BMC Genomics

RESEARCH

Open Access

Genomic insights into the mechanisms of body size evolution in Serpentes

Tian Xia[1†], Shengyang Zhou[1†], Zhihao Zhang[1†], Xiaoyang Wu[1], Xibao Wang[1], Jianqun Ding[1], Lei Zhang[1], Guolei Sun[1], Xiufeng Yang[1], Xiaodong Gao[1] and Honghai Zhang[1*]

Abstract

Background Body size is a critical trait that influences an animal's physiology, behavior, and ecology. However, the molecular mechanisms underlying its evolution remain poorly understood, particularly in snakes. Snakes exhibit an extremely wide range of body sizes and strong ecological adaptability. Among snake species, the maximum body mass exceeds the minimum by over 200,000-fold, while the maximum body length surpasses the minimum by more than 110-fold.

Results Through phylogenomic and comparative genomic analyses of 26 snake genomes, we identified 77 body size-associated genes (BSAGs) related to body length or body mass, highlighting key genetic drivers of body size evolution. Functional enrichment analyses revealed that metabolic pathways, particularly fatty acid metabolism and oxidoreductase activity, underwent significant expansion and positive selection, suggesting metabolic adaptations crucial for meeting the energetic demands of increased body size. Immune system-related genes, including those involved in antigen processing and presentation, similarly showed signatures of expansion and adaptive evolution, highlighting strengthened immune defenses in large-bodied snakes. Additionally, key candidate genes, such as *YAP1*, *PLAG1*, *MGAT1* and *SPRY1*, exhibited both strong selection signals and correlation signals, and are functionally involved in developmental pathways critical for growth regulation.

Conclusions Our findings reveal a complex interplay of sensory, immune, metabolic, and growth-related genetic adaptations driving large body size evolution in snakes. This study provides novel insights into the molecular underpinnings of snake body size diversification and advances our understanding of their evolutionary history.

Keywords Serpentes, body size evolution, coparative genomics, positive selection

†Tian Xia, Shengyang Zhou and Zhihao Zhang contributed equally to this work.

*Correspondence:
Honghai Zhang
zhanghonghai67@126.com
[1]College of Life Sciences, Qufu Normal University, Jingxuan West Street No. 57, Qufu 273165, China

Background

Body size, a key organismal trait, profound influences ecological adaptation through trade-offs between advantages and constraints across environments [1]. This trait correlates with critical evolutionary parameters including species distribution [2], habitat selection [3], reproductive maturity [4], and extinction risk [5]. It plays a pivotal role in shaping individual behavior, vulnerability to extinction, and evolutionary trajectories, including molecular evolutionary rates [6]. Ecological consequences of body size are exemplified by reduced dispersal capacity in smaller species, which may limit their geographic ranges [2], and size-associated biological traits affecting environmental threat resilience [7]. The frequency distributions of animal body size is typically right-skewed, indicating that the majority of species are relatively small [8]. For instance, 99% of the world's major lizard groups, such as the family Lacertidae, predominantly consist of smaller species [9]. Body size, as a complex quantitative trait, exhibits continuous variation and plays a vital role in shaping an organism's ecology and evolution [10].

As of January 2025, a total of 4,177 snake species have been recorded globally (http://www.reptile-database.org/), occupying a wide range of environments, including terrestrial, arboreal, fossorial, and aquatic habitats [11]. Despite this ecological diversity, the global distribution of body mass among living squamates (lizards and snakes) appears to show limited correlation with climatic factors. Modern snakes exhibit an exceptionally broad spectrum of body size, from the largest species, *Eunectes murinus* (10,000 mm, 34,510 g), to the smallest, *Indotyphlops veddae* (91 mm, 0.17 g) [1]. Their simplified body plan, characterized by the absence of large appendages, makes snakes an ideal model for studying body size variation and its evolutionary drivers [12]. A previous study has indicated that garter snakes experience higher mortality from predators when they have smaller body sizes [13], while other research has explored the relationship between body size evolution and reproductive success in snakes [14]. Meiri generated a comprehensive dataset on multiple key traits for all 11,744 recognised species of squamates worldwide [15], providing an unparalleled resource for studying size variation in both snakes and lizards.

With the rapid advancements in genome sequencing technology, researchers have increasingly sought to uncover the molecular mechanisms underlying major phenotypic traits. Recent improvements in sequencing technology have provided unprecedented opportunities to elucidate the genetic basis of body size variation. Studies have identified genes and signaling pathways associated with body size differences across various taxa, including domestic dogs [16], carnivores [17], and cetaceans [18]. In addition, genome-wide association studies (GWAS) in mice have revealed thousands of loci variants linked to body size development [10]. However, despite the remarkable diversity in snake body sizes, the molecular mechanisms underlying this variation remain poorly understood.

In this study, we performed a comparative genomic analysis using 26 high-quality snake genomes. First, we applied phylogenetic generalized least squares (PGLS) methods to identify body-size-associated genes (BSAGs). Next, we characterized gene family expansions, positively selected genes (PSGs), and rapidly evolving genes (REGs) in three large-bodied snake lineages. The findings from this study are expected to provide novel insights into the molecular mechanisms driving body size evolution in snakes (Serpentes).

Methods

Genome and coding protein collection

The whole genome sequences of Serpentes species were retrieved from the NCBI database. Most published genomes consist only of raw assembly sequences without accompanying annotation information. To address this limitation, we utilized a high-quality annotated reference genome, *Candoia aspera* (GCF_035149785.1), to perform genome alignments using LAST (v.956) [19] and MULTIZ (v.10.6) [20]. Through a series of alignment and filtering steps, we extracted the protein sequences for all Serpentes species based on the annotation file of the reference genome. Genome completeness was assessed using BUSCO (Benchmarking Universal Single-Copy Orthologs, v.5.2.2) with the "vertebrata_odb10" library and default parameters. The analysis utilized the following versions of third-party components: Python (v.3.8.3), Augustus (v.3.3.2) [21], and HMMER (v.3.2.1) [22].

Phenotype data collection

Phenotypic data, including the maximum body length and body mass of representative Serpentes species in our dataset, were obtained from SquamBase [15] and used for subsequent analyses. All phenotype data were derived from adult individuals. For our study, we selected 26 high-quality snake genomes (BUSCO) spanning eight Afrophidia families—Viperidae, Elapidae, Boidae, Colubridae, Dipsadidae, Pythonidae, Natricinae, and Lamprophiidae—to capture a broad range of body size diversity. The selected species exhibit body weights ranging from 75.9 g to 23,442.2 g and body lengths ranging from 660 mm to 5,740 mm (Table S1). We selected species from our dataset with both log length and log mass values greater than 3.5 as large-bodied snakes (*Liasis olivaceus*, *Ophiophagus hannah* and *Python bivittatus*), which will serve as the target clade for subsequent analyses.

Phylogenetic tree construction and gene family identification

High-confidence "one-to-one" orthologous gene clusters across all snake species were identified using the OrthoFinder (v.2.4.0) pipeline [23], which employs the all-against-all DIAMOND algorithm for orthology inference. Phylogenetic relationships were reconstructed using RAxML (v.8.2.12) [24] with the parameters "GTRGAMMA -f a -x 12345 -N 100 -p 12345," generating the best maximum-likelihood topology based on 1,000 bootstrap replicates. A Computational Analysis of gene Family Evolution (CAFÉ v.5) was performed with default parameters and the orthogroup results from OrthoFinder to identify gene families expansion and contraction. The phylogeny used in CAFE and subsequent analyses was dated using Timetree (https://timetree.org/) [25]. To compare gene family expansions and contractions relative to the most recent common ancestor (MRCA) of the selected snakes, we summarized the changes in gene and gene family numbers for each species and calculated the number of expanded to contracted gene families. Gene families with an exact p value ≤ 0.05 at any node were classified as "significantly expanded" or "significantly contracted." We collected the orthogroups (OGs) inferred from *L. olivaceus*, *O. hannah* and *P. bivittatus* and used to count shared OGs among these three large-bodied snakes.

PGLS scanning the body-size-associated genes (BSAGs)

The PGLS method, implemented in the "caper" package in R [26], was employed to examine the relationship between evolutionary rates of individual genes and phenotypic traits, specifically body length and body mass. The ultrametric tree for 26 snake species used in the PGLS analysis was derived from a species tree constructed using orthologous supergenes. Evolutionary rates (ω) were estimated using the free-ratios model (model = 1) implemented in the codeml program of PAML (v.4.10.6) [27], and the root-to-tip ω for each species was calculated by averaging the ω values along the branches from the ancestral Serpentes node to the terminal branch, following the approach described by Montgomery et al. [28]. To ensure normality in the regression analysis, all root-to-tip ω values and phenotypic trait data were log10-transformed. A Brownian motion model was applied, and the phylogenetic signal (λ) was estimated using the maximum likelihood (ML) method. The λ parameter provides a quantitative measure of phylogenetic signal, with values close to 1 indicating a strong phylogenetic effect. Genes that were significantly associated with either body length or body mass ($p < 0.05$) were classified as BSAGs.

Selective pressure analysis

All one-to-one orthologues identified from the Orthofider results were analyzed to detect positive selection and rapid evolution in target branches using the codeml program in PAML. Positive selection was assessed by calculating the ratio (ω) of nonsynonymous substitution rate (dN) to synonymous substitution rate (dS), with $\omega > 1$ indicating positive selection, $\omega < 1$ indicating purifying (negative) selection, and $\omega = 1$ suggesting neutral evolution. Three large-bodied snakes were designated as foreground branches, with the remaining species serving as background branches. Branch-site models (model A vs. null model A) were used to identify PSGs. To test for rapid evolution, one-ratio and two-ratio models under the branch model framework were applied. We defined genes as REGs if their ω of the foreground was higher than that of the background branches. Statistical significance was evaluated using a chi-square test with a threshold of p value < 0.05.

Gene ontology and KEGG enrichment analyses

Functional similarity of BSAGs, PSGs and REGs were quantitatively assessed using the semantic similarity of Gene Ontology (GO) terms and Kyoto Encyclopedia of Genes and Genomes (KEGG) pathways. To obtain GO annotations, protein sequences were analyzed for functional domains using the InterPro database through InterProScan (v.5.16-93) [29]. Pathway annotation of the protein sequences was subsequently performed using the KEGG database. GO enrichment and KEGG enrichment analyses were conducted to identify biological processes associated with gene family expansion or contraction, as well as positive selection.

Results

Genome-scanning of BSAGs

After evaluating genome integrity and compiling body size data, we selected 26 high-quality genomes (BUSCO > 90%) representing a wide range of body size diversity within Serpentes (Table 1, Table S1). A total of 20,278 orthogroups were identified from ten snake genomes using OrthoFinder, with 13,875 orthogroups present in all species, of which 12,494 consisted entirely of single-copy genes (Table S2). A PGLS analysis identified 66 genes significantly associated with body length and 43 genes significantly associated with body mass. Among these, 32 genes were shared between the two groups. After removing the shared duplicate genes, a total of 77 unique genes were identified, which we define as BSAGs (Table S3). Functional enrichment analyses revealed that BSAGs were significantly enriched in GO terms annotated with metabolic processes, such as "fatty acid metabolic process (GO: 0006631), oxidoreductase activity (GO:0016627) and carbohydrate

Table 1 List of genomic information for snakes

Family	Scientific name	Genome Version	Coverage	Assemble Level	BUSCO (%)
Boidae	*Candoia aspera*	GCF_035149785.1	37.0x	Chromosome	96.5
Colubridae	*Ahaetulla prasina*	GCF_028640845.1	140.0x	Chromosome	90
Colubridae	*Arizona elegans*	GCA_022577455.1	31.0x	Scaffold	93.5
Colubridae	*Pantherophis alleghaniensis*	GCA_030052795.1	69.0x	Scaffold	91.9
Colubridae	*Pantherophis obsoletus*	GCA_012654085.1	92.0x	Scaffold	91.6
Colubridae	*Pantherophis spiloides*	GCA_037575465.1	35.0x	Contig	93.5
Colubridae	*Pantherophis vulpinus*	GCA_037215915.1	25.4x	Scaffold	93.3
Colubridae	*Ptyas mucosa*	GCA_012654045.1	59.0x	Scaffold	92.6
Dipsadidae	*Thermophis baileyi*	GCA_003457575.1	185.0x	Scaffold	93.6
Elapidae	*Aipysurus laevis*	GCA_040207615.1	40.0x	Scaffold	91.9
Elapidae	*Bungarus multicinctus*	GCA_023653725.1	80.0x	Chromosome	90.8
Elapidae	*Hydrophis cyanocinctus*	GCA_019473425.1	318.0x	Chromosome	90
Elapidae	*Hydrophis elegans*	GCA_033807725.1	60.0x	Scaffold	91.5
Elapidae	*Hydrophis major*	GCA_033807585.1	30.0x	Chromosome	92
Elapidae	*Laticauda colubrina*	GCA_015471245.1	43.0x	Scaffold	92.1
Elapidae	*Ophiophagus hannah*	GCA_000516915.1	28.0x	Scaffold	91.9
Lamprophiidae	*Oxyrhabdium leporinum*	GCA_032468155.1	85.0x	Scaffold	91.6
Natricidae	*Natrix natrix*	GCA_029891615.1	75.0x	Scaffold	91
Natricidae	*Rhabdophis nuchalis*	GCA_039707465.1	40.0x	Chromosome	91.6
Pythonidae	*Liasis olivaceus*	GCA_030867105.1	33.5x	Chromosome	95.8
Pythonidae	*Morelia viridis*	GCA_027559625.1	105.0x	Scaffold	94.8
Pythonidae	*Python bivittatus*	GCF_000186305.1	20.0x	Scaffold	94.1
Viperidae	*Azemiops feae*	GCA_023970755.1	104.0x	Contig	91.7
Viperidae	*Bothrops alternatus*	GCA_034064705.1	28.0x	Contig	92.3
Viperidae	*Crotalus ruber*	GCA_041893435.1	20.0x	Scaffold	93.3
Viperidae	*Sistrurus catenatus*	GCA_039880765.1	31.0x	Chromosome	91.6

Table 2 List of genes related to growth and development GO terms

GO terms	Genes
positive regulation of developmental growth (GO:0048639)	*PIM1*
	YAP1
developmental growth (GO: 0048589)	*HELT*
	PLAG1
	PRKG1
	YAP1
positive regulation of developmental process (GO:0051094)	*CYLD*
	LRRTM2
	SPI1
	ANGPT2

binding (GO:0030246) etc.". Although some pathways were not significantly enriched, we still identified several genes associated with important developmental pathways. For instance, five genes (*HELT*, *PLAG1*, *PRKG1*, *SPRY1* and *YAP1*) in the GO term "developmental growth (GO: 0048589)" that play key roles in regulating the growth and development of an organism. Meanwhile, *YAP1* and *PIM1*, identified as BSAGs, are associated with the GO term "positive regulation of developmental growth (GO:0048639)," which plays a key role in maintaining normal organ and body development. In addition to the aforementioned genes, the *CYLD*, *LRRTM2*, *SPI1* and *ANGPT2* genes are also associated with the GO term "positive regulation of developmental process (GO:0051094)," which pertains to the positive regulation of growth and development (Table 2).

Gene family evolution

We then examined the expansion and contraction of gene families in three large-bodied snake species within the Serpentes lineage. In large-bodied snakes, we found that 32, 51, and 41 gene families showed significant expansions, while 34, 110, and 87 gene families exhibited significant contractions in *L. olivaceus*, *O. hannah*, and *P. bivittatus*, respectively (Fig. 1). GO enrichment analysis of the expanded gene families in these three large-bodied snake species revealed ten shared GO terms (Table 3; Fig. 2a), including those related to olfactory receptor activity (GO:0004984), immune response (GO:0006955), adaptive immunity (e.g., antigen processing and presentation (GO:0019882) and MHC class II protein complex (GO:0042613)), and metabolic processes (e.g., oxidoreductase activity (GO:0016491) and G protein-coupled receptor activity (GO:0007186)).

Positive selection in large-bodied snakes

We identified 29 genes with significant positive selection signatures (false discovery rate [FDR]-adjusted $P < 0.05$)

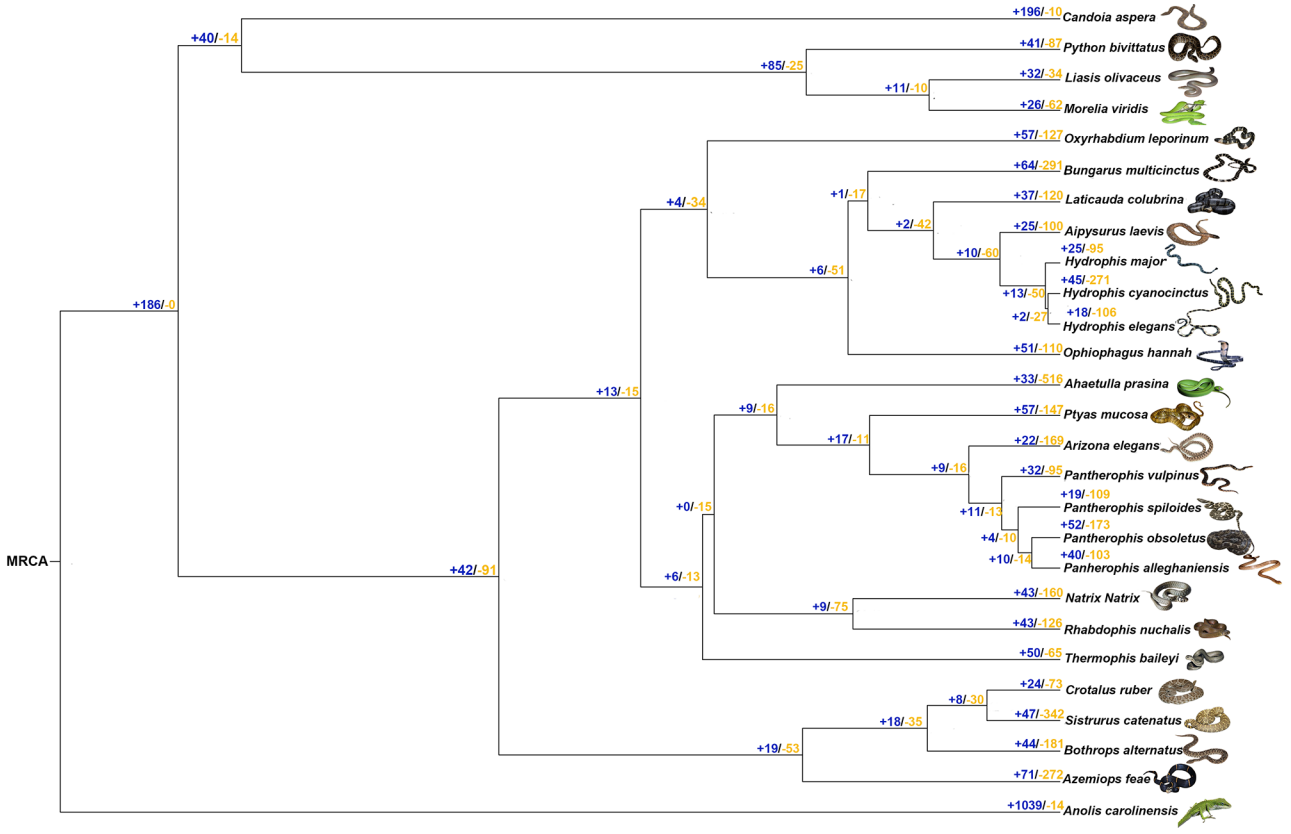

Fig. 1 Phylogenetic tree of 26 snakes. The number of expanded (blue) and contracted (yellow) gene families is shown at each branch. MRCA, means the most recent common ancestor

Table 3 GO enrichment results of expanded gene families shared among three large-bodied snakes

GO	type	property
GO:0004930	molecularfunction	G protein-coupled receptor activity
GO:0004984	molecularfunction	olfactory receptor activity
GO:0006355	biologicalprocess	regulation of transcription, DNA-templated
GO:0006955	biologicalprocess	immune response
GO:0007186	biologicalprocess	G protein-coupled receptor signaling pathway
GO:0016020	cellularcomponent	membrane
GO:0016021	cellularcomponent	integral component of membrane
GO:0016491	molecularfunction	oxidoreductase activity
GO:0019882	biologicalprocess	antigen processing and presentation
GO:0042613	cellularcomponent	MHC class II protein complex

in three large-bodied snake species (Table S3). Correlations between ω values and both body length and body mass were calculated, revealing 12 genes with significant correlations. Among them, 2 genes (*MGAT1*and *SPRY1*) showed significant correlations with both body length and body mass (Fig. 3). GO enrichment analysis revealed that the positively selected genes were significantly associated with terms such as single-stranded DNA binding (GO:0003697), sodium channel activity (GO:0005272), negative regulation of cell population proliferation (GO:0008285), and regulation of signal transduction (GO:0009966). KEGG pathway enrichment analysis further showed that these positively selected genes were significantly enriched in pathways including fatty acid biosynthesis (ko00062), fatty acid metabolism (ko01212), osteoclast differentiation (ko04380), and the Wnt signaling pathway (ko04310) (Table 4). Additionally, several genes are involved in transmembrane signaling and ion transport, emphasizing the importance of signal transduction in these snakes. In terms of biological processes, these genes play critical roles in metabolism, particularly carbohydrate metabolism and phosphatidylinositol phosphate biosynthesis.

Rapidly evolving genes (REGs) analyses

We identified 75 rapidly evolving genes in three large-bodied snakes using branch model (Table S4). Functional enrichment analysis of the REGs revealed several key categories of molecular functions and biological processes (Table 4; Fig. 2b). The molecular functions include various binding activities, such as nucleic acid, RNA, and protein binding, as well as catalytic activities exemplified

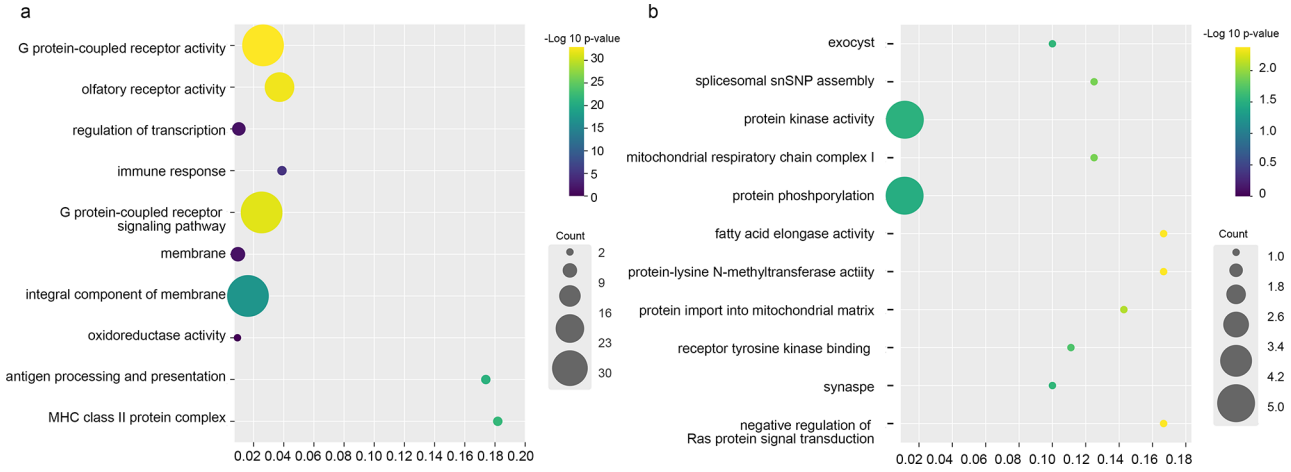

Fig. 2 GO term enrichment results for the expanded gene families (**a**) and REGs (**b**) of large-bodied snakes

by GTPase and protein kinase activities. They also contribute to transcriptional regulation, encompassing both positive and negative regulation, as well as protein phosphorylation and glycosylation. Cellular processes, such as signal transduction regulation and protein transport, are also prominently represented, highlighting their essential role in cellular communication and homeostasis.

Discussion

The regulation of body size is a critical aspect of an animal's life history, influencing its physiology, behavior, and morphology [30, 31]. However, the mechanisms underlying body size regulation are highly complex, and the factors driving body size evolution in different directions remain poorly understood [32]. The unique anatomy of snakes is thought to have evolved through early colonization and specialization, potentially linked to their divergence from other reptiles [33]. Modern snakes exhibit a remarkable range of body sizes [34], reflecting diverse ecological and evolutionary pressures. While the molecular evolution of body size has been extensively studied in mammals [18, 35], the body size evolution of snakes has received comparatively little attention. Our phylogenomic analyses of 26 snake genomes, combined with comparative genomic approaches, suggest a complex interplay of molecular adaptations underlying body size evolution. We successfully reconstructed the phylogenomic relationships of 26 high-quality snake genomes, indicating significant evolutionary patterns that align with previous studies [36, 37]. We provide molecular evidence highlighting distinct evolutionary pressures acting on genes associated with body size, including both body length and body mass, in snake clades with contrasting body size. These findings offer new insights into the molecular basis of body size evolution in snakes and improve our understanding of their evolutionary history.

In this study, we utilized PGLS to identify subtle variations among Serpentes and to uncover correlations between genes and phenotypes across their phylogeny. The root-to-tip dN/dS method has been shown to be a powerful tool for detecting gene-phenotype associations, as it accounts more comprehensively for the evolutionary history of a locus and is thus better suited for regression analyses using phenotypic data from extant Serpentes [38]. The PGLS approach used here parallels methodologies validated in carnivore studies, where phylogenetically controlled regression effectively disentangled size-associated genomic signals [39]. Thus, we aimed to investigate the evolutionary mechanisms that may underlie body size variation and to identify important candidate genes potentially involved in body size changes in Serpentes. Our identification of 77 BSAGs associated with both body length and mass in snakes appears consistent with evolutionary patterns observed in other vertebrates, where overlapping genes frequently regulate correlated size traits. The enrichment of BSAGs in metabolic processes (e.g., fatty acid metabolism, oxidoreductase activity) echoes findings in cetaceans, where body size evolution similarly involves metabolic pathway adaptation [40]. YAP1, a key downstream effector of the Hippo signaling pathway, plays a crucial role in regulating cell proliferation, apoptosis, and gene expression during limb bud regeneration in *Xenopus laevis* and has been identified as a BSAG in this study, highlighting its critical role in organ and body growth [41]. Our study suggests a significant association between the *PLAG1* gene and body size regulation in snakes, which aligns with its conserved role as a key regulator of body size in mammal [42]. PLAG1 is a zinc finger transcription factor primarily known for its role in tumorigenesis, but it also plays a significant role in normal physiology, particularly in regulating growth and reproduction. The phenotype of *plag1* knockout mice indicates that *PLAG1* plays a crucial role

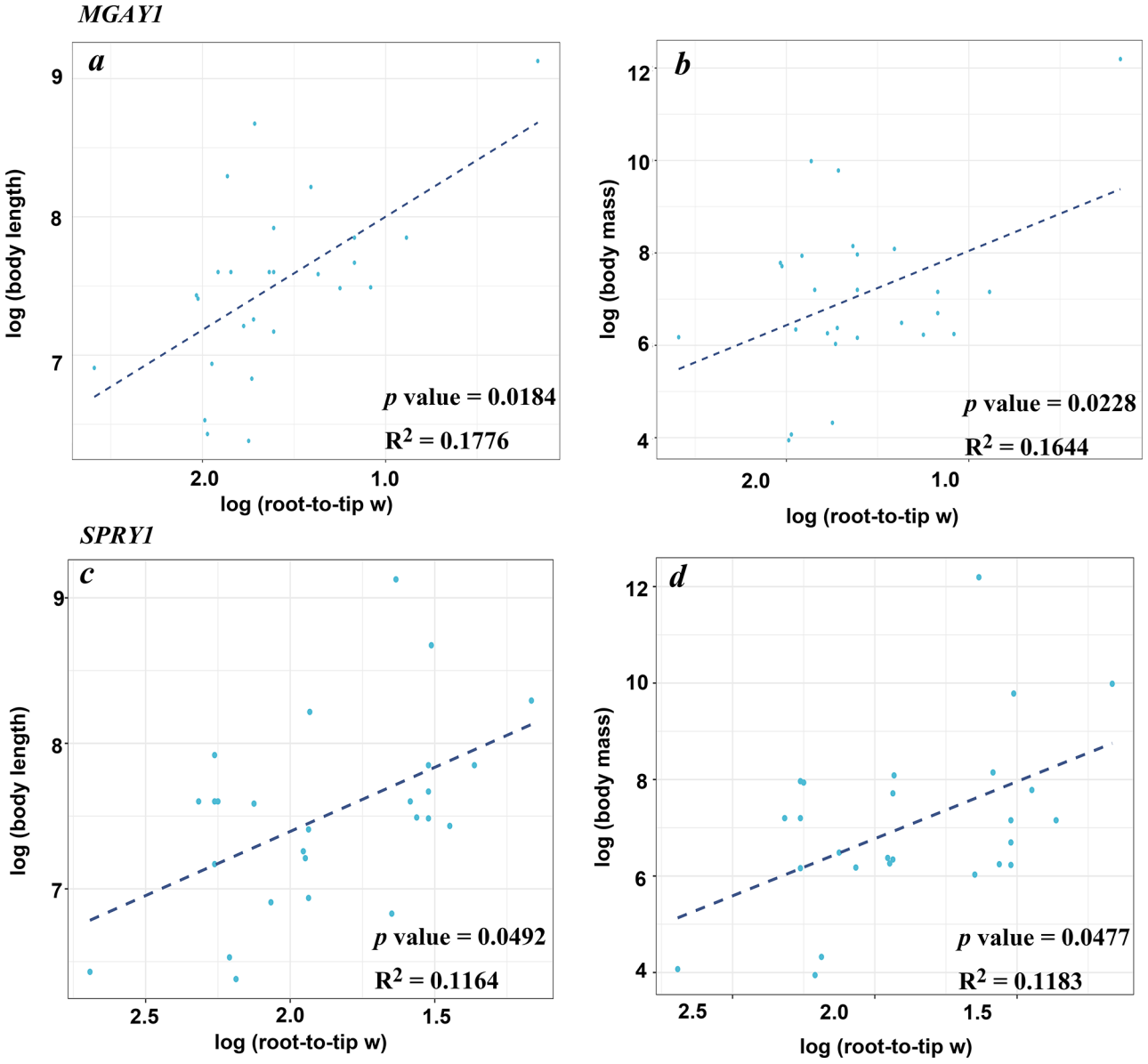

Fig. 3 Regression analyses between root-to-tip (ω) against body length (**a** and **c**) and body mass (**b** and **d**) by PGLS in R for two BASGs: *MGAP1* and *SPRY1*

in postnatal growth and reproduction, as its deficiency leads to growth retardation and reduced fertility [43].

Our study identified several genetic adaptations and molecular mechanisms that likely drive the evolution of large body size in snakes (*Liasis olivaceus*, *Ophiophagus hannah*, and *Python bivittatus*). These adaptations span a range of biological functions, including sensory perception, immune response, metabolic regulation, and growth, demonstrating the multifaceted challenges and opportunities associated with increased body size.

Gene ontology (GO) enrichment and pathway analyses highlighted significant expansions in gene families associated with olfactory receptor activity (GO:0004984). These findings may point to a role for sensory adaptations, potentially enabling large-bodied snakes to enhance their foraging strategies and thrive in diverse ecological environments. The olfactory receptor gene family, a multigene family in vertebrates, plays a crucial role in detecting and discriminating odor molecules [44]. Previous research has shown that marine snakes possess a reduced number of olfactory receptor genes compared to terrestrial species, reflecting specific adaptations to their aquatic habitats [45]. This highlights how olfactory receptor diversification may have facilitated niche-specific ecological diversification in large-bodied snakes.

In addition to sensory adaptation, immune-related processes were significant in the enriched results of expanded gene families and positively selected genes.

Table 4 Kegg enrichment results of PSGs and REGs

Genes	Pathway	ID	Input number	Background number	*P*-Value
PSGs	Oxytocin signaling pathway	ko04921	2	153	0.0040
	Biosynthesis of unsaturated fatty acids	ko01040	1	27	0.0170
	Fatty acid elongation	ko00062	1	27	0.0170
	N-Glycan biosynthesis	ko00510	1	50	0.0307
	Fatty acid metabolism	ko01212	1	57	0.0349
	VEGF signaling pathway	ko04370	1	59	0.0360
	Acute myeloid leukemia	ko05221	1	66	0.0402
	Pathways in cancer	ko05200	2	530	0.0415
	Chronic myeloid leukemia	ko05220	1	76	0.0460
	B cell receptor signaling pathway	ko04662	1	82	0.0495
REGs	ErbB signaling pathway	ko04012	2	85	0.0098
	Leukocyte transendothelial migration	ko04670	2	112	0.0163
	Yersinia infection	ko05135	2	121	0.0188
	Fluid shear stress and atherosclerosis	ko05418	2	139	0.0243
	Retrograde endocannabinoid signaling	ko04723	2	148	0.0272
	Protein processing in endoplasmic reticulum	ko04141	2	165	0.0331
	Tight junction	ko04530	2	170	0.0350
	Protein export	ko03060	1	23	0.0401
	Transcriptional misregulation in cancer	ko05202	2	186	0.0411
	Kaposi sarcoma-associated herpesvirus infection	ko05167	2	186	0.0411
	Biosynthesis of unsaturated fatty acids	ko01040	1	27	0.0467
	Fatty acid elongation	ko00062	1	27	0.0467

Specifically, expanded gene families were enriched in immune-related GO terms such as antigen processing and presentation (GO:0019882) and the MHC class II protein complex (GO:0042613), essential for adaptive immunity. Moreover, positively selected genes were associated with immune response-related pathways, including the RIG-I-like receptor signaling pathway (ko04622) and the intestinal immune network for IgA production (ko04672). Body size influences the extent of a host's exposure to parasites and affects responses to infections through its impact on metabolic rates and other related factors [46]. Hence, expansions and selection pressures on immune-related genes may reflect evolutionary strategies for resolving the challenges associated with increased body size.

Metabolic adaptation emerged as another critical feature in the evolution of large body size. GO enrichment of expanded gene families suggested a significant association with oxidoreductase activity (GO:0016491) and G protein-coupled receptor activity (GO:0007186), supporting their role in improving energetic efficiency to meet the physiological demands of larger body sizes [47]. In addition, PSGs were enriched in metabolic pathways such as glycolysis/gluconeogenesis (ko00010) and fatty acid biosynthesis (ko00061), which are directly linked to energy production and storage. Functional enrichment analysis of REGs further highlighted involvement in carbohydrate metabolism and phosphatidylinositol phosphate biosynthesis, processes essential for energy regulation. Meiri et al. [48], observed that smaller species gain or lose heat more rapidly than larger species, suggesting that larger-bodied snakes may face greater challenges in thermoregulation. Additionally, larger snakes tend to consume relatively larger prey, which increases the energetic costs associated with handling, digestion, and mechanical effort [49]. The convergence of results across gene family expansions, positively selected genes, and REGs suggests that metabolic adaptations may have been important for supporting the energetic demands and physiological maintenance of larger snakes.Our analyses also identified key developmental processes linked to body size regulation. Positively selected genes were enriched in signaling pathways critical to regulating growth, including Hedgehog signaling (ko04340) and Hippo signaling (ko04392), both known to control cell proliferation, tissue growth, and size regulation in vertebrates [50–52].

In addition to the pathway-level adaptations identified in our study, several individual genes stood out as likely contributors to body size regulation in snakes. Notably, we identified *MGAT1* and *SPRY1* as BSAGs that are not only functionally important but also under significant selective pressure, highlighting their potential roles in the evolution of large-bodied snakes. *MGAT1* encodes a protein involved in the N-glycan synthesis process, which is crucial for mediating cell-cell and cell-matrix interactions essential to multicellular development. As a novel transcriptional target of the conserved Wnt/β-catenin

signaling pathway, *MGAT1* plays a critical role in vertebrate early development, including axis formation, cellular proliferation, and morphogenesis [53]. Similarly, *SPRY1* was identified as a regulator of body size and growth, encoding a protein that acts as an inhibitor of receptor tyrosine kinase (RTK) signaling. RTK signaling plays a major role in pathways regulating cell proliferation, differentiation, and morphogenesis [54, 55]. *SPRY1* has been implicated in multiple developmental processes, including skeletal muscle development, kidney morphogenesis, and overall tissue homeostasis [56]. There is evidence linking *SPRY1* to body size and growth regulation, as it modulates signaling pathways (such as FGF, MAPK/ERK) that are critical for cell growth and differentiation [57]. The selection observed on *SPRY1* and *MGAT1* suggests that they may have played important roles in mediating developmental processes necessary to support larger body sizes in these snakes. These genes exemplify how body size evolution in snakes may arise from a combination of broad pathway-level changes and fine-scale selection on individual genes that are functionally critical for growth regulation. Their dual roles as BSAGs and targets of selection may hint at the intricate genetic architecture potentially contributing to large-bodied adaptations in snakes.

Taken together, our findings suggest that the evolution of large snake body sizes is driven by an intricate interplay of sensory, immune, metabolic, and growth-related adaptations. Gene family expansions, positively selected genes, and rapidly evolving genes all contributed to these traits, reflecting evolutionary responses to the unique ecological and physiological challenges posed by increased size. These genetic adaptations may help to explain how large-bodied snakes have diversified and thrived across complex and dynamic habitats.

Furthermore, due to the current limitations in the availability of genomic data, our analyses did not include species with extremely small body sizes, such as blind snakes (Scolecophidia), or taxa with even larger body sizes than those analyzed, such as additional python species. This taxonomic sampling limitation may affect the comprehensiveness of our conclusions regarding body size evolution across the entire snake phylogeny. As more high-quality snake genomes become available in the future, additional insights into the molecular mechanisms underlying body size evolution may emerge. The continued expansion of genomic resources could enhance our understanding of the molecular underpinnings of snake body size evolution, allowing for more robust testing of the hypotheses suggested in this study.

Conclusions

To our knowledge, our results provide the first direct evidence for genotype-phenotype correlations underlying body size evolution in snakes. Through comparative genomic analyses, we found that snake body size evolution is shaped by gene family expansions, positive selection, and rapid changes in pathways related to sensory adaptation, immune response, metabolism, and growth regulation, highlighting coordinated genetic adaptations across these processes. We also identified four critical BSAGs. *MGAT1* and *SPRY1* showed evidence of selection and play important roles in growth regulation via Wnt/β-catenin and RTK signaling pathways. *PLAG1* was associated with growth and reproduction, while *YAP1* regulates cell proliferation and organ growth. These findings offer new insights into the genetic mechanisms underlying body size evolution in snakes. Our work not only fills a gap in understanding reptilian body size evolution but also highlights the role of genetic changes at multiple scales, ranging from pathway-level adaptations to the selection of individual genes, in shaping ecologically driven phenotypic diversity. This framework offers insights for comparative genomics and advances efforts to unravel the genomic basis of complex traits across vertebrates.

Abbreviations

BUSCO	Benchmarking universal single-copy orthologs
PAML	Phylogenetic analysis by maximum likelihood
CAFE	Computational analysis of gene family evolution
dN	Nonsynonymous substitutions
dS	Synonymous substitutions
GO	Gene ontology
KEGG	Kyoto encyclopedia of genes and genomes

Supplementary Information

The online version contains supplementary material available at https://doi.org/10.1186/s12864-025-11601-1.

Supplementary Material 1

Supplementary Material 2

Supplementary Material 3

Supplementary Material 4

Acknowledgements

Not applicable.

Author contributions

TX, SYZ and ZHZ conceived and designed this project. TX and HHZ wrote the main manuscript text, XYW, XBW, LZ and JQD prepared figures, XDG, GLS and XFY prepared tables. All authors reviewed the manuscript.

Funding

This research was supported by the National Natural Science Foundation of China (32270444, 32200407 and 32200349), the Natural Science Foundation of Shandong Province (ZR2023ZD47), the Natural Science Foundation of Shandong Province (ZR2022QC153) and the Postdoctoral Fellowship Program of CPSF (GZC20231395).

Data availability
All genome sequences used in this study are available on the NCBI database (https://www.ncbi.nlm.nih.gov/) under the accession numbers reported in Table 1.

Declarations

Ethics approval and consent to participate
Not applicable.

Consent for publication
Not applicable.

Competing interests
The authors declare no competing interests.

Received: 15 December 2024 / Accepted: 14 April 2025
Published online: 29 April 2025

References

1. Feldman A, Sabath N, Pyron RA, Mayrose I, Meiri S. Body sizes and diversification rates of lizards, snakes, amphisbaenians and the Tuatara. Glob Ecol Biogeogr. 2016;25(2):187–97.
2. Gaston KJ, Blackburn TM. Range size-body size relationships: evidence of scale dependence. Oikos. 1996;75(3):479–85.
3. Jaffe AL, Slater GJ, Alfaro ME. The evolution of Island gigantism and body size variation in tortoises and turtles. Biol Lett. 2011;7(4):558–61.
4. Iverson SJB. Patterns of survival, growth and maturation in turtles. Oikos. 1995;72(3):343–8.
5. Slavenko A, Tallowin OJS, Itescu Y, Raia P, Meiri S. Late quaternary reptile extinctions: size matters, insularity dominates. Global Ecology & Biogeography; 2016.
6. Wollenberg KC, Vieites DR, Glaw F, Vences M. Speciation in little: the role of range and body size in the diversification of Malagasy mantellid frogs. BMC Evol Biol. 2011;11(1):217–217.
7. Marcel C, Mace GM, Jones KE, Jon B, Bininda-Emonds ORP, Wes S, Orme CDL, Andy P. Multiple causes of high extinction risk in large mammal species. Science. 2005;309(5738):1239–41.
8. Orme CD, Isaac NJ, Purvis A. Are most species small? Not within species-level phylogenies. Proc Biol Sci. 2002;269(1497):1279–87.
9. Blackwell P, Meiri S. Evolution and ecology of Lizard body sizes. Global Ecol Biogeogr. 2010;17(6):724–34.
10. Kemper KE, Visscher PM, Goddard ME. Genetic architecture of body size in mammals. Genome Biol. 2012;13(4):244.
11. Hsiang AY, Field DJ, Webster TH, Behlke AD, Davis MB, Racicot RA, Gauthier JA. The origin of snakes: revealing the ecology, behavior, and evolutionary history of early snakes using genomics, phenomics, and the fossil record. BMC Evol Biol. 2015;15(1):87.
12. Maritz B, Kgaditse M, Alexander GJ. Snake body size frequency distributions are robust to the description of novel species. Ecosphere. 2016;7(5):7.
13. Taylor Q, Cox CL. Evidence of predation risk increases with body size in a diminutive snake. J Zool. 2019;307(2):141–8.
14. Madsen T, Shine R, COSTS OF REPRODUCTION INFLUENCE THE EVOLUTION OF SEXUAL SIZE DIMORPHISM IN SNAKES. Evolution. 1994;48(4):1389–97.
15. Shai M. SquamBase—A database of squamate (Reptilia: Squamata) traits. Glob Ecol Biogeogr. 2024;33(4):e13812.
16. Bannasch DL, Baes CF, Leeb T. Genetic variants affecting skeletal morphology in domestic dogs. Trends Genet. 2020;36(8):598–609.
17. Huang X, Sun D, Wu TZ, Liu X, Xu SX, Yang G. Genomic insights into body size evolution in Carnivora support Peto's paradox. BMC Genomics 2021, 22(1).
18. Sun Y, Liu Y, Sun X, Lin Y, Yin D, Xu S. Insights into body size variation in cetaceans from the evolution of body-size-related genes. 2019, 19(1):157.
19. Kiełbasa SM, Wan R, Sato K, Horton P, Frith MC. Adaptive seeds tame genomic sequence comparison. Genome Res. 2011;21(3):487–93.
20. Blanchette M, Kent WJ, Riemer C, Elnitski L, Smit AF, Roskin KM, Baertsch R, Rosenbloom K, Clawson H, Green ED, et al. Aligning multiple genomic sequences with the threaded blockset aligner. Genome Res. 2004;14(4):708–15.
21. Stanke M, Keller O, Gunduz I, Hayes A, Waack S, Morgenstern B. AUGUSTUS:: ab initio prediction of alternative transcripts. Nucleic Acids Res. 2006;34:W435–9.
22. Finn RD, Clements J, Eddy SR. HMMER web server: interactive sequence similarity searching. Nucleic Acids Res. 2011;39:W29–37.
23. Emms DM, Kelly S. OrthoFinder: solving fundamental biases in whole genome comparisons dramatically improves orthogroup inference accuracy. Genome Biol. 2015;16(1):157.
24. Stamatakis A. RAxML version 8: a tool for phylogenetic analysis and post-analysis of large phylogenies. Bioinformatics. 2014;30(9):1312–3.
25. Kumar S, Suleski M, Craig JM, Kasprowicz AE, Sanderford M, Li M, Stecher G, Hedges SB. TimeTree 5: an expanded resource for species divergence times. Mol Biol Evol 2022, 39(8).
26. Orme D, Freckleton RP, Thomas GH, Petzoldt T, Fritz SA, Isaac N. CAPER: comparative analyses of phylogenetics and evolution in R. Methods Ecol Evol. 2013;3:145–51.
27. Yang Z. PAML 4: phylogenetic analysis by maximum likelihood. Mol Biol Evol. 2007;24(8):1586–91.
28. Montgomery SH, Capellini I, Venditti C, Barton RA, Mundy NI. Adaptive evolution of four microcephaly genes and the evolution of brain size in anthropoid primates. Mol Biol Evol. 2011;28(1):625–38.
29. Jones P, Binns D, Chang HY, Fraser M, Li WZ, McAnulla C, McWilliam H, Maslen J, Mitchell A, Nuka G, et al. InterProScan 5: genome-scale protein function classification. Bioinformatics. 2014;30(9):1236–40.
30. Peters R. The ecological implications of body size. 1983, 22(1):291.
31. Harvey PH, Purvis A. Understanding the ecological and evolutionary reasons for life history variation: mammals as a case study. Blackwell Publishing Ltd.; 2009.
32. Wikelski M. Evolution of Body Size in Galapagos Marine Iguanas. *Proceedings: Biological Sciences* 2005.
33. Bellairs AD, Underwood G. The origin of snakes. Biol Rev Camb Philos Soc. 1951;26(2):193–237.
34. Feldman A, Sabath N, Pyron RA, Mayrose I, Meiri S. Body sizes and diversification rates of lizards, snakes, amphisbaenians and the Tuatara. Global Ecol Biogeogr. 2016;25(2):187–97.
35. Kim H, Song KD, Kim HJ, Park W, Kim J, Lee T, Shin DH, Kwak W, Kwon YJ, Sung S, et al. Exploring the genetic signature of body size in Yucatan miniature pig. PLoS ONE. 2015;10(4):16.
36. Burbrink FT, Grazziotin FG, Pyron RA, Cundall D, Donnellan S, Irish F, Keogh JS, Kraus F, Murphy RW, Noonan B, et al. Interrogating Genomic-Scale data for Squamata (Lizards, snakes, and Amphisbaenians) shows no support for key traditional morphological relationships. Syst Biol. 2019;69(3):502–20.
37. Roberts JR, Bernstein JM, Austin CC, Hains T, Mata J, Kieras M, Pirro S, Ruane S. Whole snake genomes from eighteen families of snakes (Serpentes: Caenophidia) and their applications to systematics. J Hered. 2024;115(5):487–97.
38. Muntané G, Farré X, Rodríguez JA, Pegueroles C, Hughes DA, de Magalhães JP, Gabaldón T, Navarro A. Biological processes modulating longevity across primates: A phylogenetic Genome-Phenome analysis. Mol Biol Evol. 2018;35(8):1990–2004.
39. Huang X, Sun D, Wu T, Liu X, Xu S, Yang G. Genomic insights into body size evolution in Carnivora support Peto's paradox. BMC Genomics. 2021;22(1):429.
40. Sun 孙迪 D, Chai 柴思敏 S, Huang 黄鑫 X, Wang 王滢莹 Y, Xiao 肖琳琳 L, Xu 徐士霞 S. Yang 杨光 G: novel genomic insights into body size evolution in cetaceans and a resolution of Peto's paradox. Am Nat. 2022;199(2):E28–42.
41. Hayashi S, Tamura K, Yokoyama H. Yap1, transcription regulator in the Hippo signaling pathway, is required for Xenopus limb bud regeneration. Dev Biol. 2014;388(1):57–67.
42. Karim L, Takeda H, Lin L, Druet T, Arias JAC, Baurain D, Cambisano N, Davis SR, Farnir F, Grisart B, et al. Variants modulating the expression of a chromosome domain encompassing PLAG1 influence bovine stature. Nat Genet. 2011;43(5):405–13.
43. Juma AR, Damdimopoulou PE, Grommen SV, Van de Ven WJ, De Groef B. Emerging role of PLAG1 as a regulator of growth and reproduction. J Endocrinol. 2016;228(2):R45–56.
44. Gilad Y, Man O, Glusman G. A comparison of the human and chimpanzee olfactory receptor gene repertoires. Genome Res. 2005;15(2):224–30.
45. Kishida T, Go Y, Tatsumoto S, Tatsumi K, Kuraku S, Toda M. Loss of olfaction in sea snakes provides new perspectives on the aquatic adaptation of amniotes. *Proceedings of the Royal Society B-Biological Sciences* 2019, 286(1910).
46. Brace AJ, Lajeunesse MJ, Ardia DR, Hawley DM, Adelman JS, Buchanan KL, Fair JM, Grindstaff JL, Matson KD, Martin LB. Costs of immune responses

are related to host body size and lifespan. J Exp Zool Ecol Integr Physiol. 2017;327(5):254–61.
47. Bronikowski A, Vleck D. Metabolism, body size and life span: A case study in evolutionarily divergent populations of the garter snake (Thamnophis elegans). Integr Comp Biol. 2010;50(5):880–7.
48. Meiri S. Evolution and ecology of Lizard body sizes. Glob Ecol Biogeogr. 2008;17(6):724–34.
49. Hampton PM. Ontogenetic prey size selection in snakes: predator size and functional limitations to handling minimum prey sizes. Zoology. 2018;126:103–9.
50. Vukicevic S, Sampath KT. Bone morphogenetic proteins: from laboratory to clinical practice. Springer; 2012.
51. Han H-S, Kang G, Kim JS, Choi BH, Koo S-H. Regulation of glucose metabolism from a liver-centric perspective. Exp Mol Med. 2016;48(3):e218–218.
52. Cohn MJ. Developmental mechanisms of vertebrate limb evolution. Novartis Found Symp. 2001;232:47–57. discussion 57–62.
53. Akiva I, Iyison NB. MGAT1 is a novel transcriptional target of Wnt/β-catenin signaling pathway. BMC Cancer 2018, 18 (1):60.
54. Edwin F, Anderson K, Ying C, Patel TB. Intermolecular interactions of sprouty proteins and their implications in development and disease. Mol Pharmacol. 2009;76(4):679–91.
55. Mason JM, Morrison DJ, Basson MA, Licht JD. Sprouty proteins: multifaceted negative-feedback regulators of receptor tyrosine kinase signaling. Trends Cell Biol. 2006;16(1):45–54.
56. Basson MA, Akbulut S, Watson-Johnson J, Simon R, Carroll TJ, Shakya R, Gross I, Martin GR, Lufkin T, McMahon AP, et al. Sprouty1 is a critical regulator of GDNF/RET-Mediated kidney induction. Dev Cell. 2005;8(2):229–39.
57. Taniguchi K, Ayada T, Ichiyama K, Kohno R-i, Yonemitsu Y, Minami Y, Kikuchi A, Maehara Y, Yoshimura A. Sprouty2 and Sprouty4 are essential for embryonic morphogenesis and regulation of FGF signaling. Biochem Biophys Res Commun. 2007;352(4):896–902.

Publisher's note