Comparative and evolutionary analysis of the reptilian hedgehog gene family (*Shh*, *Dhh*, and *Ihh*)

Tian Xia [1] , **Honghai Zhang** [Corresp., 1] , **Lei Zhang** [1] , **Xiufeng Yang** [1] , **Guolei Sun** [1] , **Jun Chen** [2] , **Dajie Xu** [1] , **Chao Zhao** [1]

[1] College of Life Science, Qufu Normal University, Qufu, Shandong, China

[2] College of Marine Life Science, Ocean University of Qingdao, Qingdao, Shandong, China

Corresponding Author: Honghai Zhang
Email address: zhanghonghai67@126.com

The hedgehog signaling pathway plays a vital role in human and animal patterning and cell proliferation during the developmental process. The hedgehog gene family of vertebrate species includes three genes, *Shh*, *Dhh*, and *Ihh*, which possess different functions and expression patterns. Despite the importance of hedgehog genes, genomic evidence of this gene family in reptiles is lacking. In this study, the available genomes of a number of representative reptile species were explored by utilizing adaptive evolutionary analysis methods to characterize the evolutionary patterns of the hedgehog gene family. Altogether, 33 *Shh*, 25 *Dhh*, and 20 *Ihh* genes were obtained from reptiles, and six avian and five mammalian sequences were added to the analysis. Phylogenetic analysis by maximum likelihood (ML) of the *Shh*, *Dhh*, and *Ihh* genes revealed a similar topology, which is approximately consistent with the traditional taxonomic group. No shared positive selection site was identified by the PAML site model or the three methods in the Data Monkey Server. Branch model and Clade model C (CmC) analyses revealed that the *Dhh* and *Ihh* genes experienced different evolutionary forces in reptiles and other vertebrates, while the *Shh* gene was not significantly different from the other genes in terms of selection pressure. The different evolutionary rates of the *Dhh* and *Ihh* genes suggest that these genes may be potential contributors to the discrepant sperm and body development of different clades. The different adaptive evolutionary history of the *Shh*, *Dhh*, and *Ihh* genes among reptiles may be due to their different functions in regulating cellular events of development from the embryonic stages to adulthood. Overall, this study has provided meaningful information regarding the evolution of the hedgehog gene family in reptiles and a theoretical foundation for further analyses on the functional and molecular mechanisms that have shaped the reptilian hedgehog genes.

1 **Comparative and evolutionary analysis of the reptilian**

2 **hedgehog gene family (*Shh*, *Dhh*, and *Ihh*)**

3 Tian Xia[1], Honghai Zhang[1]*, Lei Zhang[1], Xiufeng Yang[1], Guolei Sun[1], Jun Chen[2], Dajie Xu[1],

4 Chao Zhao[1].

5 [1] College of Life Science, Qufu Normal University, Qufu, Shandong, China

6 [2] College of Marine Life Science, Ocean University of China, Qingdao, Shandong, China

7 Corresponding Author:

8 Honghai Zhang[1]

9 57 Jingxuan West Road, Qufu, Shandong, 273165, China

10 E-mail adress: zhanghonghai67@126.com

12
13
14
15
16
17
18
19
20
21
22
23
24
25
26
27
28
29
30
31
32
33
34
35
36
37
38
39
40
41
42
43
44

45
46
47
48

Abstract

49
50 The hedgehog signaling pathway plays a vital role in human and animal patterning and cell
51 proliferation during the developmental process. The hedgehog gene family of vertebrate species
52 includes three genes, *Shh*, *Dhh*, and *Ihh*, which possess different functions and expression
53 patterns. Despite the importance of hedgehog genes, genomic evidence of this gene family in
54 reptiles is lacking. In this study, the available genomes of a number of representative reptile
55 species were explored by utilizing adaptive evolutionary analysis methods to characterize the
56 evolutionary patterns of the hedgehog gene family. Altogether, 33 *Shh*, 25 *Dhh*, and 20 *Ihh* genes
57 were obtained from reptiles, and six avian and five mammalian sequences were added to the
58 analysis. Phylogenetic analysis by maximum likelihood (ML) of the *Shh*, *Dhh*, and *Ihh* genes
59 revealed a similar topology, which is approximately consistent with the traditional taxonomic
60 group. No shared positive selection site was identified by the PAML site model or the three
61 methods in the Data Monkey Server. Branch model and Clade model C (CmC) analyses revealed
62 that the *Dhh* and *Ihh* genes experienced different evolutionary forces in reptiles and other
63 vertebrates, while the *Shh* gene was not significantly different from the other genes in terms of
64 selection pressure. The different evolutionary rates of the *Dhh* and *Ihh* genes suggest that these
65 genes may be potential contributors to the discrepant sperm and body development of different
66 clades. The different adaptive evolutionary history of the *Shh*, *Dhh*, and *Ihh* genes among reptiles
67 may be due to their different functions in regulating cellular events of development from the
68 embryonic stages to adulthood. Overall, this study has provided meaningful information
69 regarding the evolution of the hedgehog gene family in reptiles and a theoretical foundation for
70 further analyses on the functional and molecular mechanisms that have shaped the reptilian
71 hedgehog genes.

Introduction

72
73 The hedgehog signaling pathway is described as "one of the most enigmatic" pathways, although
74 we know that it is essential for patterning and cell proliferation in humans and animals (Hooper
75 & Scott 2005). Hedgehog genes were initially identified during genetic screenings of *Drosophila*
76 *melanogaster*, with only a small number of genes in this family being isolated (Nüsslein-Volhard
77 & Wieschaus 1980). Hedgehog orthologues from vertebrates, such as *Mus musculus* (mouse) and
78 *Danio rerio* (zebrafish), were cloned in 1993 (Echelard et al. 1993; Krauss et al. 1993).
79 Duplication of the vertebrate genome contributed to the expansion of the hedgehog gene family,
80 which can be classified into three subgroups: the sonic hedgehog (*Shh*), Indian hedgehog (*Ihh*),
81 and desert hedgehog (*Dhh*) genes (Kumar et al. 1996).
82 The different functions of these hedgehog genes result mainly from their discrepant expression
83 patterns (Markku & Jussi 2008). Namely, the *Shh* gene plays a core role in the development and
84 patterning of the skeletal and nervous systems (Ingham & Mcmahon 2001a), and all systems
85 have undergone remarkable morphological changes in primates, especially in humans (Steve et
86 al. 2006). The *Ihh* gene regulates the formation of vascular and endochondral ossification, and
87 *Dhh* is indispensable for the formation of the peripheral nervous system (Nagase et al. 2010), as
88 well as the differentiation of peritubular myoid cells and the formation of the embryonic testis
89 cord (Humphrey Hung-Chang et al. 2002). Vertebrate hedgehog genes are expressed in different
90 tissues during the developmental process. *Shh* gene expression is highest in the growing organs

91   of the foregut, including the lungs, pancreas, esophagus, liver, and proximal stomach (Litingtung
92   et al. 1998; Motoyama et al. 1998). *Ihh* gene expression occurs in hindgut-induced tissues,
93   especially the distal stomach, coelom, and intestines (Zacharias et al. 2010). The *Dhh* gene is
94   crucial for the maintenance of male spermatogenesis and the germline (Szczepny et al. 2010) and
95   is also expressed in Schwann cells, taking part in the formation of the nerve sheath (Mirsky et al.
96   2010). Each gene possesses diverse expression patterns and a specific biological function,
97   making them play diverse roles in different vertebrate taxa (Shimeld 1999).
98   Previous research on the molecular evolution of the hedgehog gene family in invertebrates
99   revealed positive selection signals, which appeared to be associated with the divergence of the
100 two major bilaterian groups, Deuterostomia and Ecdysozoa (Gunbin et al. 2007). Comparative
101 genomics was used to characterize the evolution of the hedgehog genes in 45 avian and three
102 reptilian genomes, suggesting that hedgehog paralogous genes in vertebrates evolved
103 independently within homologous linkage groups at different evolutionary rates (Pereira et al.
104 2014). Previous research has focused on the functional divergence of the hedgehog pathway
105 during evolution and its integration with other signaling pathways to control and regulate cell
106 growth, differentiation, and survival (Beachy et al. 2004; Ingham & McMahon 2001). Recently,
107 the function of the hedgehog pathway became a research hotspot, which resulted in an
108 accumulation of knowledge regarding human disorders, including cancer and birth defects
109 (Mcmahon et al. 2003; Nieuwenhuis & Hui 2005). Moreover, research on the hedgehog gene
110 family has been extensively studied in cnidaria (Matus et al. 2008), zebrafish (Avaron et al.
111 2006), amphioxi (Shimeld 1999), and avians (Pereira et al. 2014).
112 Reptiles have a unique physiological feature as the sole poikilothermic amniotes (Zimmerman et
113 al. 2010). They play essential roles in their ecosystem as prey, predators, grazers, and commensal
114 species, serving as the perfect model to investigate the biological and evolutionary histories
115 underlying speciation. Despite the prominent role of reptiles in the lengthy evolutionary history
116 of vertebrates, the reptilian hedgehog signaling pathway has received little attention. Previously,
117 two reptiles, the American alligator and green turtle, were used in the evolutionary analysis of
118 the avian hedgehog gene family (Pereira et al. 2014). With the rapid development of high-
119 throughput sequencing, the available genomes of a number of reptiles have been published,
120 which provides a new opportunity for the investigation of evolutionary patterns and structural
121 characteristics of reptilian hedgehog genes. Given the current lack of comprehensive studies on
122 the hedgehog gene family molecular evolution across the reptilian phylogeny, here we performed
123 a comparative evolutionary analysis on available reptilian whole-genome sequences representing
124 three reptilian orders. Our detection of natural selection acting on three members of the
125 hedgehog gene family in reptiles unraveled adaptive evolution at the molecular level.

126 **Materials & Methods**
127 **Species and sequences**
128 The *Shh*, *Ihh*, and *Dhh* gene sequences were retrieved from the National Center for
129 Biotechnology Information (NCBI) database (www.ncbi.nlm.nih.gov). Meanwhile, the *Shh*, *Ihh*,
130 and *Dhh* amino acid sequences of multiple species, including *Pelodiscus sinensis* (Chinese soft-
131 shelled turtle), *Pseudonaja textilis* (eastern brown snake), *Notechis scutatus* (mainland tiger
132 snake), *Terrapene Mexicana* (Mexican box turtle), *Pogona vitticeps* (central bearded dragon),
133 *Protobothrops mucrosquamatus* (Taiwan habu), *Anolis carolinensis* (green anole), *Alligator*
134 *mississippiensis* (American alligator), and *Chrysemys picta* (painted turtle) were treated as
135 queries to detect the genome sequences of *Crocodylus porosus* (Australian saltwater crocodile),
136 *Malaclemys terrapin* (diamondback terrapin), *Cuora mccordi* (McCord's box turtle),

137 *Chelonoidis abingdonii* (Abingdon island giant tortoise), *Gopherus agassizii* (Agassiz's desert
138 tortoise), *Platysternon megacephalum* (big-headed turtle), *Crotalus viridis* (western rattlesnake),
139 *Vipera berus* (European adder), *Salvator merianae* (Argentine black and white tegu), *Lacerta*
140 *viridis* (green lizard), *Protobothrops flavoviridis* (habu), *Lacerta bilineata* (western green lizard),
141 *Paroedura picta* (panther gecko), *Pantherophis guttatus* (corn snake), *Ophiophagus hannah*
142 (king cobra), *Crotalus pyrrhus* (white speckled rattlesnake), *Hydrophis cyanocinctus* (Asian
143 annulated sea snake), *Hydrophis hardwickii* (Hardwick's sea snake), *Thermophis baileyi*
144 (Bailey's Snake), and *Ophisaurus gracilis* (Anguidae lizard) (Gilbert et al. 2014; Green et al.
145 2014; Hara et al. 2018; Li et al. 2018; Pasquesi et al. 2018; Roscito et al. 2018; Shibata et al.
146 2018; Song et al. 2015; Tollis et al. 2017; Ullate-Agote et al. 2014; Vonk et al. 2013). Reptiles
147 were divided into three disparate orders, including Crocodylia, Testudines, and Squamata. The
148 available information on reptilian genomes was integrated (Table S1). The commands BLASTN
149 and TBLASTN in BLAST v2.7.1 were used for orthologous gene searches with an E-value of
150 1e–10 in the aforementioned reptilian genomes. Identification of the hedgehog genes was
151 determined by a previously established method (Monie et al. 2008). Additionally, the hedgehog
152 gene sequences of birds and other vertebrates were obtained for further analysis. Information
153 regarding accession numbers and sequences of hedgehog genes is provided in Table S1.
154 **Phylogenetic analysis**
155 Multiple sequence alignment of the three hedgehog genes was analyzed by the Multiple
156 Sequence Comparison by Log-Expectation (MUSCLE) (www.ebi.ac.uk/Tools/msa/muscle).
157 Phylogenetic trees were constructed to estimate the differentiation among the three hedgehog
158 genes, using RAxML v8.2.12 with 1,000 bootstrap replications and a GTR+Γ sequence evolution
159 model. Relevant phylogenetic relationships among reptiles were obtained using Timetree v3.0
160 (www.timetree.org). Moreover, a comparison of the *Shh*, *Dhh*, and *Ihh* gene trees with the
161 reptilian species tree was conducted.
162 **Test for positive selection**
163 Based on the corresponding topology of the aforementioned species, the phylogenetic analysis
164 by maximum likelihood (PAML) in Codeml v4.9d with codon-based likelihood models was used
165 to estimate nonsynonymous/synonymous substitution ratios (ω = dN/dS) to quantify natural
166 selection for the hedgehog gene. Different ω values represent different types of selection, namely,
167 $\omega < 1$ represents purifying selection, $\omega = 1$ represents neutral selection, and $\omega > 1$ represents
168 positive selection (Kimura 1979). P < 0.05 was applied to estimate whether the alternative
169 selective hypothesis was significant or not. Previous studies have focused on the function of
170 hedgehog genes during the developmental process, mainly in the nervous system, bones, skin,
171 and pancreas (Ingham et al. 2011; Varjosalo & Taipale 2008). The hedgehog proteins control the
172 growth, survival, and fate of cell, and pattern of vertebrate body plan (Varjosalo & Taipale 2008).
173 Reptiles have different physiological characteristics, which indicate that there are different types
174 of development. In order to determine branch-specific evolutionary rates, we performed the
175 branch model (two-ratio vs. one-ratio model) in CODEML program to evaluate ω ratio for each
176 branch. In addition, to detect the probabilities of sites under positive selection in each linage,
177 branch-site model was used in which the ω ratio could vary among sites in divergent clades of
178 reptilian species. To further contrast the evolutionary rates of hedgehog genes in response to
179 divergent clades, we implemented Clade Model C, which allows different evolution along the
180 phylogeny. Hence, we partitioned all species according to the appearance of the limb: limbless
181 species and species with limbs. The best-suited model allowing discrepant evolutionary rates was
182 compared with the null model M2a_rel, which has an unconstrained ω.

183 Additionally, three evolutionary models (M7, M8a, and M8) in PAML were used to estimate
184 site-specific selection. The likelihood ratio test (LRT) results of two nested models were
185 compared to detect dramatic events of positive selection with two degrees of freedom (Yang et al.
186 2000). Amino acids with selection pressures were detected using a Bayes empirical Bayes
187 approach (BEB) by counting the posterior probability with site model M8 (Ziheng et al. 2005).
188 Moreover, the Data Monkey Server (www.datamonkey.org) was used to test for positive
189 selection in reptilian hedgehog genes as a supplemental method. Finally, hedgehog genes were
190 analyzed through three distinct methods in the HyPhy package of the Data Monkey Server,
191 including the fixed-effect likelihood (FEL), random effect likelihood (REL), and single
192 likelihood ancestor counting (SLAC) methods, which were used to detect positive selection
193 (Pond & Frost 2005). Default settings for significance levels of p = 0.1 for SLAC and FEL, and
194 Bayes factor > 50 for REL, were used for screening positive selection sites. Sites estimated to be
195 under positive selection were determined by at least two of the four methods

196 **Results**

197 **Hedgehog gene sequences**

198 The hedgehog genes of *C. porosus, C. abingdonii, C. viridis, C. mccordi, G. agassizii, L.*
199 *bilineata, L. viridis, M. terrapin, P. guttatus, P. picta, P. megacephalum, P. flavoviridis, S.*
200 *merianae, O. hannah, C. pyrrhus, H. cyanocinctus, H. hardwickii, T. baileyi, O. gracilis*, and *V.*
201 *berus* have not been previously reported. No intact *Ihh* gene or intact *Dhh* gene was identified in
202 Crocodylia. After all partial genes were removed, 33 *Shh*, 20 *Ihh*, and 25 *Dhh* intact genes were
203 obtained (Table S2). Additionally, six avian and five vertebrate sequences of each hedgehog
204 gene were obtained with high quality. The percent identity matrix ranged from 67.3%–100.0%
205 for the *Shh* genes, 77.9%–99.8% for the *Ihh* genes, and 70.6%–99.8% for the *Dhh* genes (Table S3).
206 The similarity between the newly obtained sequences and the previously obtained query
207 sequences ranged from 67.3%–100.0%, suggesting that the newly identified sequences of the
208 three hedgehog genes were credible.

209 **Phylogenetic analysis**

210 A total of 44 *Shh*, 36 *Dhh*, and 31 *Ihh* genes were obtained for building the phylogenetic tree.
211 The ML tree of the 111 hedgehog genes revealed that the 44 *Shh* genes were divided into 6
212 primary clades, Serpentes, Sauria, Crocodylia, Testudines, Aves, and Mammalia (Figure 1),
213 while the *Dhh* genes were divided into four major clades, Squamata, Testudines, Aves, and
214 Mammalia (Figure S1), and the *Ihh* genes were divided into four different clades, Serpentes,
215 Sauria, Aves, and Mammalia (Figure S2). The phylogenetic tree of reptiles, birds, and other
216 vertebrates downloaded from Timetree divided reptiles into four clades according to their
217 catalogs. Then, the hedgehog gene tree was compared to the species trees, revealing that the
218 phylogenetic proximity of the *Shh, Dhh*, and *Ihh* genes tended to be consistent with the
219 traditional taxonomic group.

220 **Identification of selection pressure for *Shh, Dhh,* and *Ihh***

221 The site model analysis did not detect positive selection in the three hedgehog genes (Table S4).
222 The best nucleotide substitution bias models of the three hedgehog genes were then executed in
223 Data Monkey Server with three methods. Again, no positive selection site was found by the FEL
224 and SLAC methods with a significance level of 0.1, which supports the M8 model findings.
225 However, the REL method did identify six sites in the *Dhh* gene, and the SLAC method detected
226 one site, but neither of these methods could confirm identical sites.
227 The branch model was next used to evaluate the evolutionary rates of each specific branch; the
228 targeted branches were assigned as foreground and the others were set as background branches.

229   We set different orders as independent lineages, according to the topology of reptiles. The branch
230   model analysis revealed that *Dhh and Ihh* had significantly different ω values among the diverse
231   lineages of reptiles, indicating heterogeneous selective pressures on different lineages, while *Shh*
232   had no obvious difference among lineages (Table 1). The branch-site model analysis found
233   positive selection sites for the three hedgehog genes, but none of the lineages of *Dhh* and *Ihh*
234   genes reached the significant level. Meanwhile, the analysis revealed that the *Shh* gene of
235   Testudines underwent positive selection and the sites were significantly different ($p = 0.018$)
236   (Table 2). Furthermore, CmC detected that the partitions of *Dhh* and *Ihh* genes were markedly
237   better suited relative to the M2a_rel model ($p < 0.01$), which supports the findings of different ω
238   rates among the partition of limbs. Comparatively, the LRTs between these two models for all
239   *Shh* gene clades were not significant (p = 0.074, Table 3).

Discussion

241   Reptiles are a group of ancient vertebrates that embody a momentous position in the long and
242   complex evolutionary history of vertebrates. As is widely known, reptiles cannot maintain a
243   constant body temperature and experience seasonal shifts in behavior that are correlated with the
244   ambient temperature (Zimmerman et al. 2010). There is also a complex relationship between
245   ontogenesis and animal body temperature. In this study, the molecular evolution of the *Shh*, *Dhh*,
246   and *Ihh* genes in reptiles was investigated. Collectively, three hedgehog gene sequences from 20
247   reptilian genomes were retrieved, although there are likely to be more. These genes were
248   identified in *C. porosus, C. abingdonii, C. viridis, C. mccordi, G. agassizii, L. bilineata, L.*
249   *viridis, M. terrapin, P. guttatus, P. picta, P. megacephalum, P. flavoviridis, S. merianae, O.*
250   *hannah, C. pyrrhus, H. cyanocinctus, H. hardwickii, T. baileyi, O. gracilis*, and *V. berus* for the
251   first time. Combined with formerly predicted hedgehog genes from other reptiles, birds, and
252   mammals, these data have provided a new perspective on the reptilian intercellular signaling
253   pathway at the molecular level. Although the genome sequencing data have been growing and its
254   quality is getting better, it is unavoidable that some genes will not be successfully identified due
255   to the complexity of the gene sequence. Unfortunately, we failed to identify intact *Dhh* in
256   Crocodilian and intact *Ihh* sequence in Crocodilian and Testudines. We believe that some of the
257   genes may have been lost but this hypothesis remains to be established. Moreover, the homology
258   analysis revealed that the similarity between the reptilian *Shh*, *Dhh*, and *Ihh* gene sequences
259   ranged from 67.3%–100.0%, which indicates that the sequences were species-specific among the
260   subclasses of reptiles.
261   The phylogenetic relationship among the *Shh*, *Dhh*, and *Ihh* genes was also explored, and the
262   results revealed that these genes could be divided into six, four, and four clades, respectively. No
263   gene duplication was detected for these genes. When comparing the hedgehog gene trees with
264   the species tree, the topology of the hedgehog genes tended to be consistent with the traditional
265   taxonomic group. Thus, we speculate that the hedgehog genes were highly conserved due to their
266   critical importance in development and that the driving force behind the evolution of these genes
267   may contribute to the differentiation of reptilian species. Additionally, the phylogenetic analysis
268   of the three hedgehog genes' coding sequences found an overall (*Dhh*, (*Shh*, *Ihh*)) topology
269   (Figure 2) that is consistent with the evolutionary history of the three hedgehog genes, as *Shh*
270   and *Ihh* is considered to be more closely related to each other than to the *Dhh* (Markku & Jussi
271   2008).
272   Site models performed in PAML found that the *Shh*, *Dhh*, and *Ihh* sequences were under diverse
273   selective pressures, with no positively selected residues. Moreover, the data did not find evidence
274   for positive selection, despite the use of three methods implemented in the Data Monkey server

275   to verify positive selection. As expected, no positively selected site was detected in any of the
276   three hedgehog genes in reptiles by at least two methods. Although these methods failed to
277   recognize residues under positive selection, several residues under negative selection were found.
278   These results are consistent with the findings of previous studies exploring the adaptive
279   evolution of the hedgehog gene family in vertebrates (Pereira et al. 2014). The purifying
280   selection plays an essential role in the long term stability of biological structures by deleting
281   harmful mutations (Karlsson et al. 2014). Thus, we speculate that the highly conserved
282   characteristic of the hedgehog genes may be due to their crucial role in controlling multiple
283   different developmental processes.
284   Squamate reptiles, such as snakes and snake-like lizards, are the most extreme reptilian species
285   in terms of shape, given their elongated and limbless bodies (Gans 1975). Thus, for the branch
286   model, another branch was added in addition to the traditional classification, which included the
287   Serpentes species and limbless *O. gracilis*. The LRT between the null and alternate likelihood
288   models indicated that all clades of *Dhh* and *Ihh* genes fit the two-ratio model better, implying
289   that all lineages of these two genes may have evolved at different rates. *Shh* plays a crucial role
290   in regulating vertebrate organogenesis, such as the development of digits on limbs. A different
291   evolutionary rate of *Shh* was not found, as expected, which is likely because the mechanism of
292   limb reduction in reptiles is regulated by other related genes, such as *HoxA13* (Singarete et al.
293   2015). The overall branches of *Shh* rejected the two-ratio model, suggesting that the reptilian *Shh*
294   gene evolved at a relatively stable rate. However, six positive selection sites were detected in
295   Testudines using the branch-site model, suggesting that this *Shh* gene underwent positive
296   selection in Testudines. *Shh* expression in snake embryos indicates that almost all species in the
297   palatal odontogenic group have an intact domain of *Shh* expression (Vonk et al. 2008). Previous
298   research indicated that in turtle embryos, the impairment of *Shh* signaling at an early-stage
299   arrests odontoblast development, thus contributing to the tooth loss of turtles (Tokita et al. 2013).
300   We speculate that the diversity of tooth patterns in turtles might affect the selective pressures of
301   *Shh* in Testudines.
302   *Dhh* plays a role in spermatogenesis. It was previously found that *Dhh*-null male mice have small
303   gonads, do not produce sperm, and their gender is reversed, as these males develop as phenotypic
304   females (Bitgood et al. 1996). In contrast, *Ihh* is one of the key genes associated with body
305   development and is required for embryonic bone formation (Shi et al. 2015). Therefore, the
306   different evolutionary rates of the *Dhh* and *Ihh* genes between limb and limbless clades may
307   indicate different body development patterns. Moreover, the proteins encoded by *Shh*, *Dhh*, and
308   *Ihh* are believed to have similar physiological effects, and that their differences during
309   development may arise from varying patterns of expression (Mcmahon et al. 2003). Therefore,
310   the evolutionary differences in the three hedgehog genes among reptiles may be due to their
311   different roles in the hedgehog signaling pathway.

312   **Conclusions**
313   In summary, this study thoroughly investigated the structure and evolution of three hedgehog
314   genes in reptiles for the first time. Characterization of these hedgehog members in reptiles
315   indicates that the gene family was conserved in vertebrates. The selection pressure analysis of
316   hedgehog genes indicated that three members had evolved, along with their vital function in
317   body development. The data revealed that the three hedgehog genes underwent significant
318   purifying selection. Interestingly, it is likely that the degeneration of snakes' limbs was a result
319   of the regulation of multiple genes, rather than *Shh* alone. Additionally, *Dhh* and *Ihh* genes had
320   faster evolutionary rates within different clades. Together, the findings of this study have

321 provided extensive information regarding the molecular evolutionary history of the hedgehog
322 gene family in reptiles and new insights for the study on reptilian adaptive evolution. Inferring
323 gene function from public genome databases is not conclusive. Therefore, further functional
324 experimentation is required to confirm our observations substantially.

Acknowledgments

326 This research was supported by the Special Fund for Forest Scientific Research in the Public
327 Welfare (No. 201404420) and the National Natural Science Foundation of China (Nos.
328 31872242, 31672313, and 31372220). We would like to thank all those involved in the
329 implementation of this study for their assistance, guidance, and support.

References

331 Avaron F, Hoffman L, Guay D, and Akimenko MA. 2006. Characterization of two new zebrafish
332 members of the hedgehog family: Atypical expression of a zebrafish indian hedgehog
333 gene in skeletal elements of both endochondral and dermal origins. *Developmental*
334 *Dynamics* 235:478-489.
335 Beachy PA, Karhadkar SS, and Berman DM. 2004. Tissue repair and stem cell renewal in
336 carcinogenesis. *Nature* 432:324-331.
337 Bitgood MJ, Shen L, and McMahon AP. 1996. Sertoli cell signaling by Desert hedgehog regulates
338 the male germline. *Curr Biol* 6:298-304.
339 Echelard Y, Epstein DJ, St-Jacques B, Shen L, Mohler J, Mcmahon JA, and Mcmahon AP. 1993.
340 Sonic hedgehog, a member of a family of putative signaling molecules, is implicated in
341 the regulation of CNS polarity. *Cell* 75:1417.
342 Gans C. 1975. Tetrapod Limblessness: Evolution and Functional Corollaries. *American Zoologist*
343 15:455-467.
344 Gilbert C, Meik JM, Dashevsky D, Card DC, Castoe TA, and Schaack S. 2014. Endogenous
345 hepadnaviruses, bornaviruses and circoviruses in snakes. *Proc Biol Sci* 281:20141122.
346 Green RE, Braun EL, Armstrong J, Earl D, Nguyen N, Hickey G, Vandewege MW, St John JA,
347 Capella-Gutierrez S, Castoe TA, Kern C, Fujita MK, Opazo JC, Jurka J, Kojima KK,
348 Caballero J, Hubley RM, Smit AF, Platt RN, Lavoie CA, Ramakodi MP, Finger JW, Jr., Suh
349 A, Isberg SR, Miles L, Chong AY, Jaratlerdsiri W, Gongora J, Moran C, Iriarte A,
350 McCormack J, Burgess SC, Edwards SV, Lyons E, Williams C, Breen M, Howard JT,
351 Gresham CR, Peterson DG, Schmitz J, Pollock DD, Haussler D, Triplett EW, Zhang G, Irie
352 N, Jarvis ED, Brochu CA, Schmidt CJ, McCarthy FM, Faircloth BC, Hoffmann FG, Glenn
353 TC, Gabaldon T, Paten B, and Ray DA. 2014. Three crocodilian genomes reveal ancestral
354 patterns of evolution among archosaurs. *Science* 346:1254449.
355 Gunbin KV, Afonnikov DA, and Kolchanov NA. 2007. The evolution of the Hh-signaling pathway
356 genes: a computer-assisted study. *In Silico Biol* 7:333-354.
357 Hara Y, Takeuchi M, Kageyama Y, Tatsumi K, Hibi M, Kiyonari H, and Kuraku S. 2018.
358 Madagascar ground gecko genome analysis characterizes asymmetric fates of duplicated
359 genes. *BMC Biol* 16:40.
360 Hooper JE, and Scott MP. 2005. Communicating with Hedgehogs. *Nature Reviews Molecular Cell*
361 *Biology* 6:306.
362 Humphrey Hung-Chang Y, Wendy W, and Blanche C. 2002. Desert Hedgehog/Patched 1
363 signaling specifies fetal Leydig cell fate in testis organogenesis. *Genes Dev* 16:1433-
364 1440.
365 Ingham PW, and McMahon AP. 2001. Hedgehog signaling in animal development: paradigms
366 and principles. *Genes Dev* 15:3059-3087.
367 Ingham PW, Nakano Y, and Seger C. 2011. Mechanisms and functions of Hedgehog signalling
368 across the metazoa. *Nature Reviews Genetics* 12:393-406.

369 Karlsson EK, Kwiatkowski DP, and Sabeti PC. 2014. Natural selection and infectious disease in
370   human populations. *Nat Rev Genet* 15:379-393.
371 Kimura M. 1979. The neutral theory of molecular evolution. *Scientific American* 241:98-100.
372 Krauss S, Concordet JP, and Ingham PW. 1993. A functionally conserved homolog of the
373   Drosophila segment polarity gene hh is expressed in tissues with polarizing activity in
374   zebrafish embryos. *Cell* 75:1431-1444.
375 Kumar S, ., Balczarek KA, and Lai ZC. 1996. Evolution of the hedgehog gene family. *Genetics*
376   142:965.
377 Li JT, Gao YD, Xie L, Deng C, Shi P, Guan ML, Huang S, Ren JL, Wu DD, Ding L, Huang ZY, Nie
378   H, Humphreys DP, Hillis DM, Wang WZ, and Zhang YP. 2018. Comparative genomic
379   investigation of high-elevation adaptation in ectothermic snakes. *Proceedings of the*
380   *National Academy of Sciences of the United States of America* 115:8406-8411.
381 Litingtung Y, ., Lei L, ., Westphal H, ., and Chiang C, . 1998. Sonic hedgehog is essential to
382   foregut development. *Nature Genetics* 20:58-61.
383 Markku V, and Jussi T. 2008. Hedgehog: functions and mechanisms. *Genes Dev* 22:2454-2472.
384 Matus DQ, Magie CR, Pang K, Martindale MQ, and Thomsen GH. 2008. The Hedgehog gene
385   family of the cnidarian, Nematostella vectensis, and implications for understanding
386   metazoan Hedgehog pathway evolution. *Developmental Biology* 313:501-518.
387 Mcmahon AP, Ingham PW, and Tabin CJ. 2003. Developmental roles and clinical significance of
388   Hedgehog signaling. *Current Topics in Developmental Biology* 53:1-114.
389 Mirsky R, Parmantier E, Mcmahon AP, and Jessen KR. 2010. Schwann cell-derived desert
390   hedgehog signals nerve sheath formation. *Annals of the New York Academy of Sciences*
391   883:196-202.
392 Monie TP, Gay NJ, and Gangloff M. 2008. Bioinformatic analysis of Toll-like receptor sequences
393   and structures. *Methods in Molecular Biology* 517:69-79.
394 Motoyama J, ., Liu J, ., Mo R, ., Ding Q, ., Post M, ., and Hui CC. 1998. Essential function of Gli2
395   and Gli3 in the formation of lung, trachea and oesophagus. *Nature Genetics* 20:54-57.
396 Nüsslein-Volhard C, and Wieschaus E. 1980. Mutations affecting segment number and polarity
397   in Drosophila. *Nature* 287:795-801.
398 Nagase T, Nagase M, Machida M, and Yamagishi M. 2010. Hedgehog signaling: a biophysical or
399   biomechanical modulator in embryonic development? *Annals of the New York Academy*
400   *of Sciences* 1101:412-438.
401 Nieuwenhuis E, ., and Hui C-C. 2005. Hedgehog signaling and congenital malformations. *Clinical*
402   *Genetics* 67:193-208.
403 Pasquesi GIM, Adams RH, and Card DC. 2018. Squamate reptiles challenge paradigms of
404   genomic repeat element evolution set by birds and mammals. *Nat Commun* 9:2774.
405 Pereira J, Johnson WE, O'Brien SJ, Jarvis ED, Zhang GJ, Gilbert MTP, Vasconcelos V, and
406   Antunes A. 2014. Evolutionary Genomics and Adaptive Evolution of the Hedgehog Gene
407   Family (Shh, Ihh and Dhh) in Vertebrates. *Plos One* 9:35.
408 Pond SLK, and Frost SDW. 2005. Datamonkey: rapid detection of selective pressure on
409   individual sites of codon alignments. *Bioinformatics* 21:2531-2533.
410 Roscito JG, Sameith K, Pippel M, Francoijs KJ, Winkler S, Dahl A, Papoutsoglou G, Myers G, and
411   Hiller M. 2018. The genome of the tegu lizard Salvator merianae: combining Illumina,
412   PacBio, and optical mapping data to generate a highly contiguous assembly. *Gigascience*
413   7.
414 Shi Y, Chen J, Karner CM, and Long F. 2015. Hedgehog signaling activates a positive feedback
415   mechanism involving insulin-like growth factors to induce osteoblast differentiation. *Proc*
416   *Natl Acad Sci U S A* 112:4678-4683.

417   Shibata H, Chijiwa T, Oda-Ueda N, Nakamura H, Yamaguchi K, Hattori S, Matsubara K, Matsuda
418   Y, Yamashita A, Isomoto A, Mori K, Tashiro K, Kuhara S, Yamasaki S, Fujie M, Goto H,
419   Koyanagi R, Takeuchi T, Fukumaki Y, Ohno M, Shoguchi E, Hisata K, and Satoh N. 2018.
420   The habu genome reveals accelerated evolution of venom protein genes. *Sci Rep*
421   8:11300.
422   Shimeld SM. 1999. The evolution of the hedgehog gene family in chordates: insights from
423   amphioxus hedgehog. *Dev Genes Evol* 209:40-47.
424   Singarete ME, Grizante MB, Milograna SR, Nery MF, Kin K, Wagner GP, and Kohlsdorf T. 2015.
425   Molecular evolution of HoxA13 and the multiple origins of limbless morphologies in
426   amphibians and reptiles. *Genet Mol Biol* 38:255-262.
427   Song B, Cheng S, Sun Y, Zhong X, Jin J, Guan R, Murphy RW, Che J, Zhang Y, and Liu X. 2015.
428   A genome draft of the legless anguid lizard, Ophisaurus gracilis. *Gigascience* 4:17.
429   Steve D, Anderson JR, Vallender EJ, Gilbert SL, Li Z, Chemnick LG, Ryder OA, Weimin L, and
430   Lahn BT. 2006. Sonic Hedgehog, a key development gene, experienced intensified
431   molecular evolution in primates. *Human Molecular Genetics* 15:2031-2037.
432   Szczepny A, Hime GR, and Loveland KL. 2010. Expression of hedgehog signalling components in
433   adult mouse testis. *Developmental Dynamics An Official Publication of the American*
434   *Association of Anatomists* 235:3063-3070.
435   Tokita M, Chaeychomsri W, and Siruntawineti J. 2013. DEVELOPMENTAL BASIS OF
436   TOOTHLESSNESS IN TURTLES: INSIGHT INTO CONVERGENT EVOLUTION OF
437   VERTEBRATE MORPHOLOGY. *Evolution* 67:260-273.
438   Tollis M, DeNardo DF, Cornelius JA, Dolby GA, Edwards T, Henen BT, Karl AE, Murphy RW, and
439   Kusumi K. 2017. The Agassiz's desert tortoise genome provides a resource for the
440   conservation of a threatened species. 12:e0177708.
441   Ullate-Agote A, Milinkovitch MC, and Tzika AC. 2014. The genome sequence of the corn snake
442   (Pantherophis guttatus), a valuable resource for EvoDevo studies in squamates. *Int J*
443   *Dev Biol* 58:881-888.
444   Varjosalo M, and Taipale J. 2008. Hedgehog: functions and mechanisms. *Genes Dev* 22:2454-
445   2472.
446   Vonk FJ, Casewell NR, Henkel CV, Heimberg AM, Jansen HJ, McCleary RJ, Kerkkamp HM, Vos
447   RA, Guerreiro I, Calvete JJ, Wuster W, Woods AE, Logan JM, Harrison RA, Castoe TA, de
448   Koning AP, Pollock DD, Yandell M, Calderon D, Renjifo C, Currier RB, Salgado D, Pla D,
449   Sanz L, Hyder AS, Ribeiro JM, Arntzen JW, van den Thillart GE, Boetzer M, Pirovano W,
450   Dirks RP, Spaink HP, Duboule D, McGlinn E, Kini RM, and Richardson MK. 2013. The king
451   cobra genome reveals dynamic gene evolution and adaptation in the snake venom
452   system. *Proc Natl Acad Sci U S A* 110:20651-20656.
453   Yang Z, ., Nielsen R, ., Goldman N, ., and Pedersen AM. 2000. Codon-substitution models for
454   heterogeneous selection pressure at amino acid sites. *Genetics* 155:431-449.
455   Zacharias WJ, Li X, Madison BB, Kretovich K, Kao JY, Merchant JL, and Gumucio DL. 2010.
456   Hedgehog Is an Anti-Inflammatory Epithelial Signal for the Intestinal Lamina Propria.
457   *Gastroenterology* 138:2368-2377.e2364.
458   Ziheng Y, Wong WSW, and Rasmus N. 2005. Bayes empirical bayes inference of amino acid
459   sites under positive selection. *Molecular Biology & Evolution* 22:1107-1118.
460   Zimmerman LM, Vogel LA, and Bowden RM. 2010. Understanding the vertebrate immune
461   system: insights from the reptilian perspective. *Journal of Experimental Biology* 213:661-
462   671.

463

Figure 1

Phylogenetic tree based on CDS sequences of Shh genes.

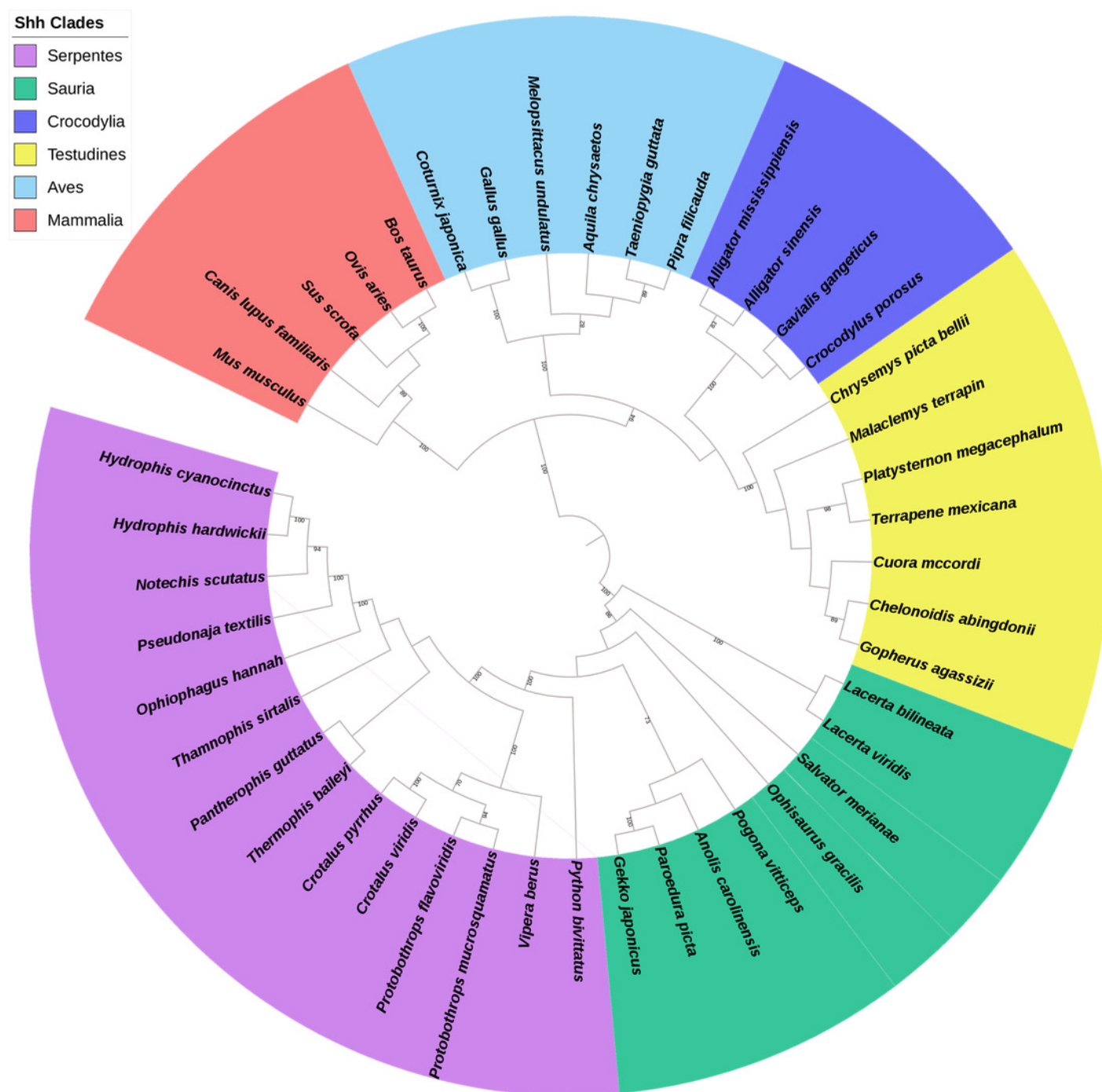
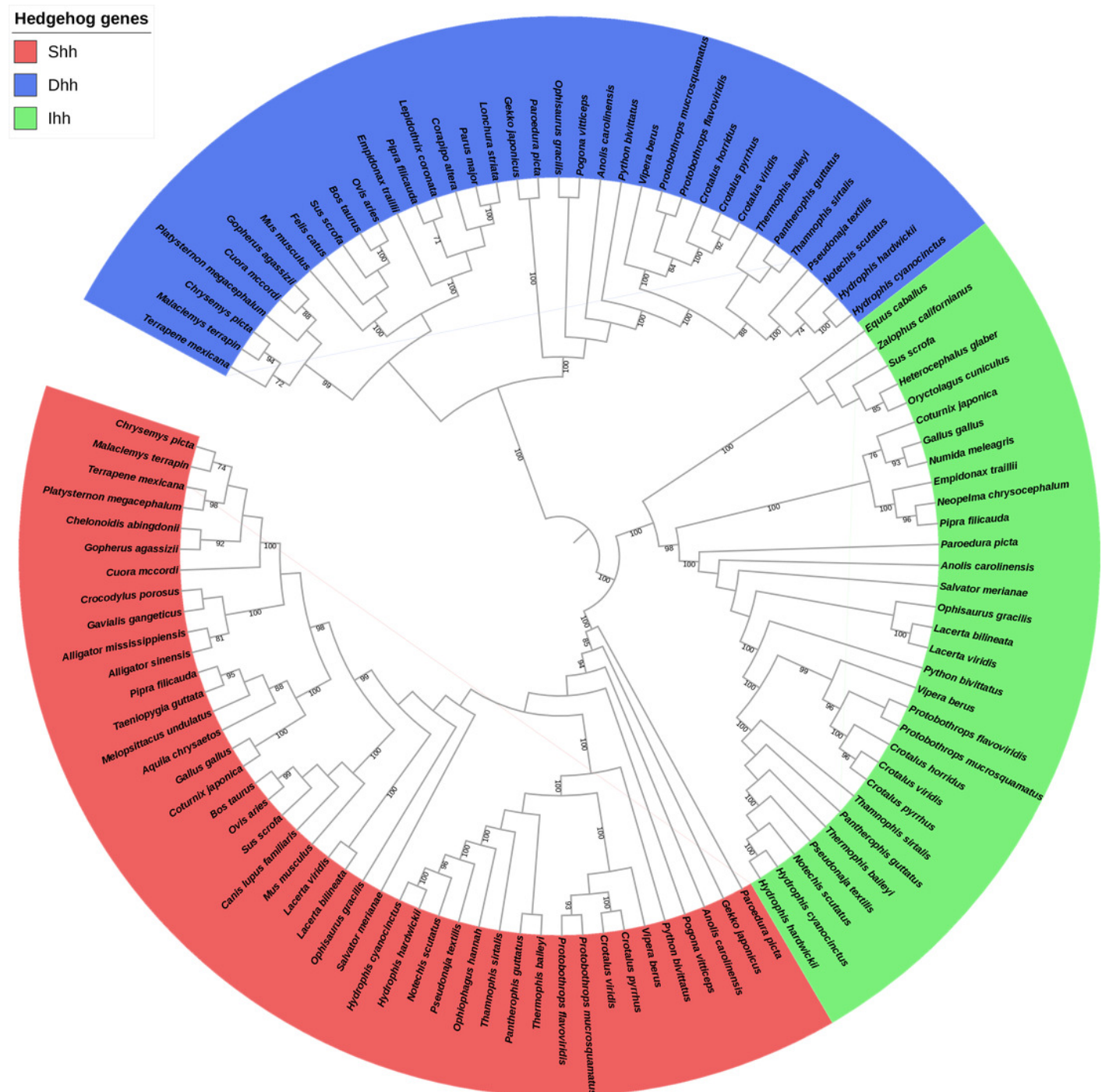



Figure 2

Phylogenetic tree based on CDS sequences of three hedgehog genes.

Test for positive selection in divergent clades of *Shh*, *Dhh*, and *Ihh* genes with branch model.

Gene	Model Compared	np	LnL	LRT P-value	ω for branch
Shh	M0	88	-11647.017474		ω1=0.07001
	M2(clade of Crocodylia)	89	-11646.585976	0.352900523	ω1=0.05660; ω2= 0.07081
	M2(clade of Testudines)	89	-11646.736015	0.453094143	ω1=0.09345; ω2= 0.06965
	M2(clade of Sauria)	89	-11645.645632	0.097638458	ω1=0.06189; ω2= 0.07429
	M2(clade of Serpentes)	89	-11646.656047	0.395226992	ω1=0.07728; ω2=0.06869
	M2(Limbless Squamates)	89	-11646.431808	0.279134287	ω1=0.07844; ω2= 0.06817
Dhh	M0	72	-9913.616846		ω1= 0.07078
	M2(clade of Serpentes)	73	-9925.890310	0.000000725[a]	ω1=0.10761; ω2= 0.06381
	M2(clade of Sauria)	73	-9924.330432	0.000003675[a]	ω1=0.10124; ω2= 0.06045
	M2(Limbless Squamates)	73	-9926.514254	0.00000038[a]	ω1=0.10077; ω2= 0.06309
	M2(Testudines)	73	-9928.261444	0.000000062[a]	ω1= 0.11848 ω2= 0.06720
Ihh	M0	62	-8482.453248		ω1= 0.08385
	M2(clade of Serpentes)	63	-8475.468795	0.00018586[a]	ω1=0.13019; ω2= 0.07412
	M2(Sauria)	63	-8471.679842	0.000003453[a]	ω1=0.13072; ω2= 0.06886
	M2(Limbless Squamates)	63	-8470.418614	0.000000929[a]	ω1=0.13521; ω2= 0.06857

1   Note: [a] Means significant level. Limbless Squamates mean Serpentes and *Ophisaurus gracilis*, representing

2   the limbless clades.

Test for positive selection in divergent clades of *Shh*, *Dhh,* and *Ihh* genes with branch-site model.

off
on
off

Gene	Lineage	Models compared	np	LnL	p-value	ω values	Positively Selected Sites(BEB Analysis)
Shh	Serpentes	Model A	91	-11462.09307	1	ω0=0.03960;ω1=1.000;ω2=1.000	85 I 0.970* 86 F 0.524 87 K 0.980* 95 D 0.907 96 R 0.948 181 I 0.938 186 K 0.842 187 A 0.931 371 L 0.567
		Model A null	90	-11462.09307		ω0=0.03960;ω1=1.000 ;ω2=1.000	
	Sauria	Model A	91	-11460.7336	1	ω0=0.04246; ω1=1.000 ;ω2=1.000	4 L 0.990** 9 L 0.989* 20 A 0.706 136 E 0.531 204 W 0.882 232 Q 0.998** 240 A 0.600 244 R 0.523 301 Q 0.936 308 P 0.620 310 A 0.630 361 A 0.959*
		Model A null	90	-11460.7336		ω0=0.04246 ω1=1.000 ;ω2=1.000	
	Crocodilian	Model A	91	-11479.21629	0.511596661	ω0=0.04535;ω1=1.000;ω2=3.93515	206 N 0.847
		Model A null	90	-11479.43171		ω0= 0.04546; ω1=1.000 ;ω2=1.000	
	Testudines	Model A	91	-11475.91887	0.017704046[a]	ω0=0.04593;ω1=1.000;ω2=11.54603	2 L 0.641 14 G 0.576 277 H 0.974* 313 S 0.601 358 F 0.602 410 S 0.616
		Model A null	90	-11478.73149		ω0=0.04569;ω1=1.000 ;ω2=1.000	
	Limbless Squamates	Model A	91	-11463.08743	1	ω0=0.03868; ω1=1.000 ;ω2=1.000	85 I 0.833 87 K 0.709 96 R 0.682 186 K 0.658 187 A 0.925 244 R 0.895 258 Q 0.795 261 Q 0.524 295 H 0.828 360 I 0.780 370 L 0.995**
		Model A null	90	-11463.08743		ω0=0.03868; ω1=1.000 ;ω2=1.000	
Dhh	Serpentes	Model A	75	-9889.30068	1	ω0=0.05652; ω1=1.000 ;ω2=1.000	67 V 0.965* 200 R 0.741 224 Q 0.960* 240 K 0.540 243 L 0.958* 270 R 0.987* 275 M 0.999** 286 C 0.901 346 Y 0.902
		Model A null	74	-9889.30068		ω0=0.05652; ω1=1.000 ;ω2=1.000	
	Sauria	Model A	75	-9871.673179	1	ω0=0.05115; ω1=1.000 ;ω2=1.000	1 L 0.991** 3 L 0.773 7 C 0.916 10 P 0.822 27 S 0.928 33 P 0.508 62 S 0.858 63 E 1.000** 64 R 0.942 186 R 0.835 196 T 0.992** 201 S 0.625 204 K 0.932 206 S 0.934 212 H 0.521 226 A 0.908 253 V 0.951* 280 V 0.984* 309 Q 0.744 313 T 0.982* 364 A 0.802 369 V 0.658 374 Q 0.963* 382 D 0.692

Gene	Clade	Model	np	lnL		ω values		Positively selected sites
		Model A null	74	-9871.673179		ω0=0.05115; ;ω2=1.000	ω1=1.000	
	Limbless Squamates	Model A	75	-9885.126495	1	ω0=0.05492; ;ω2=1.000	ω1=1.000	2 T 0.943 67 V 0.926 112 V 0.805 200 R 0.982* 209 K 0.889 224 Q 0.924 226 A 0.593 243 L 0.896 270 R 0.970* 275 M 0.999** 286 C 0.860 313 T 0.578 346 Y 0.819 369 V 0.531
		Model A null	74	-9885.126495		ω0=0.05492; ;ω2=1.000	ω1=1.000	
	Testudines	Model A	75	-9897.057753	1	ω0=0.05873; ;ω2=1.000	ω1=1.000	21 V 0.917 25 Q 0.569 35 Q 0.915 57 K 0.503 91 T 0.923 97 R 0.520 242 T 0.884 276 D 0.993** 296 D 0.693 356 G 0.918
		Model A null	74	-9897.057753		ω0=0.05873; ;ω2=1.000	ω1=1.000	
Ihh	Serpentes	Model A	65	-8316.257473	1	ω0=0.04207; ;ω2=1.000	ω1=1.000	13 L 0.680 182 T 0.528 206 A 0.627 232 V 0.996** 261 T 0.638 331 G 0.794 338 H 0.975* 342 R 0.707 364 F 0.581
		Model A null	64	-8316.257473		ω0=0.04207; ;ω2=1.000	ω1=1.000	
	Sauria	Model A	65	-8303.182547	1	ω0=0.03904; ;ω2=1.000	ω1=1.000	2 E 0.983* 3 V 0.590 6 C 0.736 9 S 0.607 11 R 0.639 217 A 0.702 223 V 0.598 226 R 0.987* 235 Q 0.577 242 A 0.554 257 A 0.883 259 M 0.795 262 I 0.840 268 Q 0.989* 271 Q 0.977* 281 L 0.858 282 Q 0.920 283 P 0.837 307 L 0.741 336 Y 0.865 337 H 0.616344 W 0.991**
		Model A null	64	-8303.182547		ω0=0.03904; ;ω2=1.000	ω1=1.000	
	Limbless Squamates	Model A	65	-8310.549535	1	ω0= 0.03920; ;ω2=1.000	ω1=1.000	6 C 0.679 9 S 0.541 11 R 0.544 13 L 0.982* 206 A 0.542 223 V 0.716 227 T 0.629 232 V 0.994** 261 T 0.509 283 P 0.629 331 G 0.728 338 H 0.954* 342 R 0.506
		Model A null	64	-8310.549535		ω0= 0.03920 ; ;ω2=1.000	ω1=1.000	

1   Note: ᵃ Means significant level. Limbless Squatma mean Serpentes and *Ophisaurus gracilis*, representing the limbless clades.

Test for positive selection in divergent clades of *Shh, Dhh,* and *Ihh* genes with CmC model.

Gene	Model	np	LnL	k	Site class 0 (all branched)	Site class 1 (all branches)	Site class 2 (backgroud branches and different clade vary)	df	p-value
Shh	M2a_rel (null)	91	-11291.52935	2.53263	p0=0.73101 ω0=0.01124	p1=0.01911 ω1=1.00000	p2=0.24989 ω2=0.23489		
	CmC	93	-11288.927085	2.53001	p0=0.73326 ω0=0.01159	p1=0.01964 ω1=1.00000	p2=0.24711 ωClade 0 (background)= 0.34939 ωClade 1=0.24768 ωClade 2 =0.21839	2	0.074106651
Dhh	M2a_rel (null)	75	-9740.01743	3.64268	p0=0.62809 ω0= 0.00910	p1=0.00000 ω1=1.00000	p2=0.37191 ω2=0.19189		
	CmC	77	-9719.900358	3.71887	p0=0.62023 ω0=0.00795	p1=0.00000 ω1=1.00000	p2=0.37977 ωClade 0 (background)= 0.13580 ωClade 1=0.25351 ωClade 2 =0.17494	2	0.000000002 [a]
Ihh	M2a_rel (null)	65	-8252.834363	3.53731	p0=0.71133 ω0=0.01260	p1=0.03695 ω1=1.00000	p2=0.25172 ω2=0.25547		
	CmC	67	-8247.531941	3.57165	p0=0.70435 ω0=0.01177	p1=0.03482 ω1=1.00000	p2=0.26083 ωClade 0 (background)= 0.24841 ωClade 1=0.37843 ωClade 2 =0.20881	2	0.004979628 [a]

1   Note: Clade 1 means limbless Squamates clade, clade 2 means limbs clade.