung development,[77] has also been lost in snakes (Figure 3C; Table S3). Six snake-specific amino acid residue deletions were detected in AP3B1 (Figure S4I), which controls alveolar epithelium development.[78] SD-CNE-associated genes are significantly enriched in MGI terms related to abnormal lung development and morphology and small lung (Figure S3D). We also identified SSSV insertions in the potential regulatory regions of the 13 lung development-related genes (e.g., *CRISPLD2*, *FGF18*, *TITF1*, *MAN2A1*, *GRHL2*, *SEC24B*, *NFIB*, and *PDE4D*) (Table S3). Comparative transcriptome analysis further revealed 2,195 differentially expressed genes between the left and right lungs in *P. dhumnades* embryos 40 dpo, including 388 SSSV-associated genes and 678 genes with SD-CNEs. Some of these genes, such as *FGF18, NFIB*, and *FOXF1* have been identified as essential for lung development[79,80] (Figure S4J). These differentially expressed genes are enriched in functions including lung development (GO:0030324), and the top 20 significantly differentially expressed genes are enriched in several GO terms involved in alveolar development (Table S3). We also noticed that the differentially expressed genes between left and right lungs in adults (168 genes) are much fewer than that of embryos, emphasizing the importance of differentially expressed patterns of genes in development stages. Furthermore, we also observed that the identities (relative to the ancestral sequence) of some CNEs around the above potential asymmetric genes decreased with increasing lung asymmetry (Figure S4K). The identification of differentially expressed genes, associated CNEs, and losses of symmetric development-related genes provided insights into the asymmetric development of snake lungs. However, additional studies are needed to elucidate the underlying mechanism.

We performed weighted gene co-expression network analysis (WGCNA), and the results indicated that the genes related to lizard lung function were different from those in snakes. Three hub genes (*SFTPC*, *CHI3L1*, and *TITF1*) and four lung development-related genes (*SAV1*, *SEC24B*, *PDE4D*, and *WWP1*) were differentially expressed in snakes compared with the green anole (Figure 3C). Of these, *WWP1* was identified as a PSG; *SFTPC*, which was highly expressed in snake lung, was associated with SD-CNEs; and *TITF1*, *SAV1*, *SEC24B*, and *PDE4D* were identified



(legend on next page)

as SSSV-associated genes. The regulatory region of thyroid transcription factor 1 (*TITF1*), an important transcription factor responsible for lung saccule development,[81] contained an additional HOXA5-binding site via an SSSV insertion (Figure S4L). Furthermore, SAV1 (protein salvador homolog 1) regulates lung size by restricting lung epithelial cell differentiation,[82] and nuclear factor I-B (NFIB) can promote ciliated cell differentiation in the lung.[79] In addition, two respiratory gaseous exchange regulators (*NR4A2* and *PHX2A*) were identified as PSG and REG, respectively. Thus, the development of asymmetrical lungs might have led to alterations in the respiratory function of the lungs of snakes, including the evolution of an enhanced lung saccule, as well as ciliated and epithelial cell differentiation to maintain lung function.

Evolution of the sensory system

In addition to limb loss, the ability of snakes to perceive visual and auditory stimuli has decreased, and these changes are associated with a fossorial lifestyle.[83,84] Snakes are successful predators and have a keen ability to perceive chemical cues from their prey through their olfactory system. Snakes occur in various environments; their lifestyles range from fossorial to arboreal, and they have evolved various adaptations to enhance their fitness in diverse environments, such as vision loss and enhanced thermoreception. Understanding the genetic architecture of these adaptations will enhance our understanding of sensory system evolution in snakes and the processes underlying the evolution of new traits.

Visual adaptation

All extant snakes have distinct visual systems compared with other surface-living squamates. Eye reduction during the early evolution of snakes is similar to that observed in other squamates with reduced visual function.[85,86] Compared with lizards, snake eyes lack the large ciliary muscles, scleral ossicles, and scleral cartilage, and they have smaller conus papillaris, annular pads, and anterior pads.[86] These changes suggest that snakes have a weakened ability to focus images on the retina.[86] According to the comparative genomic analysis, we found that photoreceptor-related coding genes (e.g., *OPN4A*, *OPN3*, *OPSP*, *OPSB*, *GUC1A*, *CRY1*, *CRY2*, and *GRK1*) have been lost in all investigated snake species (Figure 4A; Table S3). Furthermore, the excitatory amino acid transporter 2 (*SLC1A2*) gene, which is important for the light stimulus response,[87] and three beta-crystallin genes (*CRBB3*, *CRBB1*, and *CRBA1*), which are important for lens development,[88] were also absent. The expression of 15 photoreceptor cell maintenance and photoreceptor cell component-related genes was significantly up-regulated in snake eyes compared with the eyes of green anoles (Figure 4A). Of these, the *ESRRB* gene had two SSSV deletions (Table S3), and six genes (*PDE6G*, *AGTPBP1*, two copies of *CLTC*s, *MKKS*, and *NXNL1*) were surrounded by SD-CNEs, indicating that regulatory elements may have evolved to regulate the expression of these genes in snakes. These findings indicate that early snakes might have undergone a low-light fossorial period before invading well-lit terrestrial habitats. If this, indeed, occurred, the structure and function of the eye must have adapted to well-lit habitats.

We identified five SSSVs (one insertion and four deletions) located in the CDSs of five eye development-related genes (*FAT1*, *TWIST2*, *CC2D2A*, *CACNA1C*, and *FBN1*) of which CC2D2A showed 32 snake-specific amino acid residue deletions (Figure S4M). Knockout of *CC2D2A* is reported to result in microphthalmia.[89] Here, 30 SSSVs (14 deletions and 16 insertions) were located within the potential regulatory regions of 22 eye development-related genes (e.g., *RP1*, *AKT3*, *FRIZZLED-5*, *KERA*, *BFSP2*, *CRYBG3*, *RORB*, *SDK2*, and *CPAMD8*) (Figure S3G; Table S3). We further sequenced the transcriptome of red cornsnake embryo's eyes (10, 30, and 50 dpo) and found that five eye development-related genes (*CLIC4*, *MAN2A1*, *CRYBG3*, *GRHL2*, and *CPAMD8*) with SSSVs are significantly lower expressed in snakes compared with that of human embryos. It is likely that the SSSVs have changed the expression of these genes during snake eye development (Figure S5A). Furthermore, SD-CNEs located around coding genes are significantly enriched in MGI terms related to eye development (Figure S3D). These findings suggest that SSSVs and SD-CNEs may have played a role in the evolution of snake eye morphology.

The eyelids protect the eyes of terrestrial vertebrates against desiccation. In snakes, however, the eyelids are fused into an immovable transparent scale (i.e., spectacle) over the cornea,[41,86] rendering them incapable of blinking. Our results showed that SD-CNE-annotated coding genes are enriched in eyelid development-related MGI terms (Figure S3D). Three snake-specific amino acid residue deletions and one nearby SD-CNE were observed in histone-lysine N-methyltransferase 2C (KMT2C), an important regulator of eyelid development (Figure S5B),[90] indicating that functional alterations in this protein might contribute to the snake-specific features of eyelid development. These results shed light on the potential genetic mechanism underpinning the evolution of fused eyelids in snakes.

Low-frequency hearing

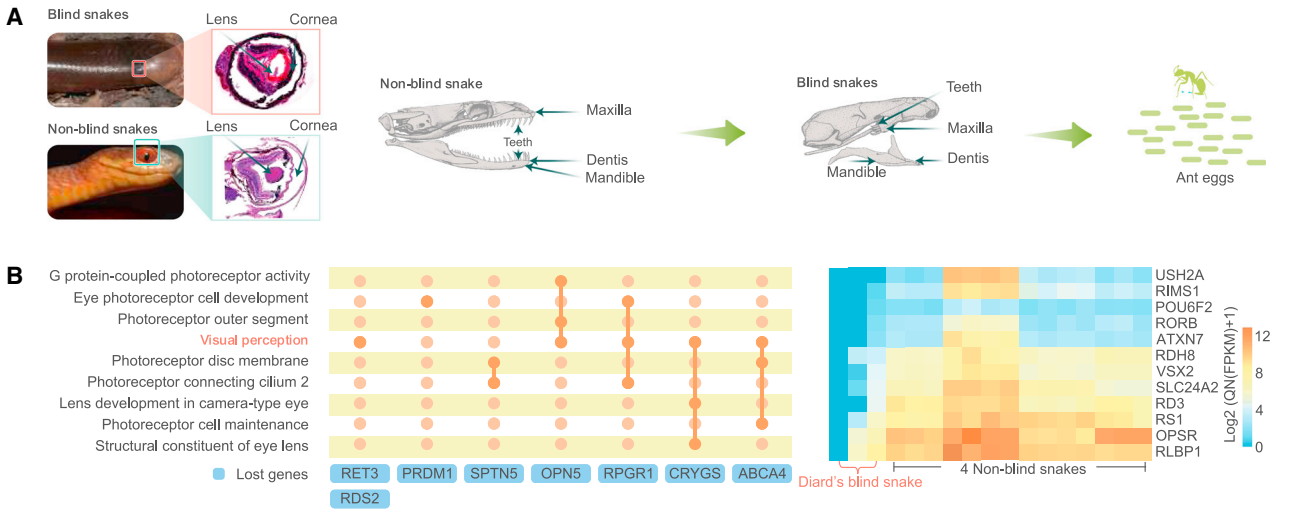
Snakes have lost external ears, middle ears, and tympanums, which are elaborate structures required to collect and transmit sound energy. Although snakes are not truly deaf, their sound frequency range is limited to 50–1,000 Hz,[41] which is much narrower than that of humans (20–20,000 Hz) and lizards

Figure 4. Genome features associated with snake-specific sense organ evolution, highlighting PSGs, REGs, newly evolved genes, SD-CNE-associated genes, lost genes, and SSSV-associated genes

(A) Lost and highly expressed genes in snake eyes. Diagram of the snake eye showing lost genes in blue. Heatmap of highly expressed genes (QN, quantile normalized).

(B) Diagram of the snake inner ear. Twelve SD-CNE-associated genes and four SSSV-associated genes are involved in ear development. Expression levels (scaled fragments per kilobase per million mapped reads [FPKM]) of sound perception-related PSGs and REGs in 10 keeled slug snake (Pber) tissues are shown in circles. Circle size is positively correlated to the expression level.

(C) Diagram of the taste transduction process. Solid lines indicate direct interactions, and dotted lines indicate indirect interactions. See also Table S3.



(legend on next page)

(<100–7,700 Hz).[91,92] However, the connection of the middle ear bone to the jawbone in snakes provides acute sensitivity to substrate vibrations (Figure 4B),[93] and the single-hair-cell type in snakes implies that functional changes in the inner ear have occurred to enhance their ability to sense vibrations.[94,95] Our results indicated that the rate of evolution of regulators that play important roles in the development of the ears is increased, and many coding genes (e.g., *FGFR1*, *OTX1*, and *MAFB*) are surrounded by SD-CNEs (Figure 4B). We also found that otospiralin (OTOS), which controls the sensory perception of sound,[88] has been lost in snakes (Table S3). Down-regulation of this protein causes hair cell degeneration and deafness in humans,[96] indicating that its absence may affect hair cell morphology and hearing. Moreover, the expression levels of three REGs (*ASIC2*, *POU3F4*, and *CCDC50*) and four PSGs (*ASIC2*, *S26A4*, *TMIE*, and *GRM7*) related to the sensory perception of sound are either up-regulated or normal in the inner ear (Figure 4B). Among all SSSVs, a total of 31 are located in potential regulatory regions of ear development-associated genes (e.g., *PDZD7*, *OTOP1*, *MYO15A*, and *MCOLN3*) (Table S3). *PDZD7* plays a key role in the development of auditory receptor cells and sound perception, the loss of which causes hearing loss in mice.[97] Here, we identified an EBF1-binding site in the regulatory region of *PDZD7*, which contains an SSSV (Figure S5C). Because EBF1 is an important transcription factor in inner ear development,[98] the newly evolved binding site may alter the expression of *PDZD7* during inner ear development in snakes. Our observations suggest that gene loss and the evolution of new regulators might contribute to inner ear specialization and low-frequency hearing.

Olfaction and taste

Despite their poor vision and hearing, snakes are excellent predators in many habitats because of their well-developed olfactory system. Enrichment analyses confirmed that many newly evolved genes in snakes are involved in olfactory receptor activity (Figure S3E), including *VMN2R26*, *VMN2R1*, *OR14A16*, and *OR14J1* (Table S3). One copy of *VMN2R26* shared by snakes, and the outgroup was a REG (Table S3). The olfactory signal amplifier UBR3[99] also showed signs of rapid evolution in snakes. These genomic innovations may underlie the evolution of the unique olfactory system of snakes.

Snakes swallow prey whole and consume large prey relative to their body mass,[41] which may have driven the evolution of their enhanced taste system to detect prey and their efficient digestive systems to digest large food items. Evolutionary analysis revealed several REGs involved in taste transduction (i.e., *TAS2R40*, *5HT1A*, and *GNG13*), and *5HT1A* is a PSG. A gene that regulates purinergic neurotransmission, *ENTPD2*, which is partially responsible for salty taste transduction, is a PSG (Figure 4C). Transcriptomic analysis indicated that all taste genes, except for *TAS2R40*, are expressed in the tongue and brain of at least three different snake families, which suggests that most of these receptors are still functional (Figure S5D). *PKD2L1*, *TAS1R2*s, and *TAS1R3* are also expressed in the tongues of snakes from two different families (Figure S5D). Whether snakes have a sense of taste is unclear.[100,101] However, our integrated data analyses suggest that the snake ancestor did have taste receptors and possibly had an enhanced sense of taste.

Specialized ant-eating adaptation and infra-red sensing

Ant-eating diet in blind snakes

Blind snakes are highly specialized burrowing snakes with small narrow heads, substantially reduced eyes and teeth, and a specialized diet (ant and ant egg).[3,4,84,102] According to comparative morphological analyses of the eyes of Diard's blind snake with those of two visually oriented predatory species (Asian vine snake and keeled slug snake), we found that the retina, lens, cornea, and spectacle of the Diard's blind snake were nearly fused, and the lens was loosely organized (Figures 5A and S5E).

To understand the genetic features that underpin these traits, we conducted comparative genomic and transcriptomic analyses and found that 41 genes (e.g., *RPGRIP1*, *RBP3*, *ABCA4*, and *PRDM1*) have been lost in the two studied blind snakes (Figure 5B; Table S4). These genes are significantly enriched in GO terms related to eye photoreceptor development and visual perception (Figure 5B; Table S4). Furthermore, the expression of 12 visual perception-related genes was significantly down-regulated in Diard's blind snake compared with that in four non-blind snake species (Figure 5B). We also found that the blind snake SD-CNEs (BSD-CNEs) are located in the potential regulatory regions of 39 eye development-related genes (Figure S5F). The gene regression and rapidly evolving regulatory elements may provide a genetic explanation for the evolution of "blindness" in these snakes.

Although blind snakes have reduced eye function, their well-developed olfactory systems allow pheromone trail-following for communication, reproduction, and foraging.[103] Evolutionary analyses revealed that the gene controlling lateral olfactory tract development (*SLIT2*)[104] is a PSG (Table S4). Furthermore, BSD-CNE-related genes are enriched in olfactory bulb development (Figure S5F). These results provide insights into the genetic basis of blindness-enhanced olfaction.

Figure 5. Evolution of blind and infrared-sensitive snakes

(A) Specialized traits of blind snakes.

(B) Gene regression underlies eye degeneration in blind snakes. Lost genes are shown in green. Heatmap showing the significant down-regulation of the expression of genes in blind snake eyes (QN, quantile normalized).

(C) BSD-CNE-associated genes enriched in odontogenesis and bone-development-related GO terms.

(D) Amino acids sequences alignment (left) and enzymatic activity (right) of CHIA. The amino acids in red are positively selected sites in blind snakes. The error bar represents mean ± SD of enzymatic activity. The enzymatic activity is significantly higher in Diard's blind snake (Student's t-test p value = 8.78e-09).

(E) Infrared sensing-related genes in infrared-sensitive snakes. PSGs, REGs, and variation-shared genes are indicated in purple, blue, and red, respectively. Average expression levels (represented by FPKM) of these genes in seven tissues of the brown-spotted pitviper are shown. See also Figure S5 and Tables S4 and S5.

Moreover, REGs involved in the growth of multicellular organisms and BSD-CNE-associated genes enriched in odontogenesis and bone development-related GO terms, both of which may explain the small head and reduced teeth in blind snake (Figure 5C). Regarding to the digestion of ant chitin, two positively selected sites in the GH18 domain of the acidic mammalian chitinase (CHIA)-coding gene, which is a well-known chitin and chitotriose decomposer,[105,106] were detected (Figure 5D; Table S4). Further experiments showed that the enzymatic activity of CHIA is approximately 40% higher in the Diard's blind snake than in the Burmese python (Figure 5D), suggesting that the two substitutions could contribute to improving the digestion of chitin in ants and ant eggs, as has been observed in insectivorous mammals.[107,108]

Convergent evolution of the infrared-sensing in snakes

Three lineages of snakes—pitvipers, pythons, and boas—are known to possess specialized temperature-sensitive infrared-sensing organs (e.g., pit organs and labial pits) with a trigeminal nerve.[5] These well-equipped predators perceive infrared radiation, which involves the *TRPA1* gene.[109] Evolutionary analyses revealed that these three lineages contained several PSGs and REGs involved in the heat response and pain detection (Figure 5E). In addition to *TRPA1*, an essential activator of the heat response, *MAP2K7*,[110] has evolved rapidly in the python, boa, and pitvipers. Moreover, comparative transcriptome analysis confirmed that *TRPV4* was highly expressed in the pit organ of the brown-spotted pitviper compared with other tissues (Figure S5H). Furthermore, *ISL1* is crucial for trigeminal nerve development,[111] and it has evolved rapidly in both boa constrictor and Burmese python (Table S5). In addition to coding genes, we further investigated the evolution of non-coding elements in infrared-sensitive snakes. Infrared-sensitive SD-CNEs were located around *PMP22* and *NFIB* (Figure S5I), which were also highly expressed in the pit organs of brown-spotted pitvipers (Figure 5E). *PMP22* plays a key role in cellular responses to heat,[112] and *NFIB* plays a key role in the principal sensory nucleus of the trigeminal nerve.[113] These results indicate that both coding and non-coding elements may play an important role in infrared perception.

DISCUSSION

The assembly of high-quality snake genomes has the potential to enhance our understanding of the genetic basis underpinning the evolution of various snake phenotypes. Our evolutionary and comparative multi-omics analyses revealed many genetic variants correlated with specific traits in snakes. In addition, we found several candidate and target genetic elements for future studies of developmental biology. Aside from the reference-quality genomic sequences, the results of our study provide insights into snake evolution and diversification and enhance our knowledge of reptile biology and developmental biology.

Limitations of the study

Studies aimed at characterizing genetic changes in the genome, especially in non-model species, such as snakes, face major challenges in determining whether certain changes are directly responsible for corresponding changes in traits. First, variation in a trait may be determined by different genetic changes, which makes it difficult to determine the primary cause. Second, although functional experiments can narrow down the list of candidate changes, such experiments are difficult to perform on snakes and lizards. Although we could have attempted to conduct experiments by precisely simulating the naturally occurring mutations in model organisms using CRISPR-Cas9, these experiments are either time-consuming, sometimes unreliable, especially when there are large differences in the genetic backgrounds among distantly related species, or difficult to perform when the number of potential changes is large. Despite these limitations, our study identified several genetic changes that might underlie evolutionary changes in the phenotypes of snakes, and our findings will aid future studies aimed at clarifying various aspects of the evolution and development of snakes, reptiles, and vertebrates.

STAR★METHODS

Detailed methods are provided in the online version of this paper and include the following:

- KEY RESOURCES TABLE
- RESOURCE AVAILABILITY
  - Lead contact
  - Materials availability
  - Data and code availability
- EXPERIMENTAL MODEL AND STUDY PARTICIPANT DETAILS
  - Source organisms
- METHOD DETAILS
  - Selection of species
  - Genome sequencing
  - Genome size estimation and assembly
  - Genome quality evaluation
  - RNA extraction and sequencing
  - Genome annotation
  - Chromosome evolution analysis
  - Progressive Cactus alignment
  - Analysis of conserved non-coding elements (CNEs)
  - Phylogenetic tree construction
  - Divergence time estimation
  - Identification of structural variants
  - Orthologous gene identification
  - Identification of newly evolved genes
  - Lost gene identification
  - Positively selected gene (PSG) identification
  - Rapidly evolving gene (REG) identification
  - Relaxed selection analysis
  - Transcriptomic analysis of genes related to snake traits
  - Blind snake adaptive evolution analysis
  - Evolutionary analysis of infrared sensing
  - Functional experiment of the *PTCH1* gene
  - Chitinase enzymatic activity measurement
- QUANTIFICATION AND STATISTICAL ANALYSIS

SUPPLEMENTAL INFORMATION

Supplemental information can be found online at https://doi.org/10.1016/j.cell.2023.05.030.

ACKNOWLEDGMENTS

We thank Peng Shi and Yong Shao (Kunming Institute of Zoology, Chinese Academy of Sciences) for the suggestions on data analysis and figure revisions. This study was supported by the Strategic Priority Research Program of Chinese Academy of Sciences (CAS) (XDB31000000); the National Natural Science Foundation of China (32220103004; 32000296); the Second Tibetan Plateau Scientific Expedition and Research Program (STEP) (2019QZKK0501); the International Partnership Program of Chinese Academy of Sciences (151751KYSB20190024); the Sichuan Science and Technology Program (2021JDJQ0002).

AUTHOR CONTRIBUTIONS

J.-T.L. initiated and designed the study. C.P., D.-D.W., J.-L.R., Z.-L.P., Z.M., W.W., Y.L., Z.W., C.D., K.J., and Y.Q. conducted this work. C.P. and J.-L.R. wrote the manuscript. J.-T.L., D.-D.W., C.L.P., and Z.-Y.Z. edited the manuscript. All authors read and approved the final manuscript.

DECLARATION OF INTERESTS

The authors declare no competing interests.

Received: August 23, 2022
Revised: April 6, 2023
Accepted: May 19, 2023
Published: June 19, 2023

REFERENCES

1. Zug, G.R., Vitt, L., and Caldwell, J.P. (2001). Herpetology: an Introductory Biology of Amphibians and Reptiles (Elsevier/Academic Press).
2. Uetz, P. (2022). The reptile database. Zoological Museum Hamburg. http://reptile-database.org/.
3. Avila, R.W., Ferreira, V.L., and Souza, V.B. (2006). Biology of the blindsnake *Typhlops brongersmianus* (Typhlopidae) in a semideciduous forest from central Brazil. Herpetol. J. *16*, 403–405.
4. Mizuno, T., and Kojima, Y. (2015). A blindsnake that decapitates its termite prey. J. Zool. *297*, 220–224.
5. Goris, R.C. (2011). Infrared organs of snakes: an integral part of vision. J. Herpetol. *45*, 2–14.
6. Darbaniyan, F., Mozaffari, K., Liu, L., and Sharma, P. (2021). Soft matter mechanics and the mechanisms underpinning the infrared vision of snakes. Matter *4*, 241–252.
7. Margres, M.J., Rautsaw, R.M., Strickland, J.L., Mason, A.J., Schramer, T.D., Hofmann, E.P., Stiers, E., Ellsworth, S.A., Nystrom, G.S., Hogan, M.P., et al. (2021). The tiger rattlesnake genome reveals a complex genotype underlying a simple venom phenotype. Proc. Natl. Acad. Sci. USA *118*, e2014634118.
8. Myers, E.A., Strickland, J.L., Rautsaw, R.M., Mason, A.J., Schramer, T.D., Nystrom, G.S., Hogan, M.P., Yooseph, S., Rokyta, D.R., and Parkinson, C.L. (2022). De novo genome assembly highlights the role of lineage-specific gene duplications in the evolution of venom in fea's viper (*Azemiops feae*). Genome Biol. Evol. *14*, evac082.
9. Yin, W., Wang, Z.-J., Li, Q.-Y., Lian, J.-M., Zhou, Y., Lu, B.-Z., Jin, L.-J., Qiu, P.-X., Zhang, P., Zhu, W.-B., et al. (2016). Evolutionary trajectories of snake genes and genomes revealed by comparative analyses of five-pacer viper. Nat. Commun. *7*, 13107.
10. Vonk, F.J., Casewell, N.R., Henkel, C.V., Heimberg, A.M., Jansen, H.J., McCleary, R.J., Kerkkamp, H.M., Vos, R.A., Guerreiro, I., Calvete, J.J., et al. (2013). The king cobra genome reveals dynamic gene evolution and adaptation in the snake venom system. Proc. Natl. Acad. Sci. USA *110*, 20651–20656.
11. Castoe, T.A., De Koning, A.P.J., Hall, K.T., Card, D.C., Schield, D.R., Fujita, M.K., Ruggiero, R.P., Degner, J.F., Daza, J.M., Gu, W., et al. (2013). The Burmese python genome reveals the molecular basis for extreme adaptation in snakes. Proc. Natl. Acad. Sci. USA *110*, 20645–20650.
12. Li, J.-T., Gao, Y.-D., Xie, L., Deng, C., Shi, P., Guan, M.-L., Huang, S., Ren, J.L., Wu, D.-D., Ding, L., et al. (2018). Comparative genomic investigation of high-elevation adaptation in ectothermic snakes. Proc. Natl. Acad. Sci. USA *115*, 8406–8411.
13. Schield, D.R., Card, D.C., Hales, N.R., Perry, B.W., Pasquesi, G.M., Blackmon, H., Adams, R.H., Corbin, A.B., Smith, C.F., Ramesh, B., et al. (2019). The origins and evolution of chromosomes, dosage compensation, and mechanisms underlying venom regulation in snakes. Genome Res. *29*, 590–601.
14. Yan, C., Wu, W., Dong, W., Zhu, B., Chang, J., Lv, Y., Yang, S., and Li, J.T. (2022). Temperature acclimation in hot-spring snakes and the convergence of cold response. Innovation (Camb) *3*, 100295.
15. Suryamohan, K., Krishnankutty, S.P., Guillory, J., Jevit, M., Schröder, M.S., Wu, M., Kuriakose, B., Mathew, O.K., Perumal, R.C., Koludarov, I., et al. (2020). The Indian cobra reference genome and transcriptome enables comprehensive identification of venom toxins. Nat. Genet. *52*, 106–117.
16. Peng, C., Ren, J.-L., Deng, C., Jiang, D., Wang, J., Qu, J., Chang, J., Yan, C., Jiang, K., Murphy, R.W., et al. (2020). The genome of Shaw's sea snake (*Hydrophis curtus*) reveals secondary adaptation to its marine environment. Mol. Biol. Evol. *37*, 1744–1760.
17. Armstrong, J., Hickey, G., Diekhans, M., Fiddes, I.T., Novak, A.M., Deran, A., Fang, Q., Xie, D., Feng, S., Stiller, J., et al. (2020). Progressive Cactus is a multiple-genome aligner for the thousand-genome era. Nature *587*, 246–251.
18. Zhang, C., Rabiee, M., Sayyari, E., and Mirarab, S. (2018). ASTRAL-III: polynomial time species tree reconstruction from partially resolved gene trees. BMC Bioinformatics *19* (*Supplement 6*), 153.
19. Figueroa, A., McKelvy, A.D., Grismer, L.L., Bell, C.D., and Lailvaux, S.P. (2016). A species-level phylogeny of extant snakes with description of a new colubrid subfamily and genus. PLoS One *11*, e0161070.
20. Zaher, H., Murphy, R.W., Arredondo, J.C., Graboski, R., Machado-Filho, P.R., Mahlow, K., Montingelli, G.G., Quadros, A.B., Orlov, N.L., Wilkinson, M., et al. (2019). Large-scale molecular phylogeny, morphology, divergence-time estimation, and the fossil record of advanced caenophidian snakes (Squamata: Serpentes). PLoS One *14*, e0216148.
21. Hsiang, A.Y., Field, D.J., Webster, T.H., Behlke, A.D., Davis, M.B., Racicot, R.A., and Gauthier, J.A. (2015). The origin of snakes: revealing the ecology, behavior, and evolutionary history of early snakes using genomics, phenomics, and the fossil record. BMC Evol. Biol. *15*, 87.
22. Klein, C.G., Pisani, D., Field, D.J., Lakin, R., Wills, M.A., and Longrich, N.R. (2021). Evolution and dispersal of snakes across the Cretaceous-Paleogene mass extinction. Nat. Commun. *12*, 5335.
23. Zheng, Y., and Wiens, J.J. (2016). Combining phylogenomic and supermatrix approaches, and a time-calibrated phylogeny for squamate reptiles (lizards and snakes) based on 52 genes and 4162 species. Mol. Phylogenet. Evol. *94*, 537–547.
24. Scotese, C.R. (2001). Atlas of Earth History (University of Texas at Arlington, Department of Geology).
25. Roscito, J.G., Sameith, K., Pippel, M., Francoijs, K.-J., Winkler, S., Dahl, A., Papoutsoglou, G., Myers, G., and Hiller, M. (2018). The genome of the tegu lizard *Salvator merianae*: combining Illumina, PacBio, and optical mapping data to generate a highly contiguous assembly. GigaScience *7*, giy141.
26. Lind, A.L., Lai, Y.Y.Y., Mostovoy, Y., Holloway, A.K., Iannucci, A., Mak, A.C.Y., Fondi, M., Orlandini, V., Eckalbar, W.L., Milan, M., et al. (2019). Genome of the Komodo dragon reveals adaptations in the cardiovascular and chemosensory systems of monitor lizards. Nat. Ecol. Evol. *3*, 1241–1252.
27. Shaffer, H.B., Minx, P., Warren, D.E., Shedlock, A.M., Thomson, R.C., Valenzuela, N., Abramyan, J., Amemiya, C.T., Badenhorst, D., Biggar,

K.K., et al. (2013). The western painted turtle genome, a model for the evolution of extreme physiological adaptations in a slowly evolving lineage. Genome Biol. *14*, R28.

28. Wan, Q.-H., Pan, S.-K., Hu, L., Zhu, Y., Xu, P.-W., Xia, J.-Q., Chen, H., He, G.-Y., He, J., Ni, X.-W., et al. (2013). Genome analysis and signature discovery for diving and sensory properties of the endangered Chinese alligator. Cell Res. *23*, 1091–1105.
29. Hillier, L.W., Miller, W., Birney, E., Warren, W., Hardison, R.C., Ponting, C.P., Bork, P., Burt, D.W., Groenen, M.A.M., Delany, M.E., et al. (2004). Sequence and comparative analysis of the chicken genome provide unique perspectives on vertebrate evolution. Nature *432*, 695–716.
30. Pasquesi, G.I.M., Adams, R.H., Card, D.C., Schield, D.R., Corbin, A.B., Perry, B.W., Reyes-Velasco, J., Ruggiero, R.P., Vandewege, M.W., Shortt, J.A., et al. (2018). Squamate reptiles challenge paradigms of genomic repeat element evolution set by birds and mammals. Nat. Commun. *9*, 2774.
31. Rhie, A., McCarthy, S.A., Fedrigo, O., Damas, J., Formenti, G., Koren, S., Uliano-Silva, M., Chow, W., Fungtammasan, A., Kim, J., et al. (2021). Towards complete and error-free genome assemblies of all vertebrate species. Nature *592*, 737–746.
32. Weber, C.C., Boussau, B., Romiguier, J., Jarvis, E.D., and Ellegren, H. (2014). Evidence for GC-biased gene conversion as a driver of between-lineage differences in avian base composition. Genome Biol. *15*, 549.
33. Duret, L., and Galtier, N. (2009). Biased gene conversion and the evolution of mammalian genomic landscapes. Annu. Rev. Genomics Hum. Genet. *10*, 285–311.
34. Sanderson, M.J. (2003). R8s: inferring absolute rates of molecular evolution and divergence times in the absence of a molecular clock. Bioinformatics *19*, 301–302.
35. Chen, L., Qiu, Q., Jiang, Y., Wang, K., Lin, Z., Li, Z., Bibi, F., Yang, Y., Wang, J., Nie, W., et al. (2019). Large-scale ruminant genome sequencing provides insights into their evolution and distinct traits. Science *364*, eaav6202.
36. Roscito, J.G., Sameith, K., Parra, G., Langer, B.E., Petzold, A., Moebius, C., Bickle, M., Rodrigues, M.T., and Hiller, M. (2018). Phenotype loss is associated with widespread divergence of the gene regulatory landscape in evolution. Nat. Commun. *9*, 4737.
37. Pyron, R.A., Burbrink, F.T., and Wiens, J.J. (2013). A phylogeny and revised classification of Squamata, including 4161 species of lizards and snakes. BMC Evol. Biol. *13*, 93.
38. Morinaga, G., and Bergmann, P.J. (2020). Evolution of fossorial locomotion in the transition from tetrapod to snake-like in lizards. Proc. Biol. Sci. *287*, 20200192.
39. Pees, M., Kiefer, I., Thielebein, J., Oechtering, G., and Krautwald-Junghanns, M.E. (2009). Computed tomography of the lung of healthy snakes of the species *Python regius*, *boa constrictor*, *Python reticulatus*, *Morelia viridis*, *Epicrates cenchria*, and *Morelia spilota*. Vet. Radiol. Ultrasound *50*, 487–491.
40. van Soldt, B.J., Metscher, B.D., Poelmann, R.E., Vervust, B., Vonk, F.J., Müller, G.B., and Richardson, M.K. (2015). Heterochrony and early left-right asymmetry in the development of the cardiorespiratory system of snakes. PLoS One *10*, e116416.
41. Lillywhite, H.B. (2014). How Snakes Work: Structure, Function and Behavior of the World's Snakes (Oxford University Press).
42. Cohn, M.J., and Tickle, C. (1999). Developmental basis of limblessness and axial patterning in snakes. Nature *399*, 474–479.
43. Apesteguía, S., and Zaher, H. (2006). A Cretaceous terrestrial snake with robust hindlimbs and a sacrum. Nature *440*, 1037–1040.
44. Kvon, E.Z., Kamneva, O.K., Melo, U.S., Barozzi, I., Osterwalder, M., Mannion, B.J., Tissières, V., Pickle, C.S., Plajzer-Frick, I., Lee, E.A., et al. (2016). Progressive loss of function in a limb enhancer during snake Evolution. Cell *167*, 633–642.e11.
45. Sandell, L.L., Sanderson, B.W., Moiseyev, G., Johnson, T., Mushegian, A., Young, K., Rey, J.P., Ma, J.X., Staehling-Hampton, K., and Trainor, P.A. (2007). RDH10 is essential for synthesis of embryonic retinoic acid and is required for limb, craniofacial, and organ development. Genes Dev. *21*, 1113–1124.
46. Makino, S., Masuya, H., Ishijima, J., Yada, Y., and Shiroishi, T. (2001). A Spontaneous mouse mutation, mesenchymal dysplasia (*mes*), is caused by a deletion of the most C-terminal cytoplasmic domain of patched (*ptc*). Dev. Biol. *239*, 95–106.
47. Goodrich, L.V., Milenković, L., Higgins, K.M., and Scott, M.P. (1997). Altered neural cell fates and medulloblastoma in mouse patched mutants. Science *277*, 1109–1113.
48. Damerla, R.R., Cui, C., Gabriel, G.C., Liu, X., Craige, B., Gibbs, B.C., Francis, R., Li, Y., Chatterjee, B., San Agustin, J.T., et al. (2015). Novel Jbts17 mutant mouse model of Joubert syndrome with cilia transition zone defects and cerebellar and other ciliopathy related anomalies. Hum. Mol. Genet. *24*, 3994–4005.
49. Chidambaram, A., Goldstein, A.M., Gailani, M.R., Gerrard, B., Bale, S.J., DiGiovanna, J.J., Bale, A.E., and Dean, M. (1996). Mutations in the human homologue of the Drosophila *patched* gene in Caucasian and African-American nevoid basal cell carcinoma syndrome patients. Cancer Res. *56*, 4599–4601.
50. Palmer, K., Fairfield, H., Borgeia, S., Curtain, M., Hassan, M.G., Dionne, L., Yong Karst, S.Y., Coombs, H., Bronson, R.T., Reinholdt, L.G., et al. (2016). Discovery and characterization of spontaneous mouse models of craniofacial dysmorphology. Dev. Biol. *415*, 216–227.
51. van den Boogaard, M.-J.H., Dorland, M., Beemer, F.A., and van Amstel, H.K.P. (2000). *MSX1* mutation is associated with orofacial clefting and tooth agenesis in humans. Nat. Genet. *24*, 342–343.
52. Vastardis, H., Karimbux, N., Guthua, S.W., Seidman, J.G., and Seidman, C.E. (1996). A human MSX1 homeodomain missense mutation causes selective tooth agenesis. Nat. Genet. *13*, 417–421.
53. Secor, S.M., and Diamond, J. (1995). Adaptive responses to feeding in Burmese pythons: pay before pumping. J. Exp. Biol. *198*, 1313–1325.
54. Secor, S.M., and Diamond, J. (1998). A vertebrate model of extreme physiological regulation. Nature *395*, 659–662.
55. Wang, T., and Rindom, E. (2021). The physiological response to digestion in snakes: A feast for the integrative physiologist. Comp. Biochem. Physiol. A Mol. Integr. Physiol. *254*, 110891.
56. Perry, B.W., Andrew, A.L., Mostafa Kamal, A.H., Card, D.C., Schield, D.R., Pasquesi, G.I.M., Pellegrino, M.W., Mackessy, S.P., Chowdhury, S.M., Secor, S.M., et al. (2019). Multi-species comparisons of snakes identify coordinated signalling networks underlying post-feeding intestinal regeneration. Proc. Biol. Sci. *286*, 20190910.
57. Kitazawa, T., and Kaiya, H. (2019). Regulation of gastrointestinal motility by motilin and ghrelin in vertebrates. Front. Endocrinol. *10*, 278.
58. Ryu, S., Huh, I.-S., Cho, E.-Y., Cho, Y., Park, T., Yoon, S.C., Joo, Y.H., and Hong, K.S. (2016). Association study of 60 candidate genes with antipsychotic-induced weight gain in schizophrenia patients. Pharmacopsychiatry *49*, 51–56.
59. Schroeter, J.C., Fenn, C.M., and Small, B.C. (2015). Elucidating the roles of gut neuropeptides on channel catfish feed intake, glycemia, and hypothalamic NPY and POMC expression. Comp. Biochem. Physiol. A Mol. Integr. Physiol. *188*, 168–174.
60. Costantini, V.J.A., Vicentini, E., Sabbatini, F.M., Valerio, E., Lepore, S., Tessari, M., Sartori, M., Michielin, F., Melotto, S., Bifone, A., et al. (2011). GSK1614343, a novel ghrelin receptor antagonist, produces an unexpected increase of food intake and body weight in rodents and dogs. Neuroendocrinology *94*, 158–168.
61. Vergnes, L., Lee, J.M., Chin, R.G., Auwerx, J., and Reue, K. (2013). Diet1 functions in the FGF15/19 enterohepatic signaling axis to modulate bile acid and lipid levels. Cell Metab. *17*, 916–928.

62. Zhang, H., Ables, E.T., Pope, C.F., Washington, M.K., Hipkens, S., Means, A.L., Path, G., Seufert, J., Costa, R.H., Leiter, A.B., et al. (2009). Multiple, temporal-specific roles for HNF6 in pancreatic endocrine and ductal differentiation. Mech. Dev. *126*, 958–973.

63. MacDonald, R.J., Stary, S.J., and Swift, G.H. (1982). Two similar but nonallelic rat pancreatic trypsinogens. Nucleotide sequences of the cloned cDNAs. J. Biol. Chem. *257*, 9724–9732.

64. Boot, R.G., Verhoek, M., Donker-Koopman, W., Strijland, A., Van Marle, J., Overkleeft, H.S., Wennekes, T., and Aerts, J.M.F.G. (2007). Identification of the non-lysosomal glucosylceramidase as β-glucosidase 2. J. Biol. Chem. *282*, 1305–1312.

65. Toomey, C.B., Kelly, U., Saban, D.R., and Bowes Rickman, C.B. (2015). Regulation of age-related macular degeneration-like pathology by complement factor H. Proc. Natl. Acad. Sci. USA *112*, E3040–E3049.

66. Leal, F., and Cohn, M.J. (2018). Developmental, genetic, and genomic insights into the evolutionary loss of limbs in snakes. Genesis *56*, e23077.

67. Choorapoikayil, S., Willems, B., Ströhle, P., and Gajewski, M. (2012). Analysis of *her1* and *her7* mutants reveals a spatio temporal separation of the somite clock module. PLoS One *7*, e39073.

68. Van Eeden, F.J., Granato, M., Schach, U., Brand, M., Furutani-Seiki, M., Haffter, P., Hammerschmidt, M., Heisenberg, C.-P., Jiang, Y.-J., Kane, D.A., et al. (1996). Mutations affecting somite formation and patterning in the zebrafish, *Danio rerio*. Development *123*, 153–164.

69. Kume, T., Jiang, H., Topczewska, J.M., and Hogan, B.L. (2001). The murine winged helix transcription factors, Foxc1 and Foxc2, are both required for cardiovascular development and somitogenesis. Genes Dev. *15*, 2470–2482.

70. Skuntz, S., Mankoo, B., Nguyen, M.-T.T., Hustert, E., Nakayama, A., Tournier-Lasserve, E., Wright, C.V.E., Pachnis, V., Bharti, K., and Arnheiter, H. (2009). Lack of the mesodermal homeodomain protein MEOX1 disrupts sclerotome polarity and leads to a remodeling of the cranio-cervical joints of the axial skeleton. Dev. Biol. *332*, 383–395.

71. Mankoo, B.S., Skuntz, S., Harrigan, I., Grigorieva, E., Candia, A., Wright, C.V.E., Arnheiter, H., and Pachnis, V. (2003). The concerted action of Meox homeobox genes is required upstream of genetic pathways essential for the formation, patterning and differentiation of somites. Development *130*, 4655–4664.

72. Kawamura, A., Koshida, S., Hijikata, H., Ohbayashi, A., Kondoh, H., and Takada, S. (2005). Groucho-associated transcriptional repressor ripply1 is required for proper transition from the presomitic mesoderm to somites. Dev. Cell *9*, 735–744.

73. Chan, T., Kondow, A., Hosoya, A., Hitachi, K., Yukita, A., Okabayashi, K., Nakamura, H., Ozawa, H., Kiyonari, H., Michiue, T., et al. (2007). *Ripply2* is essential for precise somite formation during mouse early development. FEBS Lett. *581*, 2691–2696.

74. Supp, D.M., Brueckner, M., Kuehn, M.R., Witte, D.P., Lowe, L.A., McGrath, J., Corrales, J., and Potter, S.S. (1999). Targeted deletion of the ATP binding domain of left-right dynein confirms its role in specifying development of left-right asymmetries. Development *126*, 5495–5504.

75. Tian, T., Zhao, L., Zhang, M., Zhao, X., and Meng, A. (2009). Both *foxj1a* and *foxj1b* are implicated in left-right asymmetric development in zebrafish embryos. Biochem. Biophys. Res. Commun. *380*, 537–542.

76. Wallmeier, J., Frank, D., Shoemark, A., Nöthe-Menchen, T., Cindric, S., Olbrich, H., Loges, N.T., Aprea, I., Dougherty, G.W., Pennekamp, P., et al. (2019). De novo mutations in *FOXJ1* result in a motile ciliopathy with hydrocephalus and randomization of left/right body asymmetry. Am. J. Hum. Genet. *105*, 1030–1039.

77. Pelletier, G.J., Brody, S.L., Liapis, H., White, R.A., and Hackett, B.P. (1998). A human forkhead/winged-helix transcription factor expressed in developing pulmonary and renal epithelium. Am. J. Physiol. *274*, L351–L359.

78. Lyerla, T.A., Rusiniak, M.E., Borchers, M., Jahreis, G., Tan, J., Ohtake, P., Novak, E.K., and Swank, R.T. (2003). Aberrant lung structure, composition, and function in a murine model of Hermansky-Pudlak syndrome. Am. J. Physiol. Lung Cell. Mol. Physiol. *285*, L643–L653.

79. Hsu, Y.-C., Osinski, J., Campbell, C.E., Litwack, E.D., Wang, D., Liu, S., Bachurski, C.J., and Gronostajski, R.M. (2011). Mesenchymal nuclear factor I B regulates cell proliferation and epithelial differentiation during lung maturation. Dev. Biol. *354*, 242–252.

80. Mahlapuu, M., Enerbäck, S., and Carlsson, P. (2001). Haploinsufficiency of the forkhead gene *Foxf1*, a target for sonic hedgehog signaling, causes lung and foregut malformations. Development *128*, 2397–2406.

81. Zhang, Y., Rath, N., Hannenhalli, S., Wang, Z., Cappola, T., Kimura, S., Atochina-Vasserman, E., Lu, M.M., Beers, M.F., and Morrisey, E.E. (2007). GATA and Nkx factors synergistically regulate tissue-specific gene expression and development in vivo. Development *134*, 189–198.

82. Lee, J.H., Kim, T.S., Yang, T.H., Koo, B.K., Oh, S.P., Lee, K.P., Oh, H.J., Lee, S.H., Kong, Y.Y., Kim, J.M., et al. (2008). A crucial role of WW45 in developing epithelial tissues in the mouse. EMBO J. *27*, 1231–1242.

83. Yi, H., and Norell, M.A. (2015). The burrowing origin of modern snakes. Sci. Adv. *1*, e1500743.

84. Simões, B.F., Sampaio, F.L., Jared, C., Antoniazzi, M.M., Loew, E.R., Bowmaker, J.K., Rodriguez, A., Hart, N.S., Hunt, D.M., Partridge, J.C., et al. (2015). Visual system evolution and the nature of the ancestral snake. J. Evol. Biol. *28*, 1309–1320.

85. Senn, D.G., and Northcutt, R.G. (1973). The forebrain and midbrain of some squamates and their bearing on the origin of snakes. J. Morphol. *140*, 135–151.

86. Caprette, C.L., Lee, M.S.Y., Shine, R., Mokany, A., and Downhower, J.F. (2004). The origin of snakes (Serpentes) as seen through eye anatomy. Biol. J. Linn. Soc. *81*, 469–482.

87. Harada, T., Harada, C., Watanabe, M., Inoue, Y., Sakagawa, T., Nakayama, N., Sasaki, S., Okuyama, S., Watase, K., Wada, K., et al. (1998). Functions of the two glutamate transporters GLAST and GLT-1 in the retina. Proc. Natl. Acad. Sci. USA *95*, 4663–4666.

88. Gaudet, P., Livstone, M.S., Lewis, S.E., and Thomas, P.D. (2011). Phylogenetic-based propagation of functional annotations within the Gene Ontology consortium. Brief. Bioinform. *12*, 449–462.

89. Veleri, S., Manjunath, S.H., Fariss, R.N., May-Simera, H., Brooks, M., Foskett, T.A., Gao, C., Longo, T.A., Liu, P., Nagashima, K., et al. (2014). Ciliopathy-associated gene *Cc2d2a* promotes assembly of subdistal appendages on the mother centriole during cilia biogenesis. Nat. Commun. *5*, 4207.

90. Lee, S., Lee, D.-K., Dou, Y., Lee, J., Lee, B., Kwak, E., Kong, Y.-Y., Lee, S.-K., Roeder, R.G., and Lee, J.W. (2006). Coactivator as a target gene specificity determinant for histone H3 lysine 4 methyltransferases. Proc. Natl. Acad. Sci. USA *103*, 15392–15397.

91. Fettiplace, R. (1987). Electrical tuning of hair cells in the inner ear. Trends Neurosci. *10*, 421–425.

92. Manley, G.A. (2002). Evolution of structure and function of the hearing organ of lizards. J. Neurobiol. *53*, 202–211.

93. Christensen, C.B., Christensen-Dalsgaard, J., Brandt, C., and Madsen, P.T. (2012). Hearing with an atympanic ear: good vibration and poor sound–pressure detection in the royal python, *Python regius*. J. Exp. Biol. *215*, 331–342.

94. Miller, M.R. (1978). Scanning electron microscope studies of the papilla basilaris of some turtles and snakes. Am. J. Anat. *151*, 409–435.

95. Miller, M.R., and Beck, J. (1990). Further serial transmission electron microscopy studies of auditory hair cell innervation in lizards and in a snake. Am. J. Anat. *188*, 175–184.

96. Delprat, B., Boulanger, A., Wang, J., Beaudoin, V., Guitton, M.J., Ventéo, S., Dechesne, C.J., Pujol, R., Lavigne-Rebillard, M., Puel, J.-L., et al. (2002). Downregulation of otospiralin, a novel inner ear protein, causes hair cell degeneration and deafness. J. Neurosci. *22*, 1718–1725.

97. Zou, J., Zheng, T., Ren, C., Askew, C., Liu, X.-P., Pan, B., Holt, J.R., Wang, Y., and Yang, J. (2014). Deletion of *PDZD7* disrupts the Usher

syndrome type 2 protein complex in cochlear hair cells and causes hearing loss in mice. Hum. Mol. Genet. *23*, 2374–2390.

98. Sajan, S.A., Rubenstein, J.L.R., Warchol, M.E., and Lovett, M. (2011). Identification of direct downstream targets of Dlx5 during early inner ear development. Hum. Mol. Genet. *20*, 1262–1273.
99. Tasaki, T., Sohr, R., Xia, Z., Hellweg, R., Hörtnagl, H., Varshavsky, A., and Kwon, Y.T. (2007). Biochemical and genetic studies of UBR3, a ubiquitin ligase with a function in olfactory and other sensory systems. J. Biol. Chem. *282*, 18510–18520.
100. Kroll, J.C. (1973). Taste buds in the oral epithelium of the blind snake, *Leptotyphlops dulcis* (Reptilia: Leptotyphlopidae). Southwest. Nat. *17*, 365–370.
101. Emerling, C.A. (2017). Genomic regression of claw keratin, taste receptor and light-associated genes provides insights into biology and evolutionary origins of snakes. Mol. Phylogenet. Evol. *115*, 40–49.
102. Kley, N.J. (2006). Morphology of the lower jaw and suspensorium in the Texas blindsnake, *Leptotyphlops dulcis* (Scolecophidia: Leptotyphlopidae). J. Morphol. *267*, 494–515.
103. Gehlbach, F.R., Watkins, J.F., and Kroll, J.C. (1971). Pheromone trail-following studies of typhlopid, leptotyphlopid, and colubrid snakes. Behaviour *40*, 282–294.
104. Nguyen-Ba-Charvet, K.T., Plump, A.S., Tessier-Lavigne, M., and Chédotal, A. (2002). Slit1 and slit2 proteins control the development of the lateral olfactory tract. J. Neurosci. *22*, 5473–5480.
105. Boot, R.G., Blommaart, E.F., Swart, E., Ghauharali-van der Vlugt, K., Bijl, N., Moe, C., Place, A., and Aerts, J.M. (2001). Identification of a novel acidic mammalian chitinase distinct from chitotriosidase. J. Biol. Chem. *276*, 6770–6778.
106. Hartl, D., He, C.H., Koller, B., Da Silva, C.A., Homer, R., Lee, C.G., and Elias, J.A. (2008). Acidic mammalian chitinase is secreted via an ADAM17/epidermal growth factor receptor-dependent pathway and stimulates chemokine production by pulmonary epithelial cells. J. Biol. Chem. *283*, 33472–33482.
107. Janiak, M.C., Chaney, M.E., and Tosi, A.J. (2018). Evolution of acidic mammalian chitinase genes (*CHIA*) is related to body mass and insectivory in primates. Mol. Biol. Evol. *35*, 607–622.
108. Tabata, E., Itoigawa, A., Koinuma, T., Tayama, H., Kashimura, A., Sakaguchi, M., Matoska, V., Bauer, P.O., and Oyama, F. (2022). Noninsect-based diet leads to structural and functional changes of acidic chitinase in carnivora. Mol. Biol. Evol. *39*, msab331.
109. Gracheva, E.O., Ingolia, N.T., Kelly, Y.M., Cordero-Morales, J.F., Hollopeter, G., Chesler, A.T., Sánchez, E.E., Perez, J.C., Weissman, J.S., and Julius, D. (2010). Molecular basis of infrared detection by snakes. Nature *464*, 1006–1011.
110. Foltz, I.N., Gerl, R.E., Wieler, J.S., Luckach, M., Salmon, R.A., and Schrader, J.W. (1998). Human mitogen-activated protein kinase kinase 7 (MKK7) is a highly conserved c-Jun N-terminal kinase/stress-activated protein kinase (JNK/SAPK) activated by environmental stresses and physiological stimuli. J. Biol. Chem. *273*, 9344–9351.
111. Sun, Y., Dykes, I.M., Liang, X., Eng, S.R., Evans, S.M., and Turner, E.E. (2008). A central role for Islet1 in sensory neuron development linking sensory and spinal gene regulatory programs. Nat. Neurosci. *11*, 1283–1293.
112. Okamoto, Y., Pehlivan, D., Wiszniewski, W., Beck, C.R., Snipes, G.J., Lupski, J.R., and Khajavi, M. (2013). Curcumin facilitates a transitory cellular stress response in trembler-J mice. Hum. Mol. Genet. *22*, 4698–4705.
113. Kumbasar, A., Plachez, C., Gronostajski, R.M., Richards, L.J., and Litwack, E.D. (2009). Absence of the transcription factor *Nfib* delays the formation of the basilar pontine and other mossy fiber nuclei. J. Comp. Neurol. *513*, 98–112.
114. Kajitani, R., Toshimoto, K., Noguchi, H., Toyoda, A., Ogura, Y., Okuno, M., Yabana, M., Harada, M., Nagayasu, E., Maruyama, H., et al. (2014). Efficient de novo assembly of highly heterozygous genomes from whole-genome shotgun short reads. Genome Res. *24*, 1384–1395.
115. Mcglothlin, J.W., Chuckalovcak, J.P., Janes, D.E., Edwards, S.V., Feldman, C.R., Brodie, E.D., Pfrender, M.E., and Brodie, E.D. (2014). Parallel evolution of tetrodotoxin resistance in three voltage-gated sodium channel genes in the garter snake *Thamnophis sirtalis*. Mol. Biol. Evol. *31*, 2836–2846.
116. Tang, C.-Y., Zhang, X., Xu, X., Sun, S., Peng, C., Song, M.-H., Yan, C., Sun, H., Liu, M., Xie, L., et al. (2023). Genetic mapping and molecular mechanism behind color variation in the Asian vine snake. Genome Biol. *24*, 46.
117. Aird, S.D., Arora, J., Barua, A., Qiu, L., Terada, K., and Mikheyev, A.S. (2017). Population genomic analysis of a pitviper reveals microevolutionary forces underlying venom chemistry. Genome Biol. Evol. *9*, 2640–2649.
118. Zhang, Z.-Y., Lv, Y., Wu, W., Yan, C., Tang, C.-Y., Peng, C., and Li, J.-T. (2022). The structural and functional divergence of a neglected three-finger toxin subfamily in lethal elapids. Cell Rep. *40*, 111079.
119. Wang, Z., Peng, C., Wu, W., Yan, C., Lv, Y., and Li, J.-T. (2023). Developmental regulation of conserved non-coding element evolution provides insights into limb loss in squamates. Sci. China Life Sci. https://doi.org/10.1007/s11427-023-2362-5.
120. Andrade, P., Pinho, C., Pérez i de Lanuza, G., Afonso, S., Brejcha, J., Rubin, C.-J., Wallerman, O., Pereira, P., Sabatino, S.J., Bellati, A., et al. (2019). Regulatory changes in pterin and carotenoid genes underlie balanced color polymorphisms in the wall lizard. Proc. Natl. Acad. Sci. USA *116*, 5633–5642.
121. Alföldi, J., Di Palma, F.D., Grabherr, M., Williams, C., Kong, L., Mauceli, E., Russell, P., Lowe, C.B., Glor, R.E., Jaffe, J.D., et al. (2011). The genome of the green anole lizard and a comparative analysis with birds and mammals. Nature *477*, 587–591.
122. Yan, C., Zhang, Z.-Y., Lv, Y., Wang, Z., Jiang, K., and Li, J.-T. (2022). Genome of *Laudakia sacra* provides new insights into high-altitude adaptation of ectotherms. Int. J. Mol. Sci. *23*, 10081.
123. Mellough, C.B., Bauer, R., Collin, J., Dorgau, B., Zerti, D., Dolan, D.W.P., Jones, C.M., Izuogu, O.G., Yu, M., Hallam, D., et al. (2019). An integrated transcriptional analysis of the developing human retina. Development *146*, dev169474.
124. Chen, S., Zhou, Y., Chen, Y., and Gu, J. (2018). fastp: an ultra-fast all-in-one FASTQ preprocessor. Bioinformatics *34*, i884–i890.
125. Marçais, G., and Kingsford, C. (2011). A fast, lock-free approach for efficient parallel counting of occurrences of *k*-mers. Bioinformatics *27*, 764–770.
126. Liu, B., Shi, Y., Yuan, J., Hu, X., Zhang, H., Li, N., Li, Z., Chen, Y., Mu, D., and Fan, W. (2013). Estimation of genomic characteristics by analyzing k-mer frequency in de novo genome projects https://doi.org/10.48550/arXiv.1308.2012.
127. Vurture, G.W., Sedlazeck, F.J., Nattestad, M., Underwood, C.J., Fang, H., Gurtowski, J., and Schatz, M.C. (2017). GenomeScope: fast reference-free genome profiling from short reads. Bioinformatics *33*, 2202–2204.
128. Ye, C., Hill, C.M., Wu, S., Ruan, J., and Ma, Z.S. (2016). DBG2OLC: efficient assembly of large genomes using long erroneous reads of the third generation sequencing technologies. Sci. Rep. *6*, 31900. https://doi.org/10.1038/srep31900.
129. Langmead, B., and Salzberg, S.L. (2012). Fast gapped-read alignment with Bowtie 2. Nat. Meth. *9*, 357–359.
130. Li, H., and Durbin, R. (2010). Fast and accurate long-read alignment with Burrows-Wheeler transform. Bioinformatics *26*, 589–595.
131. Hu, J., Wang, Z., Sun, Z., Hu, B., Ayoola, A.O., Liang, F., Li, J., Sandoval, J.R., Cooper, D.N., Ye, K., et al. (2023). An efficient error correction and accurate assembly tool for noisy long reads https://doi.org/10.1101/2023.03.09.531669.

132. Hu, J., Fan, J., Sun, Z., and Liu, S. (2020). NextPolish: a fast and efficient genome polishing tool for long-read assembly. Bioinformatics *36*, 2253–2255.

133. Ruan, J., and Li, H. (2020). Fast and accurate long-read assembly with wtdbg2. Nat. Meth. *17*, 155–158.

134. Servant, N., Varoquaux, N., Lajoie, B.R., Viara, E., Chen, C.-J., Vert, J.-P., Heard, E., Dekker, J., and Barillot, E. (2015). HiC-Pro: an optimized and flexible pipeline for Hi-C data processing. Genome Biol. *16*, 259.

135. Durand, N.C., Shamim, M.S., Machol, I., Rao, S.S.P., Huntley, M.H., Lander, E.S., and Aiden, E.L. (2016). Juicer provides a one-click system for analyzing loop-resolution Hi-C Experiments. Cell. Syst. *3*, 95–98.

136. Dudchenko, O., Batra, S.S., Omer, A.D., Nyquist, S.K., Hoeger, M., Durand, N.C., Shamim, M.S., Machol, I., Lander, E.S., Aiden, A.P., et al. (2017). De novo assembly of the *Aedes aegypti* genome using Hi-C yields chromosome-length scaffolds. Science *356*, 92–95.

137. Burton, J.N., Adey, A., Patwardhan, R.P., Qiu, R., Kitzman, J.O., and Shendure, J. (2013). Chromosome-scale scaffolding of de novo genome assemblies based on chromatin interactions. Nat. Biotechnol. *31*, 1119–1125.

138. Simão, F.A., Waterhouse, R.M., Ioannidis, P., Kriventseva, E.V., and Zdobnov, E.M. (2015). BUSCO: assessing genome assembly and annotation completeness with single-copy orthologs. Bioinformatics *31*, 3210–3212.

139. Flynn, J.M., Hubley, R., Goubert, C., Rosen, J., Clark, A.G., Feschotte, C., and Smit, A.F. (2020). RepeatModeler2 for automated genomic discovery of transposable element families. Proc. Natl. Acad. Sci. USA *117*, 9451–9457.

140. Tarailo-Graovac, M., and Chen, N. (2009). Chapter 4. Using RepeatMasker to identify repetitive elements in genomic sequences. Curr. Protoc. Bioinformatics *Chapter 4*, 4.10.1–4.10.14.

141. Stanke, M., Diekhans, M., Baertsch, R., and Haussler, D. (2008). Using native and syntenically mapped cDNA alignments to improve de novo gene finding. Bioinformatics *24*, 637–644.

142. Keilwagen, J., Wenk, M., Erickson, J.L., Schattat, M.H., Grau, J., and Hartung, F. (2016). Using intron position conservation for homology-based gene prediction. Nucleic Acids Res. *44*, e89.

143. Kovaka, S., Zimin, A.V., Pertea, G.M., Razaghi, R., Salzberg, S.L., and Pertea, M. (2019). Transcriptome assembly from long-read RNA-seq alignments with StringTie2. Genome Biol. *20*, 278.

144. Dobin, A., Davis, C.A., Schlesinger, F., Drenkow, J., Zaleski, C., Jha, S., Batut, P., Chaisson, M., and Gingeras, T.R. (2013). STAR: ultrafast universal RNA-seq aligner. Bioinformatics *29*, 15–21.

145. Haas, B.J., Salzberg, S.L., Zhu, W., Pertea, M., Allen, J.E., Orvis, J., White, O., Buell, C.R., and Wortman, J.R. (2008). Automated eukaryotic gene structure annotation using EVidenceModeler and the Program to Assemble Spliced Alignments. Genome Biol. *9*, R7.

146. Tang, S., Lomsadze, A., and Borodovsky, M. (2015). Identification of protein coding regions in RNA transcripts. Nucleic Acids Res. *43*, e78.

147. Urasaki, N., Takagi, H., Natsume, S., Uemura, A., Taniai, N., Miyagi, N., Fukushima, M., Suzuki, S., Tarora, K., Tamaki, M., et al. (2017). Draft genome sequence of bitter gourd (*Momordica charantia*), a vegetable and medicinal plant in tropical and subtropical regions. DNA Res. *24*, 51–58.

148. Zhang, Z., Schwartz, S., Wagner, L., and Miller, W. (2000). A greedy algorithm for aligning DNA sequences. J. Comput. Biol. *7*, 203–214.

149. Kim, J., Farré, M., Auvil, L., Capitanu, B., Larkin, D.M., Ma, J., and Lewin, H.A. (2017). Reconstruction and evolutionary history of eutherian chromosomes. Proc. Natl. Acad. Sci. USA *114*, E5379–E5388.

150. Krzywinski, M., Schein, J., Birol, I., Connors, J., Gascoyne, R., Horsman, D., Jones, S.J., and Marra, M.A. (2009). Circos: an information aesthetic for comparative genomics. Genome Res. *19*, 1639–1645.

151. Emms, D.M., and Kelly, S. (2015). OrthoFinder: solving fundamental biases in whole genome comparisons dramatically improves orthogroup inference accuracy. Genome Biol. *16*, 157.

152. Quinlan, A.R., and Hall, I.M. (2010). BEDTools: a flexible suite of utilities for comparing genomic features. Bioinformatics *26*, 841–842.

153. Chen, T., Liu, Y.-X., and Huang, L. (2022). ImageGP: an easy-to-use data visualization web server for scientific researchers. iMeta *1*, e5.

154. Wickham, H. (2016). ggplot2: Elegant Graphics for Data Analysis (Springer-Verlag).

155. Capella-Gutiérrez, S., Silla-Martínez, J.M., and Gabaldón, T. (2009). trimAl: a tool for automated alignment trimming in large-scale phylogenetic analyses. Bioinformatics *25*, 1972–1973.

156. Nguyen, L.-T., Schmidt, H.A., von Haeseler, A., and Minh, B.Q. (2015). IQ-TREE: a fast and effective stochastic algorithm for estimating maximum-likelihood phylogenies. Mol. Biol. Evol. *32*, 268–274.

157. Löytynoja, A., and Goldman, N. (2008). Phylogeny-aware gap placement prevents errors in sequence alignment and evolutionary analysis. Science *320*, 1632–1635.

158. Yang, Z. (2007). PAML 4: phylogenetic analysis by maximum likelihood. Mol. Biol. Evol. *24*, 1586–1591.

159. Sayyari, E., Whitfield, J.B., and Mirarab, S. (2018). DiscoVista: interpretable visualizations of gene tree discordance. Mol. Phylogenet. Evol. *122*, 110–115.

160. Beckstette, M., Homann, R., Giegerich, R., and Kurtz, S. (2006). Fast index based algorithms and software for matching position specific scoring matrices. BMC Bioinformatics *7*, 389.

161. Kim, D., Langmead, B., and Salzberg, S.L. (2015). HISAT: a fast spliced aligner with low memory requirements. Nat. Meth. *12*, 357–360.

162. Langfelder, P., and Horvath, S. (2008). WGCNA: an R package for weighted correlation network analysis. BMC Bioinformatics *9*, 559.

163. Supek, F., Bošnjak, M., Škunca, N., and Šmuc, T. (2011). REVIGO summarizes and visualizes long lists of gene ontology terms. PLoS One *6*, e21800.

164. Hiller, M., Schaar, B.T., Indjeian, V.B., Kingsley, D.M., Hagey, L.R., and Bejerano, G. (2012). A "forward genomics" approach links genotype to phenotype using independent phenotypic losses among related species. Cell. Rep. *2*, 817–823.

165. Li, B., and Dewey, C.N. (2011). RSEM: accurate transcript quantification from RNA-Seq data with or without a reference genome. BMC Bioinformatics *12*, 323.

166. Choi, Y., Sims, G.E., Murphy, S., Miller, J.R., and Chan, A.P. (2012). Predicting the functional effect of amino acid substitutions and indels. PLoS One *7*, e46688.

167. Li, B., Fillmore, N., Bai, Y., Collins, M., Thomson, J.A., Stewart, R., and Dewey, C.N. (2014). Evaluation of de novo transcriptome assemblies from RNA-Seq data. Genome Biol. *15*, 553.

168. Love, M.I., Huber, W., and Anders, S. (2014). Moderated estimation of fold change and dispersion for RNA-seq data with DESeq2. Genome Biol. *15*, 550.

169. Robinson, M.D., McCarthy, D.J., and Smyth, G.K. (2010). edgeR: a Bioconductor package for differential expression analysis of digital gene expression data. Bioinformatics *26*, 139–140.

170. Grabherr, M.G., Haas, B.J., Yassour, M., Levin, J.Z., Thompson, D.A., Amit, I., Adiconis, X., Fan, L., Raychowdhury, R., Zeng, Q., et al. (2011). Full-length transcriptome assembly from RNA-Seq data without a reference genome. Nat. Biotechnol. *29*, 644–652.

171. Li, W., and Godzik, A. (2006). Cd-hit: a fast program for clustering and comparing large sets of protein or nucleotide sequences. Bioinformatics *22*, 1658–1659.

172. Kosakovsky Pond, S.L., Poon, A.F.Y., Velazquez, R.V., Weaver, S., Hepler, N.L., Murrell, B., Shank, S.D., Magalis, B.R., Bouvier, D., Nekrutenko, A., et al. (2020). HyPhy 2.5—A customizable platform for

evolutionary hypothesis testing using phylogenies. Mol. Biol. Evol. *37*, 295–299.

173. Burbrink, F.T., Grazziotin, F.G., Pyron, P.A., Cundall, D., Donnellan, S., Frances, I., Keogh, J.S., Kraus, F., Murphy, R.W., Noonan, N., et al. (2019). Interrogating genomic-scale data for Squamata (Lizards, Snakes, and Amphisbaenians) shows no support for key traditional morphological relationships. Syst. Biol. *69*, 502–520.
174. Chen, H., Rangasamy, M., Tan, S.Y., Wang, H., and Siegfried, B.D. (2010). Evaluation of five methods for total DNA extraction from western corn rootworm beetles. PLoS One *5*, e11963.
175. Liu, Z., Zhang, L., Yan, Z., Ren, Z., Han, F., Tan, X., Xiang, Z., Dong, F., Yang, Z., Liu, G., et al. (2020). Genomic mechanisms of physiological and morphological adaptations of limestone langurs to karst habitats. Mol. Biol. Evol. *37*, 952–968.
176. Zoonomia Consortium (2020). A comparative genomics multitool for scientific discovery and conservation. Nature *587*, 240–245.
177. McLean, C.Y., Bristor, D., Hiller, M., Clarke, S.L., Schaar, B.T., Lowe, C.B., Wenger, A.M., and Bejerano, G. (2010). GREAT improves functional interpretation of *cis*-regulatory regions. Nat. Biotechnol. *28*, 495–501.
178. Hoang, D.T., Chernomor, O., von Haeseler, A., Minh, B.Q., and Vinh, L.S. (2018). UFBoot2: improving the ultrafast bootstrap approximation. Mol. Biol. Evol. *35*, 518–522.
179. Kalyaanamoorthy, S., Minh, B.Q., Wong, T.K.F., von Haeseler, A., and Jermiin, L.S. (2017). ModelFinder: fast model selection for accurate phylogenetic estimates. Nat. Meth. *14*, 587–589.
180. Jumper, J., Evans, R., Pritzel, A., Green, T., Figurnov, M., Ronneberger, O., Tunyasuvunakool, K., Bates, R., Žídek, A., Potapenko, A., et al. (2021). Highly accurate protein structure prediction with AlphaFold. Nature *596*, 583–589.
181. Zheng, R., Wan, C., Mei, S., Qin, Q., Wu, Q., Sun, H., Chen, C.-H., Brown, M., Zhang, X., Meyer, C.A., et al. (2019). Cistrome Data Browser: expanded datasets and new tools for gene regulatory analysis. Nucleic Acids Res. *47*, D729–D735.
182. Castillo-Davis, C.I., Kondrashov, F.A., Hartl, D.L., and Kulathinal, R.J. (2004). The functional genomic distribution of protein divergence in two animal phyla: coevolution, genomic conflict, and constraint. Genome Res. *14*, 802–811.
183. Kosakovsky Pond, S.L., Frost, S.D.W., and Muse, S.V. (2005). HyPhy: hypothesis testing using phylogenies *21*, 676–679.
184. Eisenberg, E., and Levanon, E.Y. (2013). Human housekeeping genes, revisited. Trends Genet. *29*, 569–574.
185. Ebert, J. (2008). Infrared sense in snakes: behavioural and anatomical examinations (*Crotalus atrox*, *Python regius*, *Corallus hortulanus*). Doctoral (Rheinische Friedrich Wilhelms Univ. Bonn.).

STAR★METHODS

KEY RESOURCES TABLE

REAGENT or RESOURCE	SOURCE	IDENTIFIER
Biological samples		
Szechwan rat snake	This paper	N/A
Red cornsnake	This paper	N/A
Hong Kong dwarf snake	This paper	N/A
Red-tailed pipe snake	This paper	N/A
Tartar sand boa	This paper	N/A
Plumbeous water snake	This paper	N/A
Common mock viper	This paper	N/A
African house snake	This paper	N/A
Black-tip worm snake	This paper	N/A
Keeled slug snake	This paper	N/A
Diard's blind snake	This paper	N/A
Shedao island pitviper	This paper	N/A
Zong's odd-scaled snake	This paper	N/A
Sunbeam snake	This paper	N/A
Ptyas dhumnades	This paper	N/A
Critical commercial assays		
Genomic kit (Cat#13343)	QIAGEN	N/A
Ultra II End Repair/dA-tailing Kit (Cat# E7546)	NEBNext	N/A
RNeasy Mini Kit	QIAGEN	N/A
Ultra™ RNA Library Prep Kit for Illumina	NEBNext	N/A
RNA Nano 6000 Assay Kit	Illumina	N/A
TruSeq PE Cluster Kit v3-cBot-HS	Illumina	N/A
TRIzol reagent	Invitrogen	N/A
TRIzol reagent	Ambion	N/A
T7 Ultra Kit	Ambion	N/A
MEGAclearTM Kit	ThermoFisher Scientific	N/A
Chitinase Assay Kit (ADS-302-F)	Jiangsu Kete Biotechnology	N/A
MEGAshortscript Kit	ThermoFisher Scientific	N/A
Deposited data		
Burmese python (*Python bivittatus*) genome assembly	Castoe et al.[11]	NCBI: GCA_000186305.2
Boa constrictor (*Boa constrictor*) genome assembly	Kajitani et al.[114]	http://platanus.bio.titech.ac.jp/platanus-assembler/platanus-1-2-1
Xizang hot-spring snake (*Thermophis baileyi*) genome assembly	Li et al.[12]	NCBI: GCA_003457575.1
Xizang hot-spring snake (*Thermophis baileyi*) genome assembly	Yan et al.[14]	National Genomic Data Center (https://bigd.big.ac.cn/gwh/): GWHBJWY00000000
Common garter snake (*Thamnophis sirtalis*) genome assembly	Vertebrate Genomes Project	NCBI: GCF_009769535.1
Common garter snake (*Thamnophis sirtalis*) genome assembly	McGlothlin et al.[115]	NCBI: GCA_001077635.2

(*Continued on next page*)

Continued

REAGENT or RESOURCE	SOURCE	IDENTIFIER
Asian vine snake (*Ahaetulla prasina*) genome assembly	Tang et al.[116]	NCBI: ASM2864084v1
Shaw's sea snake (*Hydrophis curtus*) genome assembly	Peng et al.[16]	https://doi.org/10.6084/m9.figshare.11391606.v5
Eastern brown snake (*Pseudonaja textilis*) genome assembly	University of New South Wales	NCBI: GCA_900518735.1
Indian cobra (*Naja naja*) genome assembly	Suryamohan et al.[15]	NCBI: GCA_009733165.1
King cobra (*Ophiophagus hannah*) genome assembly	Vonk et al.[10]	NCBI: GCA_000516915.1
Mainland tiger snake (*Notechis scutatus*) genome assembly	University of New South Wales	NCBI: GCA_900518725.1
Western rattlesnake (*Crotalus viridis viridis*) genome assembly	Schield et al.[13]	NCBI: GCA_003400415.2
Brown spotted pitviper (*Protobothrops mucrosquamatus*) genome assembly	Aird et al.[117]	NCBI: GCA_001527695.3
Many-banded krait (*Bungarus multicinctus*) genome assembly	Zhang et al.[118]	China National GeneBank DataBase: CNA0045869
Five-pacer viper (*Deinagkistrodon acutus*) genome assembly	Yin et al.[9]	https://ftp.cngb.org/pub/gigadb/pub/10.5524/100001_101000/100196/
Argentine giant tegu (*Salvator merianae*) genome assembly	Roscito et al.[25]	https://bds.mpi-cbg.de/hillerlab/TeguGenomeData/
Yangtze alligator (*Alligator sinensis*) genome assembly	Wan et al.[28]	NCBI: GCA_000455745.1
European glass lizard (*Pseudopus apodus*) genome assembly	Wang et al.[119]	China National GeneBank DataBase: CNP0003553
Komodo dragon (*Varanus komodoensis*) genome assembly	Lind et al.[26]	https://figshare.com/projects/Data_for_Komodo_dragon_genome_paper/61271
Common wall lizard (*Podarcis muralis*) genome assembly	Andrade et al.[120]	NCBI: GCA_004329235.1
Painted turtle (*Chrysemys picta bellii*) genome assembly	Shaffer et al.[27]	NCBI: GCA_000241765.2
Viviparous lizard (*Zootoca vivipara*) genome assembly	University of Glasgow	NCBI: GCA_011800845.1
Sand lizard (*Lacerta agilis*) genome assembly	Vertebrate Genomes Project	NCBI: GCA_009819535.1
Green anole (*Anolis carolinensis*) genome assembly	Alfoldi et al.[121]	NCBI: GCA_000090745.2
Anan's Rock Agama (*Laudakia sacra*) genome assembly	Yan et al.[122]	National Genomic Data Center (https://bigd.big.ac.cn/gwh/): PRJCA010949
Chicken (*Gallus gallus*) genome assembly	Hillier et al.[29]	NCBI: GCA_000002315.5
Human (*Homo sapiens*) genome assembly	Ensembl	Ensembl: GRCh38
Szechwan rat snake (*Euprepiophis perlacea*) genome assembly	This paper	National Genomic Data Center (https://bigd.big.ac.cn/gwh/): GWHBOZS00000000.
Red cornsnake (*Pantherophis guttatus*) genome assembly	This paper	National Genomic Data Center (https://bigd.big.ac.cn/gwh/): GWHBOZR00000000.
Hong Kong dwarf snake (*Calamaria septentrionalis*) genome assembly	This paper	National Genomic Data Center (https://bigd.big.ac.cn/gwh/): GWHBWDL00000000.
Red-tailed pipe snake (*Cylindrophis ruffus*) genome assembly	This paper	National Genomic Data Center (https://bigd.big.ac.cn/gwh/): GWHBWDM00000000.

(*Continued on next page*)

Continued

REAGENT or RESOURCE	SOURCE	IDENTIFIER
Tartar sand boa (*Eryx tataricus*) genome assembly	This paper	National Genomic Data Center (https://bigd.big.ac.cn/gwh/): GWHBWDN00000000.
Many-banded krait (*Bungarus multicinctus*) genome assembly	This paper	CNGBdb: CNA0045869.
Plumbeous water snake (*Hypsiscopus plumbea*) genome assembly	This paper	National Genomic Data Center (https://bigd.big.ac.cn/gwh/): GWHBWDO00000000.
Common mock viper (*Psammodynastes pulverulentus*) genome assembly	This paper	National Genomic Data Center (https://bigd.big.ac.cn/gwh/): GWHBWDP00000000.
African house snake (*Boaedon fuliginosus*) genome assembly	This paper	National Genomic Data Center (https://bigd.big.ac.cn/gwh/): GWHBWDQ00000000.
Black-tip worm snake (*Leptotyphlops nigroterminus*) genome assembly	This paper	National Genomic Data Center (https://bigd.big.ac.cn/gwh/): GWHBWDR00000000.
Keeled slug snake (*Pareas berdmorei*) genome assembly	This paper	National Genomic Data Center (https://bigd.big.ac.cn/gwh/): GWHBWDS00000000.
Diard's blind snake (*Argyrophis diardii*) genome assembly	This paper	National Genomic Data Center (https://bigd.big.ac.cn/gwh/): GWHBWDT00000000.
Shedao island pitviper (*Gloydius shedaoensis*) genome assembly	This paper	National Genomic Data Center (https://bigd.big.ac.cn/gwh/): GWHBWDU00000000.
Zong's odd-scaled snake (*Achalinus jinggangensis*) genome assembly	This paper	National Genomic Data Center (https://bigd.big.ac.cn/gwh/): GWHBWDV00000000.
Sunbeam snake (*Xenopeltis unicolor*) genome assembly	This paper	National Genomic Data Center (https://bigd.big.ac.cn/gwh/): GWHBWDW00000000.
Human embryonic eye transcriptome data	Mellough et al.[123]	NCBI: PRJNA384924
Transcriptome data used for annotation and analysis	This paper	National Genomic Data Center (https://bigd.big.ac.cn/gwh/): PRJCA012991
Software and algorithms		
fastp v0.19.4	Chen et al.[124]	https://github.com/OpenGene/fastp
fastp v0.23.1	Chen et al.[124]	https://github.com/OpenGene/fastp
smrtlink v7.0	Pacbio	https://www.pacb.com/support/software-downloads/
Guppy v3.2.2+9fe0a78	Oxford Nanopore Technology	https://github.com/nanoporetech/taiyaki
Jellyfish v2.2.10	Marçais and Kingsford[125]	https://www.cbcb.umd.edu/software/jellyfish/
GCE v1.02	Liu et al.[126]	https://github.com/fanagislab/GCE
findGSE v1.94	Git-hub	https://github.com/schneebergerlab/findGSE/
GenomeScope	Vurture et al.[127]	http://qb.cshl.edu/genomescope/
Platanus v1.2.4	Kajitani et al.[114]	http://platanus.bio.titech.ac.jp/platanus-assembler/platanus-1-2-4
DBG2OLC	Ye et al.[128]	https://github.com/yechengxi/DBG2OLC
Bowtie2 v2.3.2	Langmead and Salzberg[129]	https://bowtie-bio.sourceforge.net/bowtie2/index.shtml
Bowtie2 v2.3.4.3	Langmead and Salzberg[129]	https://bowtie-bio.sourceforge.net/bowtie2/index.shtml

(*Continued on next page*)

Continued

REAGENT or RESOURCE	SOURCE	IDENTIFIER
BWA v0.7.17-r1188	Li and Durbin[130]	https://github.com/lh3/bwa
NextDenovo v1.0	Hu et al.[131]	https://github.com/Nextomics/NextDenovo
NextPolish v1.01	Hu et al.[132]	https://github.com/Nextomics/NextPolish
wtdbg2 v2.3	Ruan and Li[133]	https://github.com/ruanjue/wtdbg2
HiC-Pro v2.8.0	Servant et al.[134]	https://github.com/nservant/HiC-Pro/
Juicebox	Durand et al.[135]	https://github.com/aidenlab/Juicebox
Juicer v1.5	Durand et al.[135]	https://github.com/aidenlab/juicer
3D-DNA v180922	Dudchenko et al.[136]	https://github.com/aidenlab/3d-dna
LACHESIS	Burton et al.[137]	https://github.com/shendurelab/LACHESIS
BUSCO v3.1.0	Simão et al.[138]	https://busco.ezlab.org/v3
RepeatModeler v1.0.11	Flynn et al.[139]	http://www.repeatmasker.org/
RepeatMasker v1.331	Tarailo-Graovac and Chen[140]	http://www.repeatmasker.org/
AUGUSTUS v3.3.1	Stanke et al.[141]	http://bioinf.uni-greifswald.de/augustus/
GeMoMa v1.6.1	Keilwagen et al.[142]	http://www.jstacs.de/index.php/GeMoMa
StringTie v1.3.4	Kovaka et al.[143]	https://ccb.jhu.edu/software/stringtie/
StringTie v2.0.4	Kovaka et al.[143]	https://ccb.jhu.edu/software/stringtie/
STAR v2.6.1d	Dobin et al.[144]	https://github.com/alexdobin/STAR/
PASA v2.3.3	Haas et al.[145]	https://github.com/PASApipeline/PASApipeline/releases
GeneMark-ST 5.1 March 2014	Tang et al.[146]	http://exon.gatech.edu/GeneMark/
EVidenceModeler v1.1.1	Haas et al.[145]	http://evidencemodeler.github.io/
TransposonPSI v1.0.0	Urasaki et al.[147]	http://transposonpsi.sourceforge.net/
BLAST v2.7.1+	Zhang et al.[148]	https://blast.ncbi.nlm.nih.gov/Blast.cgi
DESCHRAMBLER	Kim et al.[149]	https://github.com/jkimlab/DESCHRAMBLER
Circos v0.69-9	Krzywinski et al.[150]	http://circos.ca/
Progressive Cactus v1.0.0	Armstrong et al.[17]	https://github.com/ComparativeGenomicsToolkit/cactus/
OrthoFinder v2.2.7	Emms and Kelly[151]	https://github.com/davidemms/OrthoFinder
PhastCons v1.3	N/A	http://compgen.cshl.edu/phast/phastCons-HOWTO.html
GERP	N/A	http://mendel.stanford.edu/SidowLab/downloads/gerp/index.html
BEDTools v2.25.0	Quinlan and Hall[152]	https://BEDTools.readthedocs.io/en/latest/
ImageGP	Cheng et al.[153]	https://416h86i955.zicp.fun/Cloud_Platform/front/#/
ggplot2 v3.3.5	Wickham[154]	https://ggplot2.tidyverse.org/
phyper v3.6.0		https://www.r-project.org/
ASTRAL-III v5.6.3	Zhang et al.[18]	https://github.com/smirarab/ASTRAL
trimAl v1.4.rev22	Capella-Gutiérrez et al.[155]	http://trimal.cgenomics.org
IQ-TREE v1.6.8	Nguyen et al.[156]	http://www.iqtree.org/
prank v.150803	Löytynoja and Goldman[157]	http://wasabiapp.org/software/prank/
r8s v1.71	Sanderson[34]	https://sourceforge.net/projects/r8s/
PAML 4.9i	Yang[158]	http://abacus.gene.ucl.ac.uk/software/paml.html
DiscoVista	Sayyari[159]	https://github.com/HeJian151004/DiscoVista
PoSSuM v1.3	Beckstette et al.[160]	https://possum.cbrc.pj.aist.go.jp/PoSSuM/

(*Continued on next page*)

Continued

REAGENT or RESOURCE	SOURCE	IDENTIFIER
HISAT2 v2-2.1.0	Kim et al.[161]	https://daehwankimlab.github.io/hisat2/
preprocessCore v1.48.0	Ben Bolstad	http://www.bioconductor.org/packages/3.10/bioc/html/preprocessCore.html
WGCNA v1.70.3	Langfelder and Horvath[162]	https://horvath.genetics.ucla.edu/html/CoexpressionNetwork/Rpackages/WGCNA/
REVIGO	Supek et al.[163]	http://revigo.irb.hr/
Forward Genomics	Hiller et al.[164]	https://github.com/hillerlab/ForwardGenomics
RSEM v1.2.8	Li and Dewey[165]	https://lilab-bcb.github.io/software/
provean v1.1.5	Choi et al.[166]	https://www.jcvi.org/research/provean
TransDecoder v3.0.1	Brian Haas	https://github.com/TransDecoder/TransDecoder/
DETONATE v1.11	Li et al.[167]	https://deweylab.biostat.wisc.edu/detonate/
DESeq2 v1.26.0	Michael et al.[168]	http://bioconductor.org/packages/release/bioc/html/DESeq2.html
edgeR v3.28.1	Smyth[169]	http://www.bioconductor.org/packages/release/bioc/html/edgeR.html
Trinity v2.8.5	Grabherr et al.[170]	http://trinityrnaseq.github.io
CD-HIT v4.6	Li and Godzik[171]	http://cd-hit.org
HyPhy v2.5.20	Kosakovsky et al.[172]	http://www.hyphy.org/
Other		
Snake Multiomic Database	This paper	http://herpmuseum.cib.ac.cn/snake/
Custom scripts	This paper	https://github.com/bioinformaticspcj/Coding_gene_function_annotation; https://github.com/bioinformaticspcj/Genomic_tools_for_snake_phylogenomics

RESOURCE AVAILABILITY

Lead contact

Further information and requests for resources and reagents should be directed to and will be fulfilled by the lead contact, Jia-Tang Li (lijt@cib.ac.cn).

Materials availability

No new unique reagents were generated in this study.

Data and code availability

All the assemblies and raw sequencing data generated in this study are available in the National Genomics Data Center under the BioProject number PRJCA012991. The data are publicly available as of the date of publication. The assemblies and the corresponding annotations are also available in our Snake Multiomic Database. The custom codes are available on GitHub (https://github.com/bioinformaticspcj/Coding_gene_function_annotation, https://github.com/bioinformaticspcj/Genomic_tools_for_snake_phylogenomics). Any additional information required to reanalyze the data reported in this paper is available from the lead contact upon request.

EXPERIMENTAL MODEL AND STUDY PARTICIPANT DETAILS

Source organisms

All animal samples were collected legally and in accordance with the policy of the Animal Research and Ethics Committee of Chengdu Institute of Biology, Chinese Academy of Sciences. All animal experiments were approved by the Animal Research and Ethics Committee of Chengdu Institute of Biology, Chinese Academy of Sciences (under NO. CIBDWLL2018030). Adult male

individuals of six snakes (Szechwan rat snake, tartar sand boa, plumbeous water snake, common mock viper, African house snake, and keeled slug snake) and adult female individuals of seven snakes (red cornsnake, Hong Kong dwarf snake, red-tailed pipe snake, Diard's blind snake, Shedao island pitviper, Zong's odd-scaled snake, and Sunbeam snake) were used for genome sequencing (Table S1). Of these, nine snakes (Szechwan rat snake, tartar sand boa, plumbeous water snake, African house snake, keeled slug snake, red cornsnake, Diard's blind snake, Shedao island pitviper, and sunbeam snake) were also used for transcriptome sequencing (Table S1). One adult individual of black-tip worm snake (gender is unknown as only tissue is available) was used for genome sequencing. Adult RNA samples from eight snakes (Diard's blind snake, keeled slug snake, brown spotted pitviper, plumbeous water snake, Sichuan hot-spring keel-back, Asian vine snake, *P. dhumnades*, and sunbeam snake) and one lizard (green anole) were collected for RNA-seq. Embryos (10, 30, and 50 days post oviposition) from red cornsnake and embryos (40 days post oviposition) from *P. dhumnades* were also collected for RNA-seq (Table S1). *PTCH1* CRISPR/Cas9 editing was performed in wild-type mice (*Mus musculus*, strain: C57BL/6) (Table S1). Mouse embryos at stage E11.5 were used for limb bud transcriptomic analyses. We did not identify the sex of samples used for transcriptomic analyses.

METHOD DETAILS

Selection of species

We analyzed 28 snake genomes, of which 14 were newly assembled (Table S1) and 14 were previously published, including genomes for the Burmese python,[11] boa constrictor,[114] king cobra,[10] eastern brown snake, Indian cobra,[15] Shaw's sea snake,[16] mainland tiger snake, western rattlesnake,[13] Xizang hot-spring snake,[12,14] five-pacer viper,[9] brown-spotted pitviper,[117] common garter snake,[115] many-banded krait,[118] and Asian vine snake.[116] These snakes (covering basal and advanced snakes) were interspersed across the snake tree according to a previous study[173] (Figure S1A). Species samples as well as their sources are listed in Table S1.

Genome sequencing

Tissue samples were collected from 14 snake species (Table S1), and DNA was extracted using a QIAGEN® Genomic kit (Cat#13343, QIAGEN) for regular sequencing following the manufacturer's instructions. The sodium dodecyl sulfate method (as described in a previous study[174]) was used to extract ultra-long DNA without performing the purification step to maintain the length of DNA. The degradation and contamination of the extracted DNA were assessed on a 1% agarose gel. DNA purity was detected using a NanoDrop™ One UV-Vis spectrophotometer (Thermo Fisher Scientific, USA); the OD260/280 ratio ranged from 1.8 to 2.0, and the OD260/230 ratio ranged from 2.0 to 2.2. The DNA concentration was measured using a Qubit® 4.0 Fluorometer (Invitrogen, USA).

For Oxford Nanopore Technology (ONT) library preparation and sequencing, a total of 3–4 μg of DNA per sample was used as input material. After the sample was qualified, the PippinHT system (Sage Science, USA) was used to size-select long DNA fragments. Next, DNA end-repair and dA-ligation reactions were conducted using the NEBNext Ultra II End Repair/dA-tailing Kit (Cat# E7546). The adapters in the SQK-LSK109 Kit (Oxford Nanopore Technology, UK) were used for further ligation reactions, and the concentration of the DNA library was measured using a Qubit® 4.0 Fluorometer (Invitrogen, USA). The DNA library (ca. 700 ng) was sequenced using a Nanopore PromethION sequencer instrument (Oxford Nanopore Technology, UK) at the Genome Center of Grandomics (Wuhan, China).

For Pacbio (Continuous Long Reads, CLR) library preparation and sequencing, genomic DNA extracted from liver and muscle (Table S1) was sheared into 20-kb fragments using a g-TUBE device (Covaris, Woburn, MA). The sheared DNA was further purified and concentrated using AmpureXP beads (Agencourt, Beverly, MA). The resulting DNA was used to construct a read library with an insert size of 20 kb; it was then sequenced on a Sequel II SMRT cell.

For short-read library preparation and sequencing (Illumina reads), whole-genome shotgun sequencing of a paired-end library with an insert size of 350 bp was conducted for each snake sample. The libraries were sequenced using the Illumina NovaSeq 6000 platform following the standard manufacturer's protocol (San Diego). For BGI-Shenzhen reads sequencing, genome DNA was broken into segments. DNA segments with lengths $\geq$ 200 bp and $\leq$ 400 bp were then used to construct libraries. Sequencing of these libraries was performed on a MGISEQ-2000 platform according to the standard protocol.

To conduct quality control analyses of ONT reads, all fast5 files were converted to fastq format using Guppy v3.2.2+9fe0a78 with "-c dna_r9.4.1_450bps_fast.cfg". The reads were filtered using Guppy v3.2.2+9fe0a78 with mean_qscore_template > 7. After removing sequences containing bases of low quality and errors, the clean reads were used for genome assembly. For Pacbio reads, the Sequel raw BAM files were converted into subreads in FASTA format using PacBio smrtlink v7.0 software with default parameters. For Illumina and MGI data, fastp v0.19.4[124] was used to remove low-quality reads and reads with adapters using default parameters.

To obtain Hi-C raw reads, each sample (Table S1) was used for Hi-C library construction following the method described in a previous study,[175] which includes crosslinking, chromatin digestion with the DpnII restriction enzyme, DNA end-marking and ligation, purification, shearing, and a biotin pull-down assay. Constructed Hi-C libraries were then 150-bp paired-end sequenced on the Illumina NovaSeq 6000 platform.

Genome size estimation and assembly

We estimated the size of each genome using a K-mer frequency-based method. The K-mers were counted using Jellyfish v2.2.10[125] with the parameters "-C -m 51 -s 10000000000 -t 50." The results were used as input for GenomeScope[127] and findGSE v1.94

(https://github.com/schneebergerlab/findGSE/) to estimate the genome size. To assemble the genomes, the genome assembler NextDenovo v1.0[131] was used for the self-correction of ONT long reads with the parameters "read_cuoff = 1k, seed_cutoff = 25k". Next, the corrected reads were assembled using wtdbg2[133] with default parameters. The resulting contigs were polished using NextPolish v1.01[132] software with default parameters. Three rounds of alignment with long reads and four rounds of alignment with short reads were conducted in this step. Next, the Hi-C short reads (if available) were filtered using fastp v0.19.4 to remove low-quality sequences (quality score $\leq$ 15), adapter sequences, and sequences shorter than 30 bp. The remaining reads were input into HiC-Pro v2.8.0[134] and Bowtie2 v2.3.2[129] to obtain uniquely mapped valid reads. Finally, the clustering, ordering, and anchoring of the polished contigs were performed using LACHESIS (CLUSTER MIN RE SITES = 100; CLUSTER MAX LINK DENSITY = 2.5; CLUSTER NONINFORMATIVE RATIO = 1.4; ORDER MIN N RES IN TRUNK = 60; ORDER MIN N RES IN SHREDS = 60)[137] with the valid Hi-C data to obtain the chromosome-level assembly.

For the assembly of the Diard's blind snake genome, we first used Genomic Character Estimator (GCE) v1.02[126] to estimate the genome size with Illumina short reads. Next, we used a hybrid method as described in our previous study for genome assembly.[118] Briefly, Platanus v1.2.4[114] was used to construct contigs with Illumina short reads. DBG2OLC[128] was then used to extend the contigs with Pacbio long reads. Finally, the consensus contigs were obtained and used to assemble the chromosome-scale genome with the Hi-C data using HiC-Pro v2.8.0,[134] Juicer v1.5,[135] and 3D-DNA v180922.[136] Assembly errors detected during the Hi-C scaffolding were further corrected visually using Juicebox.[135]

Genome quality evaluation

To evaluate the quality of our genome assemblies, we used benchmarking universal single-copy orthologs (BUSCO) v3.1 in genome mode with vertebrate lineage data (vertebrata_odb9)[138] (Table S1).

RNA extraction and sequencing

To obtain transcriptomic evidence for genome annotation, we collected 36 samples (Table S1) of the newly sequenced species for RNA sequencing. RNA was extracted using TRIzol (Invitrogen) and subsequently purified using an RNeasy Mini Kit (QIAGEN). The purified RNA was then used to construct PolyA RNA-sequencing libraries with the NEBNext® Ultra™ RNA Library Prep Kit for Illumina® (NEB, USA) following the manufacturer's recommendations. Finally, RNA sequencing was conducted using the Illumina HiSeq 2000 platform.

To investigate the expression levels of genes in different species, we collected 194 samples from different tissues of the green anole (which is phylogenetically closely related to snakes[23]), keeled slug snakes, *P. dhumnades*, red corn snake, Asian vine snake, sunbeam snake, Diard's blind snake, Sichuan hot-spring keelback, brown-spotted pitviper, and mouse. Detailed information is provided in Table S1.

The total RNA of all samples was extracted using Trizol (Ambin, USA), according to the manufacturer's instructions. RNA degradation and contamination were assessed on a 1% agarose gel. A NanoPhotometer® spectrophotometer (IMPLEN, CA, USA) was used to determine RNA purity, and RNA integrity was determined using an Agilent Bioanalyzer 2100 system with an RNA Nano 6000 Assay Kit (Agilent Technologies, CA, USA). The qualified total RNA was used to construct libraries with the NEBNext® Ultra™ RNA Library Prep Kit for Illumina® (NEBNext, USA) following the manufacturer's recommendations (Novogene Company, Beijing, China). Index-coded samples were clustered on a cBot Cluster Generation System using the TruSeq PE Cluster Kit v3-cBot-HS (Illumina) according to the manufacturer's instructions. All cDNA libraries were 150-bp paired-end sequenced on a NovaSeq 6000 platform.

Genome annotation

For coding gene structural annotation, all repeats were annotated using both homology-based and *de novo* predictions. RepeatModeler v1.0.11[139] was used to construct the *de novo* transposable element library. All constructed libraries together with repeats from Repbase (20170127) were integrated into one library, which was used as the input for RepeatMasker v1.331[140] to predict repeats. The same protocol was also used to annotate repeats in the genomes of outgroup species (Figure 2).

We then annotated protein-coding genes in the repeat-masked assemblies by *de novo*-based, homology-based, and transcriptome-based (if available, Table S1) methods. AUGUSTUS v3.3.1[141] software was used for *de novo* predictions with the parameters "–gff3=on –hintsfile=hints.gff –extrinsicCfgFile=extrinsic.cfg –allow_hinted_splicesites=gcag,atac –min_intron_len=30 –softmasking=1". Whole proteomes from king cobra, Burmese python, common garter snake, brown-spotted pitviper, eastern brown snake, and Xizang spring snake were input into GeMoMa v1.6.1[142] to perform homologous gene structure prediction. For transcript-based gene prediction, RNA-seq reads were mapped to the assemblies using STAR v2.6.1d[144]; StringTie v1.3.4[143] was then used to identify the transcript position in the genome assemblies, and transcript sequences were extracted. Next, the extracted transcript sequences were mapped to the genomes using PASA v2.3.3[145] to obtain gene predictions. GeneMark-ST[146] was also used to generate gene predictions using PASA-extracted transcripts. Gene predictions from the three methods were integrated using EVidenceModeler v1.1.1[145] (weights for each: Augustus: 1; GeMoMa: 5; PASA: 10; GeneMark-ST: 10) with the parameters: "–segmentSize 1000000 –overlapSize 100000". The final gene predictions were obtained by removing genes with transposon insertions identified by TransposonPSI v1.0.0[147] (Table S1).

To annotate the functions of the coding genes of our newly assembled genomes, we used blastp v2.7.1+[148] to conduct searches (E-value $\leq$ 1e−5) against the non-redundant (NR) database in the National Center for Biotechnology Information (version:20190401), SwissProt database (version:20200709), and Kyoto Encyclopedia of Genes and Genomes (KEGG) database (version:20200712). The results of the blastp v2.7.1+ searches against SwissProt were processed using our in-house Perl script to retrieve associated GO terms from idmapping.tb.gz (ftp://ftp.pir.georgetown.edu/databases/idmapping/, version:20200709) (Table S1).

Chromosome evolution analysis

To investigate the evolution of chromosomes in snakes, we selected genomes of four lizards (used as outgroups) and 16 snakes (Figure S2A) to conduct a *k*-based synteny analysis following Shield et al.[13] First, we divided the 19 genomes into potential 100-bp markers and aligned each marker against the Diard's blind snake genome using blastn v2.7.1+. The best hit with an alignment length of at least 50 bp was retained. These markers were processed on each chromosome by requiring at least five consecutive markers supporting homology to the same Diard's blind snake chromosome. Each group of five or more consecutive markers was merged into one confirmed marker. Second, we used DESCHRAMBLER[149] to re-construct the karyotype of the common ancestor of Serpentes using the confirmed markers (Figures 2 and S2A; Table S2). Finally, the synteny between the chromosomes of selected species and the re-constructed snake ancestor was visualized using Circos v0.69-9.[150]

Progressive Cactus alignment

Progressive Cactus v1.0.0, a reference-free whole-genome alignment software,[17] has been used to align bird[17] and mammal[176] genomes. We aligned the genomes of 25 snakes and six outgroup species using Progressive Cactus v1.0.0 to obtain the whole-genome alignment (WGA). The guide tree was generated from the 4-fold degenerate third codon (4d) positions of selected species (see "phylogenetic tree construction" section). Progressive Cactus v1.0.0 software was run on the Rongjian bioinformatics big data computing cloud service platform with the parameters: "–binariesMode local –disableCaching –disableHotDeployment" to obtain the final hal file. The hal2maf command (–onlyOrthologs) implemented in the Progressive Cactus package was used to produce the raw 31-taxon genome alignment in multiple alignment format (MAF) file format. Finally, two other commands, mafDuplicateFilter and mafFilter, in the Progressive Cactus package, were applied to the MAF file to obtain the final 31-taxon alignment with default parameters.

Analysis of conserved non-coding elements (CNEs)

We used both PhastCons v1.3 (http://compgen.cshl.edu/phast/phastCons-HOWTO.html) (parameters: "expected length = 45, target coverage = 0.3, rho = 0.3") and GERP (parameters: "-d 0.01") (http://mendel.stanford.edu/SidowLab/downloads/gerp/index.html) to identify evolutionarily conserved elements in the 31-taxon alignment. The 4d-site tree (see details in the "phylogenetic tree construction" section) was used as input for phyloFit (PhastCons v1.3 package; parameters: –EM –precision HIGH –subst-mod REV) to estimate neutral branch lengths required by PhastCons. The results from PhastCons and GERP were combined by BEDTools v2.25.0[152] (parameters: "-d 10"). To obtain reliable CNEs, we rigorously filtered the elements with conserved areas that overlapped annotated coding regions using a custom Perl script and BEDTools v2.25.0. To identify CNEs of the 31 species, annotated coding regions in the painted turtle genome were used; to identify CNEs of 25 snake species, the coding region annotations of Diard's blind snake were used. Finally, CNEs with lengths $\geq$ 30 bp were retained for subsequent analyses.

To detect SD-CNEs, we followed the methods described in our previous study[16] to estimate the sequence identity between the aligned ancestral sequences and the CNE sequences of each species. Next, we calculated the global Z-scores following the method described in Roscito et al.[36] SD-CNEs were defined as those with global Z-scores < 0 (Figure S3C).

We associated the SD-CNEs to potential target genes for functional annotation using the regulatory domain concept of GREAT.[177] We defined a basal (promoter-associated) domain, which was 5 kb upstream and 1 kb downstream of the transcription start site. We also defined a distal domain that extended the basal domain up to the basal domain of the next gene or at most 300 kb in either direction. An in-house R script was used to obtain significantly enriched Gene Ontology (GO) or Mouse Genome Informatics (MGI) phenotype ontology terms for the potential target genes of CNEs' through a hypergeometric test (using the phyper v3.6.0 (https://www.r-project.org/) function with lower.tail = F). The p values were corrected using the "Benjamini and Hochberg" method implemented in the R v3.6.0 p.adjust function, and the threshold for significant enrichment was adjusted p value < 0.05 (Figure S3D).

Phylogenetic tree construction

To infer the whole-genome evolutionary tree, orthologous genomic segments from the 31-taxon alignment of each species were concatenated into one sequence using an in-house Perl script and trimmed using trimAl v1.4.rev22 (http://trimal.cgenomics.org). A total of 51,302,441 sites were used to construct the maximum likelihood (ML) phylogenetic tree with IQ-TREE v1.6.8[156,178,179] software using the parameters: "–alrt 3000 –bb 3000 –bnni" (Figure S1B). For the tree constructed using CNEs, the CNE sequences of each species were concatenated and trimmed using the same procedures to build the whole-genome tree. IQ-TREE v1.6.8 software was then used to infer the tree with the parameters "–alrt 5000 –bb 5000 –bnni" (Figure S1E).

To construct the orthologous gene tree, we first used OrthoFinder v2.2.7[151] (with default parameters) to identify one-to-one orthologs shared by 25 snakes; four lizards (common wall lizard[120] European glass lizard, Komodo dragon,[26] and Argentine giant tegu[25]; painted turtle[27]; and Yangtze alligator.[28] A total of 1,980 1:1 single-copy gene families were selected. Coding sequences (CDSs) of

each family were translated to proteins and then aligned using prank v.150803.[157] After trimming ambiguous sites, the alignments were back-translated to corresponding CDS alignments using our in-house Perl script. The CDS of each gene family was concatenated to generate a supergene sequence for tree construction using IQ-TREE v1.6.8 with the parameters: "-alrt 10000 -bb 10000 -bnni" (Figure S1C). To construct the tree on the basis of the 4d sites, we extracted the 4d sites of the 1,980 single-copy genes using an in-house Perl script. A total of 135,405 sites were concatenated and used to build the ML phylogenetic tree with IQ-TREE v1.6.8 software with the parameters "-alrt 10000 -bb 10000 -bnni" (Figure S1D).

Although most of the nodes in concatenated trees (Figures S1B–S1E) were well supported, two nodes in the whole-genome tree, 4d-site tree, and CNE tree were not 100% supported. To investigate phylogenetic discordance across genomic regions, we extracted 51,302 1-kb independent windows from the aligned genomes to infer the coalescent phylogenetic trees on the basis of the whole genome. We also constructed the phylogenetic trees using each single-copy gene. These trees were all inferred and constructed using ASTRAL-III v5.6.3[18] with default parameters.

Approximately 94.40% and 93.49% of the topologies of the genomic orthologous segment trees and gene trees, respectively, were consistent with the concatenated trees. DiscoVista[159] was used to summarize the discordances using the ASTRAL-III coalescent gene tree.

Divergence time estimation

We first used the whole-genome tree (constructed from the WGA dataset) of the 31 species to estimate the divergence times using r8s v1.71.[34] Six time points obtained from Timetree (http://www.timetree.org/) were used to calibrate the tree using the Langley-Fitch method in r8s v1.71 (Figure S1I). Next, we inferred the divergence times using MCMCTREE in PAML v4.9i.[158] The HKY85 model was used with 100,000 iterations and a burn-in of 10,000 iterations. We ran the program twice to check for convergence of the stationary distribution and confirmed the results (Figure S1J).

Using the 31-taxon whole-genome alignment, we estimated DNA evolutionary rates for each snake lineage, the common ancestor of snakes, and six outgroup taxa using r8s v1.71. After taking the body length of each snake (which was used as a dependent variable), we log-transformed each value and found that snake body length was negatively correlated with DNA evolutionary rates (R^2 = -0.57; $P < 0.03$) (Figure S2C).

Identification of structural variants

To identify snake-specific structural variants (SSSVs) in the most recent common ancestor (MRCA) of snakes, we extracted lineage-specific sequences of the snake and outgroup MRCA from the 31-taxon alignment using the findRegionsExclusivelyInGroup program in the Progressive Cactus package. The program defines lineage-specific sequences as sequences that 1) can only be found in the target lineage, 2) do not align to the non-target lineages, and 3) are located in the reconstructed genome of the MRCA of the target lineage.[17]

We defined snake lineage-specific genome sequences ($\geq$ 50 bp) as snake-specific genome insertions and outgroup lineage-specific genome sequences ($\geq$ 50 bp) as snake-specific genome deletions (Figure S2D). These insertions and deletions are all SSSVs (Figure S3A). We defined genes with regulator regions (5 kb flanking the transcriptional start and end sites) or CDSs overlapped with the SSSVs as SSSV-associated genes (Table S3). Protein-coding genes with $\geq$ 100-bp overlap with SSSVs were considered lost or newly evolved genes and used in subsequent analyses. The three-dimensional structure of CC2D2A was predicted using the AlphaFold2 package.[180] The binding sites conveyed by SSSV insertions of *TITF1* and *PDZD7* were predicted by PoSSuM Search v1.3[160] using binding position-specific scoring matrix (PSSM) of green anole. The regulatory potential (RP) scores were calculated by Cistrome Data Browser.[181]

Orthologous gene identification

To avoid the negative effects of low-quality genome annotations on the orthologous gene analysis and reduce the computational time, we only conducted analyses of orthologous genes using genomes of the wall lizard, European glass lizard, Komodo dragon, Argentine giant tegu, viviparous lizard, and sand lizard, as well as 21 snake genomes (with BUSCO scores $\geq$ 87%). The orthologous gene set was established using the reciprocal best hit (RBH) method to obtain more reliable orthogroups. Proteins of these snakes, the wall lizard, Komodo dragon, Argentine giant tegu, viviparous lizard, and sand lizard were aligned to those of European glass lizard with an E-value $\leq$ 1e−5. We extracted the longest transcript (LT) of each coding gene from all selected species to represent each gene. The LT sequences were then translated into proteins. The best hits (E-value $\leq$ 1e−5) for each gene pair were obtained, and gene families of the European glass lizard shared with at least three lizards and 10 snakes were retained. A total of 17,015 orthologous gene families were obtained with ~14,926.81 families per species (Figure S3H).

Identification of newly evolved genes

To detect newly evolved genes in the ancestral branch of snakes, we first selected genes with CDSs that overlapped with snake lineage-specific sequences longer than 100 bp on the basis of the annotation of the Xizang hot-spring snake genome. We then removed all orthologous genes found in the outgroup lineages. Raw Illumina short reads (30× coverage) of four randomly selected snakes (Zong's odd-scaled snake, keeled slug snake, red corn snake, and sunbeam snake) and six randomly selected lizards (Komodo dragon, European glass lizard, green anole, Anan's rock agama, European wall lizard, and viviparous lizard) were mapped to the

Xizang hot-spring snake genome using Burrows-Wheeler Aligner (BWA) v0.7.17-r1188[130] with default parameters. We defined snake newly evolved genes as genes that met the following two criteria: 1) less than 30% of the length of a gene's CDS was covered by at least 3× reads in all lizards, and 2) more than 50% of the length of a gene's CDS was covered by at least 3× reads in at least three snakes. A total of 142 genes met these criteria (Table S3).

Lost gene identification

To identify genes specifically lost in the snake ancestor, we selected genes in which the CDS overlapped at least 100 bp of the outgroup lineage-specific sequences according to the annotations of the European glass lizard genome. Genes with any orthologs in snakes were removed according to the orthologous gene set obtained in the "orthologous gene identification" section. We then aligned the amino acid sequences of the lost genes in European wall lizard (the outgroup species) to all 25 selected snake genomes with tblastn v2.7.1+. Genes with alignment length covering $\geq$ 60% of their sequence and having identity $\geq$ 70% were removed. Next, candidate genes were rechecked using the raw Illumina short reads of five lizards (green anole, Anan's rock agama, European wall lizard, Komodo dragon, and viviparous lizard) and 14 newly sequenced snakes (Table S1). Reads were mapped to the European glass lizard using BWA v0.7.17-r1188[130] with default parameters, and the coverage depth was calculated using BEDTools v2.25.0 (Figure S3B). Snake-specific lost genes were genes that met one of the two following sets of criteria. The first group of snake-specific lost genes comprised genes with 1) more than 30% of the length of a gene's CDS covered by at least 2× reads in at least two lizards, and 2) 100% of the length of a genes CDS covered by 0× reads in all snakes. The second group of snake-specific lost genes comprised genes with 1) more than 30% of the length of a gene's CDS covered by at least 2× reads in at least two lizards, and 2) more than 50% of the length of a gene's CDS covered by 0× reads in at least 13 snakes. Overall, 156 genes were considered lost in the MRCA of snakes (Table S3).

Enrichment analyses of snake-specific newly evolved and lost genes were conducted respectively using a hypergeometric test via an in-house R script as described in "Analysis of conserved non-coding elements (CNEs)" (Figures S3E and S3F).

Positively selected gene (PSG) identification

We ran the codeml program in the PAML package v4.9i with the branch-site model for each of the 17,015 gene families to identify potential PSGs in the ancestral branch of the snake lineage. First, we used OrthoFinder v2.2.7 to generate single-copy genes and IQ-TREE v1.6.8 to construct the phylogenetic tree of the 27 species for PAML v4.9i. Next, we conducted a likelihood ratio test (LRT) to compare the alternative model that allowed sites to be selected positively on the foreground branch with the null model where sites were subjected to either neutral or purifying selection. The p values were calculated on the basis of Chi-square statistics with one degree of freedom. PSGs with p value $\leq$ 0.05 and with at least one positive site were selected (Table S3). Enriched MGI, GO, and KEGG terms were obtained using hypergeometric tests as described above (see the "analysis of conserved non-coding elements (CNEs)" section) (Table S3).

Rapidly evolving gene (REG) identification

REGs were detected using the same orthologous gene set and tree topology used for PSG identification. We used the branch model in PAML v4.9i, which assumes a null model (model=0, NSsites=0) in which all branches evolve at the same rate, and an alternative model (model=2, NSsites=0), in which the foreground branch has a different rate of evolution. The two models were discriminated using a LRT (df = 1) to obtain the p values. Genes with dS (the rate of synonymous substitution) > 3 or $\omega 0 > 5$[182] were removed, and the remaining with p values < 0.05 were considered REGs in the foreground branch. We detected REGs at the same branches as PSGs (see above). Identified REGs and enriched MGI, GO, and KEGG terms identified by hypergeometric tests (see the "analysis of conserved non-coding elements (CNEs)" section) are listed in Table S3.

Relaxed selection analysis

HOX genes were selected from the orthologous gene set used for PSG identification. We identified genes under relaxed selection using HyPhy v2.5.20[172,183] with default parameters and the tree topology used for the identification of PSGs. The p value for each gene was corrected using the "Benjamini and Hochberg" method, and genes under significantly relaxed selection were those with adjusted p value < 0.05.

Transcriptomic analysis of genes related to snake traits

For transcriptomic comparisons with reference genomes, raw RNA-seq reads were filtered using fastp v0.23.1 with the parameters: "-g -q 5 -u 50 -n 15 -l 150 –min_trim_length 10 –overlap_diff_limit 1 –overlap_diff_percent_limit 10". High-quality reads were mapped to the corresponding reference genomes with HISAT2 v2-2.1.0[161] (default parameters). Mapped read counts and fragments per kilobase per million mapped reads (FPKM) were computed using StringTie v2.0.4 (default parameters). To compare gene expression differences across species, the FPKM was quantile-normalized using the R package preprocessCore v1.48.0. Differentially expressed genes (DEGs) were defined as those with fold-change > 2 and Student's *t*-test p value < 0.05. We also conducted WGCNA to identify the most significantly associated genes among the selected samples (including 21 RNA samples (covering seven tissues) in a previous study of many-banded kraits[118]). The quantile-normalized FPKMs were input into the WGCNA R package v1.70.3.[162] Candidate genes were identified using the following criteria: Pearson correlations between module eigengenes (ME)

and gene-expression values (module membership, MM) $\geq$ 0.8 and correlations between gene-expression values and samples (gene significance, GS) $\geq$ 0.8.

For transcriptomic analyses of lung and vertebral column development in *P. dhumnades*, raw RNA-seq reads were filtered using fastp v0.23.1 as described above (see the "RNA extraction and sequencing" section). The clean reads were *de novo*-assembled using Trinity v2.8.5[170] with the parameters "-group_pairs_distance 230 –min_contig_length 350 –min_glue 4". The raw Trinity transcripts were filtered using DETONATE v1.11[167] with contig_impact_score > 0. CD-HIT v4.6 (http://cd-hit.org) was then used to remove redundancies in the remaining transcripts with the parameter -c 0.90. To quantify the expression profiles of unigenes, we used the align_and_estimate_abundance.pl script of Trinity v2.8.5 with bowtie2 v2.3.4.3[129] and RSEM v1.2.8[165] to obtain the read counts and FPKM values. DEGs were identified using edgeR v3.28.1.[169]

Comparative transcriptomic analysis of the embryonic eye in red corn snake (10, 30, and 50 days post oviposition) and human embryos 4.7 to 8 weeks post-conception (Table S1) was conducted as described above. We used eight housekeeping genes (*PSMB4*, *VCP*, *CHMP2A*, *RAB7A*, *EMC7*, *GPI*, *PSMB2*, and *VPS29*) recommended in a previous study[184] to normalize the FPKM values and compare gene expression levels across distantly related species (red corn snake vs. human).

Blind snake adaptive evolution analysis

Comparative morphological analyses of the eyes were conducted to understand differences between Diard's blind snake and non-blind snakes (Asian vine snake and keeled slug snake) using paraffin sections (Figure S5E).

To explore the genetic basis underlying these adaptive phenotypes, we investigated the evolution of both coding and non-coding elements in the two genomes. We extracted gene families that contained orthologs shared by Diard's blind snake and the slender blind snake from the orthologous gene families obtained in the "orthologous gene identification" section. The PSGs and REGs of the most common ancestor of the two snakes were identified using the same methods described in "positively selected gene (PSG) identification" and "rapidly evolving gene (REG) identification." A total of 171 genes were determined to have undergone significant positive selection, and 432 genes in the blind snake ancestor were REGs (Table S4). Significantly enriched GO terms (p value < 0.01) of the REGs were obtained using hypergeometric tests in our custom R script as described above in "analysis of conserved non-coding elements (CNEs)" and then clustered using REVIGO[163] (Figure S5G). According to the methods described in "newly evolved and lost gene identification", we scanned the most common ancestor genome of blind snakes to identify potentially lost coding genes (Table S4). Genes showing CDS overlap of at least 100 bp with lineage-specific sequences (identified by Progressive Cactus) of 29 non-blind reptiles (including 23 snakes, four lizards, the Yangtze alligator, and the painted turtle from the 31-taxon alignment) according to Xizang hot-spring snake genome annotations and that were lost in both blind snakes according to the "orthologous gene identification" were considered candidate-specific lost genes in blind snakes. We then aligned the amino acid sequences of the blind snake lost genes in Xizang hot-spring snake to the two blind snake genomes with tblastn v2.7.1+. Genes with alignment length covering $\geq$ 60% of their sequence and having identity $\geq$ 70% were removed. The raw Illumina short reads of the two blind snakes were further mapped to the genome of the Xizang hot-spring snake to confirm that the remained genes had been lost using BWA v0.7.17-r1188[130] with default parameters. We defined blind snake-specific lost genes as genes that met one of the two following criteria: 1) more than 50% of the length of the CDSs covered by 0× reads in both blind snakes, and 2) less than 20% of the length of the CDSs covered by at least 2× reads in both blind snakes (Table S4). Blind snake lineage-diverged CNEs were detected using the "Forward Genomics" branch method[164] with a p value < 1e−5. These were associated with potential target genes as described in "analysis of conserved non-coding elements (CNEs)" (Figure S5F).

We used eye RNA samples of five snake species (Diard's blind snake, Asian vine snake, keeled slug snake, Xizang hot-spring snake, and western rattlesnake), with at least three biological replicates for each species (Table S1), to explore the functional reduction of the eyes in blind snakes. RNA isolation, library construction, and sequencing were performed by Novogene Company (China). Reference-based transcript assembly was conducted using Bowtie2 v2.3.4.3[129] and the gene expression profiles were determined using RSEM v1.2.8.[165] To identify DEGs, we first built a set of 9,579 single-copy genes using RBH and quantile-normalized the FPKM values of all genes. These genes in Diard's blind snake were compared with those in other snakes using Student's *t*-tests (p value < 0.05). Genes with down-regulated expression in all four comparisons were retained.

Evolutionary analysis of infrared sensing

Pitvipers, pythons, and some infrared-sensitive boas possess pit organs or infrared receptors and are thus capable of sensing infrared signals.[185] To investigate the genetic mechanism underpinning these specific phenotypes, we first identified 220 temperature-sensing related and 14 trigeminal development-related genes from the orthologous gene set obtained in the "orthologous gene identification" section according to the GO database. The most common ancestor of the pitviper, the Burmese python, and the boa constrictor branches were tested to identify PSGs and REGs (Table S5) among genes using the branch-site model and branch model (as described in the "positively selected gene (PSG) identification" and "rapidly evolving gene (REG) identification" sections, respectively). The ancestral protein sequences of all species were constructed using prank v.150803.[157] The genes with convergent amino acid sites in pitvipers (Shedao pitviper, five-pacer viper), Burmese python, and boa constrictor were selected using an in-house Perl script. We defined convergent sites on the basis of the following criteria: 1) the sites were the same in the target species but different in other species and the most recent common ancestor of all species and 2) sites had PROVEAN scores < -1.7. We identified the *TRPA1* gene, which contained at least one such convergent amino acid site. Using the "Forward Genomics" branch method,[164]

we conducted evolutionary analyses of CNEs to identify diverged CNEs (p value < 5e–3); We then conducted an enrichment analysis as described in "Analysis of conserved non-coding elements (CNEs)". To further investigate the function of the pit organs, we conducted a comparative transcriptomics analysis on the tissues of seven brown-spotted pitvipers (Table S1). Total RNA extraction, sequencing, and quantitation were performed following the methods described in the "RNA extraction and sequencing" and "transcriptomic analysis of genes related to snake traits" sections using the brown-spotted pitviper genome as a reference. The DEGs were identified using the following criteria in DESeq2 v1.26.0[168]: $\log_2$-transformed fold change > 2 and adjusted p value < 0.05.

Functional experiment of the *PTCH1* gene

Point mutated mice with R946del-L947del-R948del mutations of murine *PTCH1* were designed and generated by the Shanghai Model Organisms Center, Inc. (Shanghai, China). Briefly, Cas9 mRNA was *in vitro* transcribed using the mMESSAGE mMACHINE T7 Ultra Kit (Ambion, TX, USA) according to the manufacturer's instructions; the mRNA was then purified using the MEGAclear™ Kit (ThermoFisher, USA). The 5′-GGCTGAGAAGTAAGTAGCAC-3′ sequence was selected as the Cas9-targeted guide RNA (sgRNA), *in vitro* transcribed using the MEGAshortscript Kit (ThermoFisher, USA), and subsequently purified using the MEGAclear™ Kit. The transcribed Cas9 mRNA and sgRNA, as well as a 120-bp single-stranded oligodeoxynucleotide (ssODN) were co-injected into zygotes of C57BL/6J mice. Obtained F0 mice (C57) were validated by PCR and sequenced using the following primer pairs: F1, 5′-CGCCCTGAATGACCTCTGTT-3′; R1, 5′-GGAAGAGGAGAGCACGGATG-3′. F0 mice with expected point mutations were crossbred with wild-type (WT) C57BL/6J mice to produce F_1 mice. The genotypes of F_1 mice were identified by PCR and confirmed by sequencing. The ssODN sequence used for point mutation mice generation was 5′-AACATCCGGCCTCACCGGCCGGAGTGGGTC CATGACAAAGCCGACTACATGCCAGAGACCAGTAAGTAGCACTCCCATCCTGAGAGGCTGGGAGCTCTTGGAGGCTCCTTCAGTC ACTGC-3′.

1-week-old *PTCH1* WT mice (n = 13) and *PTCH1*-mutated mice with the snake deletion (n = 13) were randomly selected. The relative body mass and length (from neck to tail root) of each mouse were measured and then normalized by dividing by the averages. Independent Student's *t*-tests were conducted to compare the values between the WT mice and the *PTCH1*-mutated mice. The relative body mass of the two groups of eight-week-old mice (n = 6) was also measured using the same method.

After sacrificing the mice, the shin bone of WT and *PTCH1*-mutated mice (n = 4) was collected and fixed in 10% neutral-buffered formalin for 24 h, and the excess soft tissue was removed. The shin bone was scanned using a Skyscan1276 micro-CT instrument (BrukermicroCT, Kontich, Belgium) with the following settings: source voltage, 55 kV; source current, 200 μA; Al, 0.25 mm filter; pixel size, 6 μm; and rotation step, 0.3 degrees. The images were then reconstructed using NRecon software (Bruker microCT, Kontich, Belgium) with the following settings: ring artifact correction, 5; smoothing, 3; and beam hardening correction, 30%. Trabecular bone volume (BV), tissue volume (TV), bone volume per tissue volume (BV/TV), trabecular number (Tb.N), and trabecular thickness (Tb.Th) were analyzed using the program CTAn (BrukermicroCT, Kontich, Belgium).

Adult (approximately 9 weeks old) WT *PTCH1* mice (n = 8) and *PTCH1*-mutated mice (n = 8) were anesthetized using 10% chloral hydrate (0.04 ml/10 g). X-ray images of the whole body of both groups of mice (n = 8) were acquired using high-resolution *in vivo* x-ray micro-tomography (SkyScan 1276, Bruker, Germany). ImageJ software (version v1.53k; National Institutes of Health, USA) was used to obtain the following measurements: body length (from nose to tail root), tail length, phalanx length of all digits on the right limbs, and right limb length. We took measurements of each sample three times, and the average value was obtained. The average values of the tail length, phalanx length, and right limb length were divided by corresponding body lengths to obtain the relative values.

To investigate whether the expressions of the genes of the mutated mice were altered during limb development, we sequenced forelimb and hindlimb buds of the embryo at E11.5 in both WT and mutated mice (three for each) to obtain transcriptomic data. RNA extraction, sequencing, and quantitation were performed as described in the "RNA extraction and sequencing" and "transcriptomic analysis of genes related to snake traits" sections using the mm10 mouse genome as a reference. The DESeq2 v1.26.0 package[168] was used to identify DEGs.

Chitinase enzymatic activity measurement

The CDSs of the *CHIA* gene in Diard's blind snake and Burmese python were codon-optimized using *Escherichia coli*; they were then chemically synthesized and sub-cloned into the pET-28a (+) plasmid using the *Nde*I–*Xho*I restriction sites. The N-terminal HIS target was then added to the sequence to indicate its expression.

The recombinant expression plasmids were transformed into *E. coli* BL21 (DE3), and the cells were spread on Luria broth agar plates containing 30 μg/ml ampicillin; they were then incubated upside down at 37°C overnight. Next, the cells were harvested via centrifugation for 20 min at 6,000 rpm and 4°C, and the cell pellets were collected for further purification. The pellets were resuspended in lysis buffer, followed by disruption using an ultrasonic homogenizer. The lysates were centrifuged for 30 min at 12,000 rpm and 4°C. The supernatant was collected and incubated at 80°C for 30 min, followed by centrifugation at 12,000 rpm for 30 min and 4°C to remove denatured protein. The supernatant was applied to nickel-charged (Ni-NTA) beads for affinity purification. The column was washed twice with two column volumes of wash buffer (8 M urea, 50 mM Tris, 300 mM NaCl, 20/50 mM imidazole, pH 8.0), followed by one column volume of elution buffer to recover the target protein. Proteins were then purified by gel filtration, dialyzed into a storage buffer (50 mM Tris, 300 mM NaCl, 0.1% sarkosyl, 2 mM DTT, pH 8.0), and lyophilized.

The chitinase enzymatic activities of the snakes were measured using a Chitinase Assay Kit (ADS-302-F, Jiangsu Kote Biotechnology Co., China) using the purified chitinase proteins according to the standard protocol. Measurements of the chitinase enzymatic

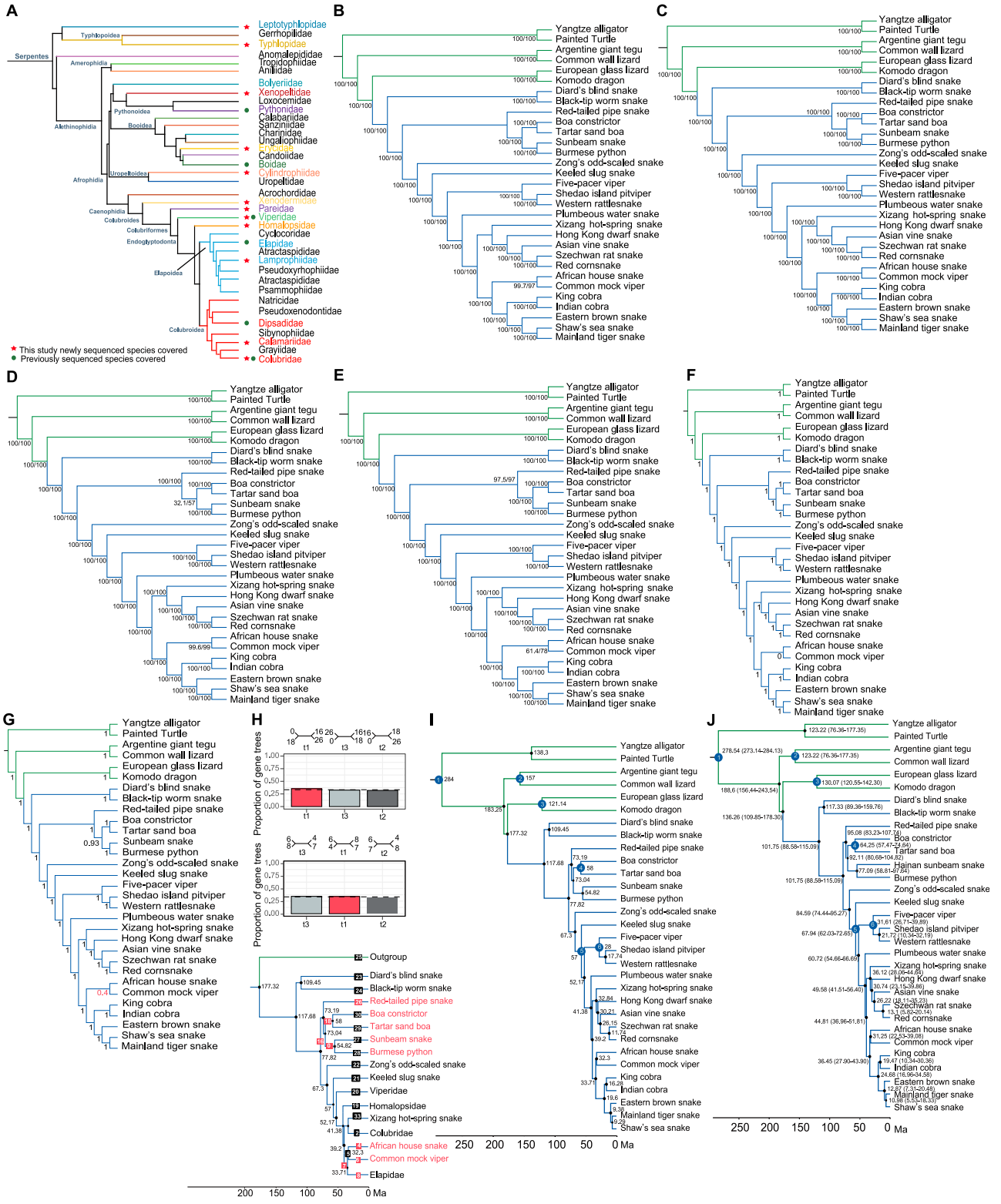
activity of each snake were taken five times. Differences in chitinase activity between groups were evaluated using Student's *t*-tests (Figure 5D).

QUANTIFICATION AND STATISTICAL ANALYSIS

The statistical methods used in this study are indicated in the figures, figure legends, and methods. Statistical analyses were performed using R software v3.6.0 and the DEseq2 v1.26.0 and edgeR v3.28.1 packages.

CellPress
OPEN ACCESS

Supplemental figures



(legend on next page)

Figure S1. Phylogenetic and divergent time tree of snakes, related to Figure 1

(A) Phylogenetic tree distributions of selected snake species, topology inferred from a previous study.[173]

(B) Maximum likelihood (ML) phylogenetic tree inferred from the 31-taxon whole-genome alignments. Ultrafast bootstraps were repeated 3,000 times, with NNI UFBoot tree optimization and SH-like approximate likelihood ratio test (SH-aLRT) performed. SH-aLRT support rate/ultrafast bootstrap support rate is indicated at each node.

(C) Inferred ML phylogenetic tree using orthologous genes. 1,980 1:1 orthologous genes were concatenated to a super gene sequence for constructing the ML phylogenetic tree using IQ-TREE. 10,000 ultrafast bootstraps were carried out and the two support rates were marked at each node.

(D) ML phylogenetic tree inferred from 4d sites. 4d sites were extracted from the orthologous genes and taken as input for IQ-TREE to infer the ML phylogenetic tree. The two supports were computed by 10,000 ultrafast bootstraps and SH-aLRT.

(E) Inferred ML phylogenetic tree of conserved non-coding elements (CNEs). All CNEs were identified and concatenated into a single sequence for the ML phylogenetic tree inferred using 5,000 ultrafast bootstraps and SH-aLRT.

(F) Coalescent phylogenetic tree inferred from 51,302 1-kb orthologous genomic segments using ASTRAL-III.

(G) Coalescent phylogenetic tree inferred from 1,980 1:1 orthologous genes using ASTRAL-III. The poorly supported nodes are indicated in red.

(H) DicoVista gene tree topologies frequency analysis. The frequency of three topologies (t1–t3) is shown, and the red is the main topology. The divergence time (million years ago [mya]) of the species is estimated using r8s and the incongruent clades are in red.

(I) Divergence time of the 31 species. r8s estimated divergence time using the whole-genome alignments. The estimated divergence time (million years ago [mya]) is labeled on each node and 6 calibrating nodes used are marked.

(J) MCMCTree estimated divergence time of the 31 species using 4d sites. Six calibrating nodes used are marked and the estimated divergence time represented in million years is labeled.

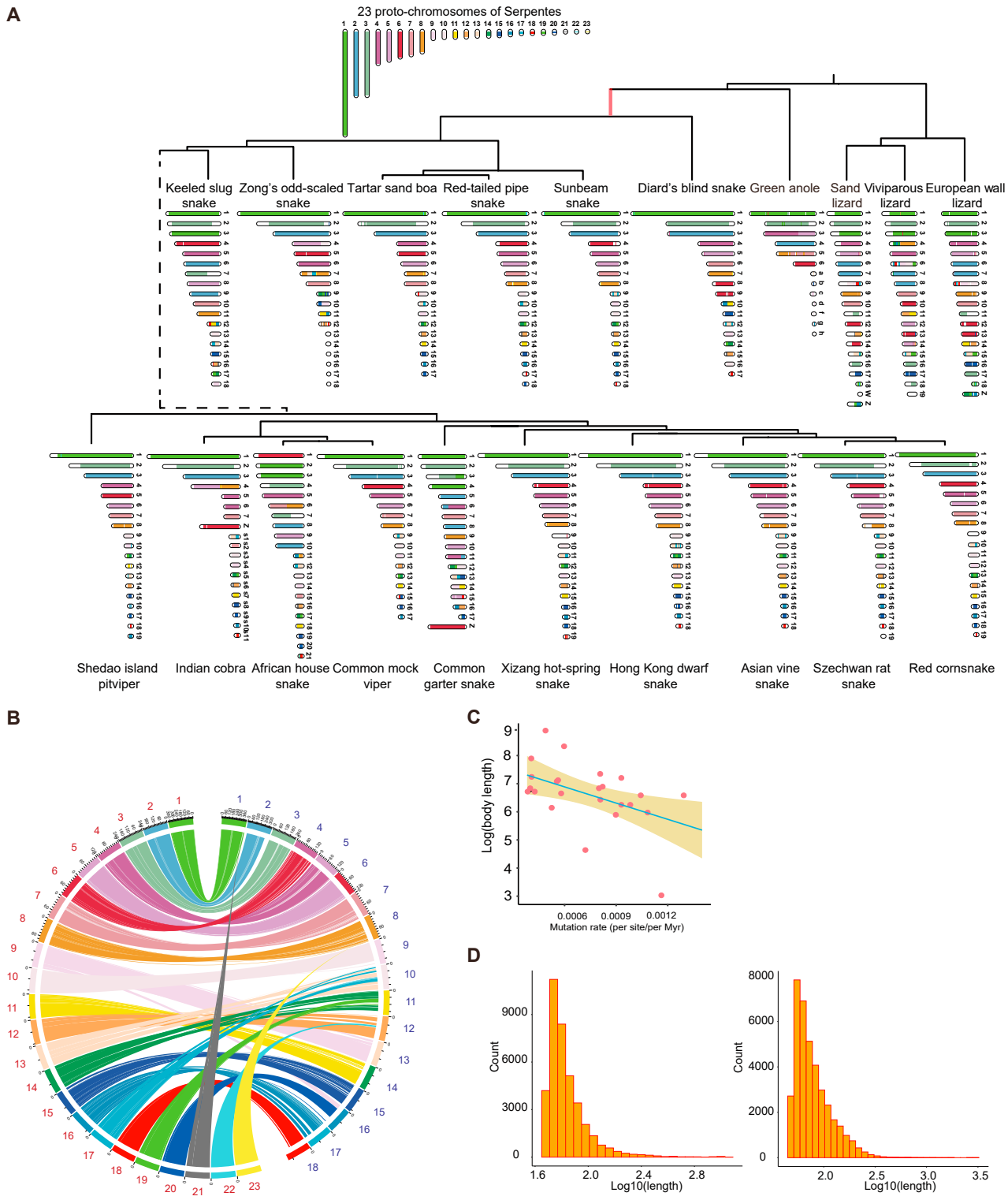


Figure S2. Snake genome features, related to Figure 2

(A) Evolution of chromosomes in snakes. In total, 23 proto-chromosomes of Serpentes were reconstructed using four lizards as outgroup.

(B) Circos plots showing conserved synteny between the hypothesized Serpentes ancestor: Serpentes (red) and Hong Kong dwarf snake (Csep-blue).

(C) Snake body lengths were significantly negatively correlated with the genome evolutionary rate (correlation coefficient = −0.50, p value = 0.011).

(D) Length distribution of snake ancestor gain and lost genome segments. Length distribution of snake ancestor unique genome segments (left). Length distribution of snake ancestor lost genome segments (right).

(legend on next page)

Figure S3. Snake-specific genome structural variations (SSSVs), snake-diverged conserved non-coding elements (SD-CNEs that were diverged in snakes but were still conserved in the outgroup) (SD-CNEs), and orthologous genes used for evolutionary analysis, related to Figure 3 and STAR Methods

(A) SSSVs distribution in different genome regions.

(B) Genomic reads coverage of the snake-specific lost gene *GHRL* across four lizards (green anole [Acar], Anan's rock agama [Lsac], Komodo dragon [Vkom], and viviparous lizard [Zviv]), and 14 newly sequenced snakes.

(C) *Z*-score cut off of SD-CNEs.

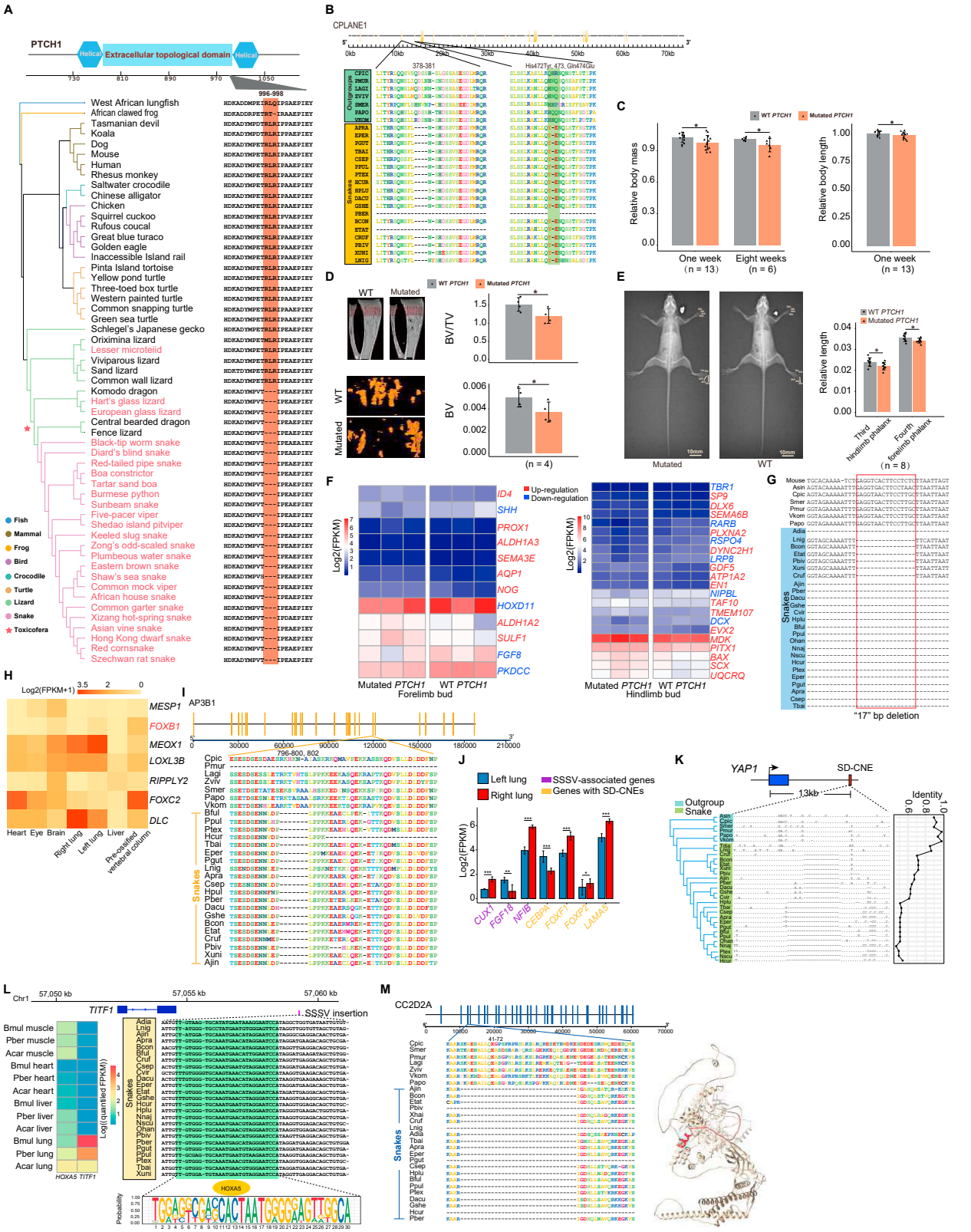
(D) The enriched MGI terms that related to snake phenotypes for genes with SD-CNE (adjusted p value < 0.05). SD-CNEs (dots) are located around the transcription start site of the genes enriched in eye, eyelid, ear, lung, mandible, maxillary, sternum, and tooth development-related terms.

(E) Enriched GO terms (adjusted p value < 0.05) for the newly evolved coding genes in snakes. Those related to dietary excess, protein digestion, and olfactory receptor activity were colored in red, blue, and brown, respectively.

(F) Enriched GO terms (adjusted p value < 0.05) for snake-specific coding genes that were lost. Red, blue, orange, and green were used to indicate the terms involved in vision, lens development, appetite, and bile acid biosynthetic process, respectively.

(G) SSSV deletion in the potential regulatory region of *RP1*.

(H) Counts of orthologous genes used for evolutionary analysis among 27 selected species.



(legend on next page)

Figure S4. Genetic bases of snake limb loss, body elongation, lung asymmetry, and eye structure simplification, related to Figures 3 and 4

(A) Multiple alignments of PTCH1. Fish, mammals, frogs, birds, crocodiles, turtles, lizards, and snakes are indicated in different colors.

(B) CPLANE1 multiple alignments. The gene structure is outlined by the black line with yellow blocks representing CDSs. The position span is obtained based on the European glass lizard reference genome. The "378–381" and "473" show the positions of snake-specific missing amino acids. The "His472Try" and "Gln474Glu" indicate the two snake-specific amino acid mutations.

(C) Relative body mass and relative body length of WT mice and *PTCH1*-mutated mice with snake deletion. Mean ± SD is shown by error bar.

(D) Microstructural characteristics of the proximal shinbone (colored in red) in 1-week-old WT mice and *PTCH1*-mutated mice with snake deletion. The trabecular bone volume (BV) and bone volume per tissue volume (BV/TV) of WT type and *PTCH1*-mutated mice. Significantly smaller *$p < 0.05$ (independent Student's t test). Mean ± SD is shown by error bar.

(E) Statistics of phalanx relative length (absolute length divided by body length) in WT mice and *PTCH1*-mutated mice with snake deletion. Mutated mice had significantly shorter fourth forelimb phalanx ($n = 8$, *$p < 0.05$, independent Student's t test) and third hindlimb phalanx ($n = 8$, *$p < 0.05$, independent Student's t test). Mean ± SD is shown by error bar.

(F) Expressions of genes related to limb development in stage E11.5 forelimb bud (left) and hindlimb (right) of WT mice and *PTCH1*-mutated mice with snake deletion.

(G) The 17-bp deletion of snake ZRS enhancer in our newly sequenced snakes.

(H) Expression levels of somite development-related genes (under positive selection) in *P. dhumnades* (40 days post oviposition).

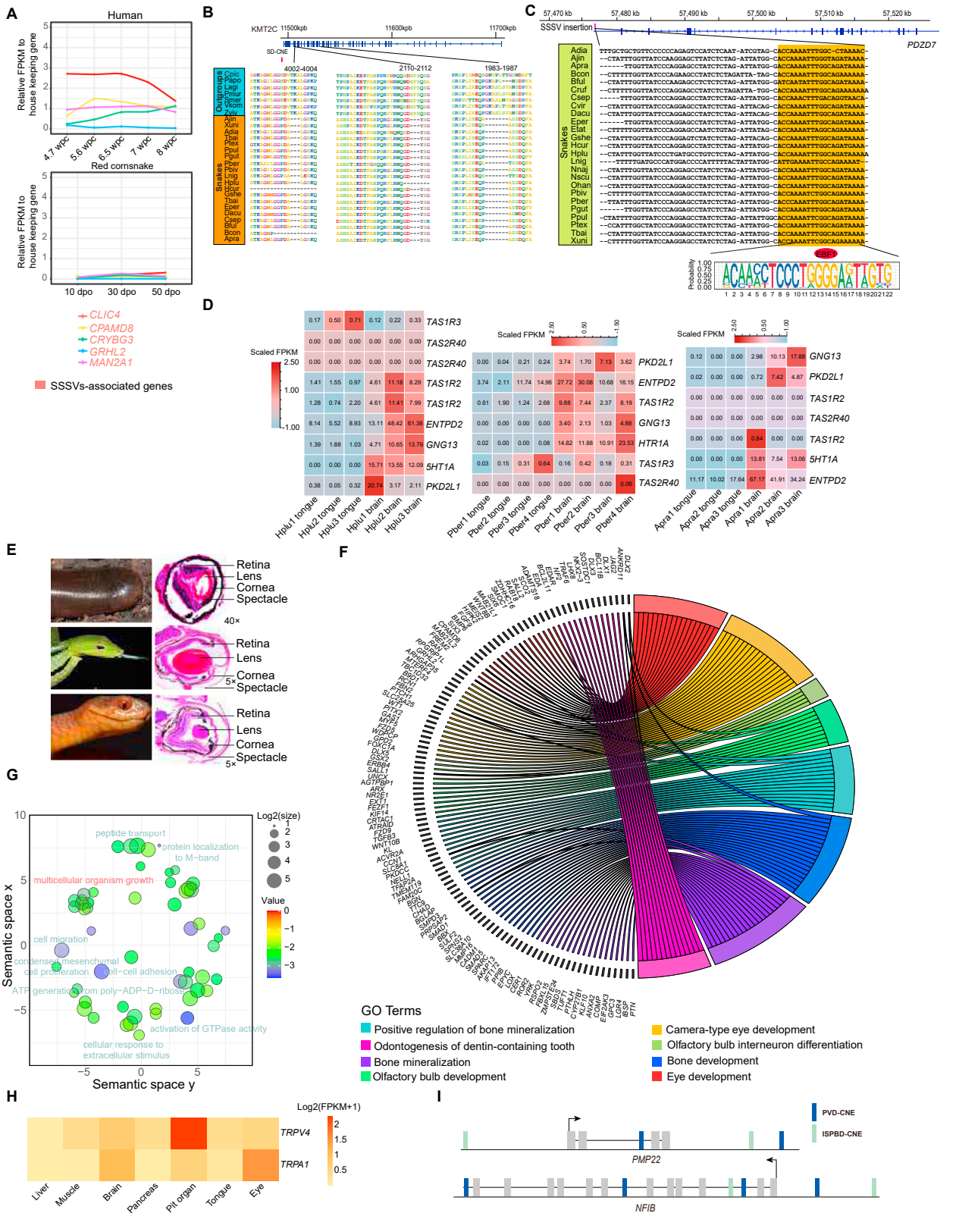
(I) AP3B1 multiple alignments of snakes and outgroup. The gene structure is outlined by the black line with orange sticks representing CDSs and the numbers indicating the position span according to the European glass lizard reference genome. The positions of snake lost amino acids residues are indicated by the interval number "796–800."and number "802".

(J) Example of differentially expressed genes (DEGs) between embryonic left and right lungs (40 days post oviposition) of *P. dhumnades*. The significance is indicated as *$p < 0.05$, **$p < 0.01$, or ***$p < 0.001$. Mean ± SD is shown by error bar.

(K) The SD-CNE located at ~13 kb downstream of *YAP1*. The identities (compared with the 31 reptile species ancestor sequence) of the SD-CNE were decreasing along the evolution of snakes.

(L) A SSSV inserted to the 5 kb upstream of the *TITF1* transcriptional start site. The SSSV was predicted to have one binding site of HOXA5 transcription factor. Heatmap (left) displays the expressions (log_{10}(quantiled FPKMs)) of *HOXA5* and *TITF1* in lung, liver, heart, and muscle of green anole (Acar), keeled slug snake (Pber), and many-banded krait (Bmul). *TITF1* and *HOXA5* were highly expressed in the lung of all species but the expressions of *TITF1* were significantly higher in the two snakes than in green anole (Student's t test p values: 0.0048 [Pber] and 0.0077 [Bmul]).

(M) Snake-specific amino acid residues loss of CC2D2A. The black line layouts the gene structure with blue sticks representing CDSs and the numbers indicating the position span based on the European glass lizard reference genome. The interval number "41–74" points out the sites of lost amino acid residues in CC2D2A based on European glass lizard sequence. The lost amino acid residues were indicated in red on 3D structure.



(legend on next page)

Figure S5. Genome evolution of snake sensory system, related to Figures 4 and 5

(A) Expression patterns of genes associated with eye development in human embryos at different embryonic stages (4.7–8 weeks post conception [wpc]) and red cornsnake embryos at 10, 30, and 50 days post oviposition (dpo).

(B) KMT2C specific amino acids residue lost in snakes and its associated snakes divergent CNE (SD-CNE). The black line portrays gene structure with blue blocks representing CDSs and the numbers indicate the exact positions on the European glass lizard reference genome. Three segments of snake-specific amino acid residue loss are manifested by internal numbers "4002–4004," "2110–2112," and "1983–1987." One SD-CNE located at 3′ regulate region of *KMT2C* is marked by a red block.

(C) Alignment of SSSV that inserted into the 5 kb upstream of *PDZD7* transcription start site. This SSSV conveys a new EBF1 binding site with a regulatory potential (RP) score: 0.78.

(D) Taste transduction involved genes were expressed in tongue and brain of three snakes (plumbeous water snake [Hplu], keeled slug snake [Pber], and Asian vine snake [Apra]). The numbers in cells indicate the scaled FPKM values.

(E) Comparison of eye structures between Diard's blind snake and Asian vine snake, keeled slug snake. The magnifications were marked as "*×" at the bottom-right corner of each histological sections.

(F) GOChord plot (produced by GOplot package) of blind snake divergent CNEs involved coding gene-enriched GO terms (p value < 0.05). Left half of GOChord displayed genes of different GO terms and the right showed the GO term descriptions. Each gene was linked to a GO term by the colored bands.

(G) REVIGO clusters of significantly overrepresented (p value < 0.01) GO terms for blind snake REGs.

(H) Expression levels of *TRPV4* and *TRPA1* in seven tissues of brown spotted pitviper.

(I) Genome locations of pitviper diverged conserved non-coding elements (PVD-CNEs), and the infrared-sensitive python and boa divergent conserved non-coding elements (ISPBD-CNEs) in *PMP22* and *NFIB*.