# Evolution trajectories of snake genes and genomes revealed by comparative analyses of five-pacer viper

Wei Yin[1,*], Zong-ji Wang[2, 3, 4, 5,*], Qi-ye Li[3, 4, 6], Jin-ming Lian[3], Yang Zhou[3], Bing-zheng Lu[7], Li-jun Jin[3], Peng-xin Qiu[7], Pei Zhang[3], Wen-bo Zhu[7], Bo Wen[8], Yi-jun Huang[7], Zhi-long Lin[8], Bi-tao Qiu[3, 9], Xing-wen Su[7], Huan-ming Yang[8,9], Guo-jie Zhang[3, 4, 10], Guang-mei Yan[7,#], Qi Zhou[2,11, #]

1. Department of Biochemistry, Zhongshan School of Medicine, Sun Yat-sen University, Guangzhou, 510089, China.
2. Life Sciences Institute, The Key Laboratory of Conservation Biology for Endangered Wildlife of the Ministry of Education, Zhejiang University, Hangzhou, 310058, China.
3. China National Genebank, BGI-Shenzhen, Shenzhen, 518083, China.
4. State Key Laboratory of Genetic Resources and Evolution, Kunming Institute of Zoology, Chinese Academy of Sciences, Kunming, China.
5. School of Bioscience & Bioengineering, South China University of Technology, Guangzhou, 510006, China.
6. Centre for GeoGenetics, Natural History Museum of Denmark, University of Copenhagen, Øster Voldgade 5–7, 1350 Copenhagen K, Denmark.
7. Department of Pharmacology, Zhongshan School of Medicine, Sun Yat-sen University, Guangzhou, 510089, China.
8. BGI-Shenzhen, Shenzhen, 518083, China.
9. James D. Watson Institute of Genome Sciences, Hangzhou 310058, China
10. Centre for Social Evolution, Department of Biology, University of Copenhagen, Universitetsparken 15, DK-2100 Copenhagen, Denmark.
11. Department of Integrative Biology, University of California, Berkeley, Berkeley, CA94720, United States

[*] These authors contributed equally to the work

[#] To whom correspondence should be addressed: zhouqi1982@zju.edu.cn or ygm@mail.sysu.edu.cn

**Abstract**

Snake's numerous fascinating features distinctive from other tetrapods necessitate a rich history of genome evolution that is still obscure. To address this, we report the first high-quality genome of a viper, *Deinagkistrodon acutus* and comparative analyses using other species from major snake and lizard lineages. We map the evolution trajectories of transposable elements (TEs), developmental genes and sex chromosomes onto the snake phylogeny. TEs exhibit dynamic lineage-specific expansion. And in the viper many TEs may have been rewired into the regulatory network of brain genes, as shown by their associated expression with nearby genes in the brain but not in other tissues. We detect signatures of adaptive evolution in olfactory, venom and thermal-sensing genes, and also functional degeneration of genes associated with vision and hearing. Many *Hox* and *Tbx* limb-patterning genes show evidence of relaxed selective constraints, and such genes' phylogenetic distribution supports fossil evidence for a successive loss of forelimbs then hindlimbs during the snake evolution. Finally, we infer that the Z and W sex chromosomes had undergone at least three recombination suppression events at the ancestor of advanced snakes, with the W chromosomes showing a gradient of degeneration from basal to advanced snakes. These results, together with all the genes identified as undergoing adaptive or degenerative evolution episodes at respective snake lineages forge a framework for our deep understanding into snakes' molecular evolution history.

Snakes have undergone massive adaptive radiation with ~3400 extant species successfully inhabiting almost all continents except for the polar regions[1]. This process has culminated in 'advanced snakes' (Caenophidia, ~3000 species), involved numerous evolutionary changes in body form, chemo-/thermo- perception, venom and sexual reproductive systems, which together distinguish snakes from majorities of other Squamates (lizards and worm lizards). Some of these dramatic changes can be tracked from fossils, which established that the ancestor of snakes had already evolved a elongated body plan, probably as an adaptation to a burrowing and crawling lifestyle, but had lost only the forelimbs[2-4]. Extant boa and python species retain rudimentary hindlimbs, while advanced snakes have completely lost them. The limblessness, accompanied by degeneration in visual and auditory perception have not compromised snakes' dominant role as top predators, largely due to the evolution of infrared sensing and/or venom, and the development of corresponding facial pit and fangs (specialized teeth for venom injection) independently in different lineages[5,6].

These extreme adaptations have sparked strong and standing interest into their genetic basis. Snakes are used as a model for studying various basic questions like mechanisms of axial patterning and limb development[3,7,8], 'birth-and-death' of venom proteins[9-11], and also sex chromosome evolution[12]. Cytogenetic findings in snakes first drove Ohno to propose that sex chromosomes in vertebrates evolved from ancestral autosomes[13], like those of insects[14] and plants[15]. Our insights into these questions have been recently advanced by the application of next-generation sequencing. Analyses of python and king cobra genomes and transcriptomes have uncovered the metabolic gene repertoire involved in feeding, and inferred massive expansion and adaptive evolution of toxin families in elapids (an 'advanced' group)[10,16]. However, comparative studies of multiple snake genomes unraveling their evolutionary trajectories since the divergence from lizards are lacking, and so far only a few specific developmental 'toolbox' (e.g., *Hox*[7,17,18] and Fgf signaling pathway[19]) genes have been studied between snakes and lizards. This hampers our comprehensive understanding into the molecular basis of stepwise or independent acquisition of snake specific traits. We bridge this gap here by deep-sequencing the genomes and transcriptomes of five-pacer viper (*Deinagkistrodon*

*acutus*, **Figure 1A**), a member of the Viperidae family. This pit viper is a paragon of infrared sensing, heteromorphic ZW sex chromosomes, and distinctive types of fangs and toxins (its common name exaggerates that victims can walk no more than five paces) from other venomous snake families[6,20]. Despite the intraspecific differences within the same family, comparative analyses to the available genomes of three other major snake family species, i.e., *Boa constrictor* (Boidae family)[21], *Python bivittatus* (Pythonidae family)[16], *Ophiophagus hannah* (Elapidae family)[10] and several reptile outgroups should recapitulate the major genomic adaptive or degenerative changes occurred in the ancestor and along individual branches of snake families, and would also promote current antivenom therapy and drug discovery (e.g., thrombolytics) out of the viper toxins[22].

**Evolution of snake genome architecture**

We sequenced a male and a female five-pacer viper to extremely high-coverage (♀ 238 fold, ♂114 fold, **Supplementary Table 1**), and estimated the genome size to be 1.43Gb based on k-mer frequency distribution[23] (**Supplementary Table 2, Supplementary Figure 1**). Fewer than 10% of the reads, which have a low quality or are probably derived from repetitive regions, were excluded from the genome assembly (**Supplementary Table 3**). We generated a draft genome using only male reads for constructing the contigs, and female long-insert (2kb~40kb) library reads for joining the contigs into scaffolds. It has an assembled size of 1.47Gb, with a slightly better quality than the genome using purely female reads. It has high continuity (contig N50: 22.42kb, scaffold N50: 2.12Mb) and integrity (gap content 5.6%, **Supplementary Table 4**), thus was chosen as the reference genome for further analyses. It includes a total of 21,194 predicted protein-coding genes, as estimated using known vertebrate protein sequences and massive transcriptome data generated in this study from eight tissues (**Figure 1A, Materials and Methods**). For comparative analyses, we also annotated 17,392 protein-coding genes in the boa genome (the SGA version from[21]). 80.84% (17,134) of the viper genes show robust expression (normalized expression level RPKM>1) in at least one tissue, comparable to the value of 70.77% in king cobra (**Supplementary Table 5**). Based on 5,353 one-to-one orthologous gene groups of four snake species (five-pacer

viper, boa[21], python[16], king cobra[10]), green anole lizard[24] and several other sequenced vertebrate genomes (**Materials and Methods**), we constructed a phylogenomic tree with high bootstrapping values at all nodes (**Figure 1B**). We estimated that advanced snakes diverged from boa and python about 66.9 (47.2~84.4) million years ago (MYA), and five-pacer viper and king cobra diverged 44.9 (27.5~65.0) MYA assuming a molecular clock. These results are consistent with the oldest snake and viper fossils from 140.8 and 84.7 MYA, respectively[25].

The local GC content of snakes (boa and five-pacer viper) shows variation (GC isochores) similar to the genomes of turtles and crocodiles, and intermediate between mammals/birds and lizard (**Figure 1C, Supplementary Figure 2**), confirming the loss of such a genomic feature in lizard among tetrapods[24]. Cytogenetic studies showed like most other snakes, the five-pacer viper karyotype has 2n=36 (16 macro- and 20 micro-chromosomes) chromosomes[26], with extensive inter-chromosomal conservation with the lizard[27]. This enables us to organize 56.50% of the viper scaffold sequences into linkage groups, based on their homology with sequences of known green anole lizard macro-chromosomes (**Supplementary Table 6**). As expected, autosomal sequences have the same read coverage in both sexes, whereas scaffolds inferred to be located on the viper Z chromosome (homologous to green anole lizard chr6) have coverage in the female that is half that in the male (**Figure 1C**). Additionally, the frequency of heterozygous variants on the Z chromosome is much lower in the female than in the male (0.005% vs 0.08%, *P*-value<2.2e-16, Wilcoxon signed rank test) due to the nearly hemizygous state of Z chromosome in female, while those of autosomes (~0.1%) are very similar between sexes. These results indicated that our assembly mostly assigns genes to the correct chromosome, and this is further supported by comparisons of 172 genes' locations with previous fluorescence *in situ* hybridization results (**Supplementary Data 1**)[27]. They also suggest that the viper sex chromosomes are highly differentiated from each other (see below).

47.47% of the viper genome consists of transposable element sequences (TEs), higher than any other snakes so far analyzed (33.95~39.59%), which cannot be explained solely by the higher assembly quality than that of the other species genome

sequences[10,16,21] **(Supplementary Table 7-8)**. These are mostly long interspersed elements (LINE, 13.84% of the genome) and DNA transposons (7.96%, **Supplementary Table 7**). Sequence divergence of individual families from inferred consensus sequences uncovered recent rampant activities in the viper lineage of LINEs (CR1), DNA transposons (hAT and TcMar) and retrotransposons (Gypsy and DIRS). In particular, there is an excess of low-divergence (<10% divergence level) CR1 and hAT elements in the viper genome only (**Figure 2A**). We also inferred earlier propagation of TEs shared by viper and king cobra, which thus probably occurred in the ancestor of advanced snakes. Together, these derived insertions resulted in an at least three-fold difference in the CR1 and hAT content between viper and more basal-branching snakes like boa and python (**Figure 2B**). While boa and python have undergone independent expansion of L2 and CR1 repeats, so that their overall LINE content is at a similar level to that of the viper and cobra (**Figure 2A, Supplementary Table 7**).

These TEs are presumably silenced through epigenetic mechanisms to prevent their deleterious effects of transposition or mediating genomic rearrangements. Indeed, very few are transcribed in all of the tissue examined, except, unexpectedly in brain (**Figure 2C**). This brain-specific expression prompted us to test whether some snake TE families might have been co-opted into brain gene regulatory networks. Focusing on highly expressed (RPKM>5) TEs' that are located within 5kb flanking regions of genes, we found that these nearby genes also show a significantly higher expression in brain than any other tissues (*P-value*<1.1e-40, Wilcoxon test, **Supplementary Figure 3**). The expression levels of individual genes are strongly correlated (*P-value*<1.35e-08, Spearman's test) with those of nearby TE's. These genes are predominantly enriched (*Q-value* < 0.05, Fisher's Chi-square test, **Supplementary Data 2**) in functional domains of 'biological process' compared to 'cellular component' and 'molecular function', and particularly enriched categories include environmental response ('response to organic substance', 'regulation of response to stimulus' and 'sensory perception of light stimulus') and brain signaling pathways ('neuropeptide signaling pathway', 'opioid receptor signaling pathway' and 'regulation of cell communication' etc.). Further experimental studies are required to elucidate how some of these TEs evolved to regulate the brain

gene expression; these results nevertheless highlighted the evolution dynamics and potential functional contribution of TEs in shaping the snake genome evolution.

**Evolution of snake genes and gene families**

To pinpoint the critical genetic changes underlying the phenotypic innovations of snakes, we next mapped protein coding genes' gain and loss (**Figure 1B**), signatures of adaptive or degenerative evolution (**Figure 3A**) measured by their ratios ($\omega$) of nonsynonymous vs. synonymous substitution rates (**Supplementary Data 3**) onto the phylogenetic tree. We inferred a total of 1,725 gene family expansion and 3,320 contraction events, and identified 610 genes that appear to have undergone positive selection (PSG) and 6149 with relaxed selective constraints (RSG) at different branches, using a likelihood model and conserved lineage-specific test[28]. Genes of either scenario were separated for analyzing their enriched gene ontology (GO) and mouse orthologs' mutant phenotype (MP) terms, assuming most of them have a similar function in snakes.

Significantly (*P-value*<0.05, Fisher's exact test) enriched MP terms integrated with their branch information illuminated the molecular evolution history of snake-specific traits (**Figure 3A**). For example, as adaptations to a fossorial lifestyle, the four-legged snake ancestor[29] had evolved an extreme elongated body plan without limbs, and also fused eyelids ('spectacles', presumably for protecting eyes against soil[30]). The latter is supported by the results for the PSG genes *Ereg*, and RSGs *Cecr2* and *Ext1* at the snake ancestor branch (**Supplementary Data 4**), whose mouse mutant phenotype is shown as prematurely opened or absent eyelids. The limbless body plan has already driven many comparisons of expression domains and coding-sequences of the responsible *Hox* genes between snakes and other vertebrates[7,17]. We here refined the analyses to within-snake lineages, focusing on sequence evolution of *Hox* and other genes involved in limb development and somitogenesis. We annotated the nearly complete sequences of 39 *Hox* genes organized in four clusters (*HoxA*-*HoxD*) of the five-pacer viper. Compared to the green anole lizard, the four studied snake species have *Hox* genes whose sizes are generally reduced, due to the specific accumulation of DNA transposons in the lizard's introns and intergenic regions (**Supplementary Figure 4**). However, snakes have

accumulated particularly higher proportions of simple tandem repeat and short interspersed element (SINE) sequences within *Hox* clusters (**Supplementary Figure 5**), either as a result of relaxed selective constraints and/or evolution of novel regulatory elements. We identified 11 *Hox* genes as RSGs and one (*Hoxa9*) as PSG (**Figure 3B**). Their combined information of gene function and affected snake lineage informed the stepwise evolution of snake body plan. In particular, *Hoxa5*[31], *Hoxa11*[32] and *Tbx5*[33] which specifically pattern the forelimbs in mouse, have been identified as RSGs in the common ancestor of all four snakes. While *Hoxc11* and *Tbx4*[34], which pattern the hindlimbs in mouse, and many other limb-patterning genes (e.g. *Gli3*, *Tbx18*, *Alx4*) were identified as RSGs that evolved independently on snake external branches (**Figure 3B, Supplementary Data 4**). These results provide robust molecular evidence supporting the independent loss of hindlimbs after the complete loss of forelimbs in snake ancestors. In the snake ancestor branch, we also identified RSGs *Hoxa11, Hoxc10* and *Lfng*, which are respectively associated with sacral formation[35], rib formation[8] and somitogenesis speed[36] in vertebrates. Their changed amino acids and the expression domains that have expanded in snakes relative to lizards[17,19] might have together contributed to the 'de-regionalization'[17] and elongation of the snake body plan. In several external branches, we identified *Hoxd13* independently as RSG. Besides its critical roles in limb/digit patterning[37], it is also associated with termination of somitogenesis signal and is specifically silenced at the snake tail relative to lizard[7]. This suggests that body elongation may have evolved more than once among snake lineages. Overall, limb/digit/tail development MP terms are significantly enriched in RSGs at both ancestral and external branches of snakes (**Figure 3A**), and we identified many such genes at different snake lineages for future targeted experimental studies (**Supplementary Table 9, Supplementary Data 4**).

Another important adaption to snakes' ancestrally fossorial and later ground surface lifestyle is the shift of their dominant source of environmental sensing from visual/auditory to thermal/chemical cues. Unlike most other amniotes, extant snake species do not have external ears, and some basal species (e.g., blindsnake) have completely lost their eyes. Consistently, we found MP terms associated with hearing/ear

and vision/eye phenotypes (e.g., abnormal ear morphology, abnormal vision, abnormal cone electrophysiology) are enriched among RSGs along all major branches of snakes starting from their common ancestor (**Figure 3A, Supplementary Figure 6, Supplementary Data 4**). Gene families that have contracted in the ancestor of four studied snake species, and specifically in the viper, are also significantly (*Q-value*< 9.08e-4, Fisher's Exact Test) enriched in gene ontologies (GO) of 'sensory perception of light stimulus (GO:0050953)' or 'phototransduction (GO:0007602)' (**Figure 1C, Supplementary Data 5**). In particular, only three (*RH1, LWS* and *SWS1*) out of 13 opsin genes' complete sequences can be identified in the viper genome, consistent with the results found in python and cobra[16]. By contrast, infrared receptor gene *TRPA1*[5] and ubiquitous taste-signaling gene *TRPM5*[38] have respectively undergone adaptive evolution in five-pacer viper and the ancestor of boa and python. Gene families annotated with the GO term 'olfactory receptor (OR) activity' have a significant (*Q-value* < 1.63e-4, Fisher's Exact Test) expansion in all snake species studied and at some of their ancestral nodes, except for the king cobra (**Supplementary Data 6**). In the boa and viper, whose genome sequences have much better quality than the other two snake genomes, we respectively annotated 369 and 412 putatively functional OR genes, based on homology search and the characteristic 7-TM (transmembrane) structure (**Materials and Methods**). Both terrestrial species have an OR repertoire predominantly comprised of class II OR families (OR1-14, presumably for binding airborne molecules, **Figure 3C**), and their numbers are much higher than the reported numbers in other Squamate genomes[39]. Some (ranging from 18 to 24) class I (OR51-56, for water-borne molecules) genes have also been found in the two species, indicating this OR class is not unique to python as previously suggested[39]. Compared with the green anole lizard, the boa and viper exhibit a significant (*P*<0.05, Fisher's exact test) size expansion of OR family 5, 11 and 14, and also a bias towards being located on the Z chromosome (**Figure 3C**), leading to higher expression of many OR genes in males than in females (see below). In particular, OR5 in the viper probably has experienced additional expansion events and become the most abundant (with 71 members) family in the genome. Intriguingly, this family is specifically enriched in birds of prey[40] relative to other birds, and in non-frugivorous bats vs. frugivorous

bats[41]. Therefore, its expansion in the five-pacer viper could have been positively selected for a more efficient detection of prey.

Besides acute environmental sensing, specialized fangs[6] and venoms[11] (e.g, hemotoxins of viper or neurotoxins of elapid) arm the venomous snakes (~650 species) to immediately immobilize the much larger prey for prolonged ingestion, which probably comprised one of the most critical factors that had led to the advanced snakes' species radiation. It has been proposed that the tremendous venom diversity probably reflects snakes' local adaption for the prey[42] and was generated by changes in the expression of pre-existing or duplicated genes[11,43]. Indeed, we found that the five-pacer viper's venom gland gene repertoire has a very different composition comparing to other viper[44] or elapid species[10] (**Figure 3D**). We have annotated a total of 35 venom genes or gene families using all the known snake venom proteins as the query. Certain gene families, including snake venom metalloproteinases (SVMP), C-type lectin-like proteins (CLPs), thrombin-like snake venom serine proteinases (TL), Kunitz and disintegrins, have more genomic copies in the five-pacer viper than other studied snakes or the green anole lizard (**Supplementary Table 10**); while characteristic elapid venom genes like three-finger toxins (3FTx) are absent from the viper genome. Most venom proteins of both the viper and king cobra have expression restricted to venom or accessary glands, and for both species this is particularly seen for those genes that originated in the ancestor of snakes or of advanced snakes (**Figure 3D**). But for elapid or viper specific venom genes, i.e., those that originated more recently, they usually express in the liver of the other species. Such cases include FactorV, FactorX of king cobra and PLA2-2A of viper (**Figure 3D**). This suggests that these venom genes may have originated from metabolic proteins and undergone neo-/sub-functionalization, with altered expression.

**Evolution of snake sex chromosomes**

Different snake species exhibit a continuum of sex chromosome differentiation: pythons and boas possess homomorphic sex chromosomes, which is assumed to be the ancestral state; the lack of differentiation between the W and Z chromosomes suggests that most regions of this chromosome pair recombine like the autosomes[45]. Advanced snakes

usually have heteromorphic sex chromosomes that have undergone additional recombination suppressions[45,46]. We found the five-pacer viper probably has suppressed recombination throughout almost the entire sex chromosome pair, as the read coverage in the female that we sequenced is half that in the male (**Figure 1C, Figure 4**). By contrast, boa's homologous chromosomal regions show a read coverage pattern that does not differ from that of autosomes. Assuming that these two species share the same ancestral snake sex-determining region, this suggests that that region is not included in our current chromosomal assembly.

In plants, birds and mammals, it has been found that recombination suppression probably occurred by a succession of events. This has led to the punctuated accumulation of excessive neutral or deleterious mutations on the Y or W chromosome by genetic drift, and produced a gradient of sequence divergence levels over time, which are termed 'evolutionary strata'[47-49]. Advanced snakes have been suggested to have at least two strata[12]. One goal of our much better genome assembly of the five-pacer viper compared to those of any other studied advanced snakes[10,12] (**Supplementary Table 4**) was to reconstruct a fine history of snake sex chromosome evolution. We assembled 77Mb Z-linked and 33Mb W-linked scaffolds (**Materials and Methods**). The reduction of female read coverage along the Z chromosome suggests there is a substantial divergence between Z- and W- linked sequences, and this would enable the assembly of two chromosomes' scaffolds in separate. Mapping the male reads confirmed that the inferred W-linked scaffold sequences are only present in female (**Supplementary Figure 7**). Their density and pairwise sequence divergence values within putative neutral regions along the Z chromosome indicate at least two 'evolution strata', with the older stratum extending 0~56Mb, and the younger one extending 56~70Mb. The boundary at 56Mb region can be also confirmed by analyses of repetitive elements on the Z chromosome (see below). Consistently, identifiable W-linked fragments are found at the highest density per megabase in the latter (**Figure 4**), suggesting that this region has suppressed recombination more recently. The older stratum includes much fewer identifiable fragments that can resolve the actual times of recombination suppression events. To study this region further, we inspected the homologous Z-linked region, whose recombination

has also been reduced, albeit to a much smaller degree than the W chromosome, after the complete suppression of recombination between Z/W in females. In addition, Z chromosome is biasedly transmitted in males. As males usually have a higher mutation rate than females, due to many more rounds of DNA replication during spermatogenesis than oogenesis ('male-driven evolution')[50], Z-linked regions are expected to have a generally higher mutation rate than any other regions in the genome. This male-driven evolution effect has been demonstrated in other snake species[12], and also been validated for the snakes inspected in this study (**Supplementary Figure 8**). As a result, we expected regions in older strata should be more diverged from their boa autosome-like homologs than those in the younger strata. This enabled us to identify another stratum (0~42Mb, stratum 2, S2 in **Figure 4**) and demarcate the oldest one (42~56Mb, S1), by estimating the sequence conservation level (measured by LASTZ alignment score, blue line) between the Z chromosomes of boa and viper. The Z-linked region in the inferred oldest stratum S1 exhibits the highest sequence divergence with the homologous W-linked region, and also the highest proportion of repetitive elements (CR1, Gypsy and L1 elements; **Figure 4** shows the example of Gypsy; other repeats are shown in **Supplementary Figure 9**). This can be explained by the effect of genetic drift[51], which has been acting on the Z-linked S1 for the longest time since it reduced recombination rate in females. As a result, the accumulated repeats of S1 also tend to have a higher divergence level from the inferred ancestral consensus sequences compared to nearby strata (**Figure 4**). Unexpectedly, a similar enrichment was found in the homologous region of S1 in boa, despite being a recombining region and exhibiting the same coverage depth between sexes (**Figure 4**, **Supplementary Figure 9**). This indicates the pattern is partially contributed by the ancestral repeats that had already accumulated on the proto-sex chromosomes of snake species. Since our current viper sex chromosomal sequences used the green anole lizard chromosome 6 as a reference, rearrangements within this chromosome make it impossible to test whether S2 encompasses more than one stratum.

We dated the three resolved strata by constructing phylogenetic trees with homologous Z- and W- linked gene sequences of multiple snake species. Combining the

published CDS sequences of pygmy rattlesnake (Viperidae family species) and garter snake (Colubridae family species)[12], we found 31 homologous Z-W gene pairs, representing the three strata. All of them clustered by chromosome (i.e., the Z-linked sequences from all the species cluster together, separately from the W-linked ones) rather than by species (**Supplementary Figure 10-12**). This indicates that all three strata formed before the divergence of advanced snakes, and after their divergence from boa and python, i.e., about 66.9 million years ago (**Figure 1C**).

We found robust evidence of functional degeneration on the W chromosome. It is more susceptible to the invasion of TEs: the assembled sequences' overall repeat content is at least 1.5 fold higher than that of the Z chromosome, especially in the LINE L1 (2.9 fold) and LTR Gypsy families (4.3 fold) (**Supplementary Table 11 and Supplementary Figure 13**). Of 1,135 Z-linked genes, we were only able to identify 137 W-linked homologs. Among these, 62 (45.26%) have probably become pseudogenes due to nonsense mutations (**Supplementary Table 12**). W-linked loci generally transcribe at a significantly (*P-value*<0.0005, Wilcoxon test) lower level, with pseudogenes transcribing even lower relative to the autosomal or Z-linked loci regardless of the tissue type (**Supplementary Figure 14-15**). Given such a chromosome-wide gene loss, like other snakes[12] or majority of species with ZW sex chromosomes[52], the five-pacer viper shows a generally male-biased gene expression throughout the Z-chromosome and probably has not evolved global dosage compensation (**Supplementary Figure 16**).

## Discussion

The 'snake-like' body plan has evolved repeatedly in other tetrapods (e.g., worm lizard and caecilians), in which limb reduction/loss seems to have always been accompanied by the body elongation. For example, several limb-patterning *Hox* genes (*Hoxc10*, *Hoxd13*) identified as RSG also have been characterized by previous work with a changed expression domain along the snake body axis[7,17]. Another RSG *Hoxa5* which is involved in the forelimb patterning[31], also participates in lung morphogenesis[53]. It might have been involved in the elimination of one of the snake lungs during evolution. Therefore, the newly identified PSGs and RSGs throughout the snake phylogeny in this work

(**Supplementary Data 3-4, Supplementary Table 9**) can provide informative clues for future experimental work to use snake as an emerging 'evo-devo' model[54] to understand the genomic architecture of developmental regulatory network of organogenesis, or the crosstalk between these networks.

Like many of its other reptile relatives, the snake ancestor is very likely to determine the sex by temperature and does not have sex chromosomes. Extant species Boa can still undergo occasional parthenogenesis and is able to produce viable WW offspring[55], consistent with its the most primitive vertebrate sex chromosome pair reported to date. In the ancestor of advanced snakes, we inferred that there were at least three recombination suppressions occurred between Z/W, leading to a generally degenerated W chromosome that we have observed in five-pacer viper. How snakes genetically determine the sex would be a rather intriguing question to study in future.

## Materials and Methods

### Genome sequencing and assembly

All animal procedures were carried out with the approval of China National Genebank animal ethics committee. We extracted genomic DNAs from blood of a male and a female five-pacer viper separately. A total of 13 libraries with insert sizes ranging from 250bp to 40kb were constructed using female DNA, and 3 libraries with insert sizes from 250 bp to 800 bp were constructed using male DNA. We performed paired-end sequencing (HiSeq 2000 platform) following manufacturer's protocol, and produced 528 Gb raw data (357 Gb for female and 171 Gb for male). We estimated the genome size based on the K-mer distribution. A K-mer refers to an artificial sequence division of K nucleotides iteratively from sequencing reads. The genome size can then be estimated through the equation G=K_num/Peak_depth, where the K_num is the total number of K-mer, and Peak_depth is the expected value of K-mer depth[56]. We found a single main peak in the male K-mer (K=17) frequency distribution and an additional minor peak in the female data, the latter of which probably results from the divergence between W and Z chromosomes **(Supplementary Figure 1)**. Based on the distribution, we estimated that the genome size of this species is about 1.43 Gb **(Supplementary Table 2)**, comparable

411 to that of other snakes (1.44 Gb and 1.66Gb for Burmese python and King cobra[10,16],
412 respectively).
413 After filtering out low-quality and duplicated reads, we performed additional filtering
414 using the following criteria: we excluded the reads from short-insert libraries (250, 500,
415 800 bp) with 'N's over 10% of the length or having more than 40 bases with the quality
416 lower than 7, and the reads from large-insert libraries (2 kb to 40 kb) with 'N's over 20%
417 of the length or having more than 30 bases with the quality lower than 7. Finally, 109.20
418 Gb (73X coverage) male reads and 148.49Gb (99X coverage) female reads were retained
419 for genome assembly **(Supplementary Table 1)** using SOAPdenovo[57]
420 (http://soap.genomics.org.cn). To assemble the female and male genomes, reads from
421 small-insert libraries of the female and male individual were used for contig construction
422 separately. Then read-pairs from small- and large-insert libraries were utilized to join the
423 contigs into scaffolds. We also used female long-insert libraries to join the male contigs
424 into the longer scaffolds. At last, small-insert libraries of female and male individuals
425 were used for gap closure for their respective genomes. The final assemblies of female
426 and male have a scaffold N50 length of 2.0 Mb and 2.1 Mb respectively, and the gap
427 content of the two genomes are both less than 6% (♀ 5.29%, ♂ 5.61%)**(Supplementary**
428 **Table 4)**.
429 To access the assembly quality, reads from small-insert libraries that passed our
430 filtering criteria were aligned onto the two assemblies using BWA[58] (Version: 0.5.9-r16)
431 allowing 8 mismatches and 1 indel per read. A total of ~97% reads can be mapped back
432 to the draft genome (**Supplementary Table 3**), spanning 98% of the assembled regions
433 excluding gaps (**Supplementary Table 13**), and most genomic bases were covered by
434 about 80X reads **(Supplementary Figure 17)**. Thus we conclude that we have assembled
435 most part of the five-pacer viper genome. To further test for potential mis-joining of the
436 contigs into scaffolds, we analyzed the paired-end information and found that 57% of the
437 paired-end reads can be aligned uniquely with the expected orientation and distance. This
438 proportion of the long insert library is significantly lower than that from the short insert
439 libraries due to a circulization step during the library construction. When such
440 paired-ends were excluded, the proportion increased to 94.98% (**Supplementary Table**

**3**). Overall, these tests suggested that the contigs and scaffolds are consistent with the extremely high density of paired-end reads, which in turn indicated the high-quality of the assembly.

Previous cytogenetic studies showed that snake genomes show extensive inter-chromosomal conservation with lizard[27,45]. Thus we used the chromosomal information from green anole lizard[24] as a proxy to assign the snake scaffolds. We first constructed their orthologous relationship combining information of synteny and reciprocal best BLAST hits (RBH). Then gene coordinates and strandedness from the consensus chromosome were used to place and orient the snake scaffolds. Furthermore, we linked scaffolds into chromosomes with 600 'N's separating the adjacent scaffolds. In total, 625 five-pacer viper scaffolds comprising 832Mb (56.50% scaffolds in length) were anchored to 5 autosomes and Z chromosome (**Supplementary Table 6**).

**Repeat and gene annotation**

We identified the repetitive elements in the genome combining both homology-based and *de novo* predictions. We utilized the 'Tetrapoda' repeat consensus library in Repbase[59] for RepeatMasker (http://www.repeatmasker.org) to annotate all the known repetitive elements in the five-pacer viper genome. In order to maximize the identification and classification of repeat elements, we further used RepeatModeler (http://www.repeatmasker.org/RepeatModeler.html) to construct the consensus repeat sequence libraries of the green anole lizard, boa and five-pacer viper, then used them as a query to identify repetitive elements using RepeatMasker. Finally, we retrieved a non-redundant annotation for each species after combining all the annotation results using libraries of 'Tetrapoda', 'green anole lizard', 'boa' and 'five-pacer viper'. For the purpose of comparison, we ran the same pipeline and parameters in all the snake and lizard genomes as shown in **Supplementary Table 7**. To provide a baseline estimate for the sequence divergence of TEs from the snake ancestral status, we first merged the genomes from boa and five-pacer viper, and constructed the putative ancestral consensus

sequences using RepeatModeler. Then TE sequences of each snake species were aligned to the consensus sequence to estimate their divergence level using RepeatMasker.

For gene annotation, we combined resources of sequence homology, *de novo* prediction and transcriptome to build consensus gene models of the reference genome. Protein sequences of green lizard, chicken and human were aligned to the reference assembly using TBLASTN (E-value <= 1E-5)[60]. Then the candidate gene regions were refined by GeneWise[61] for more accurate splicing sites and gene models. We randomly selected 1000 homology-based genes to train Augustus[62] for *de novo* prediction on the pre-masked genome sequences. We mapped RNA-seq reads of 13 samples to the genome using TopHat (v1.3.1)[63] and then assembled the transcripts by Cufflinks (v1.3.0) (http://cufflinks.cbcb.umd.edu/). Transcripts from different samples were merged by Cuffmerge. Finally, gene models from these three methods were combined into a non-redundant gene set.

We finally obtained 21,194 protein-coding genes with intact open reading frames (ORFs) **(Supplementary Table 14)**. The gene models (measured by gene length, mRNA length, exon number and exon length) are comparable to those of other vertebrates and are well supported by the RNA-Seq data **(Supplementary Figure 18 and Supplementary Table 5).** To annotate the gene names for each predicted protein-coding locus, we first mapped all the 21,194 genes to a manually collected Ensembl gene library, which consists of all proteins from *Anolis carolinensis, Gallus gallus, Homo sapiens, Xenopus tropicalis and Danio rerio.* Then the best hit of each snake gene was retained based on its BLAST alignment score, and the gene name of this best hit gene was assigned to the query snake gene. Most of the predicted genes can be found for their orthologous genes in the library at a threshold of 80% alignment rate (the aligned length divided by the original protein length), suggesting our annotation has a high quality **(Supplementary Table 15)**.

## RNA-seq and gene expression analyses

Total RNAs were isolated from four types of tissues collected from both sexes, including brain, liver, venom gland and gonad **(Supplementary Table 16).** RNA sequencing

libraries were constructed using the Illumina mRNA-Seq Prep Kit. Briefly, oligo(dT) magnetic beads were used to purify poly-A containing mRNA molecules. The mRNAs were further fragmented and randomly primed during the first strand synthesis by reverse transcription. This procedure was followed by a second-strand synthesis with DNA polymerase I to create double-stranded cDNA fragments. The cDNAs were subjected to end-repairing by Klenow and T4 DNA polymerases and A-tailed by Klenow lacking exonuclease activity. The fragments were ligated to Illumina Paired-End Sequencing adapters, size selected by gel electrophoresis and then PCR amplified to complete the library preparation. The paired-end libraries were sequenced using Illumina HiSeq 2000 (90/100 bp at each end).

We used TopHat (v1.3.1) for aligning the RNA-seq reads and predicting the splicing junctions with the following parameters: -I/--max-intron-length: 10000, --segment-length: 25, --library-type: fr-firststrand, --mate-std-dev 10, -r/--mate-inner-dist: 20. Gene expression was measured by reads per kilobase of gene per million mapped reads (RPKM). To minimize the influence of different samples, RPKMs were adjusted by a scaling method based on TMM (trimmed mean of M values; M values mean the log expression ratios)[64] which assumes that the majority of genes are common to all samples and should not be differentially expressed.

**Evolution analyses**

A phylogenetic tree of the five-pacer viper and the other sequenced genomes (*Xenopus tropicalis, Homo sapiens, Mus musculus, Gallus gallus, Chelonia mydas, Alligator mississippiensis, Anolis carolinensis, Boa constrictor, Python bivittatus* and *Ophiophagus hannah*) was constructed using the 5,353 orthologous single-copy genes. Treebest (http://treesoft.sourceforge.net/treebest.shtml) was used to construct the phylogenetic tree. To estimate the divergence times between species, for each species, 4-fold degenerate sites were extracted from each orthologous family and concatenated to one sequence for each species. The MCMCtree program implemented in the Phylogenetic Analysis by Maximum Likelihood (PAML)[65] package was used to estimate the species divergence time. Calibration time was obtained from the TimeTree database

(http://www.timetree.org/). Three calibration points were applied in this study as normal priors to constrain the age of the nodes described below. 61.5-100.5 MA for the most recent common ancestor (TMRCA) of human-mouse; 259.7-299.8 MA for TMRCA of Crocodylidae and Lepidosauria; 235- 250.4 MA for TMRCA of Aves and Crocodylidae[66].

To examine the evolution of gene families in Squamate reptiles, genes from four snakes (*Boa constrictor, Python bivittatus, Deinagkistrodon acutus, Ophiophagus hannah*) and green anole lizard were clusterred into gene families by Treefam (min_weight=10, min_density=0.34, and max_size=500) [67]. The family expansion or contraction analysis was performed by CAFE[68]. In CAFE, a random birth-and-death model was proposed to study gene gain and loss in gene families across a user-specified phylogenetic tree. A global parameter $\lambda$ (lambda), which described both gene birth ($\lambda$) and death ($\mu = -\lambda$) rate across all branches in the tree for all gene families was estimated using maximum likelihood method. A conditional p-value was calculated for each gene family, and the families with conditional p-values lower than 0.05 were considered to have a significantly accelerated rate of expansion and contraction.

For the PAML analyses, we first assigned orthologous relationships for 12,657 gene groups among all Squamata and outgroup (turtle) using the reciprocal best blast hit algorithm and syntenic information. We used PRANK[69] to align the orthologous gene sequences, which takes phylogenetic information into account when placing a gap into the alignment. We filtered the PRANK alignments by gblocks[70] and excluded genes with high proportion of low complexity or repetitive sequences to avoid alignment errors. To identify the genes that evolve under positive selection (PSGs), we performed likelihood ratio test (LRT) using the branch model by PAML[65]. We first performed a LRT of the two-ratio model, which calculates the dN/dS ratio for the lineage of interest and the background lineage, against the one-ratio model assuming a uniform dN/dS ratio across all branches, so that to determine whether the focal lineage is evolving significantly faster (p-value < 0.05). In order to differentiate between episodes of positive selection and relaxation of purifying selection (RSGs), we performed a LRT of two-ratios model against the model that fixed the focal lineage's dN/dS ratio to be 1 (p-value < 0.05) and also required PSGs with the free-ratio model dN/dS > 1 at the focal lineage. For the

identified RSGs and PSGs, we used their mouse orthologs' mutant phenotype information[71] and performed enrichment analyses using MamPhEA[72]. Then we grouped the enriched MP terms by different tissue types.

**Olfactory receptor (OR), *Hox* and venom gene annotation**

To identify the nearly complete functional gene repertoire of OR, *Hox* and venom toxin genes in the investigated species, we first collected known amino acid sequences of 458 intact OR genes from three species (green anole lizard, chicken and zebra finch)[73], all annotated *Hox* genes from *Mus musculus* and *HoxC3* from *Xenopus tropicalis*, and obtained the query sequences of a total of 35 venom gene families[74] from UniProt (http://www.uniprot.org/) and NCBI (http://www.ncbi.nlm.nih.gov/). These 35 venom gene families represent the vast majority of known snake venoms. Then we performed a TBlastN[60] search with the cutoff E-value of 1E-5 against the genomic data using these query sequences. Aligned sequence fragments were combined into one predicted gene using perl scripts if they belonged to the same query protein. Then each candidate gene region was extended for 2kb from both ends to predict its open reading frame by GeneWise[61]. Obtained sequences were verified as corresponding genes by BlastP searches against NCBI nonredundant (nr) database. Redundant annotations within overlapped genomic regions were removed.

For the OR gene prediction, these candidates were classified into functional genes and nonfunctional pseudogenes. If a sequence contained any disruptive frame-shift mutations and/or premature stop codons, it was annotated as a pseudogene. The remaining genes were examined using TMHMM2.0[75]. Those OR genes containing more at least 6 transmembrane (TM) structures were considered as intact candidates and the rest were also considered as pseudogenes. Finally, each OR sequence identified was searched against the HORDE (the Human Olfactory Data Explorer) database (http://genome.weizmann.ac.il/horde/) using the FASTA (ftp://ftp.virginia.edu/pub/fasta) and classified into the different families according to their best-aligned human OR sequence. For the venom toxin genes, we only kept these genes with RPKM higher than 1 in the five-pacer viper and king cobra venom gland tissue as final toxin gene set.

589

590 **Identification and analyses of sex-linked genes**

591 To identify the Z-linked scaffolds in the male assembly, we aligned the female and male
592 reads to the male genome separately with BWA[58] allowing 2 mismatches and 1 indel.
593 Scaffolds with less than 80% alignment coverage (excluding gaps) or shorter than 500 bp
594 in length were excluded. Then single-base depths were calculated using SAMtools[76], with
595 which we calculated the coverage and mean depth for each scaffold. The expected male
596 vs. female (M:F) scaled ratio of a Z-linked scaffold is equal to 2, and we defined a
597 Z-linked scaffold with the variation of an observed scaled ratio to be less than 20% (i.e.
598 1.6 to 2.4). With this criteria, we identified 139 Z-lined scaffolds, representing 76.93Mb
599 with a scaffold N50 of 962 kb (**Supplementary Table 17**). These Z-linked scaffolds were
600 organized into pseudo-chromosome sequence based on their homology with green anole
601 lizard. Another characteristic pattern of the Z-linked scaffolds is that there should be
602 more heterozygous SNPs in the male individual than in the female individual resulted
603 from their hemizygous state in female. We used SAMtools[76] for SNP/indel calling. SNPs
604 and indels whose read depths were too low (<10) or too high (>120), or qualities lower
605 than 100 were excluded. As expected, the frequency of heterozygous sites of Z
606 chromosome of the female individual is much lower than that of the male individual
607 (0.005% vs 0.08%), while the heterozygous rate of autosomes are similar in both sex
608 (~0.1%) (**Supplementary Table 18**). To identify the W-linked scaffolds, we used the
609 similar strategy as the Z-linked scaffold detection to obtain the coverage and mean depth
610 of each scaffold. Then we identified those scaffolds covered by female reads over 80% of
611 the length, and by male reads with less than 20% of the length. With this method, we
612 identified 33 Mb W-linked scaffolds with a scaffold N50 of 48 kb (**Supplementary Table**
613 **19**).

614 We used the protein sequences of Z/W gametologs from garter snake and pygmy
615 rattle snake[12] as queries and aligned them to the genomes of boa (the SGA assembly,
616 http://gigadb.org/dataset/100060), five-pacer viper and king cobra with BLAST[60]. The
617 best aligned (cutoff: identity>=70%, coverage>=50%) region with extended flanking
618 sequences of 5kb at both ends was then used to determine whether it contains an intact

open reading frame (ORF) by GeneWise[61] (-tfor -genesf -gff -sum). We annotated the ORF as disrupted when GeneWise reported at least one premature stop codon or frame-shift mutation. CDS sequences of single-copy genes' Z/W gametologs were aligned by MUSCLE[77] and the resulting alignments were cleaned by gblocks[70] (-b4=5, -t=c, -e=-gb). Only alignments longer than 300bp were used for constructing maximum likelihood trees by RAxML[78] to infer whether their residing evolutionary stratum is shared among species or specific to lineages.

**References**


1 Greene, H. W. *Snakes: the evolution of mystery in nature*. (Univ of California Press, 1997).

2 Tchernov, E., Rieppel, O., Zaher, H., Polcyn, M. J. & Jacobs, L. L. A fossil snake with limbs. *Science* **287**, 2010-2012 (2000).

3 Cohn, M. J. & Tickle, C. Developmental basis of limblessness and axial patterning in snakes. *Nature* **399**, 474-479, doi:10.1038/20944 (1999).

4 Apesteguia, S. & Zaher, H. A Cretaceous terrestrial snake with robust hindlimbs and a sacrum. *Nature* **440**, 1037-1040, doi:10.1038/nature04413 (2006).

5 Gracheva, E. O. *et al.* Molecular basis of infrared detection by snakes. *Nature* **464**, 1006-1011, doi:10.1038/nature08943 (2010).

6 Vonk, F. J. *et al.* Evolutionary origin and development of snake fangs. *Nature* **454**, 630-633, doi:10.1038/nature07178 (2008).

7 Di-Poi, N. *et al.* Changes in Hox genes' structure and function during the evolution of the squamate body plan. *Nature* **464**, 99-103, doi:10.1038/nature08789 (2010).

8 Guerreiro, I. *et al.* Role of a polymorphism in a Hox/Pax-responsive enhancer in the evolution of the vertebrate spine. *Proc Natl Acad Sci U S A* **110**, 10682-10686, doi:10.1073/pnas.1300592110 (2013).

9 Fry, B. G. From genome to "venome": molecular origin and evolution of the snake venom proteome inferred from phylogenetic analysis of toxin sequences and related body proteins. *Genome Res* **15**, 403-420, doi:10.1101/gr.3228405

(2005).

10 Vonk, F. J. *et al.* The king cobra genome reveals dynamic gene evolution and adaptation in the snake venom system. *Proc Natl Acad Sci U S A* **110**, 20651-20656, doi:10.1073/pnas.1314702110 (2013).

11 Fry, B. G., Vidal, N., van der Weerd, L., Kochva, E. & Renjifo, C. Evolution and diversification of the Toxicofera reptile venom system. *J Proteomics* **72**, 127-136, doi:10.1016/j.jprot.2009.01.009 (2009).

12 Vicoso, B., Emerson, J. J., Zektser, Y., Mahajan, S. & Bachtrog, D. Comparative sex chromosome genomics in snakes: differentiation, evolutionary strata, and lack of global dosage compensation. *PLoS Biol* **11**, e1001643, doi:10.1371/journal.pbio.1001643 (2013).

13 Ohno, S. *Sex Chromosomes and Sex-linked Genes*. Vol. 1967 (Springer Berlin Heidelberg, 1967).

14 Kaiser, V. B. & Bachtrog, D. Evolution of sex chromosomes in insects. *Annu Rev Genet* **44**, 91-112, doi:10.1146/annurev-genet-102209-163600 (2010).

15 Westergaard, M. The mechanism of sex determination in dioecious flowering plants. *Adv Genet* **9**, 217-281 (1958).

16 Castoe, T. A. *et al.* The Burmese python genome reveals the molecular basis for extreme adaptation in snakes. *Proc Natl Acad Sci U S A* **110**, 20645-20650, doi:10.1073/pnas.1314475110 (2013).

17 Woltering, J. M. *et al.* Axial patterning in snakes and caecilians: evidence for an alternative interpretation of the Hox code. *Dev Biol* **332**, 82-89, doi:10.1016/j.ydbio.2009.04.031 (2009).

18 Head, J. J. & Polly, P. D. Evolution of the snake body form reveals homoplasy in amniote Hox gene function. *Nature* **520**, 86-89, doi:10.1038/nature14042 (2015).

19 Gomez, C. *et al.* Control of segment number in vertebrate embryos. *Nature* **454**, 335-339, doi:10.1038/nature07020 (2008).

20 Zhang, B. *et al.* Transcriptome analysis of Deinagkistrodon acutus venomous gland focusing on cellular structure and functional aspects using expressed sequence tags. *BMC Genomics* **7** (2006).

21 Bradnam, K. R. *et al.* Assemblathon 2: evaluating de novo methods of genome assembly in three vertebrate species. *Gigascience* **2**, 10, doi:10.1186/2047-217X-2-10 (2013).

22 Lin, X. *et al.* The effect of the fibrinolytic enzyme FIIa from Agkistrodon acutus venom on acute pulmonary thromboembolism. *Acta Pharmacologica Sinica* **32**, 239-244 (2011).

23 Li, Z. *et al.* Comparison of the two major classes of assembly algorithms: overlap-layout-consensus and de-bruijn-graph. *Brief Funct Genomics* **11**, 25-37, doi:10.1093/bfgp/elr035 (2012).

24 Alfoldi, J. *et al.* The genome of the green anole lizard and a comparative analysis with birds and mammals. *Nature* **477**, 587-591, doi:10.1038/nature10390 (2011).

25 Pyron, R. A. & Burbrink, F. T. Extinction, ecological opportunity, and the origins of global snake diversity. *Evolution* **66**, 163-178, doi:10.1111/j.1558-5646.2011.01437.x (2012).

26 Yunfang, Q., Xingfu, X., Youjin, Y., Fuming, D. & Meihua., H. Chromosomal studies on six species of venomous snakes in Zhejiang. *Acta Zoologica Sinica* **273**, 218-227 (1981).

27 Srikulnath, K. *et al.* Karyotypic evolution in squamate reptiles: comparative gene mapping revealed highly conserved linkage homology between the butterfly lizard (Leiolepis reevesii rubritaeniata, Agamidae, Lacertilia) and the Japanese four-striped rat snake (Elaphe quadrivirgata, Colubridae, Serpentes). *Chromosome Res* **17**, 975-986, doi:10.1007/s10577-009-9101-7 (2009).

28 Yang, Z. Likelihood ratio tests for detecting positive selection and application to primate lysozyme evolution. *Mol Biol Evol* **15**, 568-573 (1998).

29 Martill, D. M., Tischlinger, H. & Longrich, N. R. EVOLUTION. A four-legged snake from the Early Cretaceous of Gondwana. *Science* **349**, 416-419, doi:10.1126/science.aaa9208 (2015).

30 Held, J. *How the Snake Lost Its Legs: Curious Tales from the Frontier of Evo-Devo*. (Cambridge University Press, 2014).

31 Xu, B. *et al.* Hox5 interacts with Plzf to restrict Shh expression in the developing

forelimb. *Proc Natl Acad Sci U S A* **110**, 19438-19443, doi:10.1073/pnas.1315075110 (2013).

32 Boulet, A. M. & Capecchi, M. R. Multiple roles of Hoxa11 and Hoxd11 in the formation of the mammalian forelimb zeugopod. *Development* **131**, 299-309, doi:10.1242/dev.00936 (2004).

33 King, M., Arnold, J. S., Shanske, A. & Morrow, B. E. T-genes and limb bud development. *Am J Med Genet A* **140**, 1407-1413, doi:10.1002/ajmg.a.31250 (2006).

34 Logan, M. & Tabin, C. J. Role of Pitx1 upstream of Tbx4 in specification of hindlimb identity. *Science* **283**, 1736-1739 (1999).

35 Wellik, D. M. & Capecchi, M. R. Hox10 and Hox11 genes are required to globally pattern the mammalian skeleton. *Science* **301**, 363-367, doi:10.1126/science.1085672 (2003).

36 McGrew, M. J., Dale, J. K., Fraboulet, S. & Pourquie, O. The lunatic fringe gene is a target of the molecular clock linked to somite segmentation in avian embryos. *Curr Biol* **8**, 979-982 (1998).

37 Muragaki, Y., Mundlos, S., Upton, J. & Olsen, B. R. Altered growth and branching patterns in synpolydactyly caused by mutations in HOXD13. *Science* **272**, 548-551 (1996).

38 Kaske, S. *et al.* TRPM5, a taste-signaling transient receptor potential ion-channel, is a ubiquitous signaling component in chemosensory cells. *BMC Neuroscience* **8**, 49 (2007).

39 Dehara, Y. *et al.* Characterization of squamate olfactory receptor genes and their transcripts by the high-throughput sequencing approach. *Genome Biol Evol* **4**, 602-616, doi:10.1093/gbe/evs041 (2012).

40 Khan, I. *et al.* Olfactory receptor subgenomes linked with broad ecological adaptations in Sauropsida. *Mol Biol Evol* **32**, 2832-2843, doi:10.1093/molbev/msv155 (2015).

41 Hayden, S. *et al.* A cluster of olfactory receptor genes linked to frugivory in bats. *Mol Biol Evol* **31**, 917-927, doi:10.1093/molbev/msu043 (2014).

42 Daltry, J. C., Wuster, W. & Thorpe, R. S. Diet and snake venom evolution. *Nature* **379**, 537-540, doi:10.1038/379537a0 (1996).

43 Hargreaves, A. D., Swain, M. T., Hegarty, M. J., Logan, D. W. & Mulley, J. F. Restriction and recruitment-gene duplication and the origin and evolution of snake venom toxins. *Genome Biol Evol* **6**, 2088-2095, doi:10.1093/gbe/evu166 (2014).

44 Casewell, N. R., Harrison, R. A., Wuster, W. & Wagstaff, S. C. Comparative venom gland transcriptome surveys of the saw-scaled vipers (Viperidae: Echis) reveal substantial intra-family gene diversity and novel venom transcripts. *BMC Genomics* **10**, 564, doi:10.1186/1471-2164-10-564 (2009).

45 Matsubara, K. *et al.* Evidence for different origin of sex chromosomes in snakes, birds, and mammals and step-wise differentiation of snake sex chromosomes. *Proc Natl Acad Sci U S A* **103**, 18190-18195, doi:10.1073/pnas.0605274103 (2006).

46 Ohno, S. *Sex chromosomes and sex-linked genes*. Vol. 1 (Springer, 1967).

47 Zhou, Q. *et al.* Complex evolutionary trajectories of sex chromosomes across bird taxa. *Science* **346**, 1246338, doi:10.1126/science.1246338 (2014).

48 Cortez, D. *et al.* Origins and functional evolution of Y chromosomes across mammals. *Nature* **508**, 488-493, doi:10.1038/nature13151 (2014).

49 Nicolas, M. *et al.* A gradual process of recombination restriction in the evolutionary history of the sex chromosomes in dioecious plants. *PLoS Biol* **3**, e4, doi:10.1371/journal.pbio.0030004 (2005).

50 Li, W. H., Yi, S. & Makova, K. Male-driven evolution. *Curr Opin Genet Dev* **12**, 650-656 (2002).

51 Vicoso, B. & Charlesworth, B. Evolution on the X chromosome: unusual patterns and processes. *Nat Rev Genet* **7**, 645-653, doi:10.1038/nrg1914 (2006).

52 Mank, J. E. The W, X, Y and Z of sex-chromosome dosage compensation. *Trends Genet* **25**, 226-233, doi:10.1016/j.tig.2009.03.005 (2009).

53 Boucherat, O. *et al.* Partial functional redundancy between Hoxa5 and Hoxb5 paralog genes during lung morphogenesis. *Am J Physiol Lung Cell Mol Physiol*

**304**, L817-L830 (2013).

54 Guerreiro, I. & Duboule, D. Snakes: hatching of a model system for Evo-Devo? *Int J Dev Biol* **58**, 727-732, doi:10.1387/ijdb.150026dd (2014).

55 Booth, W., Johnson, D. H., Moore, S., Schal, C. & Vargo, E. L. Evidence for viable, non-clonal but fatherless Boa constrictors. *Biol Lett* **7**, 253-256, doi:10.1098/rsbl.2010.0793 (2011).

56 Li, R. *et al.* The sequence and de novo assembly of the giant panda genome. *Nature* **463**, 311-317, doi:10.1038/nature08696 (2010).

57 Luo, R. *et al.* SOAPdenovo2: an empirically improved memory-efficient short-read de novo assembler. *Gigascience* **1**, 18, doi:10.1186/2047-217X-1-18 (2012).

58 Li, H. & Durbin, R. Fast and accurate short read alignment with Burrows-Wheeler transform. *Bioinformatics* **25**, 1754-1760, doi:10.1093/bioinformatics/btp324 (2009).

59 Jurka, J. Repbase update: a database and an electronic journal of repetitive elements. *Trends Genet* **16**, 418-420 (2000).

60 Altschul, S. F., Gish, W., Miller, W., Myers, E. W. & Lipman, D. J. Basic local alignment search tool. *J Mol Biol* **215**, 403-410, doi:10.1016/S0022-2836(05)80360-2 (1990).

61 Birney, E., Clamp, M. & Durbin, R. GeneWise and Genomewise. *Genome Res* **14**, 988-995, doi:10.1101/gr.1865504 (2004).

62 Stanke, M., Steinkamp, R., Waack, S. & Morgenstern, B. AUGUSTUS: a web server for gene finding in eukaryotes. *Nucleic Acids Res* **32**, W309-312, doi:10.1093/nar/gkh379 (2004).

63 Trapnell, C., Pachter, L. & Salzberg, S. L. TopHat: discovering splice junctions with RNA-Seq. *Bioinformatics* **25**, 1105-1111, doi:10.1093/bioinformatics/btp120 (2009).

64 Robinson, M. D. & Oshlack, A. A scaling normalization method for differential expression analysis of RNA-seq data. *Genome Biol* **11**, R25, doi:10.1186/gb-2010-11-3-r25 (2010).

65 Yang, Z. PAML 4: phylogenetic analysis by maximum likelihood. *Mol Biol Evol* **24**, 1586-1591, doi:10.1093/molbev/msm088 (2007).

66 Benton, M. J. & Donoghue, P. C. Paleontological evidence to date the tree of life. *Mol Biol Evol* **24**, 26-53, doi:10.1093/molbev/msl150 (2007).

67 Li, H. *et al.* TreeFam: a curated database of phylogenetic trees of animal gene families. *Nucleic Acids Res* **34**, D572-580, doi:10.1093/nar/gkj118 (2006).

68 De Bie, T., Cristianini, N., Demuth, J. P. & Hahn, M. W. CAFE: a computational tool for the study of gene family evolution. *Bioinformatics* **22**, 1269-1271, doi:10.1093/bioinformatics/btl097 (2006).

69 Loytynoja, A. & Goldman, N. An algorithm for progressive multiple alignment of sequences with insertions. *Proc Natl Acad Sci U S A* **102**, 10557-10562, doi:10.1073/pnas.0409137102 (2005).

70 Talavera, G. & Castresana, J. Improvement of phylogenies after removing divergent and ambiguously aligned blocks from protein sequence alignments. *Syst Biol* **56**, 564-577, doi:10.1080/10635150701472164 (2007).

71 Bult, C. J. *et al.* Mouse genome database 2016. *Nucleic Acids Res* **44**, D840-847, doi:10.1093/nar/gkv1211 (2016).

72 Weng, M. P. & Liao, B. Y. MamPhEA: a web tool for mammalian phenotype enrichment analysis. *Bioinformatics* **26**, 2212-2213, doi:10.1093/bioinformatics/btq359 (2010).

73 Steiger, S. S., Kuryshev, V. Y., Stensmyr, M. C., Kempenaers, B. & Mueller, J. C. A comparison of reptilian and avian olfactory receptor gene repertoires: species-specific expansion of group gamma genes in birds. *BMC Genomics* **10**, 446, doi:10.1186/1471-2164-10-446 (2009).

74 Mackessy, S. P. *Handbook of venoms and toxins of reptiles*. (CRC Press, 2016).

75 Krogh, A., Larsson, B., von Heijne, G. & Sonnhammer, E. L. Predicting transmembrane protein topology with a hidden Markov model: application to complete genomes. *J Mol Biol* **305**, 567-580, doi:10.1006/jmbi.2000.4315 (2001).

76 Li, H. *et al.* The Sequence Alignment/Map format and SAMtools. *Bioinformatics* **25**, 2078-2079, doi:10.1093/bioinformatics/btp352 (2009).

77 Edgar, R. C. MUSCLE: multiple sequence alignment with high accuracy and high throughput. *Nucleic Acids Res* **32**, 1792-1797, doi:10.1093/nar/gkh340 (2004).

78 Stamatakis, A. RAxML version 8: a tool for phylogenetic analysis and post-analysis of large phylogenies. *Bioinformatics* **30**, 1312-1313, doi:10.1093/bioinformatics/btu033 (2014).

835 **Figure Legend**

836 **Figure 1.** The comparative genomic landscape of five-pacer viper

837 (**A**) *Deinagkistrodon acutus* (five-pacer viper) and eight adult tissues used in this study.
838 (**B**) Circos plot showing the linkage group assignment using lizard chromosomes as
839 reference (outmost circle), normalized female vs. male mapped read coverage ratio (blue
840 line) and GC-isochore structures of five-pacer viper (red), boa (yellow) and green anole
841 lizard (green). Both snake genomes have a much higher variation of local GC content
842 than that of green anole lizard. (**C**) Phylogenomic tree constructed using fourfold
843 degenerate sites from 8006 single-copy orthologous genes. We also showed bootstrapping
844 percentages, the numbers of inferred gene family expansion (in green) and contraction
845 (red), and corresponding phylogenetic terms at each node. MRCA: most recent common
846 ancestor.

847

848 **Figure 2.** Genomic and transcriptomic variation of snake transposable elements

849 **(A)** Violin plots showing each type of TE's frequency distribution of sequence divergence
850 level from the inferred ancestral consensus sequences. Clustering of TEs with similar
851 divergence levels, manifested as the 'bout' of the violin, corresponds to the burst of TE
852 amplification. **(B)** Bar plots comparing the genome-wide TE content between four snake
853 species. TE families were annotated combining information of sequence homology and
854 *de novo* prediction. **(C)** TE's average normalized expression level (measured by RPKM)
855 across different tissues in five-pace viper.

856

857 **Figure 3.** Evolution of snake genes and gene families

858 (**A**) Phylogenetic distribution of mutant phenotypes (MP) of mouse orthologs of snakes.
859 Each MP term is shown by an organ icon, and significantly enriched for snake genes
860 undergoing positive selection (red) or relaxed selective constraints (gray) inferred by
861 lineage-specific PAML analyses. (**B**) We show the four *Hox* gene clusters of snakes,
862 with each box showing the ratio of nonsynonymous (dN) over synonymous substitution
863 (dS) rate at the snake ancestor lineage. White boxes represent genes that haven't been
864 calculated for their ratios due to the genome assembly issue in species other than

five-pacer viper. Boxes with dotted line refer to genes with dS approaching 0, therefore the dN/dS ratio cannot be directly shown. Each cluster contains up to 13 *Hox* genes with some of them lost during evolution. We also marked certain *Hox* genes undergoing positive selection (in red) or relaxed selective constraints (in green) at a specific lineage above the box. Each lineage was denoted as: S: *Serpentes* (ancestor of all snakes), H: *Henophidia* (ancestor of boa and python), B: *Boa constrictor*, P: *Python bivittatus*, C: *Colubroidea*, D: *Deinagkistrodon acutus*, O: *Ophiophagus hannah*. (**C**) Comparing olfactory receptor (OR) gene repertoire between boa, viper and lizard. Each cell corresponds to a certain OR family (shown at y-axis) gene number on a certain chromosome (x-axis). (**D**) Pie chart shows the composition of normalized venom gland transcripts of male five-pacer viper. The heatmap shows the normalized expression level (in RPKM) across different tissues of viper and king cobra. We grouped the venom genes by their time of origination, shown at the bottom x-axis.

**Figure 4.** Snake sex chromosomes have at least three evolution strata

The three tracks in the top panel shows female read depths along the Z chromosome relative to the median depth value of autosomes, Z/W pairwise sequence divergence within intergenic regions, and female read depths of W-linked sequence fragments relative to the median depth value of autosomes. Depths close to 1 suggest that the region is a recombining pseudoautosomal region (PAR), whereas depths of 0.5 are expected in a highly differentiated fully sex-linked region where females are hemizygous. The identifiable W-linked fragments are much denser at the region 56Mb~70Mb, probably because this region (denoted as stratum 3, S3) has suppressed recombination most recently. S2 and S1 were identified and demarcated by characterizing the sequence conservation level (measured by LASTZ alignment score, blue line) between the chrZs of boa and viper. At the oldest stratum S1 where recombination has been suppressed for the longest time, there is an enrichment of repetitive elements on the affected Z-linked region (Gypsy track in red, 100kb non-overlapping sliding window). And these Z-linked TEs A similar pattern was found in homologous recombining region of boa, but not in lizard.

**Supplementary Figure Legend**

**Supplementary Figure 1.** K-mer estimation of the genome size of five-pacer viper. Distribution of 17-mer frequency in the used sequencing reads from female (left) and male (right) samples. The x-axis represents the sequencing depth. The Y-axis represents the proportion of a *K*-mer counts in total *K*-mer counts at a given sequencing depth. The estimated genome size is about 1.43 Gb.

**Supplementary Figure 2.** GC isochore structure of different tetrapod genomes. We show the standard deviation (SD) of GC content calculated with different window size (3 kb to 320 kb) for different vertebrate genomes.

**Supplementary Figure 3.** Comparing expression levels of genes nearby expressed TEs in each tissue.
We show expression patterns of genes around highly expressed TE (RPKM > 5) in different tissues from both sexes, including brain, liver, venom gland and gonad. We also performed comparison between the focal tissue vs. the other tissues. We show levels of significance with asterisks. *: 0.001=< *P*-value <0.01; **: 0.0001 =< *P*<0.001; ***: *P*<0.0001.

**Supplementary Figure 4.** Comparison of *Hox* gene structure between snakes and lizard. Schematic representation of four *Hox* clusters in anole lizard, boa, Burmese python, five-pacer viper and king cobra. Each number from 1 to 13 denotes the specific *Hox* gene belonging to one cluster. We showed the length difference between each species vs. mouse by the colored lines (for intergenic regions) or boxes (for intronic regions): a 1.5~3 fold increase of length was shown by blue, a more than 3-fold increase was shown in red. Exons were shown by vertical lines, and dotted lines refer to exons with unknown boundaries, either due to assembly issues. Double-slashes refer to the gap between two different scaffolds.

**Supplementary Figure 5.** Repeat accumulation at *Hox* gene clusters.

925 Comparison of the TE and simple repeat content of *Hox* cluster genes with 5kb flanking
926 regions between snakes and lizard. We calculated the repeat density by dividing the total
927 length of specific repeat sequence vs. the length of corresponding region. This density was
928 normalized over the genome-wide repeat density and then shown by heatmap.
929
930 **Supplementary Figure 6.** Phylogenetic distribution of enriched MP terms.
931 We identified enriched mutant phenotypes (MP) of mouse orthologs of snake genes that
932 are undergoing lineage-specific positive selection (red) and relaxed selective constraints
933 (gray). And then we mapped these MP terms onto the snake phylogeny.
934
935 **Supplementary Figure 7.** Read coverage density plot of different linkage groups.
936 For each linkage group (from left to right, chrW, chrZ, chr1), male reads were plotted in
937 blue, and female reads in red. The identified chrW scaffolds in this work all show a
938 female-specific read depth pattern.
939
940 **Supplementary Figure 8.** Male-driven evolution effect in snakes.
941 For each gene, we calculated the substitution rates between anole lizard and each of boa,
942 Burmese python, five-pacer viper and king cobra at synonymous sites (dS) and
943 non-synonymous sites (dN) divided into different chromosome sets. To detect
944 branch-specific differences, we obtained for each gene the ratios of these evolutionary
945 rates between the different snake species vs. boa. Since boa's homologous chromosomal
946 region to the Z chromosomes of other advanced snakes represent the ancestral status of
947 snake sex chromosomes, a higher relative ratios of Z-linked dS of advanced snakes vs.
948 boa than those of autosomes indicate the male-driven evolution effect. We shown the
949 Wilcoxon test significant differences between the Z chromosome and the autosomes
950 (chr1-5) are marked with asterisks (***, *P*-value < 0.001).
951
952 **Supplementary Figure 9.** Repeat accumulation along the snake sex chromosomes.
953 Shown are comparisons of the distribution of Gypsy, CR1 and L1 content along the
954 chromosome in anole lizard, Boa and five-pacer viper. The TE content was calculated by

averaging the TE density of each sliding window of 100kb as well as the flanking 10 windows.

**Supplementary Figure 10.** Gene trees for Z and W linked gametologs in S1 region. Shown are maximum likelihood (ML) trees using coding regions of Z and W allelic sequences from multiple snake species, with the gene name under each tree and bootstrapping values at each node. Trees that show separate clustering of Z- or W- linked gametologs provide strong evidence that these genes suppressed recombination before the speciation.

**Supplementary Figure 11.** Gene trees for Z and W linked gametologs in S2 region.

**Supplementary Figure 12.** Gene trees for Z and W linked gametologs in S3 region.

**Supplementary Figure 13.** Comparison of repeat content between viper chrZ and chrW. TE families were determined based on the combined annotations of Repbase, RepeatModeler, and coverage in the genome was annotated using in-house scripts.

**Supplementary Figure 14.** Comparison of gene expression across different chromosomes.
We show gene expression patterns between different chromosome sets across tissues. 'pseudo-W' refers to W-linked genes that have premature stop codons or frameshift mutations

**Supplementary Figure 15.** Pairwise comparison of gene expression levels between homologous Z and W alleles.

**Supplementary Figure 16.** Gene expression along the Z chromosome and autosome chr5 in different tissues of five-pacer viper.
We show log-based male-to-female gene expression ratio along the Z chromosomes and

autosome chr5. Only genes with RPKM >= 1 in both the male and female were considered. If genes are mostly non-biased, the line is expected to centered at 0. The pattern indicates five-pacer viper lacks chromosome-wide dosage compensation.

**Supplementary Figure 17.** Frequency distribution of sequencing depth.
Distribution of sequencing depth of the assembled female (left) and male (right) genomes by reads from the female and male samples. The peak depth is 76X and 77X for the female and male reads aligned to corresponding assembly, respectively.

**Supplementary Figure 18.** Comparisons of gene parameters among the sequenced representative species.
We used the published genomes of *Gallus gallus, Homo sapiens, Anolis carolinensis, Boa constrictor* to compare with *Deinagkistrodon acutus*, without finding any obvious differences between them and five-pacer viper in the annotated genes' length and number. This indicates the high quality of gene annotation.

**Supplementary Table Legend**

**Supplementary Table 1.** Statistics of five-pacer viper genome sequencing.

**Supplementary Table 2.** Statistics of 17-mer analysis.

**Supplementary Table 3.** Statistics of reads of small-insert and large-insert libraries aligned to the male assembly.
*PE mapped* refer to reads being mapped to the genome as read pairs, and *SE mapped* represent reads being mapped to the genome as single reads.

**Supplementary Table 4.** Summary of the five-pacer viper genome assemblies.

**Supplementary Table 5.** Number of expressed genes of five-pacer viper.

**Supplementary Table 6.** Number of genes and scaffold size organized into chromosomes.

**Supplementary Table 7.** Comparison of repeat content between snakes and lizard.

**Supplementary Table 8.** Comparison of genome assembly quality between snakes and lizard.

**Supplementary Table 9.** Evolution of candidate limb-patterning genes in snakes. Candidate limb-patterning genes were collected from MGI database and published paper. Genes undergoing positive selection (P) or relaxed selective constraints (R), were identified by PAML analyses. N: there is no significant selection signal. '-' refers to genes that cannot be completely assembled for their coding sequences due to genome assembly gaps.

**Supplementary Table 10. Comparison of venom genes between snakes.** Statistics of venom gene families in the four snakes and anole lizard genomes based on homology-based prediction. AVIT: Prokineticin; C3: complement C3; CVF: Cobra Venom Factor; CRISPs: Cysteine-Rich Secretory Proteins; Hy: Hyaluronidases; Natriuretic: Natriuretic peptide; NGF: Snake Venom Nerve Growth Factors; PLA2-2A: Snake Venom Phospholipase A2 (type IIA); SVMP: Snake Venom Metalloproteinases; TL: thrombin-like snake venom serine proteinases; LAAO: Snake Venom L-Amino Acid Oxidases; PDE: phosphodiesterases; CLPs: snake C-type lectin-like proteins; VEGF: vascular endothelin growth factor; PLA2-1B: Snake Venom Phospholipase A2 (type IB); 3FTX: The three-finger toxins; ACeH: Acetylcholinesterase;

**Supplementary Table 11.** Comparison of repeat content between snake sex chromosomes and their lizard homolog.

**Supplementary Table 12.** Location of W-linked putative pseudogenes.

**Supplementary Table 13.** Fractions of bases covered by reads in the male assembly.

**Supplementary Table 14.** Characteristics of predicted protein-coding genes in the male

assembly.

**Supplementary Table 15.** Number of predicted genes that can find homologs in the Ensembl library with different aligning rate cutoff.
Alignment rate was calculated by dividing the aligned length vs. the original protein length. And we required both the query and subject to satisfy our alignment cutoff. The Ensembl library consists of all proteins from Anolis carolinensis, Gallus gallus, Homo sapiens, Xenopus tropicalis and Danio rerio.

**Supplementary Table 16.** Data production and alignment statistic of RNA-Seq aligned to male genome assembly.

**Supplementary Table 17.** Statistics of the identified Z-linked scaffolds.

**Supplementary Table 18.** Statistics of SNPs identified in the female and male individual.

**Supplementary Table 19.** Statistics of identified W-linked scaffolds**.**

**Supplementary Data Legend**

**Supplementary Data 1**

Comparing five-pacer viper's chromosomal assignment vs. reported fluorescence in situ hybridization results.

**Supplementary Data 2**

GO enrichment of nearby genes of TE highly expressed in brain.

**Supplementary Data 3**

Positively selected genes (PSGs) and genes with relaxed selective constraints (RSGs) and their affected lineage.

**Supplementary Data 4**

Mouse mutant terms' enrichment analyses of PSGs and RSGs across different snake branches.

**Supplementary Data 5**

GO enrichment of expanded and contracted gene families across different snake lineages.

**Supplementary Data 6**

Plots of GO enrichment of expanded and contracted gene families using Ontologizer.

**Acknowledgement**

We would like to thank Ren-jie Wang for sharing the snake photo, and three anonymous reviewers for their valuable comments. This project was supported by the Thousand Young Talents Program funding, and a startup funding of Life Science Institute of Zhejiang University to Q. Z., and National Major Scientific and Technological Special Project for 'Significant New Drugs Development' during the Twelfth Five-year Plan Period (No. 2009ZX09102-217, 2009-2010) to W. Y. All the raw reads, genome assembly and annotation generated in this study are deposited under the BioProject # PRJNA314370 on NCBI.

**Author Contribution**

Q.Z., W.Y., G.Y. conceived and supervised the project. B.L., P.Q., W.Z., Y.H. and X.S. provided and extracted the samples. Z.L. and B.W. performed the proteomic mass spectrometry-based experiment and data analysis. J.L., Z.W. and Y.Z. performed the genome assembly and annotation. J.L. and Q.L. designed and performed the identification of sex-linked scaffolds. P.Z. performed RNA-seq data analysis. Z.W., L.J. and Y.Z. performed the genome evolution analyses. Z.W., Y.Z., L.J. and B.Q. performed genes and gene family evolution analyses. Z.W. and L.J. performed the sex chromosome evolution analyses. Q.Z., Z.W., G.Z., G.Y. interpreted the results and wrote the manuscript. All of the authors read and approved the final manuscript.

**Figure 1.** The comparative genomic landscape of five-pace viper

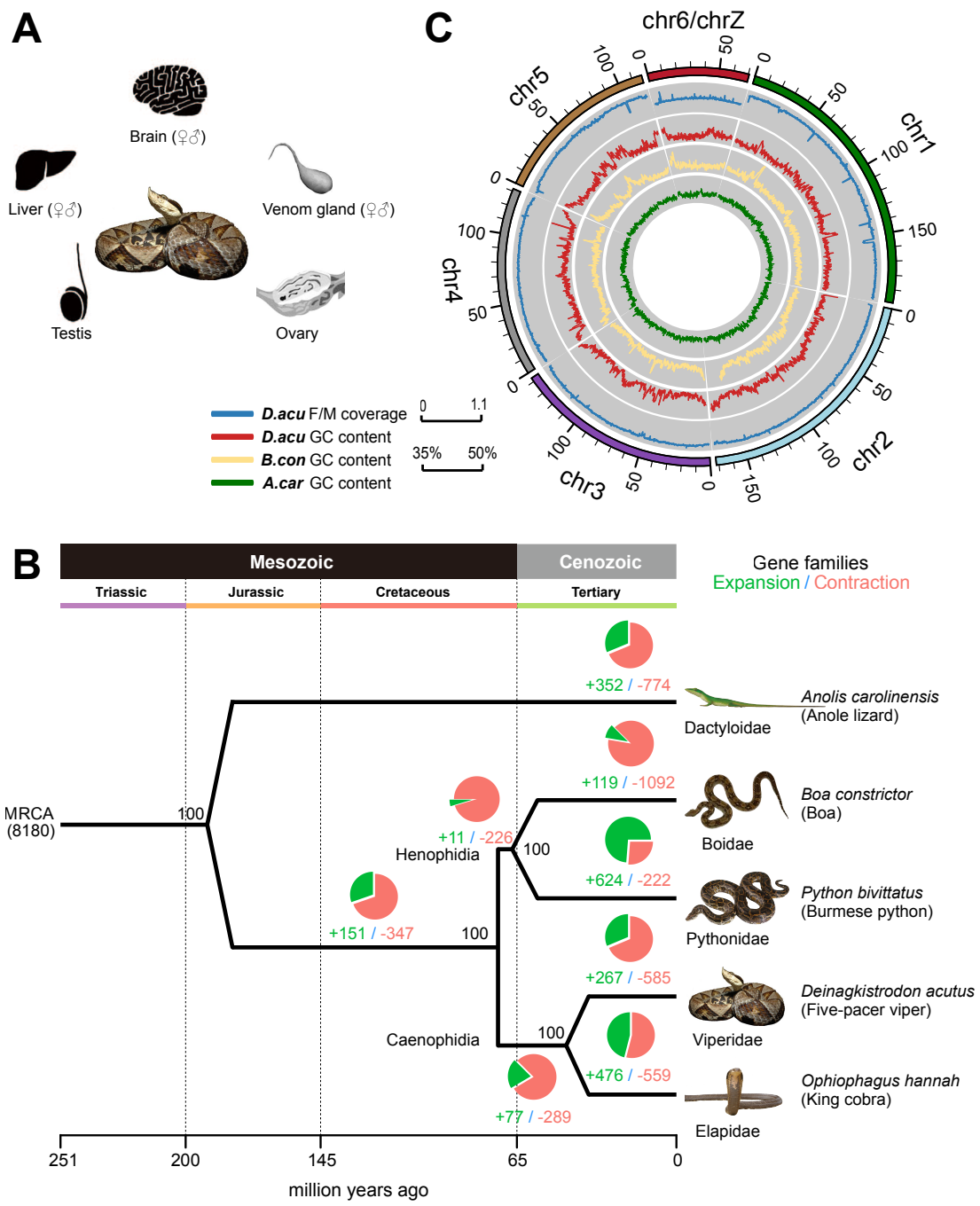
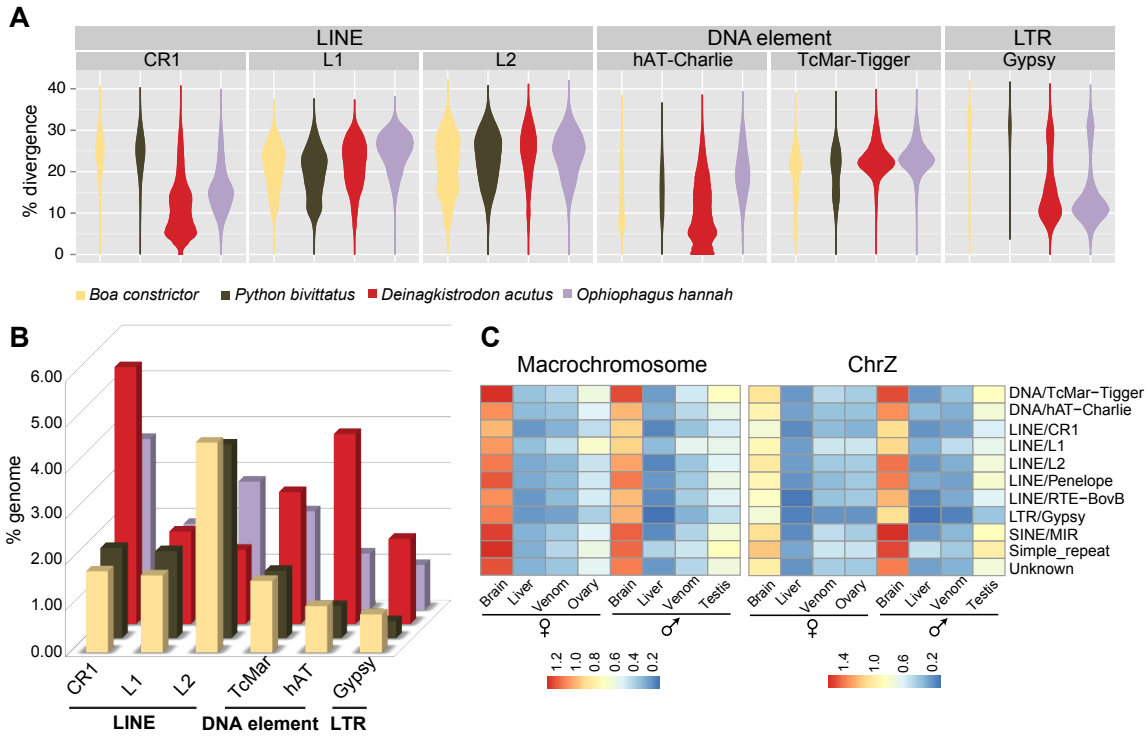
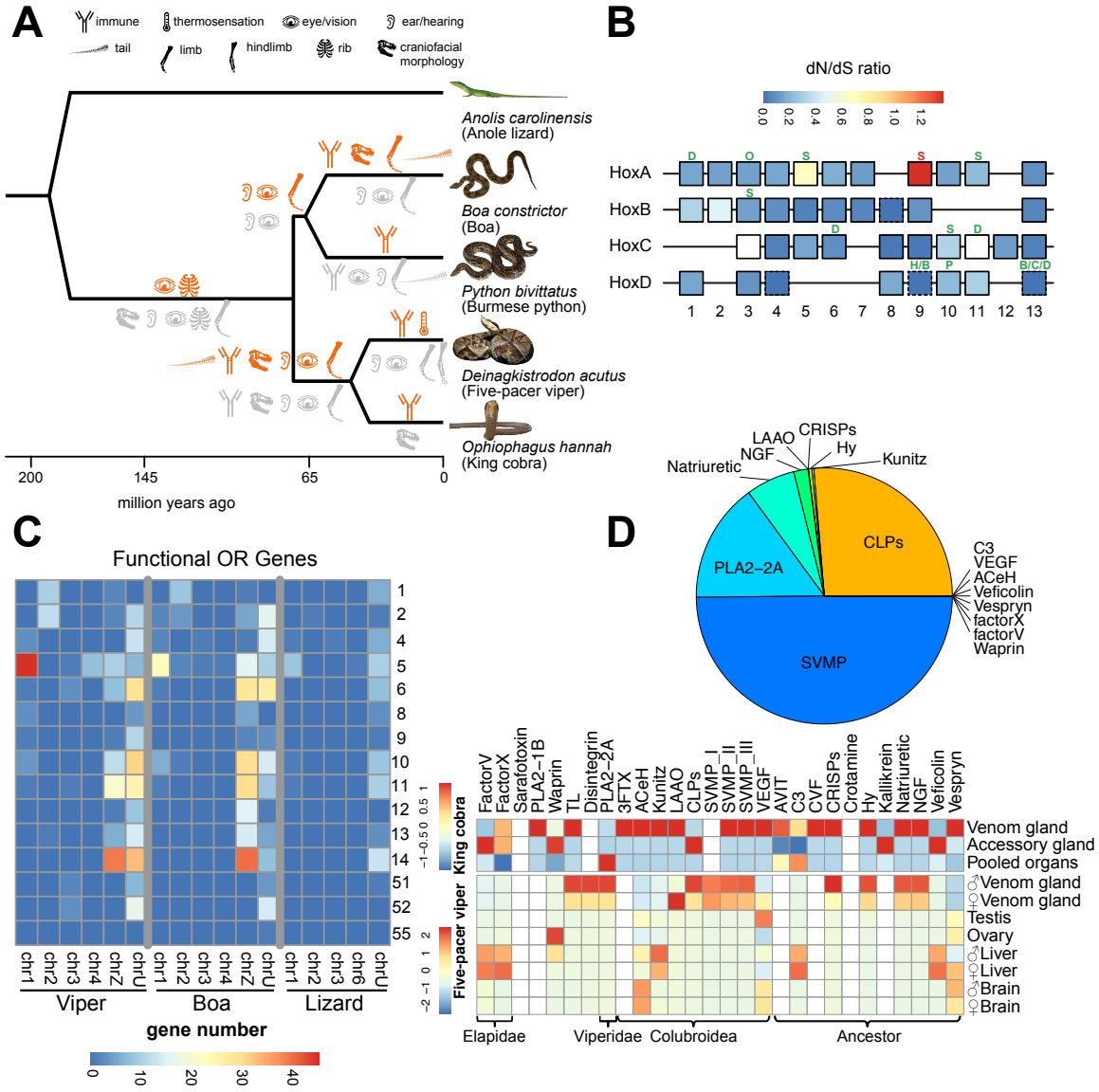


**Figure 2.** Genomic and transcriptomic variation of snake transposable elements



1117 **Figure 3.** Evolution of snake genes and gene families




1118

**Figure 4.** Snake sex chromosomes have at least three evolution strata