GENETICS

Genomic insights into evolution of parthenogenesis and triploidy in the flowerpot snake

Yunyun Lv[1,2]†, Wei Wu[1]†, Jin-Long Ren[1], Matthew K. Fujita[3], Menghuan Song[1], Zeng Wang[1,4], Ke Jiang[1], Dechun Jiang[1], Chaochao Yan[1], Changjun Peng[1], Zhongliang Peng[1,4], Jia-Tang Li[1,4,5]*

The flowerpot snake (*Indotyphlops braminus*) is the only known parthenogenetic and triploid species within Serpentes. However, the genetic basis underlying this phenomenon remains unresolved. We investigated the genomic complexities of this rare all-female triploid reptile. On the basis of the newly assembled genome, we revealed 40 chromosomes grouped into three subgenomes (A, B, and C). Comparative genomic analysis with related diploid species revealed a chromosome fusion event in ancestral genomes. This event shaped the unique genetic landscape of the flowerpot snake. We examined gene expression specificity in ovarian tissues and identified pathways essential for DNA replication and repair. Our findings suggest a potential mechanism of homologous chromosome pairing during meiosis in allopolyploid parthenogenesis. This research provides insights into the evolutionary adaptations and genetic mechanisms underlying parthenogenesis in reptilian species, challenging traditional views on reproductive strategies and genomic evolution in asexual organisms.

INTRODUCTION

Sexual reproduction is the dominant form of reproduction in the animal kingdom, but parthenogenetic species, such as reptiles, are common in certain animal groups (*1*). Despite previous studies of the parthenogenetic characteristics of numerous lizards and snakes, including their karyotypes (*2*), cellular mechanisms (*3*), and comparative fitness (*4*), the evolution of genomes in parthenogenetic species remains largely unexplored. Detailed studies on the genomic composition, gene expression patterns, and population genomics of parthenogenetic species are needed to establish whether they represent an evolutionary dead-end as a result of the accumulation of deleterious mutations, leading to their eventual extinction (*5*, *6*). In addition, many parthenogenetic reptile species are hybrid polyploids, the formation of which requires overcoming incompatibilities between different genomes and heterogeneity in gene expression (*7*). However, how parthenogenetic species overcome these obstacles remains unknown. Premeiotic endoreplication and sister chromosome pairing during meiosis have been proposed as a mechanism of meiosis in parthenogenetic polyploid species (*3*, *8*, *9*), but whether this mechanism is unique and applicable to all parthenogenetic species remains unexplored.

In this study, we focused on the snake species *Indotyphlops braminus* (commonly known as the flowerpot snake or Brahminy blindsnake; Fig. 1A and fig. S1), a triploid parthenogenetic species in snakes with a global distribution (fig. S2) (*10*). Although it was found as early as 1987 that all individuals in its natural populations have a female triploid karyotype (*11*), a detailed analysis of its genomic composition has yet to be conducted. In addition, we analyzed the genomic characteristics of two sexual diploid species, the closely related *Argyrophis diardii* (Diard's blindsnake, fig. S1) and the distantly related *Ptyas dhumnades* (Cantor's rat snake, fig. S1), to investigate the unique evolutionary characteristics of the flowerpot snake and advance evolutionary theories regarding the evolutionary mechanisms of parthenogenesis. We collected ovarian tissues from flowerpot snakes and Diard's blindsnakes during their reproductive seasons (fig. S3), as well as other tissues from the flowerpot snake (table S1), to investigate gene expression specificity. Last, we obtained genome data from various locations (table S2) to study the mechanism of homologous chromosome pairing in the flowerpot snake and revealed a previously unexplored evolutionary mechanism in parthenogenetic species.

RESULTS AND DISCUSSION

The specificity of the genome composition of the flowerpot Snake

We assembled the genomes of the flowerpot snake, Diard's blindsnake, and Cantor's rat snake using PacBio high-fidelity (HiFi) sequencing reads. The genome assembly sizes (flowerpot snake: ~5.40 Gb, Diard's blindsnake: ~1.86 Gb, and Cantor's rat snake: ~1.68 Gb) fall within the estimated size ranges for these three species (figs. S4 to S6) based on different models. We confirmed that the assembled genomes were free of contamination by analyzing the guanine-cytosine (GC) distribution and sequencing depth distribution in 10-kb windows (figs. S7 to S9). The GC distribution and sequencing depth distribution were random.

For the flowerpot snake, we assembled 40 chromosomes based on chromosome territories and interaction intensity (fig. S10). Each chromosome showed high continuity with relatively few gaps (table S3). However, this chromosome number was inconsistent with the previously reported count on the basis of karyotype observations (42-chromosome karyotype) (*2*). We investigated the cause of the discrepancy between the assembled chromosome count and the karyotype by examining the average sequencing depth [Fig. 1B and fig. S11 (top)] and SNP density (number/5 Mb) on each chromosome (fig. S11, bottom). The results revealed that chromosomes 17 and 37 had roughly twice the coverage depth compared with other

[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, China. [2]Key Laboratory of Sichuan Province for Fishes Conservation and Utilization in the Upper Reaches of the Yangtze River, College of Life Science, Neijiang Normal University, Neijiang, China. [3]Amphibian and Reptile Diversity Research Center and Department of Biology, University of Texas at Arlington, Arlington, TX, USA. [4]University of Chinese Academy of Sciences, Beijing, China. [5]Southeast Asia Biodiversity Research Institute, Chinese Academy of Sciences, Nay Pyi Taw, Myanmar.
*Corresponding author. Email: lijt@cib.ac.cn
†These authors contributed equally to this work.

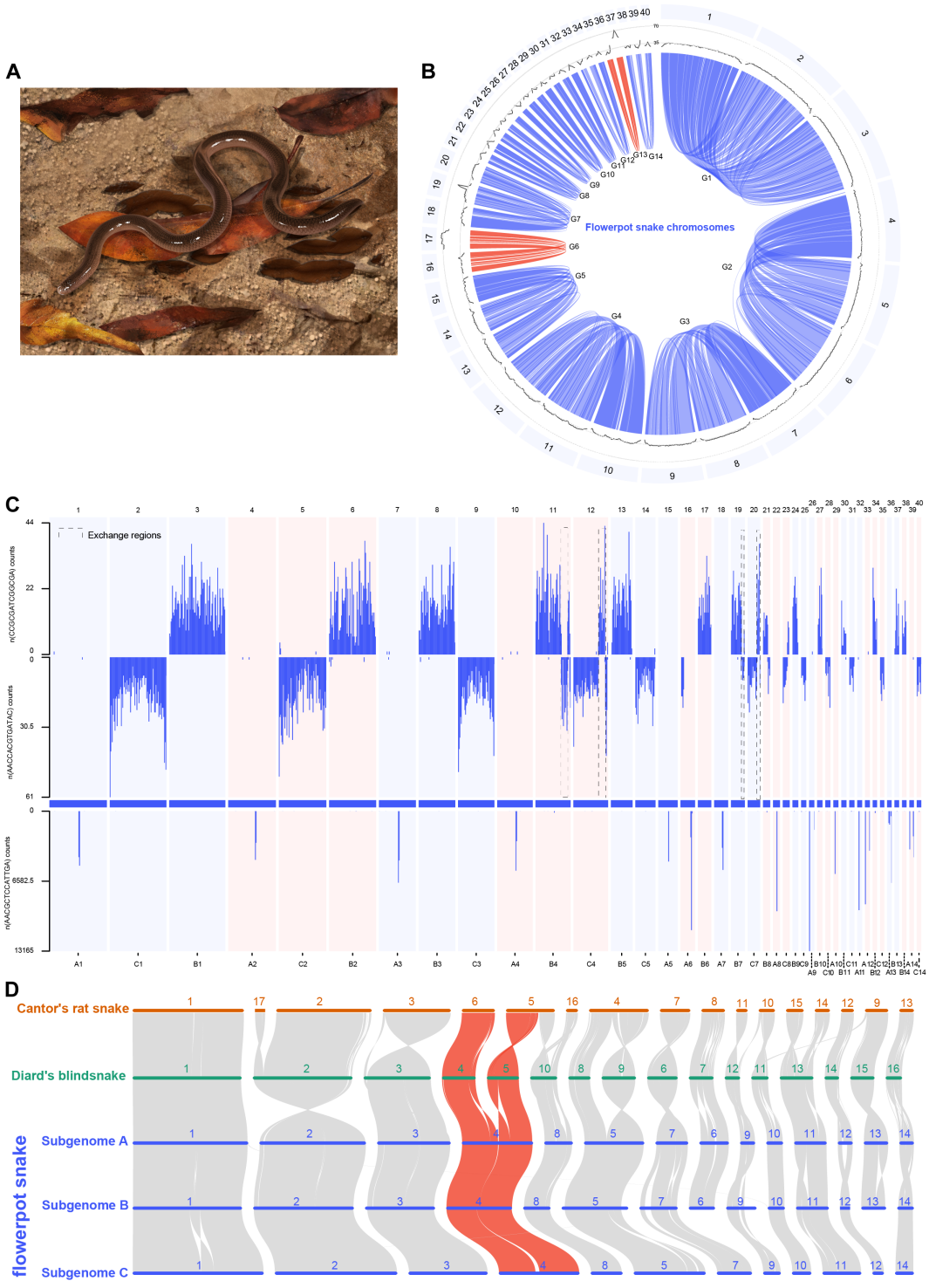

Fig. 1. Features of the allotriploid genome of the flowerpot snake and comparisons with Cantor's rat snake and Diard's blindsnake. (**A**) Flowerpot snake. (**B**) Genome map of the flowerpot snake consists of 40 chromosomes; the inner layer displays the grouping of these chromosomes (G1 to G14), while the outer layer shows the coverage depth for each chromosome. The red collinearity corresponds to two pairs of homologous chromosomes (G6 and G13), each containing only two chromosomes, but the 17th and 37th chromosomes exhibit coverage depths approximately twice that of other chromosomes. (**C**) Copy number distribution of three nucleotide repeat sequences, RS1 (AACGCTCCATTGA), RS2 (AACCACGTGATAC), and RS3 (CCGGATCGGCGA), across the chromosomes of the flowerpot snake (measured every 5 Mb), illustrating regions with interchromosomal exchange (B4 versus C4 and B7 versus C7). (**D**) Chromosomal collinearity comparisons between the flowerpot snake, Cantor's rat snake, and Diard's blindsnake indicate that chromosome 4 in the flowerpot snake's subgenome stemmed from chromosome fusion events. The blue text represents the three subgenomes of the flowerpot snake post-split (subgenome A, subgenome B, and subgenome C).

chromosomes, and their single-nucleotide density was similar to the average density observed across the other chromosomes. This suggests that there were two copies of chromosomes 17 and 37 in the assembled flowerpot snake genome, which explains the discrepancy in the chromosome count, as there were two fewer chromosomes in the assembled genome than in the karyotype-based observation.

We compared the syntenic relationships among the 40 chromosomes of the flowerpot snake and grouped them into 12 sets (Fig. 1, B and C), with each set highlighted by a light blue and light red background color. Ten of these groups had three chromosomes each (G1 to G5, G7 to G12, and G14), and the other two groups had two chromosomes each (G6 and G13). To identify the origins of different subgenomes (chromosomes derived from different ancestors), we focused on three chromosomes from subgenome G1. We identified 15–base pair (bp) short repetitive sequences (RS) with copy number variations across these chromosomes and found the greatest copy number variation for RS1 (AACGCTCCATTGA), RS2 (AACCACGTGATAC), and RS3 (CCGCGATCGGCGA). We then counted the occurrences of RS1, RS2, and RS3 within each chromosome using 5-Mb windows and observed a preferential distribution of these sequences across different chromosomes (Fig. 1C). Approximately one-third of the chromosomes showed preferences for RS1, RS2, and RS3. On the basis of these findings, we partitioned the flowerpot snake genome into three subgenomes (A, B, and C; Fig. 2D and table S3), with subgenome A containing 14 chromosomes, subgenome B containing 14 chromosomes, and subgenome C containing 12 chromosomes.

However, there were two copies each of chromosomes 17 (corresponding to subgenome B06) and 37 (corresponding to subgenome B13), meaning that the 42 chromosomes constituting the current flowerpot snake genome correspond to 14 chromosomes in subgenome A (A01 to A14), 16 chromosomes in subgenome B (B01 to B14 and copies of B06, B13), and 12 chromosomes in subgenome C (C01 to C05, C07 to C12, and C14).

To further confirm the reliability of the subgenome partitioning, we identified short segments (15 bp) with copy number variations across different chromosomes and performed clustering analysis based on transposon enrichment (fig. S12). Last, we enriched these differential segments in the respective subgenomes (fig. S13). The results were fully consistent with our previous findings (Fig. 1C), confirming that the flowerpot snake genome is composed of a hybrid triploid with three subgenomes. However, the number of chromosomes in the three subgenomes was not entirely consistent, suggesting that variation may be generated during the genetic process, which is potentially associated with variation in the underlying genetic mechanisms.

In addition, the distribution of these RS across the flowerpot snake genome revealed regions of exchange between homologous chromosomes. For example, in group G4, interchromosomal exchange between chromosome 11 (B4) and chromosome 12 (C4) was detected. Similarly, in group G7, chromosome 19 (B7) and chromosome 20 (C7) have also undergone segmental exchanges (Fig. 1C). This phenomenon was also observed in the distribution of transposons across chromosomes (fig. S13).

In addition to the 40 chromosomes of the flowerpot snake genome, we successfully assembled 16 chromosomes of Diard's blindsnake genome (fig. S14 and table S4) and 17 chromosomes of Cantor's rat snake genome (fig. S15 and table S5). Through genomic statistics (table S6), genome quality assessments (tables S7 and S8), and an evaluation of the integrity of the annotated protein-coding genes (tables S9), we achieved high-quality chromosome-level assemblies and confirmed high protein integrity for all three snake species.

On the basis of the subgenome partitioning results of the flowerpot snake genome, along with those of Diard's blindsnake and Cantor's rat snake, we conducted a comparative analysis of chromosome collinearity. Our results indicate that a chromosomal fusion event occurred in the common ancestor of the flowerpot snake subgenomes, specifically on the fourth chromosome (Fig. 1D). When using the Diard's blindsnake genome as a reference, this fusion event corresponds to chromosomes 4 and 5. In the case of Cantor's rat snake, it involves chromosome 6 and a portion of chromosome 5.

Low synonymous substitution rate in the subgenomes

In addition to comparing the genomes of Diard's blindsnake, Cantor's rat snake, and flowerpot snake, we analyzed other published snake genomes (*12*) and that of the Aeolian wall lizard (*Podarcis raffonei*) as an outgroup species (table S10). We established the evolutionary relationships among these species by constructing a phylogenetic tree using 2863 single-copy orthologous genes. The resulting tree was consistent with the snake phylogeny (*12*); Diard's blindsnake genome and the three subgenomes of the flowerpot snake were positioned at the base of the snake phylogeny, and Cantor's rat snake was grouped with other colubrids (Fig. 2A). We also determined that the three ancestral subgenomes of the flowerpot snake diverged rapidly around 41 million years ago (Ma ago). Analysis of the synonymous substitutions rate (Ks) in its subgenomes revealed low performance (Fig. 2B). Conversely, the nonsynonymous substitution rate relative to the synonymous substitution rate (Ka/Ks) was high, indicating that there was a relatively low degree of evolutionary constraint; this may facilitate the accumulation of more deleterious mutations. In contrast, Ks, representing the silent mutation rate, was neutral, suggesting that the rate of evolution of the flowerpot snake genome is slow and the accumulation of deleterious mutations is gradual.

These findings suggest that different snake species may use distinct mechanisms to mitigate the accumulation of deleterious mutations over evolutionary time. While the low level of Ks is not exclusive to the asexual flowerpot snake subgenomes (Fig. 2B), similar patterns have been observed in sexual species, including Cantor's rat snake, *Euprepiophis perlacea* (Szechwan rat snake), and *Thermophis baileyi* (hot-spring snake). One possible mechanism may involve sexual reproduction, which purges deleterious mutations over successive generations, thereby maintaining a relatively low accumulation of harmful mutations. Another potential mechanism, applicable to both sexual and asexual reproduction, might involve a reduced overall mutation rate for both nonsynonymous and synonymous substitutions, thus limiting the buildup of detrimental mutations. Therefore, evaluating parthenogenetic species solely based on the evolutionary patterns observed in sexually reproducing species may be insufficient. This indicates that the hypothesis that parthenogenetic species represent an evolutionary dead end should be reconsidered.

Effects of ILS on allotriploid formation

After estimating that the ancestral subgenomes of the flowerpot snake likely formed during a rapid speciation event around 41 Ma ago, we hypothesize that incomplete lineage sorting (ILS) may have occurred among the ancestral subgenomes (Fig. 2A). By analyzing gene tree topologies among flowerpot snake subgenomes and the genomes of Diard's blindsnake and Cantor's rat snake, we identified

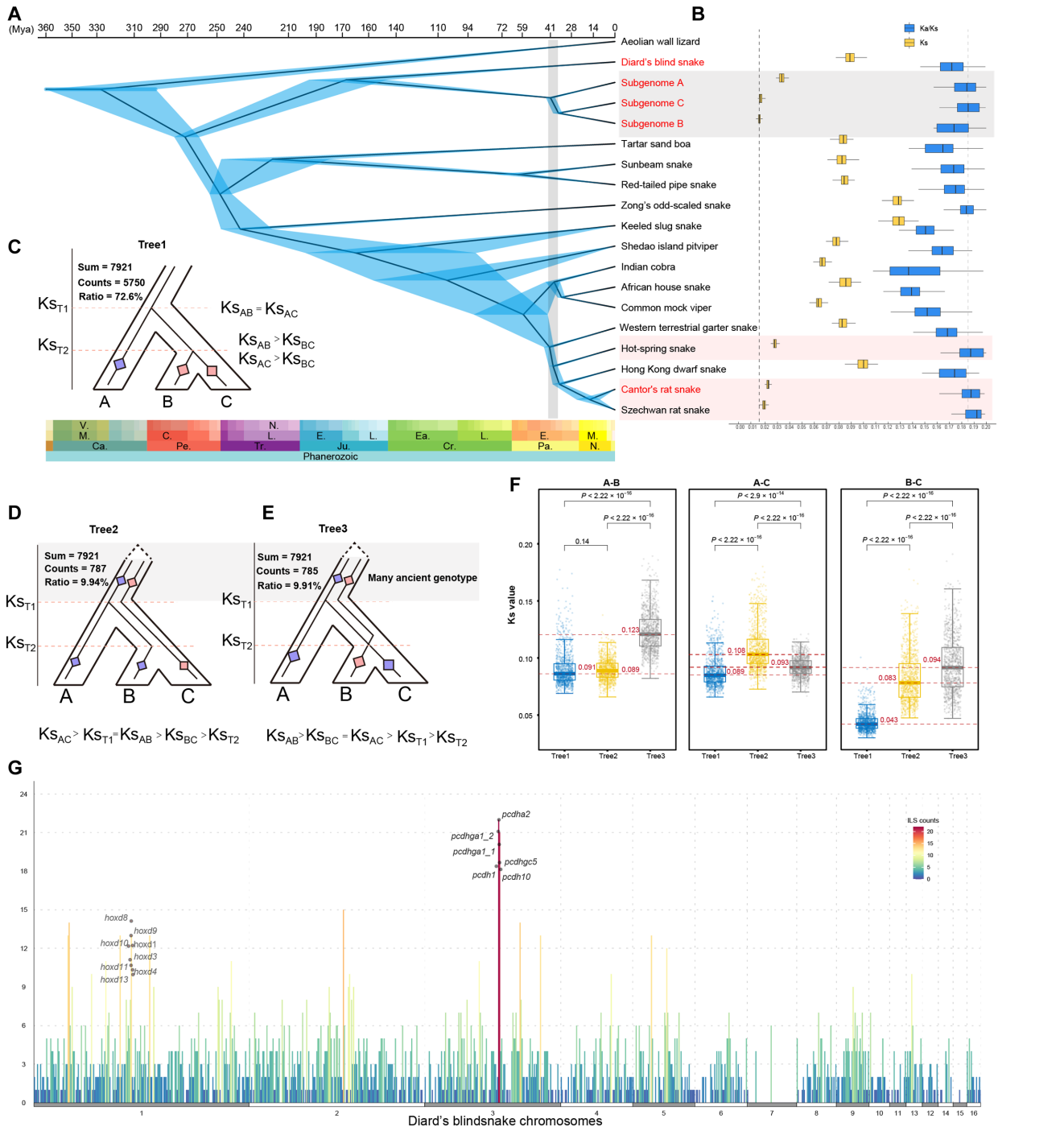

Fig. 2. Genetic evolution of the flowerpot snake. (**A**) Divergence times between the flowerpot snake and other snake species. The red font indicates the genomes or subgenomes newly assembled in this study; the shaded areas represent the estimated origin of the three subgenomes of the flowerpot snake (subgenome A, subgenome B, and subgenome C), which are likely originated through a rapid speciation event around 41 Ma ago. (**B**) Variation in average Ks and Ka/Ks ratios. The shaded regions indicate the levels of Ks and Ka/Ks for the three subgenomes of the flowerpot snake (subgenome A, subgenome B, and subgenome C). (**C**) Gene topology consistent with the species tree topology. Ks_{T1} represents the average Ks values between Subgenome A and subgenome B and between subgenome A and subgenome C (Ks_{AB} and Ks_{AC}). Ks_{T2} represents the average Ks value between subgenome B and subgenome C (Ks_{BC}), which is lower than Ks_{AB} and Ks_{AC}. (**D** and **E**) Two topologies inconsistent with the species tree topology. Ks_{T1} and Ks_{T2} correspond to the average Ks values within the tree topology depicted in panel C, where Ks_{AC}, Ks_{BC}, and Ks_{AB} represent the average Ks values of genes from different subgenomes under the current tree topology. These calculations were made on the basis of (F). (**F**) Average Ks distribution across different tree topologies. (**G**) Using the chromosomes of Diard's blindsnake as a reference, we divided the conserved syntenic segments between Diard's blindsnake and the flowerpot snake into 500-bp lengths. We calculated the number of segments that corresponded to Tree2 (D) or Tree3 (E) in 5-Mb windows to identify changes in ITS across the genome.

several distinct topologies: Tree1 {[A, (B and C)] (Fig. 2C), outgroup}, which aligned with the species tree, was recovered in 5750 instances; Tree2 {[C, (A and B)], outgroup} (Fig. 2D) and Tree3 {[B, (A and C)], outgroup} (Fig. 2E) were recovered in 787 and 785, instances, respectively, each accounting for approximately 9.9% of all trees. A comparison of Ks values across these topologies revealed that Tree2 and Tree3 exhibited higher Ks values than Tree1 (Fig. 2F). This discrepancy indicates an increased accumulation of synonymous mutations, which reflects the influence of ILS during the rapid formation of the subgenomes in the flowerpot snake. Although most genotypes evolved into species-specific variants, some retained ancestral genotypes, which resulted in different gene tree topologies. ILS may reduce genetic heterogeneity among species, thereby facilitating polyploid formation by minimizing genetic differences between subgenomes.

By aligning whole-genome syntenic regions and constructing gene trees, we identified 24,333 conserved regions, each 500 bp in length. Among these, 22,007 regions (90%) had the Tree1 species tree topology [A, (B and C)], while 1137 (4.7%) and 1189 (4.9%) regions had the Tree2 [C, (A and B)] and Tree3 [B, (A and C)] topologies, respectively. These findings indicate that ILS affects both coding and noncoding regions, although its impact was more pronounced on coding regions; ILS events were nearly 5% more common in these areas relative to the noncoding regions. Using Diard's blindsnake as a reference, we calculated the number of ILS events within each 5-Mb region and identified ILS hotspots, particularly on chromosome 3 (Fig. 2G), which harbors a high density of protocadherin (*PCDH*) genes involved in cell-cell communication and adhesion (*13*, *14*). We also found that *HoxD* genes from the Hox family were enriched in a region with high ILS on chromosome 1 (Fig. 2G). The homeobox (*Hox*) gene family is known to play a crucial role in organismal development and morphogenesis (*15*, *16*). These results indicate that the conserved genotypes of important functional genes have been preserved despite divergence from the ancestral subgenomes of the flowerpot snake. This conservation of sequences may help alleviate genome incompatibility when different species undergo polyploidy formation, thus promoting the establishment of polyploids.

Pseudogenization of immune-related genes and sexually selected genes

On the basis of the functional annotations and synteny analysis of the Diard's blindsnake genome, we identified 1,559 unitary pseudogenes (UPs) in subgenome A, 1,528 UPs in subgenome B, and 1,412 UPs in subgenome C (Fig. 3A). The number of UPs identified was not higher in the flowerpot snake subgenomes than in those of the 14 other snake species. This may be attributed to the relatively recent formation of its polyploidy.

We further compared the Ka/Ks ratios between UPs and normal genes and found that the Ka/Ks of UPs was higher than that of normal genes (Fig. 3B). This trend was consistent across the three flowerpot snake subgenomes anqd in the other snakes. However, the average Ka/Ks of UPs in the flowerpot snake subgenomes exhibited greater discrepancies compared to that of normal genes. To explore potential functional patterns, we performed functional enrichment analysis on the UPs in the flowerpot snake subgenomes to identify convergent trends. The enrichment analysis revealed functional similarities among the UPs in the three subgenomes (Fig. 3C). For example, the Gene Ontology (GO) term GO:0016503 (pheromone receptor activity) was common across all three subgenomes. In addition, we found that subgenome B and subgenome A shared two GO terms, while subgenome B and subgenome C shared three GO terms. These shared GO terms were primarily associated with immune responses (e.g., regulation of immune effector processes and regulation of cytokine production) and female sexual selection (e.g., pheromone receptor activity, regulation of female receptivity, positive regulation of female receptivity, and female mating behavior).

In addition, we identified UPs of the *TTLL3* and *TTLL8* genes in the flowerpot snake, while these genes exhibit normal gene structures in Cantor's rat snake. The TTLL3 and TTLL8 proteins are known to participate in microtubule glycosylation and play a key role in controlling flagellar movement, sperm motility, and male fertility(Fig. 3D) (*17*). Synteny analysis revealed mutations in the first exon of *TTLL3* in all three subgenomes of the flowerpot snake, indicating that pseudogenization occurred during the formation of the ancestral subgenomes (Fig. 3, E to G). We found no corresponding gene regions for *TTLL8* in subgenomes B and C and detected a frameshift mutation in the last exon of subgenome A (Fig. 3H).

Previous studies have shown that intraspecific genetic incompatibility can lead to hybrid lethality in plants (*18*), which was attributed to genetic dominance and autoimmunity (*19*), suggesting that immune responses increase genetic incompatibility. In this study, UPs in the three subgenomes were enriched in immune-related functions, implying that the flowerpot snake may reduce inter-subgenome immune responses through immune gene defects, thereby facilitating the formation of hybrid genomes. Furthermore, other major defective genes are primarily associated with sexual selection, possibly due to the lack of selective pressure following the formation of the flowerpot snake species through parthenogenesis.

Expression specificity in the parthenogenetic ovary

We performed a comprehensive collinearity analysis of the genomes of Diard's blindsnake and the three subgenomes of the flowerpot snake and identified a total of 11,173 sets of collinear genes. After normalizing gene expression, our comparison showed no significant differences in mean expression levels among the three subgenomes of the flowerpot snake (Fig. 4A). This finding suggests that gene expression in the flowerpot snake is not notably influenced by subgenomic preference.

Furthermore, we performed a tissue expression correlation analysis (fig. S16) and gene coexpression analysis across seven tissues (heart, kidney, liver, lung, muscle, ovary, and skin) from the flowerpot snake, as well as the ovary of Diard's blindsnake (figs. S17 and S18), and found that genes belonging to the blue module were the most significantly associated with the parthenogenetic ovary tissue (Fig. 4B). In contrast, the sexual ovary (represented by that of Diard's blindsnake) was significantly correlated with the brown module (Fig. 4B). Notably, we detected a substantial correlation between these two modules (Fig. 4B), indicating that their expression patterns were similar.

Further investigation of the expression trends of these modules across various tissues revealed that genes within the blue module had notably higher expression levels in the parthenogenetic ovary compared with the sexual ovary. In contrast, genes classified within the brown module had reduced expression levels in the parthenogenetic ovary than in the sexual ovary (Fig. 4C). In addition, enrichment analysis indicated that genes in the blue module were predominantly involved in key biological processes such as DNA replication, DNA repair, and cell division (Fig. 4D).

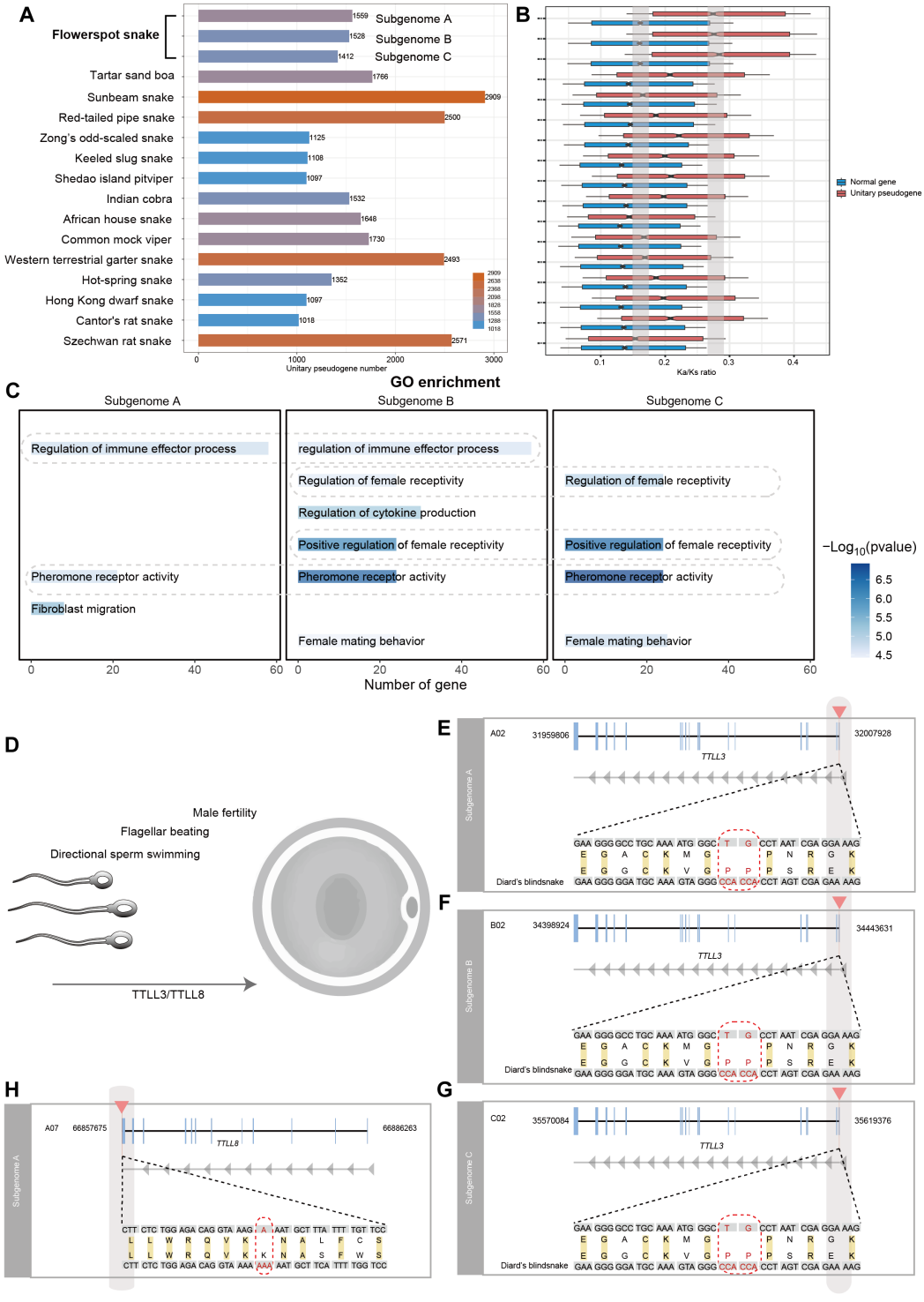

Fig. 3. UPs in the flowerpot snake. (**A**) UPs among the subgenomes of the flowerpot snake and other snake species using Diard's blindsnake as a reference. (**B**) Comparisons of Ka/Ks ratios between UPs and normal genes. The shaded areas indicate the levels of the Ka/Ks ratio for the subgenomes of the flowerpot snake. The *y* axis is shared with (A). (**C**) GO term enrichment analysis of UPs within the subgenomes of the flowerpot snake. The dashed boxes represent GO terms shared among different subgenomes. (**D**) Functional importance of TTLL3 and TTLL8 proteins in fertilization. (**E** to **H**) Pseudogene mutation sites of TTLL3 and TTLL8 within the subgenomes of the flowerpot snake. The red triangles at the top indicate the locations of mutations, with corresponding detailed views of the mutations below. The red dashed boxes highlight specific site variations.

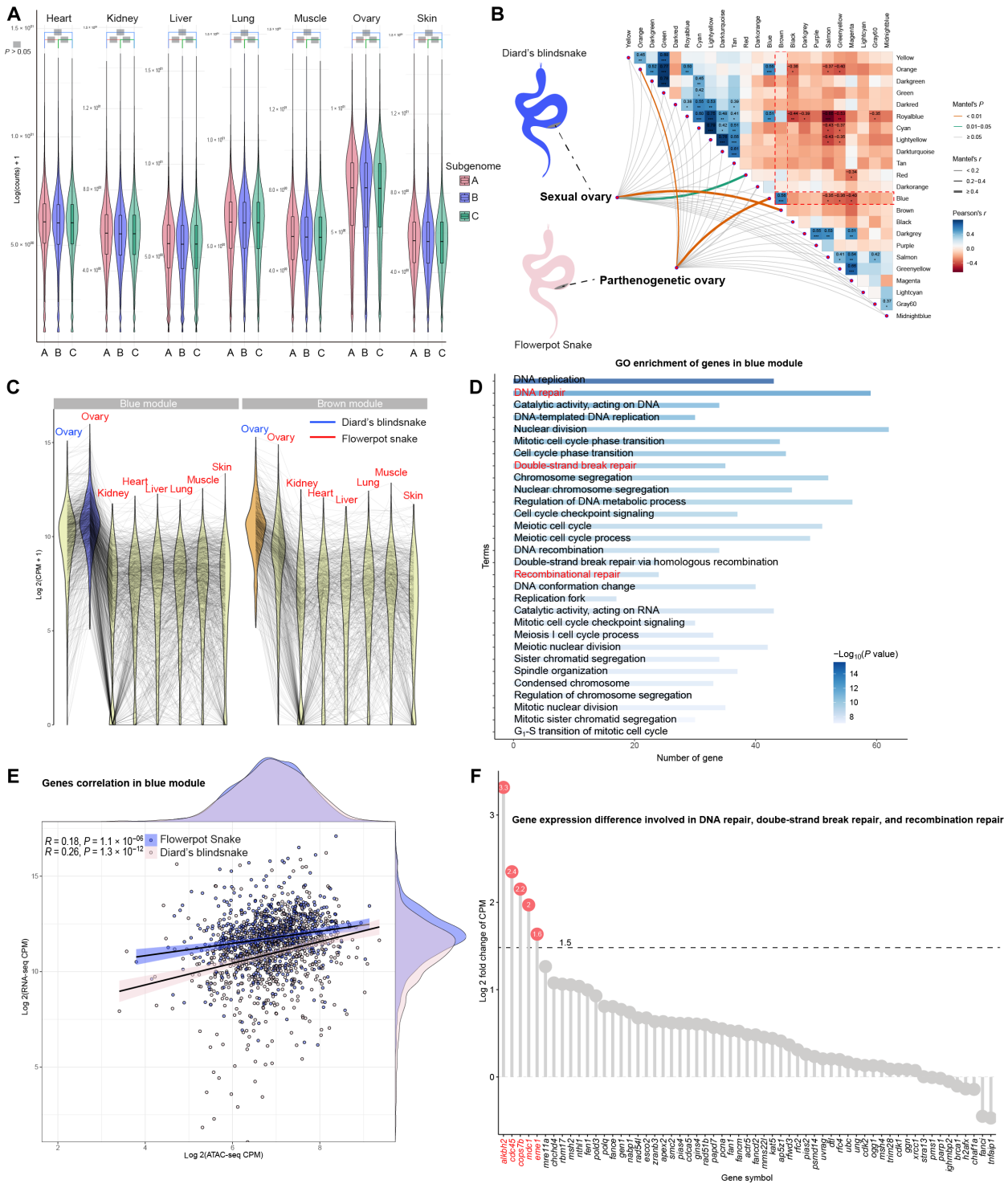

Fig. 4. Specificity of gene expression in the ovary of the flowerpot snake. (**A**) Subgenomic expression patterns showing no bias. The gray squares represent statistically insignificant differences between the two groups (*P* value > 0.05). (**B**) Correlations between gene coexpression modules and the parthenogenetic versus sexual ovary. The brown lines indicate coexpression modules most closely associated with the ovaries of both the flowerpot snake and Diard's blindsnake (the blue module and brown module, respectively). The red dashed box highlights the correlation between the blue module and brown module. The numbers in the heatmap squares represent correlation coefficients, with an asterisk *P value < 0.05, **P value < 0.01, and ***P value < 0.001. (**C**) Expression analysis of genes within the blue and brown modules, including genes from Diard's blindsnake (sexual ovary indicated in blue font) and those from the parthenogenetic ovary (indicated in red font). (**D**) GO term enrichment analysis of genes in the blue module. The red-colored GO terms are those associated with DNA repair. (**E**) Comparison of RNA-seq reads and ATAC-seq counts within the blue and brown modules. (**F**) Largest differences in gene expression related to DNA repair, double-strand break repair, and recombinational repair between the flowerpot snake and Diard's blindsnake.

Moreover, we integrated assay for transposase-accessible chromatin sequencing (ATAC-seq) data from the ovaries of Diard's blindsnake and the flowerpot snake to infer the relationship between chromatin accessibility and gene expression within the blue module. Our analysis revealed statistically significant correlations between gene expression and chromatin accessibility levels in the blue module across both Diard's blindsnake and flowerpot snake ovaries (Fig. 4E). This suggests that the expression of genes in the ovaries of both species is closely linked to chromatin accessibility dynamics.

Of particular interest in this study are functional processes associated with the first meiotic division, including DNA repair, double-strand break repair, and recombinational repair. The high expression of genes associated with these functional processes suggests that the first meiotic division is highly likely to occur in the flowerpot snake. Furthermore, the increased expression of these genes may play a critical role in minimizing errors during genome replication.

Alternatively, the enhancement of repair-related genes may indicate the presence of DNA lesions during the first meiotic division of the flowerpot snake. This may be caused by mismatched chromosome pairing in the first meiotic division as observed in parthenogenetic triploid geckos (*3*). We characterized expression differences of these relevant genes (Fig. 4F, genes in red color), including alkB homolog 2, α-ketoglutarate–dependent dioxygenase (*ALKBH2*), cell division cycle 45 (*CDC45*), COP9 signalosome subunit 7b (*COPS7B*), mediator of DNA damage checkpoint 1 (*MDC1*), and essential meiotic structure-specific endonuclease 1 (*EME1*), and their expression levels (log conversion with base 2) were at least 2.5 times higher in the parthenogenetic ovary than in the sexual ovary, suggesting that they play an important role in the DNA repair process.

Previous studies have reported the identification of two atypical meiotic mechanisms in reptiles that allow for reduction division in somatic cells to be bypassed (*3*). The first mechanism is premeiotic endoreplication, in which chromosomes are replicated before meiosis, and the second mechanism involves oocyte and polar body fusion, wherein sister chromatids replace homologous chromosome pairs during the first meiotic division. The premeiotic endoreplication mechanism has been confirmed in diploid, triploid, and tetraploid parthenogenetic species and effectively maintains heterozygosity in allopolyploids (*3*, *9*, *20*). In addition, an oocyte and polar body fusion mechanism has been observed in diploid lacertid lizards of the genus *Darevskia*. In this case, the initial stages resemble typical meiosis; however, diploidy is ultimately restored through the fusion of the oocyte and polar body (*21*). Nevertheless, this mechanism has recently been challenged as an observational error (*22*). In addition, this mechanism is not suitable for species with odd-numbered multiples of chromosomes such as the triploid flowerpot snake, as effective pairing cannot occur. Consequently, premeiotic endoreplication is likely the only confirmed mechanism for maintaining ploidy in reptiles.

In the case of the flowerpot snake, we identified highly expressed genes enriched in critical functional categories specific to the first meiotic division, such as DNA repair, double-strand break repair, and recombinational repair. These results suggest that the premeiotic endoreplication mechanism is suitable for the flowerpot snake. On the basis of previous studies and our gene expression analyses, we propose that the flowerpot snake reproduces via premeiotic endoreplication. This mechanism has likely evolved convergently in parthenogenetic reptiles (*3*, *22*) and fish (*8*), indicating that their reproductive strategies may share several common features. Our research on gene expression in the blue module within the ovaries of the flowerpot snake leads us to hypothesize that cell repair and cell division may be jointly regulated.

We present this hypothesis regarding the potential activation of premeiotic endoreplication. Most reptiles that reproduce through parthenogenesis have a hybrid origin (*1*), which reflects a high degree of heterozygosity among the genomes of their homologous chromosomes. During meiosis in parthenogenetic animals, the initial stages closely resemble typical meiosis, including the pairing of homologous chromosomes. However, the high level of heterogeneity among hybrid-derived homologous chromosomes—which are characterized by differences in allelic loci, structural variations, and topological discrepancies—may stem from DNA lesions. This recognition could enhance DNA repair mechanisms (*23*) and lead to the increased expression of DNA repair-related genes, as evidenced by the elevated levels observed in our study. Cell cycle checkpoints play a crucial role in coordinating DNA repair processes and can result in alterations within the cell cycle (*24*). This may lead to the duplication of DNA and enable certain oocytes to undergo premeiotic endoreplication, generating replicated sister chromatids that are entirely identical to the original chromosomes.

The original chromosomes and duplicated sister chromatids are completely uniform, and they can pair preferentially, as has been observed during the sister chromatid pairing in parthenogenetic lizards (*9*). This pairing mechanism ensures that the resulting gametes have the same genetic composition and chromosome number as the somatic cells, thereby facilitating clonal reproduction. This mechanism is thought to maintain heterozygosity within the genome, ensuring a high level of genetic variability (*9*). If the observed patterns of genetic variation were solely attributed to the sister chromatid pairing mechanism, then one would expect to see uniform chromosome counts across all subgenomes.

In the case of the flowerpot snake, for instance, subgenomes A, B, and C should each consistently contain 14 chromosomes, given the total chromosome number of 42, and there should be no chromosomal exchanges. However, chromosome counts for these subgenomes are in fact 14, 16, and 12, respectively, and chromosomal exchanges between regions have been identified. This discrepancy indicates that the sister chromatid pairing mechanism alone cannot explain the chromosomal exchanges observed between chromosomes in the flowerpot snake. Consequently, this assumption is inconsistent with the findings of our study. To explore this further, we expanded our analysis by collecting additional flowerpot snake samples from different geographic regions, allowing us to assess the extent of genetic variation and test the robustness of our conclusions.

Potential mechanism for parthenogenetic meiosis

We collected samples of the flowerpot snake from three regions: Fujian Province, China (Fujian population, Fig. 5A); Yunnan Province, China (Yunnan population, Fig. 5A); and and Hong Kong, China (Hong Kong population, Fig. 5A). In addition, we integrated short-read data (*25*) from one sample collected in India (Indian population, Fig. 5A). By comparing them against the reference genome (the sample from the Fujian population), we identified single-nucleotide polymorphisms (SNPs) and estimated the read coverage depth for each sample. After integrating SNP sites from all samples for principal components analysis (PCA) and hierarchical clustering analysis, we found substantial differences between the Yunnan flowerpot snake and samples from other regions, and these were divided into two distinct clusters (Fig. 5, B and C).

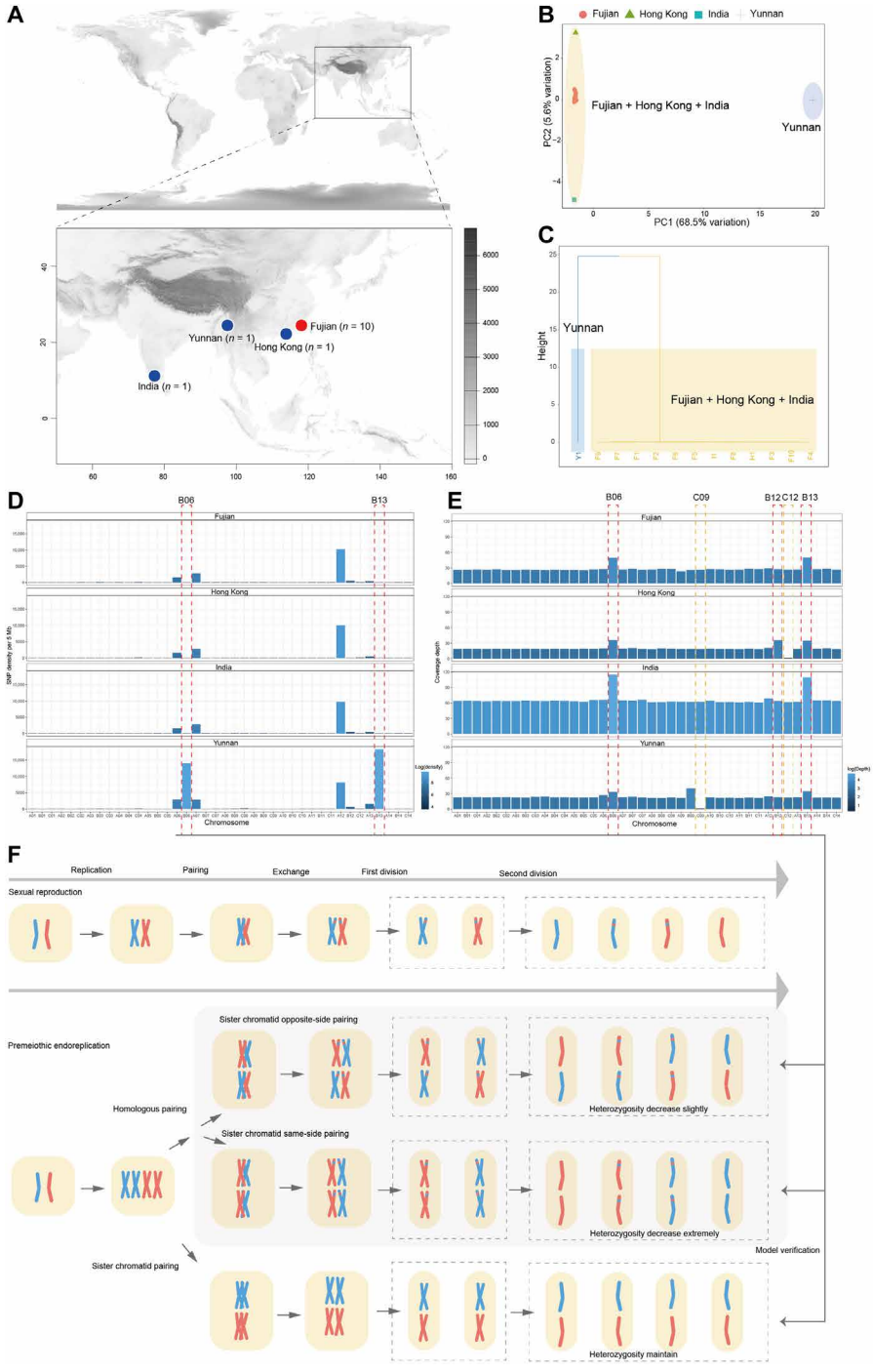

Fig. 5. Population genomics analysis identified meiotic variants in the flowerpot snake. (**A**) Short-read sequencing data were collected from Fujian, Hong Kong, and Yunnan samples in this study, while the Indian sample was obtained and sequenced from a previous study (*25*). The red dot indicates the geographic populations from which the flowerpot snake genomes were assembled. The text in parentheses specifies their respective sample sizes. (**B**) PCA of all sample SNPs shows that samples from Fujian, Hong Kong, and India cluster along PC1 (68.5% variance), marked in yellow, while Yunnan samples differ, shown in blue. (**C**) Hierarchical clustering of samples based on principal component 1 (PC1). (**D**) SNP density distribution measurements per 5 Mb across chromosomes in the flowerpot snake. Red dashed boxes highlight chromosomes B06 and B13, which show substantially higher SNP density in Yunnan compared to Fujian, Hong Kong, and India samples. (**E**) Depth coverage measurements for each chromosome show chromosomes B06, B12, and B13 with double the sequencing depth in Fujian, Hong Kong, and India samples, and slightly higher in Yunnan. Chromosome C09 has lower sequencing depth in Yunnan, and C12 is lower in Hong Kong. (**F**) Top section illustrates the meiotic process during conventional sexual reproduction, and the lower section details the meiosis process under the mechanism of premeiotic endoreplication in parthenogenesis. The first and second processes under homologous pairing depict two distinct meiotic processes, which result in slight and extreme decreases in heterozygosity, respectively. The third process, sister chromatid pairing, leads to the maintenance of heterozygosity.

Comparison of the SNP density and sequencing depth between different populations and chromosomes revealed several unexpected findings. For example, in samples from Yunnan, the SNP density of chromosomes B06 and B13 was notably higher compared with that of samples from other populations (Fig. 5D). Two copies of each of these two chromosomes were identified in our genome assembly (double B06 and B13, Fig. 1B). The high SNP density on chromosomes B06 and B13 (Yunnan population in Fig. 5D) led us to hypothesize that homologous chromosomes (C06 and C13) corresponding to B06 and B13 in Yunnan samples are present but missing in other populations. To confirm this hypothesis, we aligned the Yunnan samples after excluding the C subgenome from the reference genome. The results (fig. S19) showed that the SNP densities of B06 and B13 were similar to those of the other chromosomes in subgenome B, except for B09, confirming our hypothesis that the flowerpot snake genome from the sample collected in Yunnan still contains the C06 and C13 chromosomes that are missing in other populations. The absence of C09 in the corresponding subgenome C did not affect SNP densities in subgenomes B and A. Therefore, the 42 chromosomes constituting the flowerpot snake genome in the Yunnan population actually correspond to 14 chromosomes in subgenome A (A01 to A14), 15 chromosomes in subgenome B (B01 to B14 and a copy of B09), and 13 chromosomes in subgenome C (C01 to C08 and C10 to C14).

The absence of certain chromosomes in the C subgenome was also observed in the Hong Kong samples. Specifically, the read depth of C12 in Hong Kong was similar to that of C09 in Yunnan, indicating that these chromosomes are likely missing in their genomes. In addition, the read depth of B12 in the Hong Kong samples was nearly double that of other chromosomes, except for B06 and B13, suggesting that C06, C13, and C12 are absent in the C subgenome of the Hong Kong population. Consequently, the 42 chromosomes of the flowerpot snake in Hong Kong consist of 14 in subgenome A (A01 to A14), 17 in subgenome B (B01 to B14 plus copies of B06, B11, and B13), and 11 in subgenome C (C01 to C06, C07 to C10, C12, and C14).

Variation in SNP density and read depth in individuals from India was consistent with that in the Fujian population (Indian population and Fujian population; Fig. 5, D and E). The absence of chromosomes C06 and C13 in the C subgenome resulted in 14 chromosomes in subgenome A (A01 to A14), 16 chromosomes in subgenome B (B01 to B14 and copies of B06 and B13), and 12 chromosomes in subgenome C (C01 to C05, C07 to C12, and C14).

Our population genomic analysis revealed substantial variability in the genetic composition of the flowerpot snakes across different populations. The explanation of these variations by previously reported mechanisms of sister chromatid pairing appears inadequate (Fig. 5F, bottom). Therefore, we explored the possibility of homologous pairing. Depending on whether sister chromatids pair on the same side or opposite sides during the first meiotic division after replication, they were categorized as sister chromatid same-side pairs or sister chromatid opposite-side pairs (Fig. 5F).

In sister chromatid opposite-side pairs, chromatids from different ancestors undergo chromosome exchange during meiosis, potentially leading to a slight decrease in heterozygosity upon the restoration of ploidy in eggs (Fig. 5F). Conversely, sister chromatid same-side pairing results in identical copies in oocytes, which causes a substantial reduction in heterozygosity in some individuals (Fig. 5F). The genetic composition observed in flowerpot snake samples from different populations suggests the possibility of sister chromatid same-side pairing as a potential cytogenetic mechanism.

This mechanism may contribute to the generation of genetic variation in chromosome composition among different population groups, indicating that previous assumptions regarding sister chromatid pairing as the exclusive mechanism in observed samples might need further reconsideration. Cytological observations of homologous chromosome pairing have also been made. Previous studies of triploid parthenogenetic lizards observed that during early pachytene, most chromosomes paired as 62 bivalents, with only two forming tetravalents (*3*). Tetravalents in cytogenetic studies are considered potential signals of genome self-duplication, as observed in Atlantic salmon (*26*). Therefore, the observation of tetravalents in a triploid parthenogenetic lizard (*9*) may provide some evidence for homologous pairing, a mechanism that could generate identical copies of chromosomes from the same ancestor during a single meiotic division. However, confirming this mechanism through DNA locus comparisons alone in cytogenetic studies remains challenging.

The sister chromatid pairing mechanism plays an important role in ensuring heterogeneity between allopolyploid chromosomes. High heterogeneity can increase genetic potential, enhancing adaptability to diverse environments. However, it also raises the risk of incompatibility between allopolyploid genomes, potentially leading to genomic instability. In contrast, homologous chromosome pairing reduces the heterogeneity between allopolyploid chromosomes, promoting genomic homozygosity and decreasing genomic incompatibility. This suggests that in the formation and maintenance of allopolyploids, it is important to strike a delicate balance between heterogeneity and homozygosity. Such a balance facilitates the emergence of individuals with varied genetic compositions, which may, in turn, affect their fitness. In the case of the flowerpot snake, a parthenogenetic triploid species, our study uncovers a unique genomic evolutionary landscape. These findings provide valuable insights into the potential mechanisms driving genetic evolution in polyploid populations and contribute to a deeper understanding of the genetic dynamics of parthenogenesis and polyploidy in reptiles.

MATERIALS AND METHODS

Experimental design

We conducted multiple field surveys to sample the flowerpot snake. Because of its burrowing behavior and secretive habits, several surveys were unsuccessful. Successful collections were made annually from June to October. Ultimately, we obtained 1 sample from Hong Kong, 1 from Yunnan, and 11 from Fujian. Samples from Hong Kong and Fujian were used for genomic sequencing via both second and third-generation sequencing technologies. Initially, sequencing was performed on one Hong Kong sample using the MinION Nanopore sequencing platform [Oxford Nanopore Technologies (ONT) Inc., Oxford, UK], but because of its high error rate, distinguishing sequencing noise from biological variation was challenging. Subsequently, the high-accuracy sequencing of Fujian samples was performed using the PacBio Revio sequencing platform (Pacific Biosciences of California Inc., Menlo Park, CA, USA), followed by genome assembly using HiFi reads generated by the Circular Consensus Sequencing (CCS) method with CCS software. The behavior of Diard's blindsnake was similar to that of the flowerpot snake. Multiple samples of Diard's blindsnake were collected in July from Yunnan. Female specimens of the flowerpot snake and Diard's blindsnake were selected for ovarian sampling using a dissecting microscope to identify ovarian samples containing developing eggs. Half of these samples were subjected to transcriptome

sequencing by RNA sequencing (RNA-seq), while the other half were subjected to ATAC-seq.

The procedures used in this study were approved by the Institutional Ethics Committee of the Animal Ethical and Welfare Committee at the Chengdu Institute of Biology, Chinese Academy of Sciences (permit: CIBDWLL2023018). All methods were performed in accordance with the Code of Practice for the Care and Handling of Animals. This study adheres to the ARRIVE guidelines and complies with all relevant ethical regulations for animal use.

Construction of short DNA fragment libraries

We used 1 μg of DNA to construct short DNA fragment libraries using the MGIEasy DNA Universal Library Prep Set (MGI Tech Co. Ltd., Shenzhen, China) following the manufacturer's protocol. To ensure that each sample could be identified during subsequent sequencing, we added a unique barcode to each sample. After library construction, we measured the concentration and fragment size distribution of samples in the library using a Qubit 3.0 Fluorometer (Thermo Fisher Scientific Inc., Waltham, MA, USA) and a Bioanalyzer 2100 system (Agilent Technologies Inc., Santa Clara, CA, USA) to determine an appropriate pooling strategy for sequencing. Once libraries passed quality control checks, sequencing was performed on an Illumina NovaSeq 6000 (Illumina Inc., San Diego, CA, USA) or MGI-SEQ 2000 (MGI Tech Co. Ltd.) platform (Nanjing Sequencing Center, Novogene Biotechnology Co. Ltd., Nanjing).

Construction of the PacBio genome HiFi sequencing library

After randomly fragmenting DNA into approximately 15-kb fragments using g-TUBE (Covaris Woburn, MA, USA), we used the SMRTbell Express Template Prep Kit 2.0 (Pacific Biosciences of California Inc.) to construct SMRTbell HiFi sequencing libraries. The specific steps were as follows: (i) A total of 15 μg of DNA was used for library construction. Initially, protruding single-stranded ends were removed through enzymatic reactions, and this was followed by DNA damage repair using repair enzymes. (ii) After performing DNA damage repair, we proceeded with end repair and added A-tails to the double-stranded ends. (iii) The T-overhangs of SMRTbell adapters were ligated to the ends at 20°C for 15 hours. After ligation, the library was purified using 1X AMPure PB magnetic beads. (iv) The fragment size and the concentration of library samples were determined using the Agilent FEMTO Pulse automated pulsed-field capillary electrophoresis system (Agilent Technologies Inc.) and a Qubit 3.0 Fluorometer (Thermo Fisher Scientific Inc.). Subsequently, the BluePippin system (Sage Science Inc., Beverly, MA, USA) was used for DNA size selection to remove SMRTbell fragments below 15 kb. Following size selection, the library was purified again using 1X AMPure PB. (v) Library size and quality were evaluated using the FEMTO Pulse system and Qubit dsDNA HS assay kit. Sequencing primers and Sequel II DNA polymerase were annealed and combined with the final SMRTbell library. (vi) After library construction, sequencing was performed on the PacBio Sequel II platform (Pacific Biosciences of California Inc.) at a concentration of 120 pM, with a sequencing runtime of 30 hours.

RNA extraction and transcriptome sequencing

Total RNA was extracted using a PureLink RNA Mini Kit (Thermo Fisher Scientific Inc.) following the manufacturer's instructions. The quantity, purity, and integrity of the extracted RNA samples were measured using a Qubit 3.0 Fluorometer (Thermo Fisher Scientific Inc.), NanoDrop Spectrophotometer (Thermo Fisher Scientific Inc.), and Agilent 2100 Bioanalyzer (Agilent Technologies Inc.), respectively. Only samples with high RNA integrity values (RINs) (RIN $\geq$ 8) were selected for sequencing. To construct the cDNA library, RNA fragmentation and PCR amplification were performed according to the RNA-seq protocol using a Taq PCR core kit (Qiagen GmbH, Hilden, Germany) for amplification. RNA-seq data were generated using an Illumina HiSeq 4000 platform (Illumina Inc.) with paired-end reads of 150 bp in length.

Preparation and sequencing of the ATAC-seq library

Single-cell suspensions were prepared with single cells collected from each sample, and the obtained cell suspensions were subjected to routine procedures for nuclear isolation. The isolated nuclei were suspended in nuclear isolation buffer and washed multiple times with nuclear wash buffer. The transposition reaction was performed by mixing approximately 40,000 cell nuclei with the Tn5 enzyme (Illumina Inc.), and the resulting transposed DNA fragments were purified using the MinElute PCR Purification Kit (QIAGEN GmbH). These fragments were then amplified for 5 to 10 cycles using a NEBNext High-Fidelity 2X PCR Master Mix (New England Biolabs Inc., Ipswich, MA, USA) and barcoded primers. The amplified product was again purified using the MinElute PCR Purification Kit (QIAGEN GmbH.) before sequencing on an Illumina NovaSeq 6000 platform (Illumina Inc.) with paired-end reads of 150 bp in length.

Genome assembly on the basis of HiFi reads

For genome assembly, we used Meryl v1.4.1 (*27*) to calculate the frequency distribution of *k*-mers ($k = 21$ bp) from the Illumina short-read data. In addition, we employed the R package "findGSE" v1.94 (*28*) to estimate genome size by fitting k-mer frequencies iteratively using various models.

For the raw data obtained from the PacBio Sequel II CCS sequencing mode, we used CCS software v.6.4.0 (https://github.com/PacificBiosciences/ccs) with the command "-minPasses 3" to transform them into HiFi data. These long (>15 kb), highly accurate (99%) HiFi reads were used for subsequent assembly. For Diard's blindsnake and the flowerpot snake, we initially assembled the HiFi data using Hifiasm v.0.19.9-r616 (*29*). In addition, for the Diard's blindsnake assembly, we used Hifiasm in addition to Verkko v2.1 (*30*) and Hicanu v2.2 (*31*) software. On the basis of the contig N50 length and estimation of genome size by the Kmer analysis, we used the results from Verkko software for further analysis.

Hi-C sequencing and assisted assembly

A Hi-C library was prepared following the manufacturer's instructions and subsequently sequenced on the NovaSeq 6000 platform (Illumina Inc.) in the PE150 mode. To efficiently use the sequencing data, sequencing adapters and low-quality segments were trimmed, and all subsequent analyses were performed using clean data after the removal of low-quality reads. Trimming of raw data was performed using Trimmomatic v0.39 (*32*) following the specific steps below:

1) Low-quality reads were trimmed using a sliding window approach with a window of 4 bases; reads were trimmed if the average base quality was less than 15.

2) Reads with low-quality bases (quality < 3) were trimmed at the leading or trailing ends or reads containing "N" bases (indicating uncertain base information). The parameters used were as follows: LEADING: 3; TRAILING: 3.

3) Adapters were removed using two modes: simple alignment mode (adapter sequence matches with a score of 10 or more) and palindrome mode (for read pairs where overlapping base scores exceed 30, adapter sequences were trimmed).

4) Reads trimmed to less than 36 bases were discarded.

5) Unpaired reads that did not form paired reads were discarded.

Cluster formation, construction of interaction matrices, and visualization of interaction maps were performed using the YaHS v1.1 (*33*) scaffolding tool. The interaction maps were further visualized and corrected using Juicebox v2.20.00 (*34*) to correct assembly errors such as contig order, orientation, or internal misassemblies. The correction of the whole-genome interaction map enabled visualization of strong intrachromosomal interactions and closer linear distances corresponded to stronger interactions.

Once the assembly was complete, we aligned the Illumina short-read data to the reference genome using minimap2 v2.28 (*35*). We then utilized the "bedtools nuc" command from the bedtools v2.31.1 (*36*) to calculate the percentage of the GC content in 10-kb windows with a 5-kb step size. Similarly, we used the "bedtools coverage" command to assess the short read coverage depth within the same windows and step sizes. By analyzing the uniformity of the GC content and coverage depth, we evaluated whether contamination was present in the assembly.

The final chromosome-level genome assembly was evaluated using the BUSCO v.3.1.0 gene sets and configured in the vertebrate genome mode using related lineage data (*37*). In addition, we assessed the genome quality of the assemblies using *k*-mer analysis with a *k*-length of 21 bp (*27*).

Genome annotation

We began the genome annotation by constructing a de novo repeat library using RepeatModeler version 2.0.1 (*38*). Subsequently, we used the Repbase database (*39*) as a known library to identify repeats specific to Squamata lineages, including other snakes. These libraries were merged and searched against the assembled genomes using RepeatMasker version 4.1.1 with the parameters "-nolow -gff -poly -a -inv -e rmblast" to compile a comprehensive list of repeats (*40*).

We used three distinct approaches to predict protein-coding gene regions within the assembled genomes: (i) de novo prediction using Augustus version 3.3.3 (*41*) and Snap version 2006-07-28 (*42*), (ii) homolog-based prediction using Tblastn version 2.11.0 (*43*) and Exonerate version 2.2.0 (*44*) with proteins from seven representative species, and (iii) transcript-based annotation. Initially, a de novo transcriptome assembly was generated using Trinity version 2.11.0 (*45*). Subsequently, raw transcripts were aligned to the assembled genome using Blat version 36 (*46*) and Gmap version 2017-11-15 (*47*), followed by validation and integration in the Program to Assemble Spliced Alignments pipeline (*48*). Last, EvidenceModeler version 1.1.1 (*49*) was used to integrate results from all three strategies, removing erroneous predictions containing stop codons within coding regions. Functional annotation and characterization of protein biological functions were performed using the tool Eggnog-mapper v.267.

Determination of genetic composition and subgenome partitioning of flowerpot Snake chromosomes

We first identified segments of whole-genome collinearity using JCVI software (*50*) to assess homology relationships between chromosomes on the basis of the corresponding regions of collinearity with other chromosomes. Because of discrepancies between the number of assembled chromosomes and those observed via cytogenetic analysis, we used methods involving sequencing coverage depth and SNP density per chromosome to determine their genetic composition. For instance, if the coverage depth of a chromosome was approximately twice as high as that of other chromosomes but its SNP density was consistent with that in other chromosomes, then we inferred the existence of identical copies of that chromosome.

Using the PacBio package in PBMM2 software (https://github.com/PacificBiosciences/pbmm2), HiFi reads were aligned to the assembled genome, and the coverage command in SAMtools v1.11 (*51*) provided the average sequencing depths per chromosome. In addition, the coverage command in BEDTools v2.31.1 (*36*) recalculated sequencing depths in 5-Mb intervals across each chromosome to reduce potential biases from assembly errors. Subsequently, GATK v4.0.40 (*52*) was used for SNP calling from alignment files, and the initial SNP calling was performed using HaplotypeCaller with a --ploidy setting of 1. Filtering criteria "--QD < 2.0 || FS > 60.0 || MQ < 40.0 || SOR > 3.0 || MQRankSum < -12.5 || ReadPosRankSum < -8.0" were used to refine SNP sites. Filtered VCF files were processed using the coverage command in BEDTools to count SNPs within each 5-Mb region, which helped characterize the genetic composition of the current genome.

To further understand chromosome features, we used KMC v3.0.0 (*53*) software to count 13-bp segments (ki, i = 1, 2, 3, 4, …) within the longest set of homologous chromosomes. The discrepancy in counts among ki segments across homologous chromosomes revealed characteristic ki segments for each chromosome. Using seqkit v2.6.1 (*54*), we recorded the positions of each ki segment in all other chromosomes and assessed their densities using the coverage command in BEDTools in 5-Mb windows. This approach allowed us to determine the number and distribution of characteristic ki segments on each chromosome, thus aiding the identification of differences in subgenome composition. The distribution of a characteristic ki segment across the entire genome also facilitated the identification of potential interchromosomal exchanges.

After completing the subgenome partitioning, we renamed the chromosomes within each subgenome based on their homology relationships and chromosome lengths as A01…A14, B01…B14, and C01…C14. Here, A, B, and C represent different sources of the subgenomes, while identical numbers indicate homologous relationships; differing numbers indicate a lack of homology. Subsequently, we validated the split subgenomes using SubPhaser software v1.2.6 (*55*), which leverages the principle of differential transposon abundance across the different subgenomes.

Comparative genomics and phylogenetic analysis

In addition to the flowerpot snake, Diard's blindsnake, and Cantor's rat snake, we collected high-quality genome assemblies from 14 representative snakes to support our analysis. These included Tartar sand boa (*Eryx miliaris*), Sunbeam snake (*Xenopeltis*), Red-tailed pipe snake (*Cylindrophis ruffus*), Zong's odd-scaled snake (*Achalinus jinggangensis*), Keeled slug-eating snake (*Pareas carinatus*), Shedao island pitviper (*Gloydius shedaoensis*), Indian cobra (*Naja naja*), African house snake (*Boaedon fuliginosus*), common mock viper (*Psammodynastes pulverulentus*), Western terrestrial garter snake (*Thamnophis elegans*), hot-spring snake (*T. baileyi*), Hong Kong dwarf snake (*Calamaria septentrionalis*), Szechwan rat snake, and Aeolian wall lizard, which were used as outgroup species.

Because the distinct subgenomes of the flowerpot snake were derived from different ancestors, genes were extracted from each flowerpot snake subgenome separately. Using Orthofinder version 2.5.5 (*56*), all proteins from these species were clustered, and single-copy orthologous clusters were selected for alignment using Mafft version 7.475 (*57*). Naturally evolved fourfold degenerate synonymous sites (4dTV) from coding sequences were extracted to construct a species tree.

The species tree topology was initially constructed by Bayesian inference (BI) using MarBayes version 3.2.2 (*58*) and further confirmed by maximum likelihood (ML) using RAxML version 8.0.17 (*59*); the GTR + I + G model was used in both sets of analyses. BI included two parallel runs of 2 million generations with three chains each, sampling every 100 generations and discarding the initial 25% of the samples. ML was performed via the fast bootstrapping of 1000 replicates to calculate branch supports.

The species tree topology was calibrated using MCMCtree version 4.9 (*60*) with three time-calibrated points: node 1 (common ancestor of Szechwan rat snake, Cantor's rat snake, Hong Kong dwarf snake, Hot-spring snake, and Western terrestrial garter snake) at 32.4 to 48.7 Ma ago, node 2 (common mock viper and African house snake) at 19 to 44.2 Ma ago, and node 3 (red-tailed pipe snake and Sunbeam snake) at 52.2 to 77.8 Ma ago sourced from TimeTree (*61*). A log-normal independent molecular clock model was used to generate 1,000,000 samples, and the initial 25% was discarded as burn-in.

Credible intervals for species divergence times at each node were visualized using the R package MCMCtreeR (*62*).

Calculation of Ks and Ka/Ks for single-copy genes

To compare the genomic mutation rates and relative rates of change in protein sites among various snake species, we used PAML v4.9 (*60*) software to compute the ratio of synonymous mutations (Ks, mutations at synonymous sites that do not alter amino acids and are assumed to be unaffected by natural selection) and the ratio of nonsynonymous mutations to synonymous mutations (Ka/Ks, a measure of relative protein evolution rates) for each single-copy gene family on the basis of the free model (assuming that evolutionary rates differ among species, each node thus has an independent parameter).

From the results, we randomly sampled 500 iterations, each comprising the average Ks and Ka/Ks values for 50 single-copy gene families. This approach enabled us to obtain the average accumulation of synonymous mutations across different species and assess the relative evolution of protein sites. Ultimately, we visualized these results using ggplot2 in the R package to illustrate the variability in synonymous mutation accumulation and protein site evolution.

Construction and statistical analysis of gene trees

We explored the phylogenetic differences among the three subgenomes of the flowerpot snake, Diard's blindsnake, and Cantor's rat snake by reclustering the protein sequences of five genomes using Orthofinder, and Cantor's rat snake was designated as an outgroup. Using ETE v3 (*63*) software, we counted the occurrences of three tree topologies: Tree1 {[A, (B and C)], outgroup], Tree2 {[C, (A and B)], outgroup}, and Tree3 {[B, (A and C)], outgroup}.

Subsequently, we used Whole-Genome Duplication Integrated v0.6.5 (*64*) software to compute pairwise Ks values between coding genes under Tree1, Tree2, and Tree3 topologies. We randomly sampled 500 times, and average Ks values were calculated for 50 gene pairs for each random sample. Eventually, we visualized the results using the ggpubr package in R software.

Measurement of ILS across the whole genome and identification of ILS hotspots

We conducted whole-genome alignment of the three subgenomes of the flowerpot snake using Lastal version 1450, with the Diard's blindsnake genome serving as the reference genome, and the alignment results were merged and aligned using MULTIZ v11.2 (*65*) software. Subsequently, the alignment results were fragmented into 500-bp segments using MafFilter v1.3.1 (*66*) software, and neighbor-joining (NJ) trees were constructed for each segment. The obtained NJ trees were compared with the topologies of Tree1, Tree2, and Tree3 using ETE3 software to determine the number of different topological structures.

Using Diard's blindsnake chromosomes as a reference, we used the coverage command in Bedtools to count the number of regions within every 5-Mb interval across the entire genome that differed from the Tree1 topology. This enabled us to identify regions where ILS led to changes in tree topology. Subsequently, we used gene annotation files within these hotspot regions to determine whether functionally similar gene clusters existed in corresponding intervals.

Identification and functional analysis of UPs

Using miniprot v0.13 (*67*), we aligned functionally annotated proteins from Diard's blindsnake to the genomes of the flowerpot snake and Cantor's rat snake. For each protein, we extracted the best alignment across different genomes. Using the optimal alignment positions, we used MCScanX_h (*68*) to identify and determine whether these positions exhibited collinearity with proteins in the reference genome.

In our extracted results, we identified genes as pseudogenes if the alignment file contained frameshift mutations or premature stop codons. If a gene showed both pseudogene characteristics and collinearity, then we inferred that it evolved from a direct homologous gene during species differentiation, resulting in UPs. These sites were validated using resequencing data from different populations.

Nonmutated codon sequences were extracted to calculate the protein site relative evolutionary rate (Ka/Ks) between pseudogenes and Diard's blindsnake using the KaKs_Calculator v2.0 (*69*) on the basis of the YN model. Genes with direct homologous relationships and lacking the aforementioned mutations were considered normal genes, and their Ka/Ks values were also calculated using the KaKs_Calculator for comparison with UPs.

On the basis of the protein annotation results from Diard's blindsnake, UPs identified in the three subgenomes of the flowerpot snake were individually subjected to functional enrichment analysis using R packages to determine overlapping functionalities among UPs in different subgenomes. Last, identified UPs were searched in the STRING database to determine their association with sperm development and fertilization; these findings were then used to generate hypotheses about their potential effects on the flowerpot snake genome.

Identification and expression analysis of ovarian coexpression gene modules in the flowerpot Snake

To measure the expression levels in each tissue across different subgenomes, we used a competitive mapping approach to reduce erroneous mappings stemming from homologous regions. Initially, using subreads v2.0.2 (*70*) software, we adhered to quantification criteria by aligning transcriptome data from each tissue to each subgenome, ensuring that each read was uniquely assigned to its best alignment position on the basis of mapping quality. For Diard's blindsnake, we

used Subread software for alignment and featureCounts v2.0.6 (*71*) for quantification of aligned reads.

To identify gene expression regions across different subgenomes, we first used MMseqs v15-6f452 (*72*) software to extract reciprocal best hits among subgenomes, followed by MCScanX (*68*) software for collinearity analysis. We constructed collinearity matrix 1 (M1) with conserved collinear genes among the three subgenomes of the flowerpot snake and collinearity M2 by incorporating Diard's blindsnake protein sequences. Subsequently, using featureCounts on the basis of M1, we quantified reads mapped to each gene in every tissue, establishing expression matrix M1-E for all three subgenomes. Similarly, M2-E was constructed using M2, which encompassed the three subgenomes of the flowerpot snake and the genome of Diard's blindsnake.

Using Deseq2 (*73*), we normalized read counts in M1-E to obtain normalized expression values (M1-E-N) and assessed the distribution of gene expression across different subgenomes in various tissues. Log 2 transformation was applied to M1-E-N, followed by *t* tests to evaluate the significance of differences in average expression trends among subgenomes.

For M2-E, normalization with Deseq2 generated normalized expression values, and a trait matrix (TM) was constructed on the basis of the origin of each tissue. The ovarian samples of the flowerpot snake were designated as the "parthenogenetic ovary," the ovarian samples of the Diard's blindsnake were designated as the "sexual ovary," and other tissues were defined accordingly. Samples with consistent traits were labeled as 1, while inconsistencies were labeled as 0 in the TM. Using whole-genome coexpression network analysis (WGCNA) (*74*), we constructed expression modules on the basis of the M2-E gene expression matrix. The optimal soft threshold for network construction was determined using a scale-free R2 fit approach. Modules were delineated using a topological overlap dendrogram with a minimum module size of 30 genes, and modules were merged on the basis of a height cutoff threshold of 0.25.

Mantel's correlation analysis was performed between each coexpression module and the TM. The top two modules most significantly associated with parthenogenetic ovary and sexual ovary traits were extracted, and the average expression trends of genes within these modules were further examined. Genes within the module most significantly correlated with parthenogenetic ovary were subjected to GO term functional enrichment analysis using the R package clusterProfiler (*75*) to identify significant functional changes associated with this trait (BH corrected P value ≤ 0.05).

To determine whether the regulation of module-specific genes was influenced by chromatin accessibility, ATAC-seq data from the flowerpot snake and Diard's blindsnake ovaries were aligned to the reference genome using Chromap v0.2.6 (*76*) software. Competitive mapping was performed for the flowerpot snake across its three subgenomes to ensure that each read had a unique best alignment in each subgenome. FeatureCounts quantified reads aligned to regions containing genes from the top two modules correlated with parthenogenetic ovary and sexual ovary traits. Deseq2 was used to normalize the counts of aligned reads, and linear regression was performed in R software to assess the statistical significance of correlations (P value < 0.05) between ATAC-seq and RNA-seq measurements. The coefficient of determination (R^2) indicated the strength of the correlation. Differential expression analysis using Deseq2 was used to identify significant changes in expression levels for selected genes with log 2 fold changes greater than 1.5; these were sorted by magnitude using R software and visualized using the ggplot2 package.

Population genomic analysis

We obtained second-generation sequencing data from a flowerpot snake specimen in India from published sources. Although the study included a sample from Japan sequenced on the Nanopore platform, we excluded this dataset from our analysis because of its high error rate. We also collected 10 samples from Fujian, 1 sample from Hong Kong, and 1 sample from Yunnan. All samples underwent quality control and were aligned to the flowerpot snake genome using Burrows-Wheeler Aligner (BWA) v0.7.18 (*77*) software. The resequencing data for each sample showed a high alignment rate, ranging from approximately 99.5 to 99.9%. Subsequently, SNPs were called using the HaplotypeCaller in GATK with the --ploidy parameter set to 1. SNPs were filtered using the parameters "QD < 2.0 || FS > 60.0 || MQ < 40.0 || SOR > 3.0 || MQRankSum < -12.5 || ReadPosRankSum < -8.0" to ensure high-quality SNP calling. The remaining SNP positions across all samples were merged using VCFtools (*78*).

The integrated VCF file was imported into R and converted into a matrix using the vcfR package (*79*). PCA was performed on the SNP matrix, and hierarchical clustering analysis using the FactoMineR package (*80*) was performed on PC1 and PC2 to determine relationships among samples. The SNP density was calculated for each chromosome in 5-Mb windows on the basis of the geographic origin of the samples. Sequence depth per chromosome for each individual was obtained using the coverage tool in SAMtools. The SNP density and chromosome sequencing depth data were imported into R, averaged by sample origin and chromosome, and visualized using the ggplot2 package in R.

Statistical analysis

In our study, we conducted a series of statistical analyses to verify and interpret the genomic and transcriptomic data collected from the flowerpot snake and Diard's blindsnake. For each experiment detailed herein, specific statistical tests were performed, and values for N (the number of samples), P (P values), and the precise statistical tests performed were reported in the text and figure legends to allow the results to be verified by readers familiar with the original datasets.

Sequencing was performed using two platforms (PacBio and Illumina) across different samples. A total of 13 samples ($N = 13$; 1 from Hong Kong, $N = 1$ from Yunnan, and $N = 11$ from Fujian) were subjected to various sequencing processes, including construction of short DNA fragment libraries and HiFi sequencing. The quality of each assembled contig was evaluated, and the N50 metric was calculated to ensure a satisfactory assembly (N50 = Mb based on the HiFi read assembly).

RNA-seq produced paired-end reads of 150 bp. After normalization of the expression data, differential expression analysis was performed using the DESeq2 R package. Samples with high RNA integrity (RIN ≥ 8) were included, and a total of 32 samples were used for statistical testing. Differences in gene expression were determined using a t test, with a significance threshold of $P < 0.05$. Coexpression networks were analyzed using the WGCNA package, and modules significantly associated with ovarian traits were identified on the basis of a correlation coefficient threshold.

The genome-wide alignments for phylogenetic analysis were performed on the basis of a set of 14 reference snakes. PAML was applied to compute synonymous (Ks) and nonsynonymous (Ka) mutation rates for these genes. A total of 500 bootstrap iterations were sampled to assess the average rates, and differences among lineages were identified via a significance threshold of $P < 0.05$.

Data from 12 individual samples across various geographic locations ($N = 12$) were aligned using BWA software (*77*). SNP calling was performed using HaplotypeCaller in GATK, and stringent filtering criteria were used to yield high-quality SNPs. Population structure and genetic differentiation were assessed using PCA and hierarchical clustering, and the results were visualized on the basis of the variance explained by the first two principal components (PCA1 and PCA2).

For multiple comparisons within the gene expression analysis, Benjamini-Hochberg corrections were used to control for the false discovery rate (FDR). Functional enrichment analysis of identified gene clusters was conducted using the clusterProfiler R package, and the significance threshold was $P < 0.05$. Throughout the study, all statistical analyses were conducted using R software.

Supplementary Materials

This PDF file includes:

Figs. S1 to S19
Tables S1 to S10

REFERENCES AND NOTES

1. M. K. Fujita, S. Singhal, T. O. Brunes, J. A. Maldonado, Evolutionary dynamics and consequences of parthenogenesis in vertebrates. *Annu. Rev. Ecol. Evol. Syst.* **51**, 191–214 (2020).
2. K. Matsubara, Y. Kumazawa, H. Ota, C. Nishida, Y. Matsuda, Karyotype analysis of four blind snake species (Reptilia: Squamata: Scolecophidia) and karyotypic changes in serpentes. *Cytogenet. Genome Res.* **157**, 98–106 (2019).
3. D. Dedukh, M. Altmanova, J. Klima, L. Kratochvil, Premeiotic endoreplication is essential for obligate parthenogenesis in geckos. *Development* **149**, dev200345 (2022).
4. M. Arakelyan, T. Harutyunyan, S. A. Aghayan, M. A. Carretero, Infection of parthenogenetic lizards by blood parasites does not support the "Red Queen hypothesis" but reveals the costs of sex. *Zoology (Jena)* **136**, 125709 (2019).
5. P. H. Schiffer, E. G. J. Danchin, A. M. Burnell, C. J. Creevey, S. Wong, I. Dix, G. O'Mahony, B. A. Culleton, C. Rancurel, G. Stier, E. A. Martinez-Salazar, A. Marconi, U. Trivedi, M. Kroiher, M. A. S. Thorne, E. Schierenberg, T. Wiehe, M. Blaxter, Signatures of the evolution of parthenogenesis and cryptobiosis in the genomes of panagrolaimid nematodes. *iScience* **21**, 587–602 (2019).
6. M. O. Moreira, C. Fonseca, D. Rojas, Parthenogenesis is self-destructive for scaled reptiles. *Biol. Lett.* **17**, 20210006 (2021).
7. S. P. Otto, The evolutionary consequences of polyploidy. *Cell* **131**, 452–462 (2007).
8. M. Kuroda, T. Fujimoto, M. Murakami, E. Yamaha, K. Arai, Clonal reproduction assured by sister chromosome pairing in dojo loach, a teleost fish. *Chromosome Res.* **26**, 243–253 (2018).
9. A. A. Lutes, W. B. Neaves, D. P. Baumann, W. Wiegraebe, P. Baumann, Sister chromosome pairing maintains heterozygosity in parthenogenetic lizards. *Nature* **464**, 283–286 (2010).
10. GBIF.org. GBIF Occurrence Download (2017); 10.15468/dl.aztnnq.
11. A. H. Wynn, C. J. Cole, A. L. Gardner, Apparent triploidy in the unisexual Brahminy blind snake, *Ramphotyphlops braminus*. 2868, 1–7 (1987).
12. C. Peng, D. D. Wu, J. L. Ren, Z. L. Peng, Z. Ma, W. Wu, Y. Lv, Z. Wang, C. Deng, K. Jiang, C. L. Parkinson, Y. Qi, Z. Y. Zhang, J. T. Li, Large-scale snake genome analyses provide insights into vertebrate development. *Cell* **186**, 2959–2976.e22 (2023).
13. S. M. Buchanan, S. S. Schalm, T. Maniatis, Proteolytic processing of protocadherin proteins requires endocytosis. *Proc. Natl. Acad. Sci. U.S.A.* **107**, 17774–17779 (2010).
14. W. V. Chen, T. Maniatis, Clustered protocadherins. *Development* **140**, 3297–3302 (2013).
15. M. Mark, F. M. Rijli, P. Chambon, Homeobox genes in embryogenesis and pathogenesis. *Pediatr. Res.* **42**, 421–429 (1997).
16. A. Seifert, D. F. Werheid, S. M. Knapp, E. Tobiasch, Role of Hox genes in stem cell differentiation. *World J. Stem Cells.* **7**, 583–595 (2015).
17. S. Gadadhar, G. A. Viar, J. N. Hansen, A. Gong, A. Kostarev, C. Ialy-Radio, S. Leboucher, M. Whitfield, A. Ziyyat, A. Touré, L. Alvarez, G. Pigino, C. Janke, Tubulin glycylation controls axonemal dynein activity, flagellar beat, and male fertility. *Science* **371**, eabd4914 (2021).
18. S. D. Newcomer, J. A. Zeh, D. W. Zeh, Genetic benefits enhance the reproductive success of polyandrous females. *Proc. Natl. Acad. Sci. U.S.A.* **96**, 10236–10241 (1999).
19. E. Chae, K. Bomblies, S.-T. Kim, D. Karelina, M. Zaidem, S. Ossowski, C. Martín-Pizarro, R. A. E. Laitinen, B. A. Rowan, H. Tenenboim, S. Lechner, M. Demar, A. Habring-Müller, C. Lanz, G. Rätsch, D. Weigel, Species-wide genetic incompatibility analysis identifies immune genes as hot spots of deleterious epistasis. *Cell* **159**, 1341–1351 (2014).
20. A. A. Newton, R. R. Schnittker, Z. Yu, S. S. Munday, D. P. Baumann, W. B. Neaves, P. Baumann, Widespread failure to complete meiosis does not impair fecundity in parthenogenetic whiptail lizards. *Development* **143**, 4486–4494 (2016).
21. V. Spangenberg, M. Arakelyan, M. B. Cioffi, T. Liehr, A. Al-Rikabi, E. Martynova, F. Danielyan, I. Stepanyan, E. Galoyan, O. Kolomiets, Cytogenetic mechanisms of unisexuality in rock lizards. *Sci. Rep.* **10**, 8697 (2020).
22. D. Dedukh, M. Altmanová, R. Petrosyan, M. Arakelyan, E. Galoyan, L. Kratochvíl, Premeiotic endoreplication is the mechanism of obligate parthenogenesis in rock lizards of the genus *Darevskia*. bioRxiv 582286 [Preprint] (2024); https://doi.org/10.1101/2024.02.27.582286.
23. M. Ljungman, Activation of DNA damage signaling. *Mutat Res.* **577**, 203–216 (2005).
24. W. K. Kaufmann, R. S. Paules, DNA damage and cell cycle checkpoints. *FASEB J.* **10**, 238–247 (1996).
25. G. Khedkar, C. Kambayashi, H. Tabata, I. Takemura, R. Minei, A. Ogura, A. Kurabayashi, The draft genome sequence of the Brahminy blindsnake *Indotyphlops braminus*. *Sci. Data* **9**, 410 (2022).
26. S. Lien, B. F. Koop, S. R. Sandve, J. R. Miller, M. P. Kent, T. Nome, T. R. Hvidsten, J. S. Leong, D. R. Minkley, A. Zimin, F. Grammes, H. Grove, A. Gjuvsland, B. Walenz, R. A. Hermansen, K. von Schalburg, E. B. Rondeau, A. D. Genova, J. K. A. Samy, J. O. Vik, M. D. Vigeland, L. Caler, U. Grimholt, S. Jentoft, D. I. Våge, P. de Jong, T. Moen, M. Baranski, Y. Palti, D. R. Smith, J. A. Yorke, A. J. Nederbragt, A. Tooming-Klunderud, K. S. Jakobsen, X. Jiang, D. Fan, Y. Hu, D. A. Liberles, R. Vidal, P. Iturra, S. J. M. Jones, I. Jonassen, A. Maass, S. W. Omholt, W. S. Davidson, The Atlantic salmon genome provides insights into rediploidization. *Nature* **533**, 200–205 (2016).
27. A. Rhie, B. P. Walenz, S. Koren, A. M. Phillippy, Merqury: Reference-free quality, completeness, and phasing assessment for genome assemblies. *Genome Biol.* **21**, 245 (2020).
28. H. Sun, J. Ding, M. Piednoel, K. Schneeberger, findGSE: Estimating genome size variation within human and Arabidopsis using k-mer frequencies. *Bioinformatics* **34**, 550–557 (2018).
29. H. Cheng, G. T. Concepcion, X. Feng, H. Zhang, H. Li, Haplotype-resolved de novo assembly using phased assembly graphs with hifiasm. *Nat. Methods* **18**, 170–175 (2021).
30. M. Rautiainen, S. Nurk, B. P. Walenz, G. A. Logsdon, D. Porubsky, A. Rhie, E. E. Eichler, A. M. Phillippy, S. Koren, Telomere-to-telomere assembly of diploid chromosomes with Verkko. *Nat. Biotechnol.* **41**, 1474–1482 (2023).
31. S. Nurk, B. P. Walenz, A. Rhie, M. R. Vollger, G. A. Logsdon, R. Grothe, K. H. Miga, E. E. Eichler, A. M. Phillippy, S. Koren, HiCanu: Accurate assembly of segmental duplications, satellites, and allelic variants from high-fidelity long reads. *Genome Res.* **30**, 1291–1305 (2020).
32. A. M. Bolger, M. Lohse, B. Usadel, Trimmomatic: A flexible trimmer for Illumina sequence data. *Bioinformatics* **30**, 2114–2120 (2014).
33. C. Zhou, S. A. McCarthy, R. Durbin, YaHS: Yet another Hi-C scaffolding tool. *Bioinformatics* **39**, btac808 (2023).
34. N. C. Durand, J. T. Robinson, M. S. Shamim, I. Machol, J. P. Mesirov, E. S. Lander, E. L. Aiden, Juicebox provides a visualization system for Hi-c contact maps with unlimited zoom. *Cell Syst.* **3**, 99–101 (2016).
35. H. Li, Minimap2: Pairwise alignment for nucleotide sequences. *Bioinformatics* **34**, 3094–3100 (2018).
36. A. R. Quinlan, I. M. Hall, BEDTools: A flexible suite of utilities for comparing genomic features. *Bioinformatics* **26**, 841–842 (2010).
37. M. Manni, M. R. Berkeley, M. Seppey, F. A. Simao, E. M. Zdobnov, BUSCO Update: Novel and streamlined workflows along with broader and deeper phylogenetic coverage for scoring of eukaryotic, prokaryotic, and viral genomes. *Mol. Biol. Evol.* **38**, 4647–4654 (2021).
38. J. M. Flynn, R. Hubley, C. Goubert, J. Rosen, A. G. Clark, C. Feschotte, A. F. Smit, RepeatModeler2 for automated genomic discovery of transposable element families. *Proc. Natl. Acad. Sci. U.S.A.* **117**, 9451–9457 (2020).
39. J. Jurka, V. V. Kapitonov, A. Pavlicek, P. Klonowski, O. Kohany, J. Walichiewicz, Repbase Update, a database of eukaryotic repetitive elements. *Cytogenet. Genome Res.* **110**, 462–467 (2005).
40. M. Tarailo-Graovac, N. Chen, Using RepeatMasker to identify repetitive elements in genomic sequences. *Curr. Protoc. Bioinformatics* 4.10.1–4.10.14 (2009).
41. M. Stanke, O. Keller, I. Gunduz, A. Hayes, S. Waack, B. Morgenstern, AUGUSTUS: Ab initio prediction of alternative transcripts. *Nucleic Acids Res.* **34**, W435–W439 (2006).
42. I. Korf, Gene finding in novel genomes. *BMC Bioinformatics* **5**, 59 (2004).
43. E. M. Gertz, Y.-K. Yu, R. Agarwala, A. A. Schaffer, S. F. Altschul, Composition-based statistics and translated nucleotide searches: Improving the TBLASTN module of BLAST. *BMC Biol.* **4**, 41 (2006).

44. G. S. Slater, E. Birney, Automated generation of heuristics for biological sequence comparison. *BMC Bioinformatics* **6**, 31 (2005).
45. M. G. Grabherr, B. J. Haas, M. Yassour, J. Z. Levin, D. A. Thompson, I. Amit, X. Adiconis, L. Fan, R. Raychowdhury, Q. Zeng, Z. Chen, E. Mauceli, N. Hacohen, A. Gnirke, N. Rhind, F. di Palma, B. W. Birren, C. Nusbaum, K. Lindblad-Toh, N. Friedman, A. Regev, Full-length transcriptome assembly from RNA-Seq data without a reference genome. *Nat. Biotechnol.* **29**, 644–652 (2011).
46. W. J. Kent, BLAT—The BLAST-like alignment tool. *Genome Res.* **12**, 656–664 (2002).
47. T. D. Wu, C. K. Watanabe, GMAP: A genomic mapping and alignment program for mRNA and EST sequences. *Bioinformatics* **21**, 1859–1875 (2005).
48. B. J. Haas, A. L. Delcher, S. M. Mount, J. R. Wortman, R. K. Smith Jr., L. I. Hannick, R. Maiti, C. M. Ronning, D. B. Rusch, C. D. Town, S. L. Salzberg, O. White, Improving the Arabidopsis genome annotation using maximal transcript alignment assemblies. *Nucleic Acids Res.* **31**, 5654–5666 (2003).
49. B. J. Haas, S. L. Salzberg, W. Zhu, M. Pertea, J. E. Allen, J. Orvis, O. White, C. R. Buell, J. R. Wortman, Automated eukaryotic gene structure annotation using EVidenceModeler and the program to assemble spliced alignments. *Genome Biol.* **9**, R7 (2008).
50. H. Tang, V. Krishnakumar, X. Zeng, Z. Xu, A. Taranto, J. S. Lomas, Y. Zhang, W. C. Yim, J. Zhang, X. Zhang, JCVI: A versatile toolkit for comparative genomics analysis. *Imeta* **3**, e211 (2024).
51. H. Li, B. Handsaker, A. Wysoker, T. Fennell, J. Ruan, N. Homer, G. Marth, G. Abecasis, R. Durbin, 1000 Genome Project Data Processing Subgroup, The Sequence Alignment/Map format and SAMtools. *Bioinformatics* **25**, 2078–2079 (2009).
52. G. A. Van der Auwera, M. O. Carneiro, C. Hartl, R. Poplin, G. D. Angel, A. Levy-Moonshine, T. Jordan, K. Shakir, D. Roazen, J. Thibault, E. Banks, K. V. Garimella, D. Altshuler, S. Gabriel, M. A. DePristo, From FastQ data to high confidence variant calls: The Genome Analysis Toolkit best practices pipeline. *Curr. Protoc. Bioinformatics* **43**, 11.10.11–11.10.33 (2013).
53. M. Kokot, M. Dlugosz, S. Deorowicz, KMC 3: Counting and manipulating k-mer statistics. *Bioinformatics* **33**, 2759–2761 (2017).
54. W. Shen, S. Le, Y. Li, F. Hu, SeqKit: A cross-platform and ultrafast toolkit for FASTA/Q file manipulation. *PLOS ONE* **11**, e0163962 (2016).
55. K.-H. Jia, Z.-X. Wang, L. Wang, G.-Y. Li, W. Zhang, X.-L. Wang, F.-J. Xu, S.-Q. Jiao, S.-S. Zhou, H. Liu, Y. Ma, G. Bi, W. Zhao, Y. A. El-Kassaby, I. Porth, G. Li, R.-G. Zhang, J.-F. Mao, SubPhaser: A robust allopolyploid subgenome phasing method based on subgenome-specific k-mers. *New Phytol.* **235**, 801–809 (2022).
56. D. M. Emms, S. Kelly, OrthoFinder: phylogenetic orthology inference for comparative genomics. *Genome Biol.* **20**, 238 (2019).
57. K. Katoh, D. M. Standley, MAFFT multiple sequence alignment software version 7: improvements in performance and usability. *Mol. Biol. Evol.* **30**, 772–780 (2013).
58. F. Ronquist, M. Teslenko, P. van der Mark, D. L. Ayres, A. Darling, S. Hohna, B. Larget, L. Liu, M. A. Suchard, J. P. Huelsenbeck, MrBayes 3.2: Efficient Bayesian phylogenetic inference and model choice across a large model space. *Syst. Biol.* **61**, 539–542 (2012).
59. A. Stamatakis, RAxML version 8: a tool for phylogenetic analysis and post-analysis of large phylogenies. *Bioinformatics* **30**, 1312–1313 (2014).
60. Z. Yang, PAML 4: Phylogenetic analysis by maximum likelihood. *Mol. Biol. Evol.* **24**, 1586–1591 (2007).
61. S. Kumar, G. Stecher, M. Suleski, S. B. Hedges, TimeTree: A resource for timelines, timetrees, and divergence times. *Mol. Biol. Evol.* **34**, 1812–1819 (2017).
62. M. N. Puttick, MCMCtreeR: functions to prepare MCMCtree analyses and visualize posterior ages on trees. *Bioinformatics* **35**, 5321–5322 (2019).
63. J. Huerta-Cepas, F. Serra, P. Bork, ETE 3: Reconstruction, analysis, and visualization of phylogenomic data. *Mol. Biol. Evol.* **33**, 1635–1638 (2016).
64. P. Sun, B. Jiao, Y. Yang, L. Shan, T. Li, X. Li, Z. Xi, X. Wang, J. Liu, WGDI: A user-friendly toolkit for evolutionary analyses of whole-genome duplications and ancestral karyotypes. *Mol. Plant* **15**, 1841–1851 (2022).
65. M. Blanchette, W. J. Kent, C. Riemer, L. Elnitski, A. F. Smit, K. M. Roskin, R. Baertsch, K. Rosenbloom, H. Clawson, E. D. Green, D. Haussler, W. Miller, Aligning multiple genomic sequences with the threaded blockset aligner. *Genome Res.* **14**, 708–715 (2004).
66. J. Y. Dutheil, S. Gaillard, E. H. Stukenbrock, MafFilter: A highly flexible and extensible multiple genome alignment files processor. *BMC Genomics* **15**, 53 (2014).
67. H. Li, Protein-to-genome alignment with miniprot. *Bioinformatics* **39**, btad014 (2023).
68. Y. Wang, H. Tang, J. D. Debarry, X. Tan, J. Li, X. Wang, T. H. Lee, H. Jin, B. Marler, H. Guo, J. C. Kissinger, A. H. Paterson, MCScanX: A toolkit for detection and evolutionary analysis of gene synteny and collinearity. *Nucleic Acids Res.* **40**, e49 (2012).
69. D. Wang, Y. Zhang, Z. Zhang, J. Zhu, J. Yu, KaKs_Calculator 2.0: A toolkit incorporating gamma-series methods and sliding window strategies. *Genomics Proteomics Bioinformatics* **8**, 77–80 (2010).
70. Y. Liao, G. K. Smyth, W. Shi, The Subread aligner: Fast, accurate and scalable read mapping by seed-and-vote. *Nucleic Acids Res.* **41**, e108 (2013).
71. Y. Liao, G. K. Smyth, W. Shi, featureCounts: An efficient general purpose program for assigning sequence reads to genomic features. *Bioinformatics* **30**, 923–930 (2014).
72. M. Hauser, M. Steinegger, J. Soding, MMseqs software suite for fast and deep clustering and searching of large protein sequence sets. *Bioinformatics* **32**, 1323–1330 (2016).
73. M. I. Love, W. Huber, S. Anders, Moderated estimation of fold change and dispersion for RNA-seq data with DESeq2. *Genome Biol.* **15**, 550 (2014).
74. P. Langfelder, S. Horvath, WGCNA: An R package for weighted correlation network analysis. *BMC Bioinformatics* **9**, 559 (2008).
75. T. Wu, E. Hu, S. Xu, M. Chen, P. Guo, Z. Dai, T. Feng, L. Zhou, W. Tang, L. Zhan, X. Fu, S. Liu, X. Bo, G. Yu, clusterProfiler 4.0: A universal enrichment tool for interpreting omics data. *Innovation (Camb)* **2**, 100141 (2021).
76. H. Zhang, L. Song, X. Wang, H. Cheng, C. Wang, C. A. Meyer, T. Liu, M. Tang, S. Aluru, F. Yue, X. S. Liu, H. Li, Fast alignment and preprocessing of chromatin profiles with Chromap. *Nat. Commun.* **12**, 6566 (2021).
77. H. Li, R. Durbin, Fast and accurate short read alignment with Burrows-Wheeler transform. *Bioinformatics* **25**, 1754–1760 (2009).
78. P. Danecek, A. Auton, G. Abecasis, C. A. Albers, E. Banks, M. A. De Pristo, R. E. Handsaker, G. Lunter, G. T. Marth, S. T. Sherry, G. M. Vean, R. Durbin, 1000 Genomes Project Analysis Group, The variant call format and VCFtools. *Bioinformatics* **27**, 2156–2158 (2011).
79. B. J. Knaus, N. J. Grunwald, vcfr: a package to manipulate and visualize variant call format data in R. *Mol. Ecol. Resour.* **17**, 44–53 (2017).
80. S. Lê, J. Josse, F. Husson, FactoMineR: An R package for multivariate analysis. *J. Stat. Softw.* **25**, 1–18 (2008).

Acknowledgments

Funding: We are grateful for the funding provided to this project by the following project funds: the National Natural Science Foundation of China (32325011, 32200347, and 32220103004) and Sichuan Science and Technology Program (2023NSFSC1238). **Author contributions:** Writing—original draft: Y.L. Conceptualization: J.-T.L., Y.L., M.S., and M.K.F. Investigation: Y.L., W.W., J.-L.R., M.S., Z.W., K.J., C.Y., C.P., Z.P., and J.-T.L. Writing—review and editing: J.-T.L., W.W., and M.K.F. Methodology: Y.L., J.-T.L., and M.K.F. Resources: J.-T.L. and Z.P. Funding acquisition: J.-T.L., Y.L., and D.J. Data curation: Y.L., W.W., C.Y., J.-T.L.. Validation: Y.L., W.W., M.S., and J.-T.L. Supervision: J.-T.L. and D.J. Formal analysis: Y.L., W.W., and J.-T.L. Software: Y.L. Visualization: Y.L., W.W., and D.J. Project administration: J.T.L. and D.J. **Competing interests:** The authors declare that they have no competing interests. **Data and materials availability:** All data needed to evaluate the conclusions in the paper are present in the paper and/or the Supplementary Materials. Raw reads, genome assemblies, and annotations were deposited into National Genomics Data Center (https://ngdc.cncb.ac.cn/) with project number PRJCA030601 for *I. braminus*, PRJCA030602 for *A. diardii*, and PRJCA030603 for *P. dhumnades*. The genome assembly, mRNA sequences, protein sequences, gene annotations, and code used for analyzing and visualizing the data of the three snakes are available at Figshare (https://doi.org/10.6084/m9.figshare.27094237).

Submitted 5 October 2024
Accepted 26 February 2025
Published 2 April 2025
10.1126/sciadv.adt6477