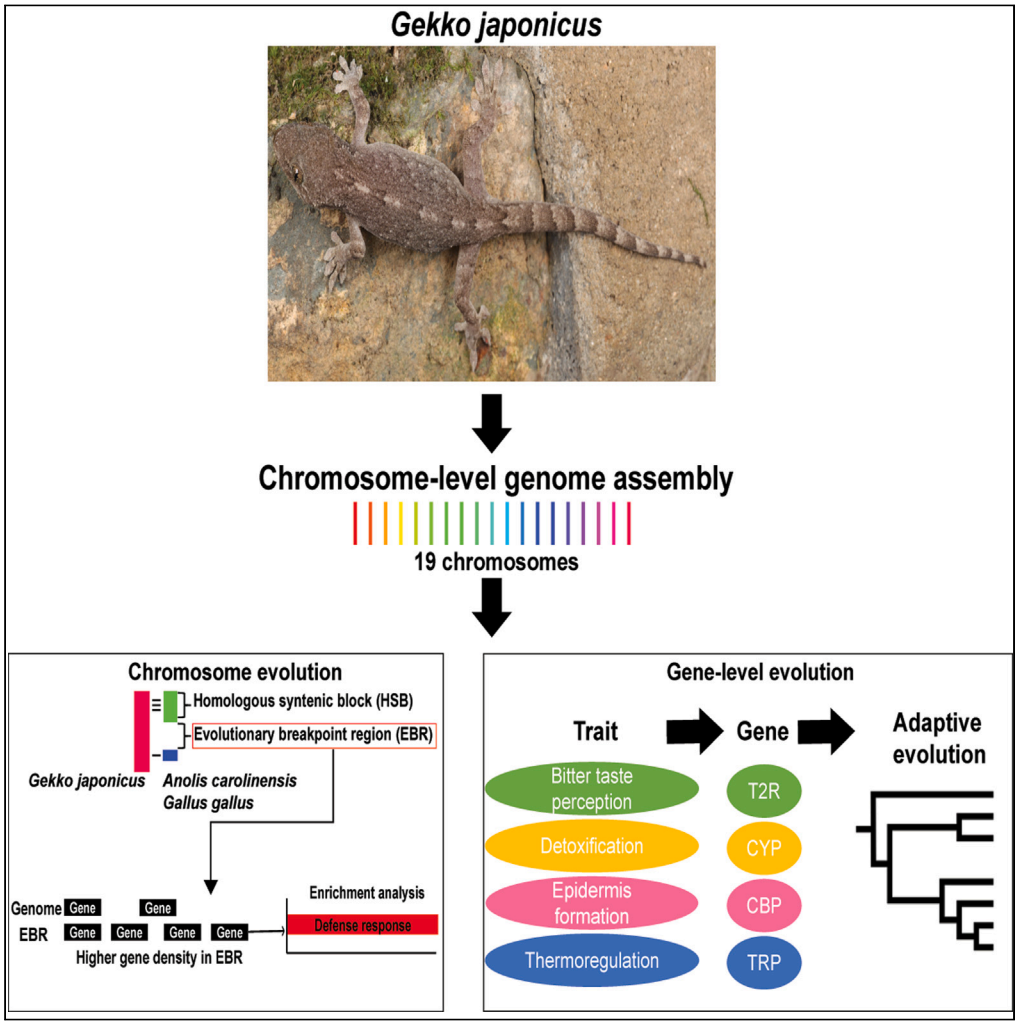
iScience

CellPress
OPEN ACCESS

Article

Insights into the adaptive evolution of chromosome and essential traits through chromosome-level genome assembly of *Gekko japonicus*



Yinwei Wang, Youxia Yue, Chao Li, ..., Kaiya Zhou, Jie Yan, Peng Li

yanjie@njnu.edu.cn (J.Y.)
lipeng@njnu.edu.cn (P.L.)

Highlights

The chromosome-level genome (2.54 Gb) of *Gekko japonicus* was assembled

Evolutionary breakpoint regions are enriched with defense response genes

Positively selected genes mainly enriched in immune, sensory, and nervous pathways

Special gene families are associated with adaptive evolution of essential traits

Wang et al., iScience *27*, 108445
January 19, 2024
https://doi.org/10.1016/j.isci.2023.108445

iScience

CellPress
OPEN ACCESS

Article

Insights into the adaptive evolution of chromosome and essential traits through chromosome-level genome assembly of *Gekko japonicus*

Yinwei Wang,[1,4] Youxia Yue,[1,4] Chao Li,[1] Zhiyi Chen,[1] Yao Cai,[2] Chaochao Hu,[1,3] Yanfu Qu,[1] Hong Li,[1] Kaiya Zhou,[1] Jie Yan,[1,*] and Peng Li[1,5,*]

SUMMARY

Gekko japonicus possesses flexible climbing and detoxification abilities under insectivorous habits. Still, the evolutionary mechanisms behind these traits remain unclarified. This study presents a chromosome-level _G. japonicus_ genome, revealing that its evolutionary breakpoint regions were enriched with specific repetitive elements and defense response genes. Gene families unique to _G. japonicus_ and positively selected genes are mainly enriched in immune, sensory, and nervous pathways. Expansion of bitter taste receptor type 2 primarily in insectivorous species could be associated with toxin clearance. Detox cytochrome P450 in _G. japonicus_ has undergone more birth and death processes than biosynthesis-type P450 genes. Proline, cysteine, glycine, and serine in corneous beta proteins of _G. japonicus_ might influence flexibility and setae adhesiveness. Certain thermosensitive transient receptor potential channels under relaxed purifying selection or positive selection in _G. japonicus_ might enhance adaptation to climate change. This genome assembly offers insights into the adaptive evolution of gekkotans.

INTRODUCTION

Gekko japonicus (Schlegel, 1836), a member of the Gekkonidae family that includes ~1,000 species, is widely distributed in East Asia.[1–3] Chromosome evolution plays an important role in speciation. Studies on the chromosomes of the Geckkonidae species have mainly focused on karyotype analysis, including standard Giemsa staining experiments. The karyotype of most species of the genus *Gecko* is now found to be 2n = 38.[4,5] We assembled a chromosome-level genome of *G. japonicus* based on Illumina HiSeq, Pacbio single molecule real-time (SMRT) sequencing, and high-throughput chromosome conformation capture (Hi-C) technologies to further investigate the chromosome evolution of *G. japonicus*. The chromosome numbers obtained were consistent with previous experimental results.[4] Evolutionary breakpoint regions (EBRs) are the regions between two neighboring homologous synteny blocks (HSBs) and are considered to be species- or lineage-specific chromosome evolution-related regions. EBRs typically have a higher guanine and cytosine (GC) content, repetitive elements, and gene density than the genome, although this is not always true. Genes enriched in EBRs, associated pathways, and repetitive elements are essential for chromosome evolution.[6] Therefore, in this study, we explored the role of EBRs in the adaptive evolution of *G. japonicus*.

Gekko japonicus has some interesting features of biological interest in ecological adaptability. It feeds mainly on insects, so how does it face the risk of toxins in insects? Taste receptor type 2 (T2R) is primarily involved in bitter taste perception, and the perception of bitter toxins by T2R may be essential for *G. japonicus*.[7,8] The Cytochrome P450 (CYP) gene family is also involved in detoxification and metabolism. Eleven clans were found in metazoans, including CYP clans 2, 3, 4, mitochondrial, 7, 19, 20, 26, 46, 51, and 74.[9] However, clan 74 has not been found in vertebrates.[10] CYP1, CYP2, CYP3, and CYP4 are considered biosynthesis type (B-type), while the others are detoxification type (D-type).[11] Exploring these two gene families may shed light on the adaptive evolution of the taste system and detoxification function in *G. japonicus*. The adaptive evolution of amniotes from aquatic to terrestrial ecosystems has made them diverse in epidermal morphology, such as the feathers of Aves, scales of Squamata, and shells of Testudines. The corneous beta proteins (CBPs) family has played an essential role in epidermis formation, as suggested by the evolutionary analysis of CBPs in *G. japonicus* in a previous study.[12] As a poikilothermal and nocturnal gecko, thermoregulation may play an essential role for *G. japonicus* in foraging, predator avoidance, and reproduction. Seasonal, diurnal temperature changes regulate the activity of gekkotans.[13,14] However, the molecular mechanisms of thermoregulation and related genes in gekkotans remain a mystery.

[1]Jiangsu Key Laboratory for Biodiversity and Biotechnology, College of Life Sciences, Nanjing Normal University, Nanjing, Jiangsu 210023, P.R. China
[2]School of Food Science, Nanjing Xiaozhuang University, Nanjing, Jiangsu 211171, P.R. China
[3]Analytical and Testing Center, Nanjing Normal University, Nanjing, Jiangsu 210023, P.R. China
[4]These authors contributed equally
[5]Lead contact
*Correspondence: yanjie@njnu.edu.cn (J.Y.), lipeng@njnu.edu.cn (P.L.)
https://doi.org/10.1016/j.isci.2023.108445

Table 1. Comparison of genome assembly in this study with previous version

Assembly	This study	Gekko_japonicus_V1.1
Assembly approach	WGS and Hi-C	WGS and BAC
Sequencing platform	PacBio, Illumina	Illumina
Scaffold N50 (bp)	172236841	707733
Contig N50 (bp)	2043270	29574
Total genes	20503	19548
Total sequence length (bp)	2535908678	2490274461
Total repeat size (%)	69.4	48.9
GC (%)	45.9	45.5
Scaffold number	146	191500
Contig number	2157	335470

Transient receptor potential (TRP) channels are a class of protein superfamily that are widely distributed in eukaryotes and play essential roles in various physiological activities, such as lysosomal function, inflammatory response, and thermoregulation.[15–17] TRP channels have seven subfamilies, including TRPA1, TRPV, TRPC, TRPM, TRPML (also called MCOLN), TRPP (also called PKD), and TRPN.[18–21] Research on TRP genes of model organisms has revealed that, among the TRP superfamily, TRPA1, TRPM2-5, TRPM8, and TRPV1-4 are thermosensitive.[22,23] The paucity of gene repertoire composition, phylogeny, and evolutionary analyses of TRPs among sauropsids is primarily attributed to the lack of genomes and the poor quality of assembly and annotation.

In this study, we explored the complex and fascinating evolutionary trajectory of *G. japonicus* based on the chromosome, genome, and gene levels, providing novel insights into species- and chromosome-specific evolution, and adaptive evolution of essential traits (e.g., bitter taste perception, detox and biosynthetic CYP, epidermis formation, and thermoregulation) in *G. japonicus*.

RESULTS

Sequencing, assembly, and annotation of the high-quality *G. japonicus* genome

We combined Illumina HiSeq, Pacbio SMRT sequencing, and Hi-C to assemble the chromosome-level genome of *G. japonicus*. The results obtained from Illumina HiSeq were first used in a k-mer analysis to estimate the genome size, which resulted in approximately 2.54 Gb (Figure S1). The initial genome assembly was first obtained using MECAT2[24] for the reads generated by Pacbio SMRT, followed by polishing the genome assembly results with subreads using GCpp version 1.9.0 (https://github.com/PacificBiosciences/gcpp) and error correction based on Illumina data using Pilon version 1.22.[25] Finally, Hi-C data were mounted using Juicer version 1.6.2[26] to obtain the final genome assembly of 19 pseudochromosomes (Figure S2; Table S1), concordant with the previous results.[4] The scaffold N50 was 172.24 Mb and the contig N50 was 2,043.27 kb, which was significantly improved compared to the earlier version (Table S2). The assembled genome size for *G. japonicus* in this study was approximately 2.54 Gb, which was close to the k-mer analysis results. We performed benchmarking universal single-copy orthologs (BUSCO) analysis, and the results revealed that the genome assembly of *G. japonicus* contained 93.8% complete sequences (Table S3), which revealed that the final assembly was highly complete. Our assembly has a lower number of scaffolds and contigs compared to previously published results, but a high N50 and a larger genome size, suggesting that the assembly results in this study are more continuous and reliable (Table 1).

We annotated 69.4% and 65.87% of the genome for repetitive elements and transposable elements, respectively, which is much higher than the previous results, most likely because we used more software and methods, including TRF version 4.09,[27] RepeatModeler version 1.0.11,[28] RepeatMasker, and RepeatProteinMask version 4.0.9,[29,30] to make the annotation more complete (Tables 1, S4, and S5).

A total of 20,503 protein-coding genes (Table S6) and 55,481 non-coding RNAs (Table S7) were annotated. 1:1 orthologous gene between the two versions was searched using OrthoFinder version 2.5.1[31] to demonstrate the improvement of gene structure annotation results in this study relative to the previously published data. Although there did not seem to be much difference in terms of length distribution plots (Figures S3 and S4), gene length (Wilcoxon signed rank test, one-tailed, $p < 2.2e\text{-}16$) and protein length (Wilcoxon signed rank test, one-tailed, $p = 0.0162$) in this study were both significantly higher than those in the previous annotation. The comparison of genes, transcripts, and exons between the two versions of the *G. japonicus* genome assembly also showed better results for the novel annotation (Tables 1 and S8).

Searching several gene annotation databases showed that 96.49% of annotated genes had functional annotations (Table S9). All these findings indicate that the genome assembly and annotation of *G. japonicus* in this study were complete and greatly improved compared to the previous genome version. The main features of the genome are shown in the Circos plot (Figure 1). Generally, the repetitive elements and genes in all chromosomes reveal an inverse relationship.

Chromosome evolution of *G. japonicus*

We selected two high-quality Sauropsidan genomes, *Anolis carolinensis* and *Gallus gallus*, based on identifying corresponding HSBs and EBRs in three carnivores,[6] to reveal the specific evolution of chromosomal regions of *G. japonicus*. They were aligned with the chromosome

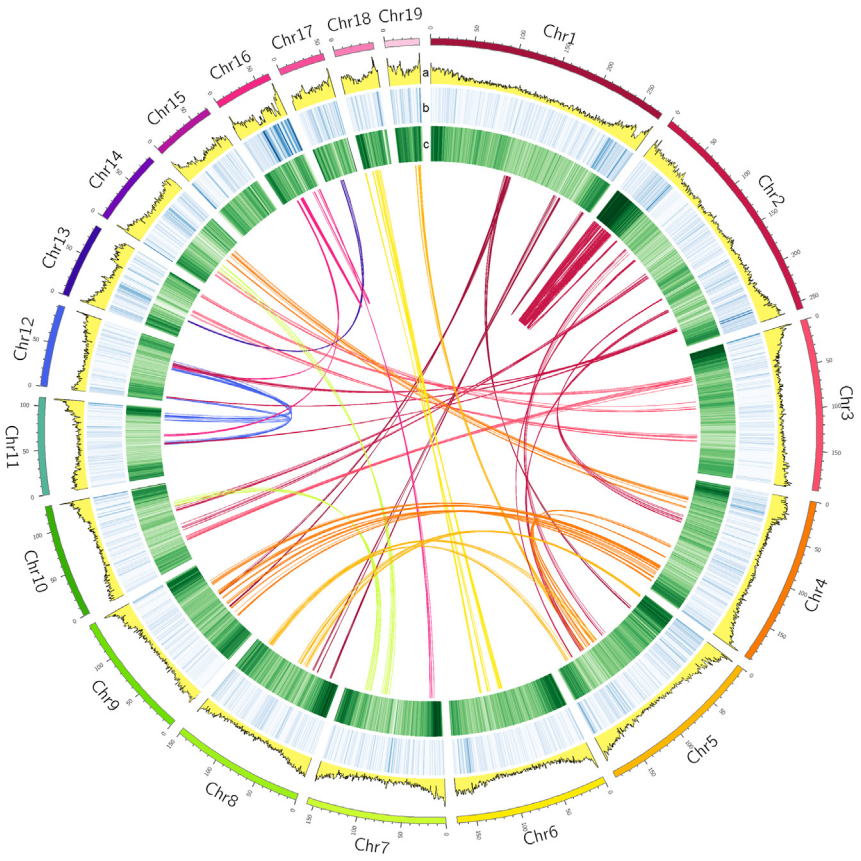


Figure 1. Circos plot of the *Gekko japonicus* genome

The outermost circle indicates the 19 chromosome information, with a tick unit of 10 Mb. Linkages between or within chromosomes show the synteny of paralogous genes. The circles labeled with a, b, and c correspond to GC content, gene density, and repeat coverage, respectively, with darker colors representing higher corresponding features.

sequences of *G. japonicus*, and corresponding HSBs were identified. A total of 39 HSBs between *G. japonicus* and *A. carolinensis* and 136 HSBs between *G. japonicus* and *G. gallus* were identified (Figure 2; Tables S10 and S11). Synteny analysis showed numerous chromosomal rearrangement events between *G. japonicus* and the two species, mainly translocations and, to a lesser extent, inversions. Except for the too-short microchromosomes, chromosomes 1–6 in *A. carolinensis* correspond to at least two chromosomes of *G. japonicus*, and most of the chromosomes of *G. japonicus* have a large segment of colinearity with only one chromosome of *A. carolinensis*, suggesting possible chromosome fission and fusion events from the ancestor of *G. japonicus* to *A. carolinensis*. More rearrangement events have occurred between *G. japonicus* and *G. gallus*, demonstrating the diversity and complexity of chromosome evolution in sauropsids.

A total of 13 EBRs were identified in *G. japonicus* by HSBs between *G. japonicus* and *A. carolinensis*. These EBRs have a total length of 106.15 Mb, accounting for 4.24% of the total chromosome size (Table S12). Sixty-three EBRs were identified in *G. japonicus* by HSBs between *G. japonicus* and *G. gallus*, with a total length of 93.83 Mb, accounting for 3.74% of the total chromosome size (Table S13). Sixty-four merged EBRs (with a total length of 188.85 Mb) were obtained by combining the information on EBRs from two alignment results, accounting for 7.54% of the total chromosome size (Table S14). The lengths of EBRs in *G. japonicus* ranged from a minimum of 16.97 kb to a maximum of 37.64 Mb, with a mean length of 2.95 Mb and a SD of 6.62 Mb. The result indicates that the length of the EBR regions was highly variable.

We compared gene density, GC content, and repeated sequence content across chromosomes between EBRs and genomes to study the variability of genomic features between EBRs and genomes (Table S15). The results revealed that EBRs have significantly higher gene density (Wilcoxon rank-sum test, $p = 0.026$) and slightly lower repeat content compared to genomes (Wilcoxon rank-sum test, $p = 0.075 < 0.1$). In contrast, GC content was not significantly higher in EBRs (Wilcoxon rank-sum test, $p = 0.76$). The results of the linear regression (Figure S5) showed an opposite linear relationship between gene density and repeated sequence content in EBRs (adjusted R-squared = 0.237, $p = 8.62e\text{-}06$), which was concordant with a high gene density and slightly lower repeated sequence content in EBRs. We also found that some repetitive elements were significantly higher in EBRs than in the rest of the genome (Table S16). These repetitive elements were mainly transposable elements, including DNA transposons (e.g., CMC, hobo-Ac-Tam3 [hAT], Kolobok, and P instability factor [PIF]), long interspersed nuclear elements (LINE) (e.g., CR1, L1, L2, and reverse transcriptase encoding element [RTE]), LTR (e.g., ERV1 and Gypsy), and short interspersed nuclear elements (SINE) (e.g., MIR, tRNA, and 5s-Deu-L2). Satellite sequences were also significantly enriched in EBRs,

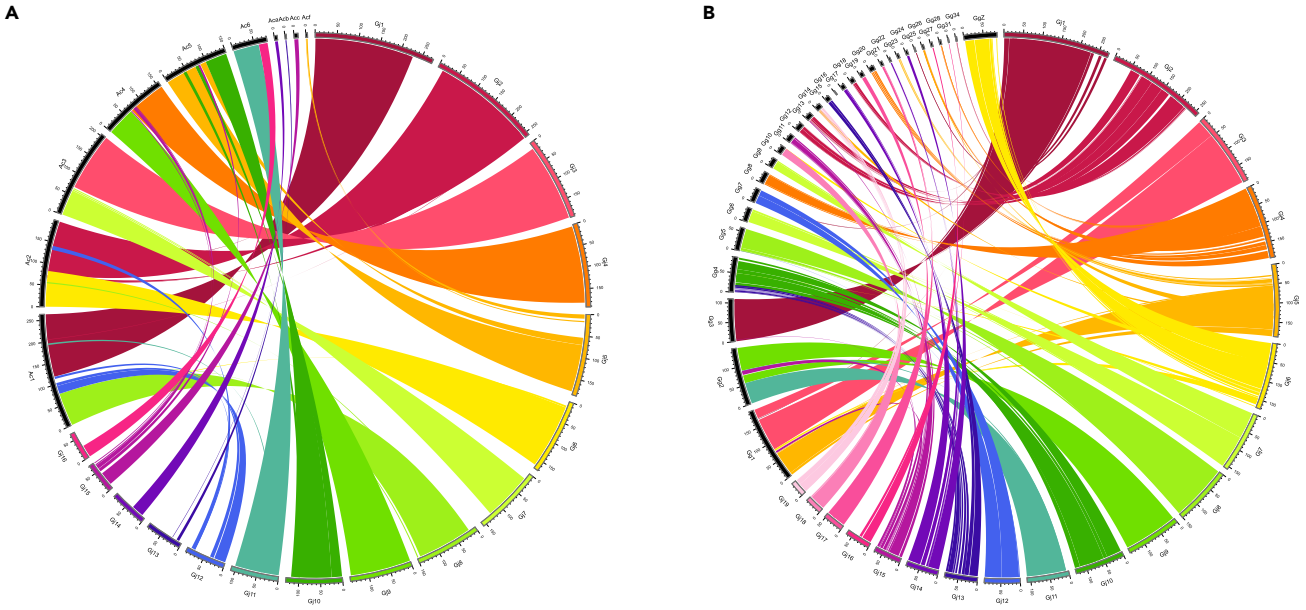


Figure 2. Synteny analysis of *G. japonicus* and two other sauropsids

"Gj", "Ac", and "Gg" are abbreviations of *G. japonicus*, *A. carolinensis*, and *G. gallus*, respectively. The numbers following them to the right correspond to the respective chromosome numbers. The circle indicates chromosome information with a tick unit of 10 Mb.

(A) Circos plot based on the results of synteny analysis between *G. japonicus* and *A. carolinensis*. The chromosomes of *G. japonicus* are marked in non-black and those of *A. carolinensis* in black, and the lines represent the synteny between the corresponding chromosomes.

(B) Circos plot based on the results of the synteny analysis between *G. japonicus* and *G. gallus*. The chromosomes of *G. japonicus* are marked in non-black, and those of *G. gallus* are marked in black. The lines represent the synteny between the corresponding chromosomes.

suggesting they may play an essential role in chromosome evolution. Enrichment analysis of genes located in EBRs showed that these genes are mainly enriched in the defense response pathway (Figure S6), and these immune-related genes may be involved in tail regeneration, damage repair, and defense against attack in *G. japonicus*.

Phylogenetic analysis and evolution of gene families

We selected a total of 18 sauropsids as representatives to elucidate the evolutionary relationships between sauropsids and the more precise phylogenetic position of *G. japonicus* and clearly show the impact of the assembly of two different versions of the genome on comparative genomics analyses. Specifically, one amphibian species and two mammals were used as outgroups. Additionally, 22 genomes including the previous version of the *G. japonicus* genome were used to construct a phylogenetic tree, estimate the divergence time of each node, and perform a gene family analysis on them (Figure 3).

The results revealed that the clade containing *G. japonicus* diverged from other squamates 191.9 million years ago (Figure 3A), consistent with the fossil record[32] and previous research.[12] The two versions of the *G. japonicus* genome showed some variability in divergence times, gene family expansion and contraction, and gene clustering analysis. This suggests that assembly and annotation with the older version may have biased the analysis results. These results highlight the importance of high-quality annotation and assembly for comparative genomics analysis. A total of 7,408 gene families were conserved in *A. carolinensis*, *Alligator sinensis*, *G. japonicus*, *Python bivittatus*, and *Pelodiscus sinensis* (Figure 3B). These gene families may constitute the core proteome of sauropsids. We performed Gene Ontology (GO) and Kyoto Encyclopedia of Genes and Genomes (KEGG) enrichment analyses for genes in *G. japonicus* that are specific, in expanded gene families, and under positive selection (Figures S7–S11). The results indicated that some of the enriched pathways were associated with sensory and nervous systems such as smell and taste, and that these pathways might be involved in activities such as predation, feeding, and escape from natural enemies. This suggests that their associated genes may have undergone adaptive evolution in *G. japonicus*. Some other pathways were immune related, such as those related to T cells, Nod-like, and tumor necrosis factor (TNF), which play essential roles in immune regulation and inflammatory responses,[33–36] and might be necessary for damage repair, tail regeneration, and defense against pathogens in *G. japonicus*.

Bitter taste perception and its evolution

Bitterness perception can detect potentially toxic compounds and thus avoid poisoning, which is critical factor for survival.[37] The T2R family is a subfamily of the taste receptor family that plays a significant role in bitter taste perception.[38] One-hundred-and-twenty-one T2R genes were identified, and the number of T2Rs in these seven species was similar to that reported in recent studies.[7,8] T2R expanded considerably in the

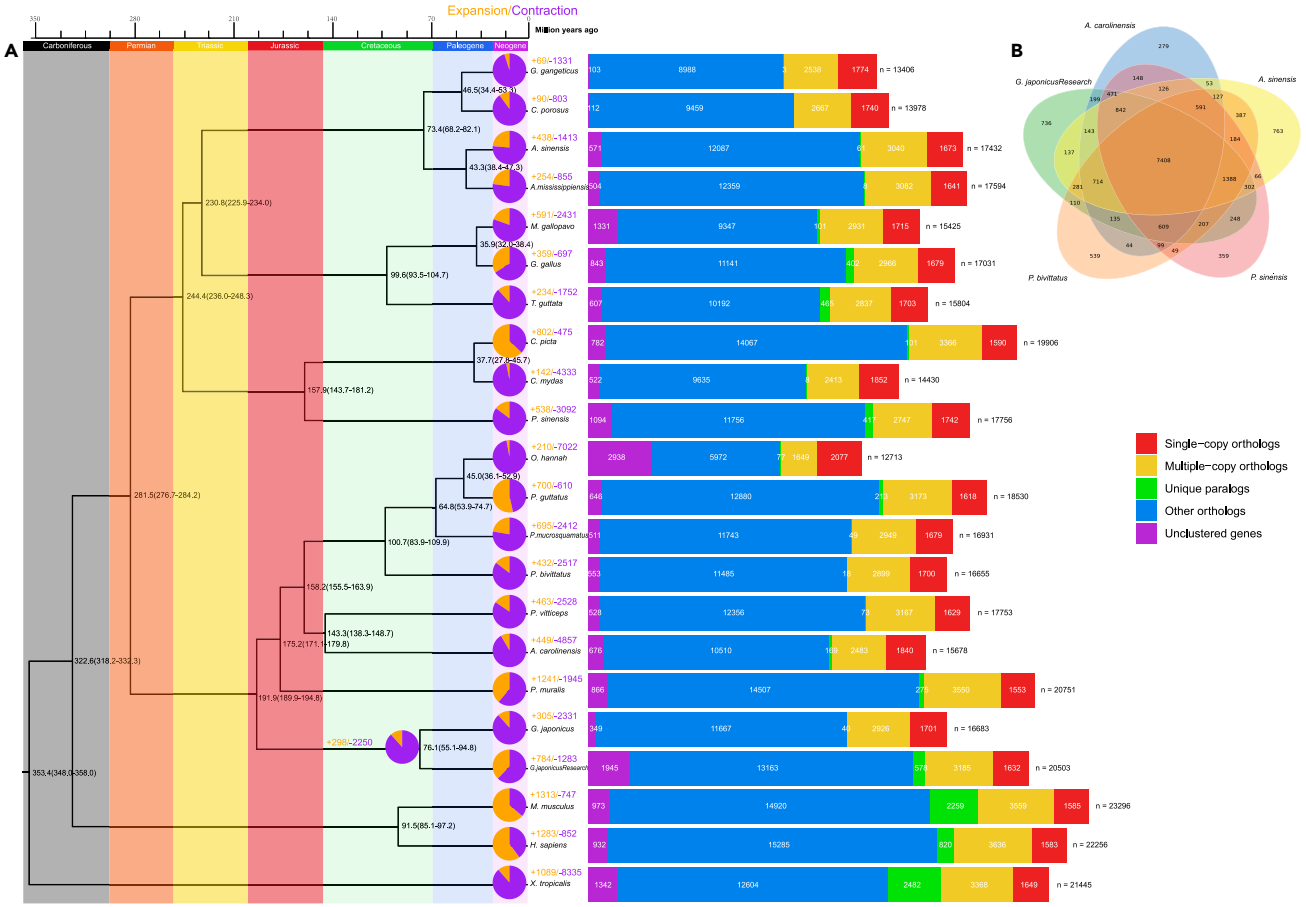


Figure 3. Phylogenetic and gene family analysis of 18 sauropsids and outgroups

"*G. japonicus*" and "*G. japonicusResearch*" correspond to the previous version and the genome assembly of this study, respectively.

(A) Phylogenetic tree for 18 sauropsids and outgroups. The bootstrap value for all branches was 100. The numbers on each node and parentheses represent the estimated divergence time (million years ago, MYA) and the 95% confidence interval. Pie charts represent the proportion of gene family expansions and contractions, with purple corresponding to expansions, orange to contractions, and colored numbers corresponding to each of them. The genes of each species are divided into five categories, the colors of which correspond to those in the legend on the right and are shown in the bar chart. At the same time, the white numbers inside represent the number of genes in the corresponding category, and "n" represents the total number.

(B) Venn diagram of the shared and unique gene families among the six sauropsids. Text and numbers represent the corresponding species and the number of gene families, respectively.

insectivorous *A. carolinensis* and *G. japonicus*, being significantly greater than in other species. Multiple independent duplications of the T2R in several species were clearly observed in the phylogenetic tree, dispersed among different clades (Figure 4A). We found that the branches under positive selection were mainly present in *A. carolinens* (13 branches) and *G. japonicus* (four branches), with only one branch associated with *S. punctatus*, which is consistent with the duplication, expansion, and rapid evolution of T2R in the insectivorous *A. carolinens* and *G. japonicus* (Figure 4A). Four pairs of site models in phylogenetic analysis by maximum likelihood (PAML)[39] and two site models in HyPhy[40] were used to detect positively selected sites (Tables S18–S20). The likelihood ratio tests (LRTs) for all four pairs of PAML models were significant (χ^2 test, $p < 0.05$), considering that the authors of the software did not recommend the positively selected sites identified by M3 and they were therefore not included in the results. The average ω value obtained from M0 was 0.54, which was a relatively high value, indicating that T2R was under relatively low purifying selection, on average, among these species. Nine overlapping sites were identified between M2 and M8, and six between fixed effects likelihood (FEL)[41] and fast, unconstrained bayesian AppRoximation (FUBAR).[42] However, there was little overlap in the positive selection sites calculated by different software, with only two identical sites between FEL and M2 or M8 (Figure 4B). A careful review of the sequence alignment results suggested that this was mainly because, in addition to the variability in the calculation methods of the models, PAML removed all sites containing gaps while HyPhy retained them. A total of 38 positively selected sites were identified by combining the results from all models, with the most in the extracellular region (ER) (17 [44.74%]), followed by the transmembrane region (TM) (15 [39.47%]), and the least in the intracellular region (IR) (6 [15.79%]), with no significant heterogeneity in this distribution pattern (χ^2 test, $\chi^2 = 5.42$, $p = 0.067$).

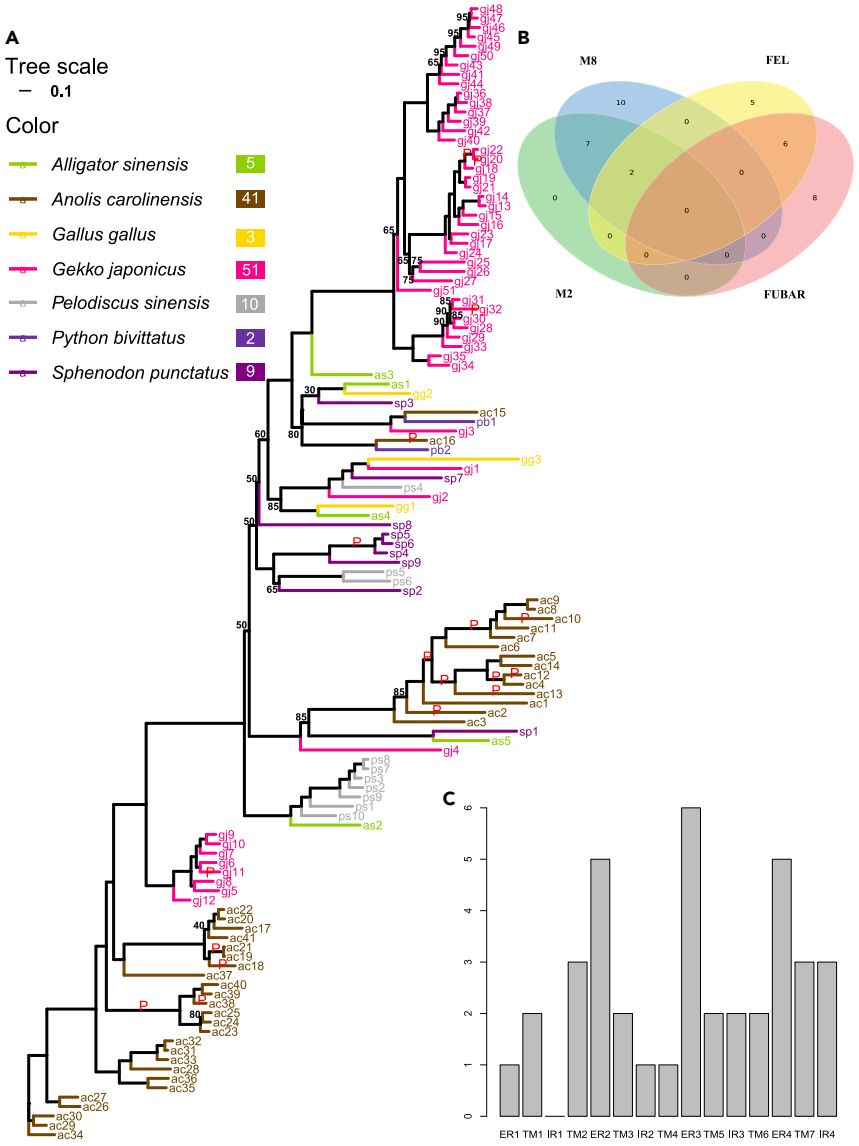


Figure 4. Phylogenetic and selection pressure analysis of the T2R gene in sauropsids

(A) Phylogenetic tree of T2R genes in seven sauropsids. The scale bar for the branch length of the tree and the corresponding color of the species are labeled in the legend, and the number to the right of the species name represents the number of T2Rs. The bootstrap values not equal to 100 are labeled by numbers at each node. Branches marked with "P" represent branches tested as under positive selection using the aBSREL model.

(B) Venn diagram of standard and unique positively selected sites in the T2R gene of seven sauropsids identified by four methods. Text and numbers represent the corresponding method and the number of positively selected sites.

(C) Bar graph of the number of positively selected sites in different regions of the T2R gene in seven sauropsids. The X axis corresponds to the different areas of the T2R gene, and the Y axis represents the number of positively selected sites in the corresponding regions.

Specifically, when T2R was divided into 15 intervals according to the ER, TM, and IR order from N-terminal to C-terminal, the ER3 region had the most positively selected sites, with 6. The least was IR1, where no positively selected sites were detected (Figure 4C). The pairwise ω values of ER were significantly higher than those of TM (two-sample independent t test, $t = 14.80$, $p < 2.2e\text{-}16$) and IR (two-sample independent t test, $t = 16.93$, $p < 2.2e\text{-}16$). The pairwise ω values of the TM region were significantly higher than those of the IR region (two-sample independent t test, $t = 12.39$, $p < 2.2e\text{-}16$).

Most of the 51 identified T2Rs of *G. japonicus* were tandemly clustered (Figure 5A). gj5~gj12 were tandemly clustered on the reverse strand and gj13~gj50 on the forward strand. These two gene clusters corresponded to two clades on the phylogenetic tree, while gj51 was close to gj50 but on the reverse strand and was on the same clade with gj13~gj50 on the phylogenetic tree. We extracted a subtree of *G. japonicus* from the T2R phylogenetic tree of seven species. We performed the same evolutionary analyses to investigate the intraspecific

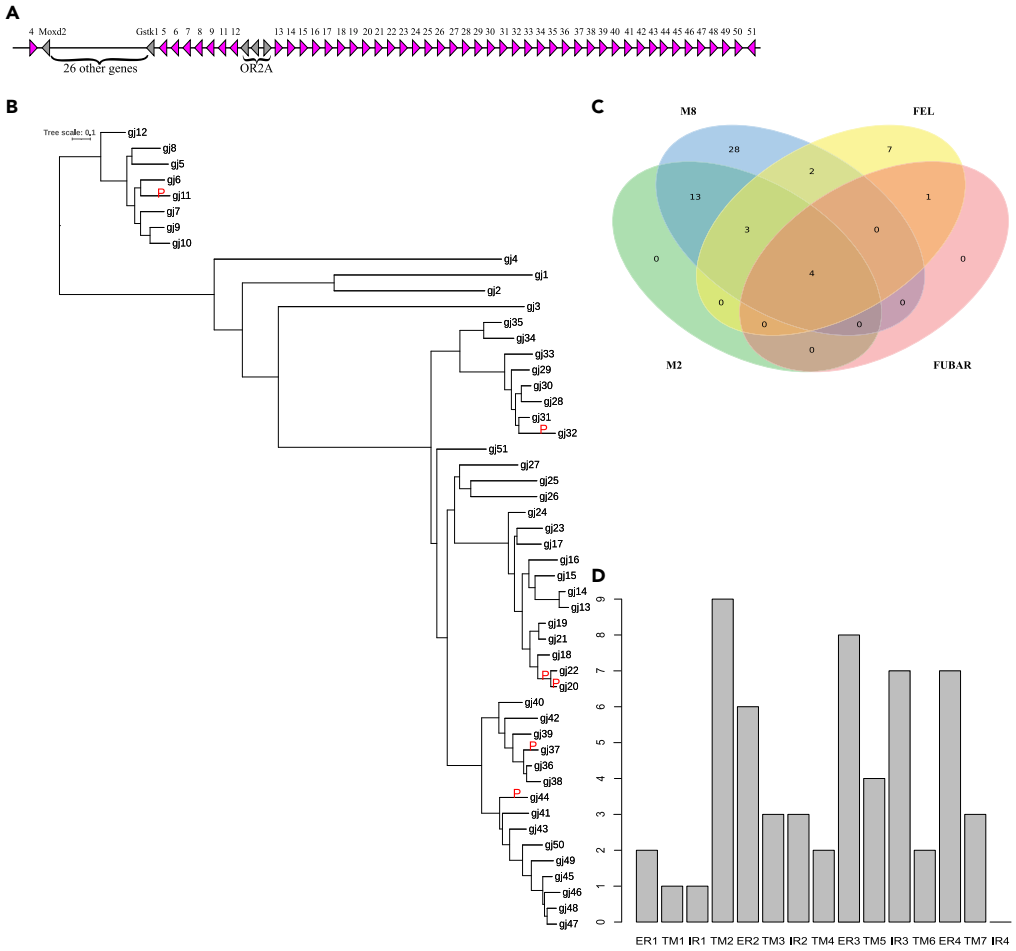


Figure 5. Location and selection pressure analysis of the T2R gene in *G. japonicus*

(A) All 51 T2R genes were clustered on chromosome 6 of *G. japonicus*. Red triangles represent T2R genes, gray triangles represent other genes, and the obtuse angle direction to the right means the forward strand, and to the left means the reverse strand.

(B) The subtree of the T2R gene in *G. japonicus*. The branch length scale bar is marked in the upper right corner, and branches marked with "P" represent branches detected as under positive selection using the aBSREL model.

(C) Venn diagram of standard and unique positively selected sites in *G. japonicus* T2R genes identified by four methods. Text and numbers represent the corresponding method and the number of positively selected sites.

(D) Bar chart of the number of positively selected sites in different regions of the *G. japonicus* T2R genes. The X axis corresponds to the different regions of the T2R genes, and the Y axis represents the number of positively selected sites in the corresponding regions.

evolution of T2R in *G. japonicus* and to compare it with the overall evolutionary pattern of sauropsids. A total of six branches under positive selection (Figure 5B) overlap with the four branches detected in the seven species dataset. The LRT tests for all four pairs of site models performed in PAML were significant (Table S21), with the mean ω value calculated using M0 being 0.70, which was higher than the mean ω value of the seven sauropsids. The positively selected sites identified using the FEL and FUBAR models overlapped with four sites identified by M8 and M2 (Figure 5C). These were located at codons 73, 92, 183, and 263 in the multiple sequence alignment, indicating the reliability of these sites. A total of 58 positively selected sites were identified, and their counts in ER, TM, and IR were 23, 24, and 11, with percentages of 41.38%, 39.66%, and 18.97%, respectively, showing no significant heterogeneity in distribution (χ^2 test, χ^2 = 5.42, p = 0.067). The distribution of the sites divided into 15 regions (Figure 5D) was consistent with the distribution of the seven species (G-test, G = 33.24, p = 0.77).

Nine positive selection sites shared by both were located at codons 13, 66, 87, 88, 90, 154, 162, 240, and 280 in the multiple sequence alignment. The mean ω values of 2.13, 0.68, and 0.44 were obtained from the YN00 analysis for the T2R alignments of the ER, TM, and IR regions, respectively. Similarly, we also found significantly higher mean ω values in the ER region than in the TM region (two-sample independent t test, t = 4.50, p = 7.3e-06) and the IR region (two-sample independent t test, t = 8.47, p < 2.2e-16). The mean ω values in the TM region are significantly higher than in the IR region (two-sample independent t test, t = 43.80, p < 2.2e-16). The aforementioned results indicated that the pattern of evolutionary pressure exerted on *G. japonicus* and exerted among the seven sauropsids was consistent overall. Specifically, the

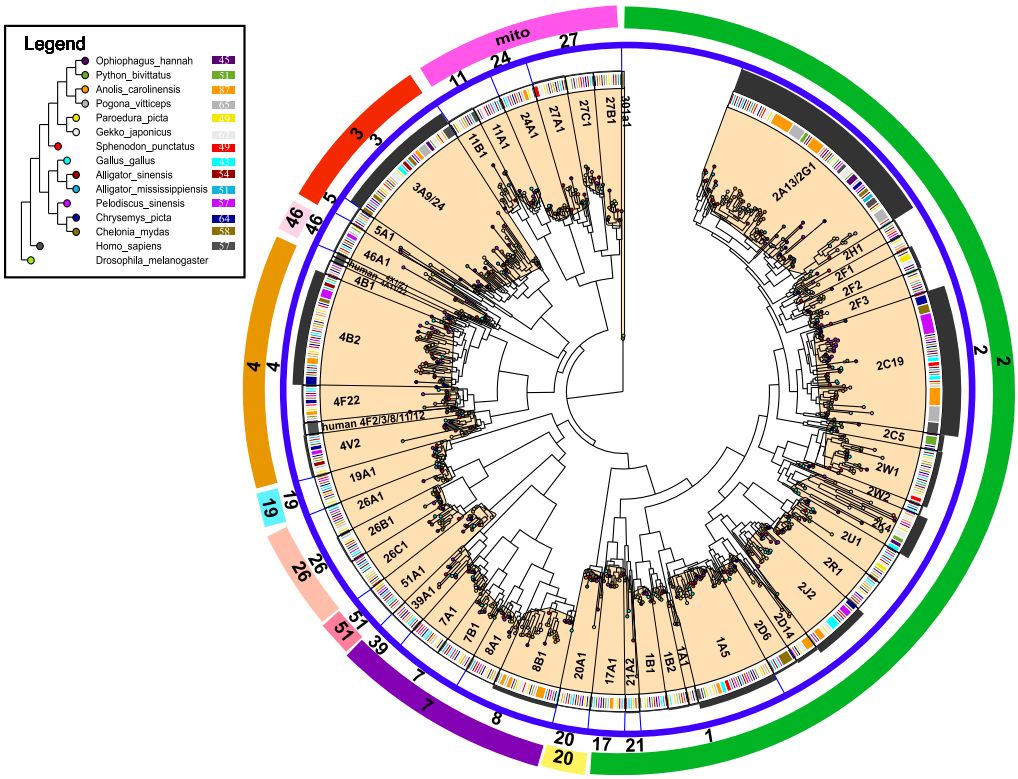


Figure 6. MIPhy analysis and phylogeny of the P450 of sauropsids

The numbers correspond to clan, gene family, and specific gene name from the outside to the inside of the circle. The width of each black circle arc fits the size of the MIphy score. The color of the endpoints in the tree corresponds to the species, as shown in the legend in the upper left corner, and the numbers in the legend indicate the size of the number of genes in the corresponding species.

ER and TM were significantly more unconstrained by purifying selection than the IR region, especially the ER region, which has very high levels of inter- and intra-species variation, indicating its importance in adaptive evolution.

Birth and death evolution of CYPs in sauropsids

The ingested toxins might be cleared by a series of detoxification genes such as CYPs, in addition to detecting bitter toxins in insects through the T2Rs mentioned earlier. Here, we identified the CYPs of 13 representative sauropsids using human functional CYPs as a reference. The sauropsid CYPs clustered well into 10 clans, and 18 gene families, but among the subfamilies. Phylogenetic relationships and clades divided by minimizing instability in phylogenetics (MIPhy)[43] for the phylogenetic tree indicated that many sauropsid CYPs were not orthologs to human CYPs and had different subfamily numbers from human CYPs or were difficult to assign to existing subfamilies of human CYPs (Figures 6 and S12).

Some subclades are independent of human CYPs and had not been reported in previous studies. We named them based on phylogenetic relationships and BLASTP search results, including CYP1B2, CYP2F2, CYP2F3, CYP2H1, CYP2W2, CYP3A9, CYP3A24, CYP4B1, and CYP4B2. The number of D-type CYP genes was significantly more variable than that of B-type CYP genes (Figure S13; Table S24), consistent with expectations. MIPhy scores were significantly higher in the D-type than in the B-type CYP genes (two-sample independent t test, $t = 3.1125$, p value = 0.00469), indicating that the clades of the D-type CYP genes had high instability. D-type CYP genes had frequent duplication and loss events, suggesting they underwent more birth and death processes than B-type CYP genes (Table S25).

We calculated ω values for each clade divided by MIPhy using the M0 model in PAML (Table S26) and found a correlation between ω values and MIPhy scores (Spearman's correlation, $\rho = 0.37$, $p = 0.012$). However, no significant differences in ω values were found between the D-type and B-type CYP genes (average ω 0.18 vs. 0.15, two-sample independent t test, $t = 1.2011$, p value = 0.2363). Still, the mean ω value was slightly high in the D-type CYP genes. Positively selected sites were found in 13 of the 46 clades using the M2a and M8 models in PAML (Table S26), including six from 24 D-type-related clades and seven from 22 B-type-related clades. In addition, 15 branches were found to be under positive selection, 12 of which were from clades of D-type CYP genes (Table S27). CYP2D14, CYP2F3, CYP2U1, CYP4F22, and CYP26C1 were under positive selection in the foreground branches of *G. japonicus*, but only CYP2D14 was significant after correction for multiple hypothesis testing.

Table 2. Comparison of the number of CBP gene families in 18 reptiles, frog and mouse

Species	Numbers (in this study)	Numbers of gene family	Reference
Alligator mississippiensis	6	23	Holthaus et al.[46]
Crocodylus porosus	5	28	
Alligator sinensis	15	–	–
Python bivittatus	5	35	Holthaus et al.[47]
Ophiophagus hannah	17	36	
Thamnophis elegans	8	–	–
Gekko japonicus	66	71	Liu et al.,[12] Hara et al.[48]
Paroedura picta	120	120	Gamble[49]
Eublepharis macularius	15	6	Li et al.[50]
Shinisaurus crocodilurus	29	7	
Anolis carolinensis	16	40	Dalla Valle et al.[45]
Podarcis muralis	27	–	–
Zootoca vivipara	24	–	–
Lacerta agilis	26	–	–
Pogona vitticeps	5	–	–
Chrysemys picta	24	89	Jurka et al.[51]
Chelonia mydas	6	37	
Pelodiscus sinensis	25	74	
Xenopus tropicalis	0	–	–
Mus musculus	0	–	–

We performed phylogenetic independent contrast (PIC) analysis of CYP1-CYP4 clades to investigate whether dietary differences are related to changes in the number of CYPs. Most clades were unstable, had frequent birth and death processes, and varied somewhat among different species. However, we found only a correlation between CYP1 and diet (Figure S14). These results demonstrate the complexity of the interaction between selection pressure and birth-death processes.

Phylogeny, structural composition, and synteny of CBPs, a critical gene family related to epidermis formation

The identified CBP genes sequences in *G. gecko*[44] and *A. carolinensis*[45] were used as reference to search proteins of selected reptiles and others (see STAR Methods for details). Sixty-six aa sequences of the CBP gene family members of *G. japonicus* were obtained (Table S29). These sequences were tentatively named CBP1 ~ CBP66 according to their positions on chromosomes. The CBP genes sequences of *G. japonicus* were principally located in the epidermal differentiation complex (EDC) region of chromosome 1. In this study, the CBP genes of 12 reptiles (in which the CBP genes family has been studied) and 6 reptiles, western clawed frog, and mouse (in which the CBP genes family has not been broadly studied) with excellent genome assembly and annotation data were identified (Table 2).

Among them, the number of CBP gene families of some species identified in this study is generally lower than those previously published. Exceptionally, the number of the CBP genes family of *Paroedura picta* is the same as that of existing studies, which was 120, and the number was the largest among the species studied so far. However, the present study identified more CBPs than reported in *Eublepharis macularius* and *Shinisaurus crocodilurus*. By comparison, the number of CBPs in Crocodilian is the least among reptiles, with an average of 9. Among the squamates, there were 10 CBP sequences on average in snakes and 38 in lizards. The average number of CBPs in Chelonia was 18. No CBP genes family has been identified in *Mus musculus* and *Xenopus tropicalis*. The difference in the number of CBPs members may be related to the characteristics of their different epidermis.

Among the 66 aa sequences of the CBP gene family in *G. japonicus*, the shortest length was 85 aa, the longest was 478 aa, and the average length was 151.73 aa. The minimum molecular weight was only 8.92 kDa, and the maximum was 41.88 kDa, with an average of 15.23 kDa. These CBP proteins are all small proteins with short sequence lengths and small molecular weights. The values of the theoretical isoelectric point range from 4.14 to 10.26. Although the data vary widely, most of them range from 7.0 to 9.0, and there are more neutral and alkaline sequences. Proline (Pro), cysteine (Cys), glycine (Gly), and serine (Ser) were the most abundant aa sequences in the 66 CBP sequences of *G. japonicus* (Figure S15). The highest contents of Pro, Cys, Gly, and Ser were 22%, 23%, 44%, and 22% respectively. In general, these proteins contain small amounts of charged aa, such as glutamic acid (Glu), aspartic acid (Asp), arginine (Arg), and lysine (Lys).

The alignment of 66 CBPs aa sequences in *G. japonicus* (Figure S16) revealed that each sequence contained a highly conserved motif known as the "core box". The motifs of 66 CBPs aa sequences in *G. japonicus* were predicted (Figure S17). Each sequence contained a

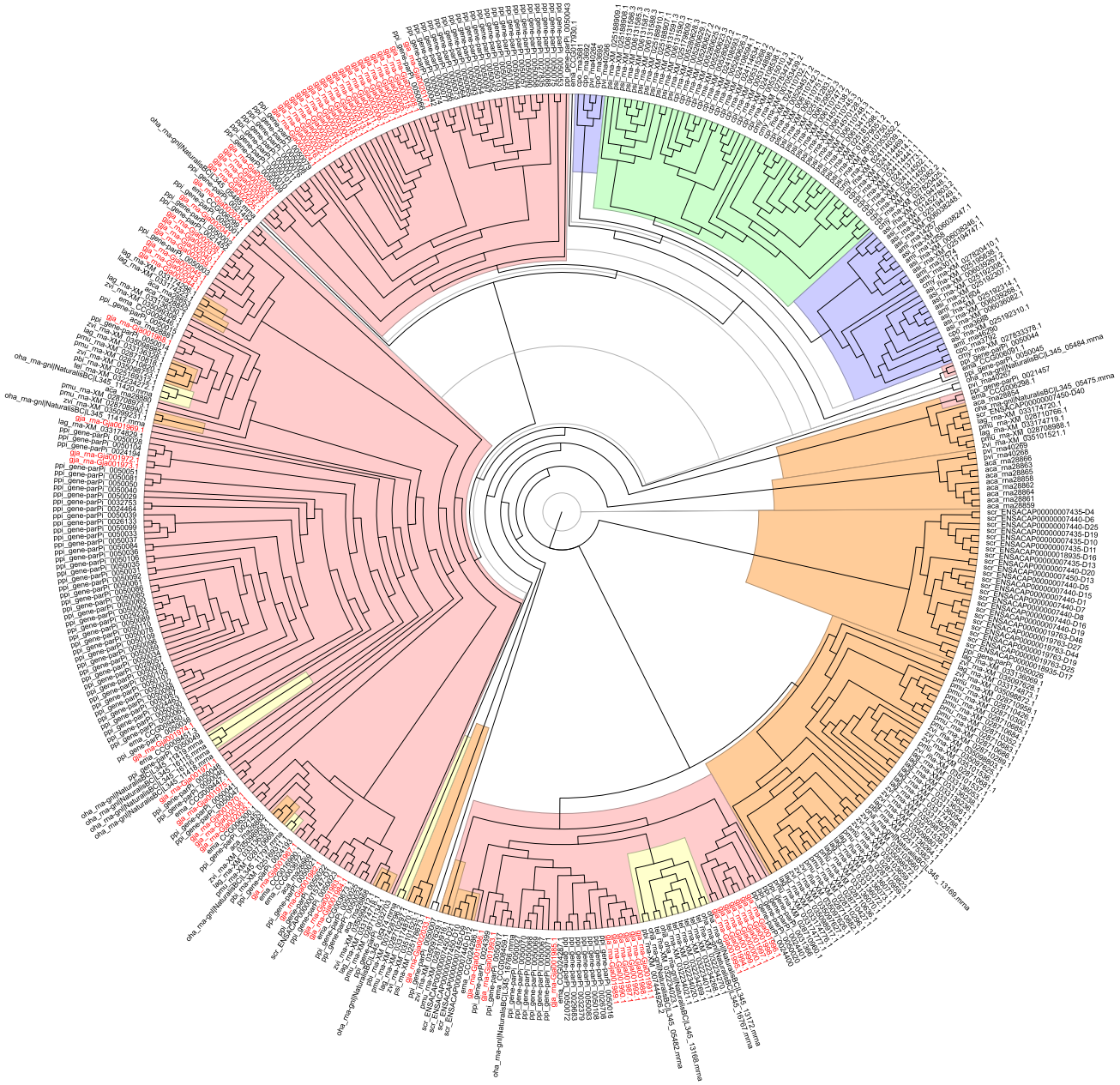


Figure 7. Phylogenetic tree of CBP gene families in 20 species

The red background indicates geckos, the orange background indicates lizards, the yellow background indicates snakes, the green background indicates turtles and crocodiles, and the blue background shows crocodiles. The CBP sequences of *G. japonicus* are shown in red font.

very conservative motif 1, consistent with sequence alignment, especially the short sequences "PPP" and "PGP" were present in each sequence. Only one sequence (rna-Gja001967.1) contained four motif regions, and three sequences (e.g., rna-Gja001974.1, rna-Gja001992.1, and rna-Gja002040.1) contained two motif regions. In addition to the conserved motifs, conserved domain analyses of these sequences were performed using the NCBI conserved domain analysis website (Figure S18). Thirty-five of the 66 sequences in the *G. japonicus* contained conserved keratin superfamily domains. However, there has been little analysis of the conserved domain of sauropsids.

A phylogenetic tree of 66 CBPs aa sequences in *G. japonicus* was constructed, and these sequences were divided into three categories after combining phylogenetic tree results with the motifs, domains, and gene structure distribution map mentioned earlier (Figure S18). Sequences with similar motif structures clustered together. The phylogenetic tree was constructed with CBP sequences of the 20 selected

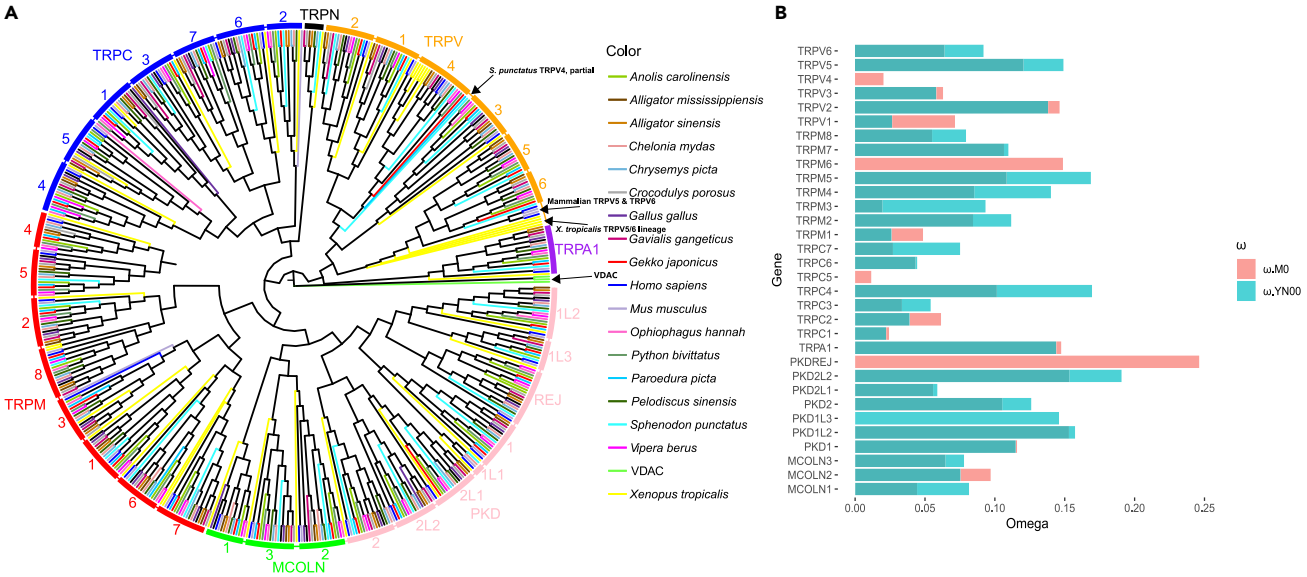


Figure 8. Phylogenetic tree of the TRP gene family and the ω values of each gene

(A) Phylogenetic tree of TRP genes in 16 sauropsids, two mammals, and one amphibian. The outer circle is divided by different TRP subfamilies and their submembers, with the same subfamilies labeled with the same color and the submembers labeled with text. Different colored branches correspond to different species, as shown in the legend on the right.

(B) Bar graph of ω values of different TRP genes. The X axis represents the ω value and the Y axis represents the corresponding gene. ω values identified with M0 and YN00 are labeled with different colors, as shown in the legend on the right.

species (Figure 7). The squamate CBP sequences were generally clustered. However, these sequences were not entirely clustered by species evolutionary relationships, suggesting that the CBP gene family may have evolved specifically in each clade.

The arrangement of CBP genes on the chromosome or scaffold of the selected 20 species (Figure S19) is similar, with most arranged in series on the same chromosome or scaffold. Besides the arrangement of CBP genes, flanking loricrin (LOR) genes also exist. However, some genomes were not assembled at the chromosomal level, and the identified CBP genes in some species were arranged on multiple scaffolds or contigs. Most of the CBPs in *G. japonicus* have a syntenic relationship with those of other reptile species (Figure S20), especially with *P. picta*, *P. muralis*, *Zootoca vivipara*, *Lacerta agilis*, and *A. carolinensis*. However, they have no syntenic relationship with Crocodilian or Chelonian. These results indicate that the CBPs in *G. japonicus* are more similar to those of the lizards in evolution.

Evolution of TRP genes in sauropsids

TRP sequences from humans, mice, chickens, and the previous genome version of *G. japonicus* as a reference were collected to identify TRP sequences of *Sphenodon punctatus*, *Vipera berus*, *Ophiophagus hannah*, *P. picta*, and *G. japonicus* (this study). The TRP genes identified in *S. punctatus*, *V. berus*, *O. hannah*, *P. picta*, and *G. japonicus* were 31, 33, 31, 31, and 31, respectively (Table S30). The number of TRP gene families and the number and composition of each subfamily were very similar among the different sauropsids, suggesting that TRPs were conservative in their evolution.

Moreover, the constructed phylogenetic tree that referred to a recent study rooted with voltage-dependent anion channel (VDAC)[52] (Figure 8A) indicated that the clade size of each subfamily was quite similar. This was because each subfamily had almost only one copy of its submember among sauropsids, with multiple copies in a few species. We found no TRPM6, PKD1L1, and TRPN in *G. japonicus* and only one copy for all other TRP subfamily members. PKD1L1 was missing in most sauropsids, with a single copy present only in *C. mydas* and *S. punctatus*. PKD1L3 and TRPN were missing in some sauropsids and were absent in all researched snakes. High-quality genome samples of representative species were required for comprehensive analyses and confirmation to determine whether these gene losses were lineage specific.

We used the branch-site model and the clade model in PAML with *G. japonicus* set as a foreground and examined the difference in TRP evolution with other sauropsids to test the changes and shifts of selection pressure on *G. japonicus*. Six TRP genes detected shifts in selection pressure, which were *TRPV5*, *TRPM3*, *TRPM5*, *TRPM8*, *MCOLN1*, and *PKD2L2* (Tables 3; S31–S62). Most of the best models with the lowest Akaike's information criterion (AIC) values were CmC and CmD. CmC and CmD were not significant only in *PKD2L2*. The lowest AIC was M2a_rel. Except for *MCOLN1*, which showed intensified purifying selection, all the other five TRP genes revealed high ω in the foreground branch, indicating the rise of ω in *G. japonicus* and the presence of purifying selection relaxation. At the same time, *TRPM3* and *PKD2L2* were

Table 3. TRP genes that are significant in LRTS of PAML analysis

Gene	Model[a]	AIC	p value[b]	q value[c]
TRPV5	BrSnull	13047.18		
	BrS	13049.18	1	1
	M2a_Rel	12959.83		
	CmC F	12957.59	0.03936	0.367359
	M3	12961.51		
	CMD F	12959.39	0.04222	0.295515
TRPM3	BrSnull	6872.747		
	BrS	6874.747	1	1
	M2a_Rel	6856.207		
	CmC F+	6852.297	0.01505	0.367359
	M3	6858.207		
	CmD F	6853.559	0.00993	0.278017
TRPM5	BrSnull	17509.39		
	BrS	17511.39	1	1
	M2a_Rel	17392.35		
	CmC	17394.27	0.773169	1
	M3	17394.1		
	CmD F	17391.44	0.03083	0.295515
TRPM8	BrSnull	9295.202		
	BrS	9297.202	1	1
	M2a_Rel	9281.566		
	CmC F	9278.971	0.03205	0.367359
	M3	9281.141		
	CmD	9280.042	0.07837	0.438873
MCOLN1	BrSnull	4802.805		
	BrS	4804.805	0.996432	1
	M2a_Rel	4748.271		
	CmC	4749.822	0.502555	1
	M3	4745.106		
	CmD B	4742.722	0.03627	0.295515
PKD2L2	BrSnull	17228.32		
	BrS	17220.49	0.00171	0.048
	M2a_Rel	17145.06		
	CmC	17146.27	0.373477	1
	M3	17145.1		
	CmD	17146.67	0.509877	1

[a]The model with the lowest AIC is marked in bold. "F" represents higher ω in the foreground, "B" represents higher ω in the background, and "+" indicates that the foreground branch ω value is greater than 1.
[b]p values less than 0.05 are marked in bold.
[c]q values less than 0.05 are marked in bold.

detected under positive selection in the foreground branch. However, after correcting for multiple hypothesis testing, only the LRT for *PKD2L2* was significant (Table S63). However, divergent selection might still exist in the other five TRP genes and should not be ignored.

We also used the branch model to effectively detect the relaxation of selection pressure in *G. japonicus*, including the paired one-ratio model (M0) and the two-ratio model (Tables S31–S61). The TRP genes with p values less than 0.05 obtained from the LRTs were *TRPM3*, *TRPM8*, and *MCOLN1* (Table S62). After correction for multiple hypothesis testing, only *TRPM3* had a q-value less than 0.05 (Table S63). Except for *MCOLN1*, which revealed intensification of purifying selection, all other genes showed relaxation of purifying selection in *G. japonicus*. These results were consistent with the results of the branch-site and clade models. We also used site models in PAML to detect

positively selected sites for TRP genes (Tables S26–S56). Still, it was hard to find positively selected sites with PAML in our data, indicating that TRP family members are generally conserved. The ω values were calculated using a one-ratio model (M0), indicating that the TRP genes were primarily under relatively strong purifying selection in sauropsids (Figure 8B), with a mean ω value of 0.085 and a SD of 0.053. The gene under the strongest purifying selection was *TRPC5* with an ω value of 0.012, and the gene under the weakest purifying selection was *PKDREJ*, with an ω value of 0.25.

Given the low quality or absence of some TRP genes in many species samples, we did not run the M0 model to calculate their mean ω values. We selected 1:1 orthologous genes between *A. carolinesis* and *G. japonicus* for the YN00 analysis to complement these results (Figure 8B). The results obtained using YN00 and those obtained by M0 for the shared genes were not different in general (Kolmogorov-Smirnov test, $p = 0.53$), indicating that the two results were consistent. However, the ω values obtained by YN00 were higher than those of M0 (paired sample t test, $t = -2.76$, $p = 0.0106$). The gene with the lowest ω value calculated using YN00 was *TRPC1* with 0.022, while the highest was *PKD2L2* with 0.19. *PKDREJ* and *PKD2L2*, two critical genes associated with reproduction, possessed high ω as expected.[53] These results were consistent with numerous previous studies of reproduction-associated proteins, in which it was shown that reproduction-associated proteins had higher evolutionary rates than other proteins, that many of them were under positive selection, and that they diverge rapidly between taxa.[54–60]

DISCUSSION

Chromosome-level genome assembly is a valuable resource that allows us to study species-specific evolution from a chromosomal perspective. Currently, there are only a few species with chromosomal-level genome assembly among reptiles, and it is challenging to perform more accurate genes identification and evolutionary analyses.

Here, we combined Illumina HiSeq sequencing, PacBio SMRT sequencing, and Hi-C technology to assemble a high-quality *G. japonicus* genome with a chromosome number of 19 and a genome size of 2.54 Gb. The genome assembly in this study was more contiguous and of higher quality than previous versions, as reflected in the high N50 and fewer scaffolds and contigs (Table 1). The larger size of the *G. japonicus* genome in the present study may also be due to its significant contiguity. Not only the assembly level but also the annotation of the genome has been significantly improved compared to the previous version, with more complete and detailed annotation results than the previous version. We have conducted comprehensive comparative genomics and evolutionary analyses with other sauropsids with such high-quality genomes.

Chromosome evolution is essential for species formation or speciation, but comprehensive bioinformatics studies on chromosome evolution of sauropsids are scarce. Murphy et al.[61] studied mammalian chromosome evolution as early as 2005. They found that nearly 20% of chromosome breakpoint regions were reused during mammalian evolution, and the gene density in EBRs was higher than that in the genome-wide level. Elsik et al.[62] identified cattle-specific EBRs in 2009 and found high segmental duplications and some repetitive elements enriched in them, mainly LINE-L1 and LINE-RTE. Moreover, Larkin et al.[63] identified EBRs from 10 amniotes species and found that genes related to adaptive function were significantly enriched in EBRs in 2009. They also found that retrotransposed genes, zinc-finger genes, and CpG islands were significantly high in EBRs, demonstrating the importance of EBRs in adaptive phenotypes, functional evolution, and species formation.

Groenen et al.[64] identified pig-specific EBRs in 2012. They found an enrichment of multiple repetitive elements such as LTR-ERV1, LINE-L2, satellite sequences, and genes located in EBRs enriched in pathways associated with taste perception. Farré et al.[65] identified bird-specific EBRs in 2016 and found four transposable elements LINE-CR1, LTR-ERVL, LTR-ERVK, and LTR-ERV1 enriched in lineage-specific EBRs. Genes located in the lineage-specific EBRs are enriched in several different GO terms.

Futhermore, Fan and colleagues[6] identified HSBs and EBRs using pairwise synteny analysis of three carnivores, including dogs, cats, and giant pandas, in 2019. They found enrichment of LINE-L1 elements in the EBRs of the three carnivores, with higher GC content, gene density, and repeat elements in the EBRs of giant pandas and dogs compared to the whole genome. In contrast, no such differences were found in cats. Some genes in the EBRs of giant pandas, dogs, and cats were enriched in olfaction-related pathways. This study identified the corresponding HSBs and EBRs by pairwise synteny analysis of *G. japonicus* with two other representative sauropsids, *A. carolinensis* and *G. gallus*. These species have relatively high-quality genome assembly and annotation at the chromosomal level and differ in chromosome number, following Fan et al.[6] The specific EBRs in *G. japonicus* have a higher gene density and fewer repetitive elements than the whole genome. This is consistent with that both are somewhat linearly negatively correlated in EBRs.

However, there was no significant difference in GC content between EBRs and the whole genome. DNA transposons (e.g., CMC, hAT, Kolobok, and PIF), LINE (e.g., CR1, L1, L2, and RTE), LTR (e.g., ERV1 and Gypsy), SINE (e.g., MIR, tRNA, and 5s-Deu-L2), and satellite sequences were significantly higher in the EBR region specific to *G. japonicus* compared to the rest of the genome. The genes located in the EBRs, which are mainly immune-related genes, significantly enriched in the defense response pathway. These genes are likely essential for various immune functions and activities in *G. japonicus*, such as damage repair and tail regeneration. In summary, EBRs resulting from non-random chromosomes breaks are essential regions and hotspots for evolution.

Specific genes located in EBRs, with highly repetitive elements, may have a powerful influence on species evolution and formation. Of course, our study was limited to single species-specific EBRs and only three species were compared. After which, as increasing high-quality reptile and bird genomes are assembled at the high-quality chromosome level, it will be possible to sample all taxa of Sauropsida evenly. Multi-species, large-sample, comprehensive sauropsids chromosome evolutionary analyses are required in the future.

Phylogenetic analyses and estimated divergence times for representative sauropsids were performed to provide a more precise phylogenetic position. We found some variability in the comparative genomic analyses between our genome assembly and the previous one. Gene family clustering, gene family expansion and contraction, and positive selection pressure analyses revealed essential functional gene sets in the *G. japonicus* genome in our assembly, which are mainly enriched in sensory systems and immunity pathways. This may also be why *G. japonicus* possesses agility, damage repair, and tail regeneration features. We thoroughly analyzed an essential part of these gene sets: T2R genes.

T2R genes, which are related to bitter taste perception, are widespread in vertebrates, vary widely among species, and lineage-specific gene duplication events have been detected in bony fishes, frogs, mammals, and lizards.[66] A recent comprehensive analysis by Zhong et al.[8] for T2Rs in squamates has demonstrated that T2Rs are extensively duplicated and expanded in several gecko and lizard lineages. Our location distribution and phylogenetic analyses suggest that this expansion will likely result from tandem duplication in multiple species. Most branches under positive selection were located within clades related to *G. japonicus* and *A. carolinensis*. This result agrees with their extensive duplication and expansion, and all of these results suggest that T2Rs are essential for the adaptive evolution in *G. japonicus*, or insectivorous lizards and geckos. Zhong et al.[8] also demonstrated that the number of T2Rs in different squamates was closely related to their dietary habits, which was consistent with previous analyses of correlations between vertebrate T2R numbers and dietary habits.[7] This is why insectivorous lizards and geckos have a high number of T2R genes, which are needed for these species to recognize poisonous substances from insects by the perception of bitter taste. The selection pressure on the ERs, TMs, and IRs of T2Rs was found to be different in some primates and mammals. The ER was favored by positive selection pressure without exceptions.

In contrast, the TMs and IRs were more favored by purifying selection.[67–69] We analyzed the selection pressure in the ERs, TMs, and IRs of sauropsids using the paired YN00 model. We still obtained the same aforementioned pattern. The formation of this pattern may be related to the fact that ERs are predicted to be involved in ligand binding, such as tastants,[70,71] and positive selection exerted on this region may facilitate adaptation to different ligand binding. Critical changes in the ER region could affect the ability of species to recognize bitter substances or toxic substances.

There were various birth and death processes in sauropsids CYPs. Numerous identified CYPs in sauropsids were not orthologous to those in humans, and they may arise through a series of birth and death processes. The MIPhy instability score and the number of gene duplications and losses suggest that birth-death processes occur more frequently in D-type CYP genes. There was no significant difference in mean ω values between the D-type CYP and the B-type CYP genes. Still, it showed a correlation with MIPhy scores. We identified a similar number of D-type and B-type CYP genes under positive selection by implementing the branch-site model.

In contrast, positive selection branches were found more often among the D-type CYP genes using the adaptive branch-site random effects Likelihood (aBSREL) model. We anticipated a correlation between the number and diet in D-type CYPs, as observed in the T2Rs discussed in the previous section. However, only a weak correlation was detected in CYP1s. These results suggest that the interaction between natural selection and the birth and death processes of CYPs in sauropsids is complex and interactive. Still, birth and death processes cannot be entirely attributed to the effect of selection pressure, as previously found in CYPs from other groups.[72]

Sixty-six CBPs were identified in *G. japonicus* in the present study, located in the EDC region of chromosome 1 and varying in length. This agrees with previous work that found CBPs clustered in the same scaffold of the *G. japonicus* genome.[12] In this regard, the formation of specific epidermis in the reptiles comprises the CBPs coded within the specific region of the chromosome. Among the 66 CBPs, Pro, Cys, Gly, and Ser were the most abundant aa. The abundance of Gly and Cys significantly affects the flexibility and adhesiveness of gecko setae.[73] These results suggest that CBPs in *G. japonicus* are essential in forming setae in toepads.

Multiple sequence alignments revealed the existence of a conserved core box region in the CBPs of *G. japonicus*, which is well conserved in sauropsids and determines the polymerization of these CBP proteins into filaments.[74] We found that the CBP genes family in *G. japonicus* can be divided into three categories in general based on phylogeny, motif, domain, and gene structure. Whether these categories correspond to different functions remains to be further studied. Among the reptile species studied, phylogenetic relationship analysis indicated that their CBP genes were not clustered exactly according to species relationships, suggesting that CBP genes are independently duplicated in multiple reptile lineages, leading to variability in the epidermis of reptiles. Therefore, CBPs in different sauropsids contribute to forming their epidermis. Synteny analysis revealed a high similarity of CBPs in *G. japonicus* and other lizards, consistent with their morphological similarity, and species relationship.

The results of TRP genes identification in sauropsids are partly consistent with those of TRP analysis in recent study on tuatara,[52] and a certain degree of difference was acceptable considering the differences in methods. For example, *TRPA1* was not identified in *A. mississippiensis* or *S. punctatus*. However, it was placed in each of the selected species in our study. In phylogeny, each member of the other TRP families is well clustered into a monophyletic group. The exception regards *TRPV5* and *TRPV6*, which are thought to be generated by independent duplication in the ancestors of the respective groups of mammals, sauropsids, amphibians, and Chondrichthyes,[75] and a low-quality *TRPV4* sequence of *S. punctatus*.

In *G. japonicus*, *TRPV5*, *TRPM5*, and *TRPM8* were under relaxed purifying selection, while *TRPM3* and *PKD2L2* were under positive selection and *MCOLN1* was under intensified purifying selection. *TRPV5* is a channel that exhibits high selectivity for calcium ions, mediates calcium ion influx into cells, and is responsive to the activation and expression of hormones such as parathyroid hormone, estrogen, and testosterone.[76] The relaxation of purified selection suggests that regulation of the expression of these hormones undergoes some adaptive changes among *G. japonicus*. *MCOLN1* (*TRPML1*) has been revealed to be associated with lysosomal function in mammals. Its abnormalities can cause lysosomal channels and function disorders, leading to lysosomal dysfunction and lysosomal storage diseases. They may also be

responsible for the pathogenesis of metabolic and common neurodegenerative diseases.[77] The importance of lysosomal function in *G. japonicus* could make it more favored for purification selection. *PKD2L2* plays an important role in reproduction and is an essential gene for spermatogenesis,[78,79] suggesting that the reproductive system of *G. japonicus* has undergone adaptive evolution.

PKDREJ and *PKD2L2*, two reproduction-related genes, have higher ω values relative to other TRP genes, consistent with previous studies finding that reproduction-related proteins have higher evolutionary rates.[54–60] *TRPM3*, *TRPM5*, and *TRPM8* are thermosensitive TRP genes,[23] suggesting that their genomic changes may be related to the evolution of thermoregulation in *G. japonicus*. Compared to other reptile species, *G. japonicus* is relatively widely distributed. Various reasons can allow a species to adapt and survive in a wide range. Some thermosensitive TRPs under relaxed purifying or positive selection in *G. japonicus* imply that one critical factor is the ability to adapt to temperature changes.

Several researchers predicted the present and future distribution of *G. japonicus* in 2020. They found that temperature seasonality significantly predicted its distribution,[80] indicating the adaptive evolution process to temperature in *G. japonicus*. These thermosensitive TRPs under relaxed purifying or positive selection may have corresponding changes in the range and magnitude of the activation temperature due to the wide environmental temperature variation. Our results provide evidence for the temperature adaptation of *G. japonicus* and reveal a possible reason for its wide distribution.

Conclusion

In this study, we assembled and annotated a high-quality *G. japonicus* genome with 19 chromosomes based on Illumina sequencing, PacBio SMRT sequencing, and Hi-C technology. To our knowledge, this study represents the first chromosomal-level assembly of the gecko genome. Comparative genomic analyses revealed the phylogenetic position of sauropsids, important functional gene sets that are primarily associated with immune and sensory systems and enriched in related pathways, chromosomal rearrangement events, and chromosome evolution in *G. japonicus*, implying important molecular mechanisms for tail regeneration, damage repair, and sensitive senses in *G. japonicus*. The evolutionary trajectories of bitter T2Rs, detox and biosynthetic CYPs, epidermis formation CBPs, and thermoregulatory TRP channels in *G. japonicus* and other representative sauropsids were identified.

These results indicated that T2Rs expanded in different lineages by tandem gene duplication. The expansion and independent duplication events of T2Rs and positively selected branches were predominantly present in insectivorous species, revealing the essential relevance of T2Rs to the clearance of bitter toxins in geckos. D-type genes in the CYP family of *G. japonicus* have frequent duplication and loss events. This suggests they underwent more birth and death processes than B-type CYP genes. Pro, Cys, Gly, and Ser are the most abundant aa in 66 CBPs of *G. japonicus*, and the abundance of Gly and Cys in CBPs implies significant effects on the flexibility and setae adhesiveness of *G. japonicus*. Some thermosensitive TRPs were under relaxed purifying or positive selection in *G. japonicus*, suggesting that one of the critical factors enhances the ability to adapt to climate change.

In summary, we assembled a high-quality chromosome-level genome of *G. japonicus*. We performed comprehensive comparative genomics and evolutionary genomic analyses of *G. japonicus*. Our results revealed the regional characterization and evolutionary contribution of EBRs, the critical relevance of T2Rs and CYPs in the detoxification mechanism of insectivorous species, and the possible role of CBPs in phenotypic adaptation evolution of toepads and TRPs in temperature perception and regional expansion in gekkotans.

Limitations of the study

We included only limited chromosomal-level genome data from reptiles in this study, and it is challenging to perform more accurate genes identification and evolutionary analyses. More substantial follow-up work (e.g., gene function verification experiment *in vivo* and *in vitro*) is needed in the future to provide greater advance over previous work.

STAR★METHODS

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

- KEY RESOURCES TABLE
- RESOURCE AVAILABILITY
  - Lead contact
  - Materials availability
  - Data and code availability
- EXPERIMENTAL MODEL AND STUDY PARTICIPANT DETAILS
  - Animal materials
- METHOD DETAILS
  - Genomic sequencing, RNA-Seq and Hi-C
  - Genome assembly and annotation
  - Synteny analysis and identification of HSBs and EBRs
  - Phylogenetic and gene family analysis
  - T2R analysis
  - CYP analysis

SUPPLEMENTAL INFORMATION

Supplemental information can be found online at https://doi.org/10.1016/j.isci.2023.108445.

ACKNOWLEDGMENTS

This work was supported by grants from the National Natural Science Foundation of China (Grants No. 32370440 to P.L. and No. 31672269 to J.Y.), Postgraduate Research & Practice Innovation Program of Jiangsu Province (KYCX22_1614) awarded to Z.Y.C., and the Priority Academic Program Development of Jiangsu Higher Education Institutions (PAPD).

AUTHOR CONTRIBUTIONS

P.L. and J.Y. conceived and designed the experiments. Y.W. and Y.Y. performed the experiments. Y.W., Y.Y., C.L., Z.C., Y.C., C.H., Y.Q., and H.L. analyzed the data and prepared all figures and tables. Y.W., Y.Y., K.Z., J.Y., and P.L. wrote and revised the manuscript. All authors discussed the results and implications and commented on the manuscript.

DECLARATION OF INTERESTS

The authors declare no competing interests.

INCLUSION AND DIVERSITY

We support inclusive, diverse, and equitable conduct of research.

Received: June 5, 2023
Revised: September 5, 2023
Accepted: November 9, 2023
Published: November 14, 2023

REFERENCES

1. Toda, M., and Yoshida, T. (2005). Issues and perspectives regarding invasive alien species of amphibians and reptiles in Japan. Bull. Herpetol. Soc. Jpn. *2005*, 139–149.
2. Luu, V.Q., Nguyen, T.Q., Le, M.D., Bonkowski, M., and Ziegler, T. (2017). A new karst dwelling species of the *Gekko japonicus* group (Squamata: Gekkonidae) from central Laos. Zootaxa *4263*, 179–193.
3. Kim, D.-I., Park, I.-K., Ota, H., Fong, J.J., Kim, J.-S., Zhang, Y.-P., Li, S.-R., Choi, W.-J., and Park, D. (2019). Patterns of morphological variation in the Schlegel's Japanese gecko (*Gekko japonicus*) across populations in China, Japan, and Korea. J. Ecol. Environ. *43*, 34.
4. Shibaike, Y., Takahashi, Y., Arikura, I., Iizumi, R., Kitakawa, S., Sakai, M., Imaoka, C., Shiro, H., Tanaka, H., Akakubo, N., et al. (2009). Chromosome evolution in the lizard genus *Gekko* (Gekkonidae, Squamata, Reptilia) in the East Asian islands. Cytogenet. Genome Res. *127*, 182–190.
5. Qin, X.-M., Li, H.-M., Zeng, Z.-H., Zeng, D.-L., and Guan, Q.-X. (2012). Genetic variation and differentiation of *Gekko gecko* from different populations based on mitochondrial cytochrome b gene sequences and karyotypes. Zoolog. Sci. *29*, 384–389.
6. Fan, H., Wu, Q., Wei, F., Yang, F., Ng, B.L., and Hu, Y. (2019). Chromosome-level genome assembly for giant panda provides novel insights into Carnivora chromosome evolution. Genome Biol. *20*, 267.
7. Li, D., and Zhang, J. (2014). Diet shapes the evolution of the vertebrate bitter taste receptor gene repertoire. Mol. Biol. Evol. *31*, 303–309.
8. Zhong, H., Shang, S., Zhang, H., Chen, J., Wu, X., and Zhang, H. (2019). Characterization and phylogeny of bitter taste receptor genes (Tas2r) in Squamata. Genetica *147*, 131–139.
9. Nelson, D.R., Goldstone, J.V., and Stegeman, J.J. (2013). The cytochrome P450 genesis locus: the origin and evolution of animal cytochrome P450s. Philos. Trans. R. Soc. Lond. B Biol. Sci. *368*, 20120474.
10. Kawashima, A., and Satta, Y. (2014). Substrate-dependent evolution of cytochrome P450: Rapid turnover of the detoxification-type and conservation of the biosynthesis-type. PLoS One *9*, e100059.
11. Nebert, D.W., and Dalton, T.P. (2006). The role of cytochrome P450 enzymes in endogenous signalling pathways and environmental carcinogenesis. Nat. Rev. Cancer *6*, 947–960.
12. Liu, Y., Zhou, Q., Wang, Y., Luo, L., Yang, J., Yang, L., Liu, M., Li, Y., Qian, T., Zheng, Y., et al. (2015). *Gekko japonicus* genome reveals evolution of adhesive toe pads and tail regeneration. Nat. Commun. *6*, 10033.
13. Hitchcock, M.A., and McBrayer, L.D. (2006). Thermoregulation in nocturnal ecthotherms: seasonal and intraspecific variation in the mediterranean gecko (*Hemidactylus turcicus*). J. Herpetol. *40*, 185–195.
14. Kim, D.-I., Park, I.-K., Kim, J.-S., Ota, H., Choi, W.-J., Kim, I.-H., and Park, D. (2019). Spring and summer microhabitat use by Schlegel's Japanese gecko, *Gekko japonicus* (Reptilia: Squamata: Gekkonidae), in urban areas. Anim. Cells Syst. *23*, 64–70.
15. Venkatachalam, K., and Montell, C. (2007). TRP channels. Annu. Rev. Biochem. *76*, 387–417.
16. Nilius, B., and Owsianik, G. (2011). The transient receptor potential family of ion channels. Genome Biol. *12*, 218.
17. Himmel, N.J., and Cox, D.N. (2020). Transient receptor potential channels: current perspectives on evolution, structure, function and nomenclature. Proc. Biol. Sci. *287*, 20201309.
18. Clapham, D.E., Runnels, L.W., and Strübing, C. (2001). The TRP ion channel family. Nat. Rev. Neurosci. *2*, 387–396.
19. Montell, C., Birnbaumer, L., Flockerzi, V., Bindels, R.J., Bruford, E.A., Caterina, M.J., Clapham, D.E., Harteneck, C., Heller, S., Julius, D., et al. (2002). A

unified nomenclature for the superfamily of TRP cation channels. Mol. Cell *9*, 229–231.
20. Yu, F.H., and Catterall, W.A. (2004). The VGL-chanome: a protein superfamily specialized for electrical signaling and ionic homeostasis. Sci. STKE *2004*, re15.
21. Clapham, D.E., Julius, D., Montell, C., and Schultz, G. (2005). International Union of Pharmacology. XLIX. Nomenclature and structure-function relationships of transient receptor potential channels. Pharmacol. Rev. *57*, 427–450.
22. Voets, T., Droogmans, G., Wissenbach, U., Janssens, A., Flockerzi, V., and Nilius, B. (2004). The principle of temperature-dependent gating in cold- and heat-sensitive TRP channels. Nature *430*, 748–754.
23. Wang, H., and Siemens, J. (2015). TRP ion channels in thermosensation, thermoregulation and metabolism. Temperature (Austin) *2*, 178–187.
24. Xiao, C.-L., Chen, Y., Xie, S.-Q., Chen, K.-N., Wang, Y., Han, Y., Luo, F., and Xie, Z. (2017). MECAT: fast mapping, error correction, and de novo assembly for single-molecule sequencing reads. Nat. Methods *14*, 1072–1074.
25. Walker, B.J., Abeel, T., Shea, T., Priest, M., Abouelliel, A., Sakthikumar, S., Cuomo, C.A., Zeng, Q., Wortman, J., Young, S.K., and Earl, A.M. (2014). Pilon: an integrated tool for comprehensive microbial variant detection and genome assembly improvement. PLoS One *9*, e112963.
26. Durand, N.C., Shamim, M.S., Machol, I., Rao, S.S.P., Huntley, M.H., Lander, E.S., and Aiden, E.L. (2016). Juicer provides a one-click system for analyzing loop-resolution Hi-C experiments. Cell Syst. *3*, 95–98.
27. Benson, G. (1999). Tandem repeats finder: a program to analyze DNA sequences. Nucleic Acids Res. *27*, 573–580.
28. Smit, A.F., and Hubley, R. (2008). RepeatModeler Open-1.0.8. http://www.repeatmasker.org/RepeatModeler.html.
29. Tarailo-Graovac, M., and Chen, N. (2009). Using RepeatMasker to identify repetitive elements in genomic sequences. Curr. Protoc. Bioinformatics *Chapter 4*, 4.10.1–4.10.14.
30. Price, A.L., Jones, N.C., and Pevzner, P.A. (2005). De novo identification of repeat families in large genomes. Bioinformatics *21* (*Suppl 1*), i351–i358.
31. Emms, D.M., and Kelly, S. (2015). OrthoFinder: solving fundamental biases in whole genome comparisons dramatically improves orthogroup inference accuracy. Genome Biol. *16*, 157.
32. Daza, J.D., Bauer, A.M., and Snively, E.D. (2014). On the fossil record of the Gekkota. Anat. Rec. *297*, 433–462.
33. Chen, G., Shaw, M.H., Kim, Y.G., and Nuñez, G. (2009). NOD-like receptors: role in innate immunity and inflammatory disease. Annu. Rev. Pathol. *4*, 365–398.
34. Platnich, J.M., and Muruve, D.A. (2019). NOD-like receptors and inflammasomes: A review of their canonical and non-canonical signaling pathways. Arch. Biochem. Biophys. *670*, 4–14.
35. Chen, G., and Goeddel, D.V. (2002). TNF-R1 signaling: a beautiful pathway. Science *296*, 1634–1635.
36. Mathew, S.J., Haubert, D., Krönke, M., and Leptin, M. (2009). Looking beyond death: a morphogenetic role for the TNF signalling pathway. J. Cell Sci. *122*, 1939–1946.
37. Behrens, M., Prandi, S., and Meyerhof, W. (2017). Taste Receptor Gene Expression Outside the Gustatory System. In Taste and Smell, D. Krautwurst, ed. (Springer International Publishing), pp. 1–34.
38. Bachmanov, A.A., and Beauchamp, G.K. (2007). Taste receptor genes. Annu. Rev. Nutr. *27*, 389–414.
39. Yang, Z. (2007). PAML 4: phylogenetic analysis by maximum likelihood. Mol. Biol. Evol. *24*, 1586–1591.
40. Kosakovsky Pond, S.L., Poon, A.F.Y., Velazquez, R., Weaver, S., Hepler, N.L., Murrell, B., Shank, S.D., Magalis, B.R., Bouvier, D., Nekrutenko, A., et al. (2020). HyPhy 2.5-A customizable platform for evolutionary hypothesis testing using phylogenies. Mol. Biol. Evol. *37*, 295–299.
41. Kosakovsky Pond, S.L., and Frost, S.D.W. (2005). Not so different after all: a comparison of methods for detecting amino acid sites under selection. Mol. Biol. Evol. *22*, 1208–1222.
42. Murrell, B., Moola, S., Mabona, A., Weighill, T., Sheward, D., Kosakovsky Pond, S.L., and Scheffler, K. (2013). FUBAR: a fast, unconstrained bayesian AppRoximation for inferring selection. Mol. Biol. Evol. *30*, 1196–1205.
43. Curran, D.M., Gilleard, J.S., and Wasmuth, J.D. (2018). MIPhy: identify and quantify rapidly evolving members of large gene families. PeerJ *6*, e4873.
44. Hallahan, D.L., Keiper-Hrynko, N.M., Shang, T.Q., Ganzke, T.S., Toni, M., Dalla Valle, L., and Alibardi, L. (2009). Analysis of gene expression in gecko digital adhesive pads indicates significant production of cysteine- and glycine-rich beta-keratins. J. Exp. Zool. B Mol. Dev. Evol. *312*, 58–73.
45. Dalla Valle, L., Nardi, A., Bonazza, G., Zucal, C., Emera, D., and Alibardi, L. (2010). Forty keratin-associated beta-proteins (beta-keratins) form the hard layers of scales, claws, and adhesive pads in the green anole lizard, *Anolis carolinensis*. J. Exp. Zool. B Mol. Dev. Evol. *314*, 11–32.
46. Holthaus, K.B., Strasser, B., Lachner, J., Sukseree, S., Sipos, W., Weissenbacher, A., Tschachler, E., Alibardi, L., and Eckhart, L. (2018). Comparative analysis of epidermal differentiation genes of crocodilians suggests new models for the evolutionary origin of avian feather proteins. Genome Biol. Evol. *10*, 694–704.
47. Holthaus, K.B., Mlitz, V., Strasser, B., Tschachler, E., Alibardi, L., and Eckhart, L. (2017). Identification and comparative analysis of the epidermal differentiation complex in snakes. Sci. Rep. *7*, 45338.
48. Hara, Y., Takeuchi, M., Kageyama, Y., Tatsumi, K., Hibi, M., Kiyonari, H., and Kuraku, S. (2018). Madagascar ground gecko genome analysis characterizes asymmetric fates of duplicated genes. BMC Biol. *16*, 40.
49. Gamble, T. (2019). Duplications in corneous beta protein genes and the evolution of gecko adhesion. Integr. Comp. Biol. *59*, 193–202.
50. Li, Y.I., Kong, L., Ponting, C.P., and Haerty, W. (2013). Rapid evolution of beta-keratin genes contribute to phenotypic differences that distinguish turtles and birds from other reptiles. Genome Biol. Evol. *5*, 923–933.
51. Jurka, J., Kapitonov, V.V., Pavlicek, A., Klonowski, P., Kohany, O., and Walichiewicz, J. (2005). Repbase Update, a database of eukaryotic repetitive elements. Cytogenet. Genome Res. *110*, 462–467.
52. Gemmell, N.J., Rutherford, K., Prost, S., Tollis, M., Winter, D., Macey, J.R., Adelson, D.L., Suh, A., Bertozzi, T., Grau, J.H., et al. (2020). The tuatara genome reveals ancient features of amniote evolution. Nature *584*, 403–409.
53. Vicens, A., Gómez Montoto, L., Couso-Ferrer, F., Sutton, K.A., and Roldan, E.R.S. (2015). Sexual selection and the adaptive evolution of PKDREJ protein in primates and rodents. Mol. Hum. Reprod. *21*, 146–156.
54. Swanson, W.J., Clark, A.G., Waldrip-Dail, H.M., Wolfner, M.F., and Aquadro, C.F. (2001). Evolutionary EST analysis identifies rapidly evolving male reproductive proteins in *Drosophila*. Proc. Natl. Acad. Sci. USA *98*, 7375–7379.
55. Swanson, W.J., and Vacquier, V.D. (2002). The rapid evolution of reproductive proteins. Nat. Rev. Genet. *3*, 137–144.
56. Clark, N.L., Aagaard, J.E., and Swanson, W.J. (2006). Evolution of reproductive proteins from animals and plants. Reproduction *131*, 11–22.
57. Walters, J.R., and Harrison, R.G. (2010). Combined EST and proteomic analysis identifies rapidly evolving seminal fluid proteins in *Heliconius* butterflies. Mol. Biol. Evol. *27*, 2000–2013.
58. Walters, J.R., and Harrison, R.G. (2011). Decoupling of rapid and adaptive evolution among seminal fluid proteins in *Heliconius* butterflies with divergent mating systems. Evolution *65*, 2855–2871.
59. Wilburn, D.B., and Swanson, W.J. (2016). From molecules to mating: Rapid evolution and biochemical studies of reproductive proteins. J. Proteomics *135*, 12–25.
60. Twort, V.G., Dennis, A.B., Park, D., Lomas, K.F., Newcomb, R.D., and Buckley, T.R. (2017). Positive selection and comparative molecular evolution of reproductive proteins from New Zealand tree weta (Orthoptera, *Hemideina*). PLoS One *12*, e0188147.
61. Murphy, W.J., Larkin, D.M., Everts-van der Wind, A., Bourque, G., Tesler, G., Auvil, L., Beever, J.E., Chowdhary, B.P., Galibert, F., Gatzke, L., et al. (2005). Dynamics of mammalian chromosome evolution inferred from multispecies comparative maps. Science *309*, 613–617.
62. Bovine Genome Sequencing and Analysis Consortium, Elsik, C.G., Tellam, R.L., Worley, K.C., Gibbs, R.A., Muzny, D.M., Weinstock, G.M., Adelson, D.L., Eichler, E.E., Guigó, R., et al. (2009). The genome sequence of taurine cattle: a window to ruminant biology and evolution. Science *324*, 522–528.
63. Larkin, D.M., Pape, G., Donthu, R., Auvil, L., Welge, M., and Lewin, H.A. (2009). Breakpoint regions and homologous synteny blocks in chromosomes have different evolutionary histories. Genome Res. *19*, 770–777.
64. Groenen, M.A.M., Archibald, A.L., Uenishi, H., Tuggle, C.K., Takeuchi, Y., Rothschild, M.F., Rogel-Gaillard, C., Park, C., Milan, D., Megens, H.J., et al. (2012). Analyses of pig

genomes provide insight into porcine demography and evolution. Nature *491*, 393–398.
65. Farré, M., Narayan, J., Slavov, G.T., Damas, J., Auvil, L., Li, C., Jarvis, E.D., Burt, D.W., Griffin, D.K., and Larkin, D.M. (2016). Novel insights into chromosome evolution in birds, archosaurs, and reptiles. Genome Biol. Evol. *8*, 2442–2451.
66. Dong, D., Jones, G., and Zhang, S. (2009). Dynamic evolution of bitter taste receptor genes in vertebrates. BMC Evol. Biol. *9*, 12.
67. Shi, P., Zhang, J., Yang, H., and Zhang, Y.P. (2003). Adaptive diversification of bitter taste receptor genes in mammalian evolution. Mol. Biol. Evol. *20*, 805–814.
68. Dong, X., Liang, Q., Li, J., and Feng, P. (2021). Positive selection drives the evolution of a primate bitter taste receptor gene. Ecol. Evol. *11*, 5459–5467.
69. Strotmann, R., Schröck, K., Böselt, I., Stäubert, C., Russ, A., and Schöneberg, T. (2011). Evolution of GPCR: change and continuity. Mol. Cell. Endocrinol. *331*, 170–178.
70. Adler, E., Hoon, M.A., Mueller, K.L., Chandrashekar, J., Ryba, N.J., and Zuker, C.S. (2000). A novel family of mammalian taste receptors. Cell *100*, 693–702.
71. Gilbertson, T.A., Damak, S., and Margolskee, R.F. (2000). The molecular physiology of taste transduction. Curr. Opin. Neurobiol. *10*, 519–527.
72. Dermauw, W., Van Leeuwen, T., and Feyereisen, R. (2020). Diversity and evolution of the P450 family in arthropods. Insect Biochem. Mol. Biol. *127*, 103490.
73. Alibardi, L. (2013). Immunolocalization of keratin-associated beta-proteins (beta-keratins) in pad lamellae of geckos suggest that glycine-cysteine-rich proteins contribute to their flexibility and adhesiveness. J. Exp. Zool. A Ecol. Genet. Physiol. *319*, 166–178.
74. Alibardi, L., Dalla Valle, L., Nardi, A., and Toni, M. (2009). Evolution of hard proteins in the sauropsid integument in relation to the cornification of skin derivatives in amniotes. J. Anat. *214*, 560–586.
75. Flores-Aldama, L., Vandewege, M.W., Zavala, K., Colenso, C.K., Gonzalez, W., Brauchi, S.E., and Opazo, J.C. (2020). Evolutionary analyses reveal independent origins of gene repertoires and structural motifs associated to fast inactivation in calcium-selective TRPV channels. Sci. Rep. *10*, 8684.
76. Na, T., and Peng, J.B. (2014). TRPV5: a Ca(2+) channel for the fine-tuning of Ca(2+) reabsorption. Handb. Exp. Pharmacol. *222*, 321–357.
77. Xu, H., and Ren, D. (2015). Lysosomal Physiology. Annu. Rev. Physiol. *77*, 57–80.
78. Chen, Y., Zhang, Z., Lv, X.Y., Wang, Y.D., Hu, Z.G., Sun, H., Tan, R.Z., Liu, Y.H., Bian, G.H., Xiao, Y., et al. (2008). Expression of *Pkd2l2* in testis is implicated in spermatogenesis. Biol. Pharm. Bull. *31*, 1496–1500.
79. Dolebo, A.T., Khayatzadeh, N., Melesse, A., Wragg, D., Rekik, M., Haile, A., Rischkowsky, B., Rothschild, M.F., and Mwacharo, J.M. (2019). Genome-wide scans identify known and novel regions associated with prolificacy and reproduction traits in a sub-Saharan African indigenous sheep (*Ovis aries*). Mamm. Genome *30*, 339–352.
80. Kim, D.-I., Park, I.-K., Bae, S.-Y., Fong, J.J., Zhang, Y.-P., Li, S.-R., Ota, H., Kim, J.-S., and Park, D. (2020). Prediction of present and future distribution of the Schlegel's Japanese gecko (*Gekko japonicus*) using MaxEnt modeling. J. Ecol. Environ. *44*, 5.
81. Xu, Z., and Wang, H. (2007). LTR_FINDER: an efficient tool for the prediction of full-length LTR retrotransposons. Nucleic Acids Res. *35*, W265–W268.
82. Cantarel, B.L., Korf, I., Robb, S.M.C., Parra, G., Ross, E., Moore, B., Holt, C., Sánchez Alvarado, A., and Yandell, M. (2008). MAKER: an easy-to-use annotation pipeline designed for emerging model organism genomes. Genome Res. *18*, 188–196.
83. Slater, G.S.C., and Birney, E. (2005). Automated generation of heuristics for biological sequence comparison. BMC Bioinf. *6*, 31.
84. Stanke, M., and Waack, S. (2003). Gene prediction with a hidden Markov model and a new intron submodel. Bioinformatics *19* (*Suppl 2*), ii215–ii225.
85. Burge, C., and Karlin, S. (1997). Prediction of complete gene structures in human genomic DNA. J. Mol. Biol. *268*, 78–94.
86. Lowe, T.M., and Eddy, S.R. (1997). tRNAscan-SE: a program for improved detection of transfer RNA genes in genomic sequence. Nucleic Acids Res. *25*, 955–964.
87. Camacho, C., Coulouris, G., Avagyan, V., Ma, N., Papadopoulos, J., Bealer, K., and Madden, T.L. (2009). BLAST+: architecture and applications. BMC Bioinf. *10*, 421.
88. Griffiths-Jones, S., Moxon, S., Marshall, M., Khanna, A., Eddy, S.R., and Bateman, A. (2005). Rfam: annotating non-coding RNAs in complete genomes. Nucleic Acids Res. *33*, D121–D124.
89. Wang, Y., Tang, H., DeBarry, J.D., Tan, X., Li, J., Wang, X., Lee, T.-h., Jin, H., Marler, B., Guo, H., et al. (2012). MCScanX: a toolkit for detection and evolutionary analysis of gene synteny and collinearity. Nucleic Acids Res. *40*, e49.
90. Quinlan, A.R. (2014). BEDTools: the Swiss-army tool for genome feature analysis. Curr. Protoc. Bioinformatics *47*, 11.12.1–11.12.34.
91. Krzywinski, M., Schein, J., Birol, I., Connors, J., Gascoyne, R., Horsman, D., Jones, S.J., and Marra, M.A. (2009). Circos: an information aesthetic for comparative genomics. Genome Res. *19*, 1639–1645.
92. Soderlund, C., Nelson, W., Shoemaker, A., and Paterson, A. (2006). SyMAP: A system for discovering and viewing syntenic regions of FPC maps. Genome Res. *16*, 1159–1168.
93. Wu, T., Hu, E., Xu, S., Chen, M., Guo, P., Dai, Z., Feng, T., Zhou, L., Tang, W., Zhan, L., et al. (2021). clusterProfiler 4.0: A universal enrichment tool for interpreting omics data. Innovation *2*, 100141.
94. Li, L., Stoeckert, C.J., Jr., and Roos, D.S. (2003). OrthoMCL: identification of ortholog groups for eukaryotic genomes. Genome Res. *13*, 2178–2189.
95. Edgar, R.C. (2004). MUSCLE: multiple sequence alignment with high accuracy and high throughput. Nucleic Acids Res. *32*, 1792–1797.
96. Stamatakis, A. (2014). RAxML version 8: a tool for phylogenetic analysis and post-analysis of large phylogenies. Bioinformatics *30*, 1312–1313.
97. Kumar, S., Stecher, G., Suleski, M., and Hedges, S.B. (2017). TimeTree: a resource for timelines, timetrees, and divergence times. Mol. Biol. Evol. *34*, 1812–1819.
98. Sanderson, M.J. (2003). r8s: inferring absolute rates of molecular evolution and divergence times in the absence of a molecular clock. Bioinformatics *19*, 301–302.
99. De Bie, T., Cristianini, N., Demuth, J.P., and Hahn, M.W. (2006). CAFE: a computational tool for the study of gene family evolution. Bioinformatics *22*, 1269–1271.
100. Keilwagen, J., Wenk, M., Erickson, J.L., Schattat, M.H., Grau, J., and Hartung, F. (2016). Using intron position conservation for homology-based gene prediction. Nucleic Acids Res. *44*, e89.
101. Krogh, A., Larsson, B., von Heijne, G., and Sonnhammer, E.L. (2001). Predicting transmembrane protein topology with a hidden Markov model: application to complete genomes. J. Mol. Biol. *305*, 567–580.
102. Ranwez, V., Douzery, E.J.P., Cambon, C., Chantret, N., and Delsuc, F. (2018). MACSE v2: toolkit for the alignment of coding sequences accounting for frameshifts and stop codons. Mol. Biol. Evol. *35*, 2582–2584.
103. Kumar, S., Stecher, G., Li, M., Knyaz, C., and Tamura, K. (2018). MEGA X: molecular evolutionary genetics analysis across computing platforms. Mol. Biol. Evol. *35*, 1547–1549.
104. Kozlov, A.M., Darriba, D., Flouri, T., Morel, B., and Stamatakis, A. (2019). RAxML-NG: a fast, scalable and user-friendly tool for maximum likelihood phylogenetic inference. Bioinformatics *35*, 4453–4455.
105. Darriba, D., Posada, D., Kozlov, A.M., Stamatakis, A., Morel, B., and Flouri, T. (2020). ModelTest-NG: a new and scalable tool for the selection of DNA and protein evolutionary models. Mol. Biol. Evol. *37*, 291–294.
106. Huerta-Cepas, J., Serra, F., and Bork, P. (2016). ETE 3: reconstruction, analysis, and visualization of phylogenomic data. Mol. Biol. Evol. *33*, 1635–1638.
107. Letunic, I., and Bork, P. (2021). Interactive Tree Of Life (iTOL) v5: an online tool for phylogenetic tree display and annotation. Nucleic Acids Res. *49*, W293–W296.
108. Yu, G., Smith, D.K., Zhu, H., Guan, Y., and Lam, T.T.Y. (2017). ggtree: an R package for visualization and annotation of phylogenetic trees with their covariates and other associated data. Methods Ecol. Evol. *8*, 28–36.
109. Katoh, K., and Standley, D.M. (2013). MAFFT multiple sequence alignment software version 7: improvements in performance and usability. Mol. Biol. Evol. *30*, 772–780.
110. Chen, K., Durand, D., and Farach-Colton, M. (2000). Notung: Dating gene duplications using gene family trees. J. Comput. Biol.: a Journal of Computational Molecular Cell Biology *7*, 96–106.
111. Paradis, E., and Schliep, K. (2019). ape 5.0: an environment for modern phylogenetics and evolutionary analyses in R. Bioinformatics *35*, 526–528.
112. Chen, C., Chen, H., Zhang, Y., Thomas, H.R., Frank, M.H., He, Y., and Xia, R. (2020). TBtools: an integrative toolkit developed for interactive analyses of big biological data. Mol. Plant *13*, 1194–1202.
113. Sela, I., Ashkenazy, H., Katoh, K., and Pupko, T. (2015). GUIDANCE2: accurate detection of unreliable alignment regions accounting for the uncertainty of multiple parameters. Nucleic Acids Res. *43*, W7–W14.
114. Storey, J.D. (2002). A direct approach to false discovery rates. J. Roy. Stat. Soc. B *64*, 479–498.

115. Simão, F.A., Waterhouse, R.M., Ioannidis, P., Kriventseva, E.V., and Zdobnov, E.M. (2015). BUSCO: assessing genome assembly and annotation completeness with single-copy orthologs. Bioinformatics *31*, 3210–3212.
116. Kanehisa, M., and Goto, S. (2000). KEGG: kyoto encyclopedia of genes and genomes. Nucleic Acids Res. *28*, 27–30.
117. Mitchell, A.L., Attwood, T.K., Babbitt, P.C., Blum, M., Bork, P., Bridge, A., Brown, S.D., Chang, H.Y., El-Gebali, S., Fraser, M.I., et al. (2019). InterPro in 2019: improving coverage, classification and access to protein sequence annotations. Nucleic Acids Res. *47*, D351–D360.
118. Harris, M.A., Clark, J., Ireland, A., Lomax, J., Ashburner, M., Foulger, R., Eilbeck, K., Lewis, S., Marshall, B., Mungall, C., et al. (2004). The Gene Ontology (GO) database and informatics resource. Nucleic Acids Res. *32*, D258–D261.
119. Marchler-Bauer, A., Derbyshire, M.K., Gonzales, N.R., Lu, S., Chitsaz, F., Geer, L.Y., Geer, R.C., He, J., Gwadz, M., Hurwitz, D.I., et al. (2015). CDD: NCBI's conserved domain database. Nucleic Acids Res. *43*, D222–D226.
120. Boeckmann, B., Bairoch, A., Apweiler, R., Blatter, M.C., Estreicher, A., Gasteiger, E., Martin, M.J., Michoud, K., O'Donovan, C., Phan, I., et al. (2003). The SWISS-PROT protein knowledgebase and its supplement TrEMBL in 2003. Nucleic Acids Res. *31*, 365–370.
121. Schott, R.K., Refvik, S.P., Hauser, F.E., López-Fernández, H., and Chang, B.S.W. (2014). Divergent positive selection in rhodopsin from lake and riverine cichlid fishes. Mol. Biol. Evol. *31*, 1149–1165.
122. Schott, R.K., Van Nynatten, A., Card, D.C., Castoe, T.A., and S W Chang, B. (2018). Shifts in selective pressures on snake phototransduction genes associated with photoreceptor transmutation and dim-light ancestry. Mol. Biol. Evol. *35*, 1376–1389.
123. Smith, M.D., Wertheim, J.O., Weaver, S., Murrell, B., Scheffler, K., and Kosakovsky Pond, S.L. (2015). Less is more: an adaptive branch-site random effects model for efficient detection of episodic diversifying selection. Mol. Biol. Evol. *32*, 1342–1353.
124. UniProt Consortium (2015). UniProt: a hub for protein information. Nucleic Acids Res. *43*, D204–D212.
125. Nelson, D.R. (2009). The cytochrome p450 homepage. Hum. Genomics *4*, 59–65.
126. Capella-Gutiérrez, S., Silla-Martínez, J.M., and Gabaldón, T. (2009). trimAl: a tool for automated alignment trimming in large-scale phylogenetic analyses. Bioinformatics *25*, 1972–1973.
127. Löytynoja, A., and Goldman, N. (2008). Phylogeny-aware gap placement prevents errors in sequence alignment and evolutionary analysis. Science *320*, 1632–1635.

STAR★METHODS

KEY RESOURCES TABLE

REAGENT or RESOURCE	SOURCE	IDENTIFIER
Biological samples		
Gekko japonicus	This paper	NA
Critical commercial assays		
DNA Mini Kit	Qiagen	NA
DNA Fragmentase kit V2.0	Annoroad	NA
DNA Library Prep Kit V2.0	Annoroad	NA
Qubit fluorimeter	Thermofifisher	NA
NanoPhotometer	IMPLEN	NA
Deposited data		
DNA-seq data	This paper	SRA: SRR22243198
PacBio sequence data	This paper	SRA: SRR22262064
RNA-seq data	This paper	SRA: SRR22574970
Hi-C sequence data	This paper	SRA: SRR22540941
Predicted gene sequences, sequence alignments and phylogenetic trees	This paper	https://doi.org/10.5281/zenodo.7211720
Gekko japonicus genome	This paper	https://doi.org/10.5281/zenodo.7394737
Software and algorithms		
MECAT2	Xiao et al.[24]	https://github.com/xiaochuanle/MECAT2
GCpp v1.9.0	PacBio	https://github.com/PacificBiosciences/gcpp
Pilon v1.22	Walker et al.[25]	https://github.com/broadinstitute/pilon
juicer v1.6.2	Durand et al.[26]	https://github.com/aidenlab/juicer
RepeatMasker v4.0.9	Tarailo-Graovac et al.[29,30]	http://repeatmasker.org/
RepeatModeler v1.0.11	Smit et al.[29,30]	http://www.repeatmasker.org/RepeatModeler/
LTR-FINDER v1.0.5	Xu et al.[81]	https://github.com/xzhub/LTR_Finder
TRF v4.09	Benson et al.[27]	https://tandem.bu.edu/trf/trf.html
Maker v3.01.02	Cantarel et al.[82]	https://github.com/Yandell-Lab/maker
Exonerate v2.2.0	Slater et al.[83]	https://www.ebi.ac.uk/about/vertebrate-genomics/software/exonerate
Augustus v3.3	Stanke et al.[84]	http://augustus.gobics.de/
Genscan v1.0	Burge et al.[85]	http://hollywood.mit.edu/GENSCAN.html
tRNAscan-se v1.3.1	Lowe et al.[86]	https://github.com/UCSC-LoweLab/tRNAscan-SE
BLAST+ v2.6.0	Camacho et al.[87]	https://ftp.ncbi.nlm.nih.gov/blast/executables/blast+/LATEST/
Rfam v14.1	Griffiths-Jones et al.[88]	https://rfam.org
OrthoFinder v2.5.1	Emms et al.[31]	https://github.com/davidemms/OrthoFinder
MCScanX	Wang et al.[89]	https://github.com/wyp1125/MCScanX
BEDTools v2.29.2	Quinlan et al.[90]	https://github.com/arq5x/bedtools2
Circos v0.69	Krzywinski et al.[91]	http://circos.ca/software/download/circos/
Symap v5.0.6	Soderlund et al.[92]	http://www.agcol.arizona.edu/software/symap/
clusterProfiler v4.0	Wu et al.[93]	https://bioconductor.org/packages/release/bioc/html/clusterProfiler.html
OrthoMcl v1.4	Li et al.[94]	https://help.rc.ufl.edu/doc/OrthoMCL
Muscle v3.8.31	Edgar et al.[95]	https://github.com/rcedgar/muscle/

(*Continued on next page*)

Continued

REAGENT or RESOURCE	SOURCE	IDENTIFIER
RAXML v8	Stamatakis et al.[96]	https://github.com/stamatak/standard-RAxML
TimeTree	Kumar et al.[97]	http://timetree.org/
PAML v4.7	Yang et al.[39]	http://abacus.gene.ucl.ac.uk/software/paml.html
R8s v1.50	Sanderson et al.[98]	https://sourceforge.net/projects/r8s/
CAFÉ v4.2.1	De Bie et al.[99]	https://github.com/hahnlab/CAFE
GeMoMa v1.7.1	Keilwagen et al.[100]	http://www.jstacs.de/index.php/GeMoMa
TMHMM v2.0	Krogh et al.[101]	https://services.healthtech.dtu.dk/service.php?TMHMM-2.0
MACSE v2.05	Ranwez et al.[102]	https://www.agap-ge2pop.org/macsee-pipelines
Mega-X v10.0.5	Kumar et al.[103]	https://www.megasoftware.net/dload_win_gui
RAxML-NG v0.9.0	Kozlov et al.[104]	https://github.com/amkozlov/raxml-ng
ModelTest-NG v0.1.7	Darriba et al.[105]	https://github.com/ddarriba/modeltest
ETE 3	Huerta-Cepas et al.[106]	http://etetoolkit.org
iTOL	Letunic et al.[107]	https://itol.embl.de
ggtree	Yu et al.[108]	https://bioconductor.org/packages/release/bioc/html/ggtree.html
HyPhy v2.5.2	Kosakovsky Pond et al.[40]	http://www.hyphy.org
MAFFT v7.475	Katoh et al.[109]	https://mafft.cbrc.jp/alignment/software/
NOTUNG v3.0.26	Chen et al.[110]	https://www.cs.cmu.edu/~durand/Notung
ape	Paradis et al.[111]	https://cran.r-project.org/web/packages/ape/index.html
TBtools v1.074	Chen et al.[112]	https://github.com/CJ-Chen/TBtools/releases
GUIDANCE2	Sela et al.[113]	http://guidance.tau.ac.il/
MIPhy	Curran et al.[43]	http://miphy.wasmuthlab.org/
Python version 3.8.1	Python Software Foundation	https://www.python.org/
R version 4.1.0	R Software Foundation	https://www.r-project.org/
q-value	Storey et al.[114]	https://github.com/StoreyLab/qvalue

RESOURCE AVAILABILITY

All of the resources are available in this manuscript.

Lead contact

Further information and requests for resources and reagents should be directed to and will be fulfifilled by the corresponding contact, Peng Li (lipeng@njnu.edu.cn).

Materials availability

This study did not generate new unique reagents.

Data and code availability

- Raw sequencing data derived from samples have been deposited at SRA database. Genome, predicted gene sequences, sequence alignments and phylogenetic trees have been deposited at Zenodo. Accession numbers are listed in the key resources table.
- No code was produced in this paper.
- Any additional information required to reanalyze the data reported in this work paper is available from the lead contact upon request.

EXPERIMENTAL MODEL AND STUDY PARTICIPANT DETAILS

Animal materials

The geckos (8♂11♀, 3.5–4.5 g) were acquired from various localities in Nanjing (32°03′N, 118°45′E), eastern China and transported to our laboratory at Nanjing Normal University in 500 mm × 300 mm × 250 mm (length × width × height) plastic cages. They were fed mealworms and given water regularly during the whole experiment. One healthy adult female gecko was chosen to prepare the muscle and visceral tissue. After weighing, 2.06 g of muscle tissue and 0.90 g of visceral tissue were put into two 1.5 mL centrifuge tubes and immediately put into a foam box filled with liquid nitrogen for preservation. The box was sent to Wuhan Frasergen Bioinformatics Co., Ltd. for further DNA extraction, RNA

extraction, Hi-C cell crosslinking experiment, Illumina Hiseq sequencing, PacBio SMRT sequencing, and Hi-C analysis. Animal welfare and experimental procedures were approved by the Institutional Animal Care and Use Committee of Nanjing Normal University [SYXK(Jiangsu) 2020-0047, IACUC-20210236]. The authors confirmed that animals did not suffer unnecessarily at any stage of experiments in this study.

METHOD DETAILS

Genomic sequencing, RNA-Seq and Hi-C

Genomic DNA (gDNA) was extracted from muscle tissue and integrity was assessed using an Agilent Bioanalyzer 2100. Concentration was measured using Qubit 3.0. After the samples were tested, library preparation was performed according to the methods and procedures of Annoroad Universal DNA Fragmentase Kit V2.0 (AN200101-L) and Annoroad Universal DNA Library Prep Kit V2.0 (AN200101-L). Clustering and pair-end sequencing were performed on the Illumina NovaSeq 6000 S4 platform using the NovaSeq 6000 S4 Reagent kit V1.5 reagent. Single standard SMRTbell libraries were constructed using the SMRTbell Template Preparation Kit following the manufacturer's protocol. The libraries had an average insert length of 20 kbp.

Sequencing was performed based on the PacBio Sequel II platform. Total RNA was extracted from visceral tissue using Trizol (Invitrogen, CA, USA). RNA purity and integrity were monitored using a NanoDrop 2000 spectrophotometer (NanoDrop Technologies, Wilmington, DE, USA) and a Bioanalyzer 2100 system (Agilent Technologies, CA, USA). RNA contamination was assessed using a 1.5% agarose gel. Oligo(dT)-attached magnetic beads were used to purify mRNA. Purified mRNA was fragmented into small pieces with fragment buffer at the appropriate temperature. Then first-strand cDNA was generated using random hexamer-primed reverse transcription, followed by a second-strand cDNA synthesis and purification using AMPure XP beads. Afterward, A-Tailing Mix and RNA Index Adapters were added by incubating to end the cDNA repair. PCR amplified the cDNA fragments obtained from previous steps, and products were purified by Ampure XP Bead to obtain the final library. Transcriptome sequencing was performed based on the Illumina NovaSeq 6000 platform.

For Hi-C, samples were cross-linked with 1% formaldehyde for 10 min at room temperature and quenched with 0.125M final concentration glycine for 5 min. The cross-linked cells were subsequently lysed. The endogenous nuclease was inactivated with 0.3% SDS. The chromatin DNA was then digested by 100U MboI (NEB), marked with biotin-14-dCTP (Invitrogen), and ligated by 50U T4 DNA ligase (NEB). After reversing cross-links, the ligated DNA was extracted through a QIAamp DNA Mini Kit (Qiagen) according to manufacturers' instructions. Purified DNA was sheared to 300–500 bp fragments and further blunt-end repaired, A-tailed and adaptor-added, followed by purification through biotin-streptavidin–mediated pull-down and PCR amplification. Finally, the Hi-C libraries were quantified and sequenced on the Illumina Nova-seq platform (San Diego, CA, USA).

Genome assembly and annotation

According to Lander-waterman theory, genome size was estimated by K-mer (17-mer) analysis using reads generated by Illumina Hiseq sequencing. The preliminary genome assembly was first obtained using MECAT2[24] based on reads generated by PacBio SMRT sequencing using GCpp version 1.9.0 (https://github.com/PacificBiosciences/gcpp) was used to polish the assembly using subreads. Pilon version 1.22[25] was used to correct the errors based on Illumina data, and finally Hi-C analysis was performed using juicer version 1.6.2[26] to obtain a chromosome-level genome assembly.

Homology prediction of repetitive elements using RepeatMasker and RepeatProteinMask version 4.0.9[29,30] based on RepBase,[51] repetitive elements were predicted based on their sequence comparison and *de novo* prediction using RepeatModeler version 1.0.11,[28] and using LTR-FINDER version 1.0.5[81] for *de novo* prediction based on repeated sequence features. In addition, the *de novo* prediction approach used TRF version 4.09[27] to find tandem repeat sequences in the genome. These results were integrated to obtain the final repeated sequence annotation.

Using the MAKER pipeline version 3.01.02[82] based on exonerate version 2.2.0[83] with homologous proteins of *P. muralis*, *A. carolinensis*, *Pogona vitticeps*, *L. agilis* and *G. japonicus* (previous version), third-generation full-length transcriptome data, second-generation transcriptome data, and the ab initio prediction software Augustus version 3.3,[84] Genscan version 1.0[85] to obtain non-redundant complete gene set. Benchmarking Universal Single-Copy Orthologs (BUSCO) version 2.0[115] was used to assess genome assembly and gene annotation. The proteins in the gene set were functionally annotated based on protein databases KEGG,[116] InterPro,[117] GO,[118] NR,[119] SwissProt and TrEMBL.[120] Non-coding RNAs were annotated based on tRNAscan-se version 1.3.1,[86] BLASTN in BLAST+ version 2.6.0[87] and Rfam version 14.1.[88] A search for 1:1 orthologous genes between two *G. japonicus* genome assemblies was performed using OrthoFinder version 2.5.1.[31] Synteny analysis was performed using BLASTP (e-value = 1e-5) and MCScanX,[89] other genomic features were calculated using BEDTools version 2.29.2[90] based on our annotations. Finally, syntenic regions and genomic features were plotted using Circos version 0.69.[91]

Synteny analysis and identification of HSBs and EBRs

OrthoFinder version 2.5.1[31] was used to identify the orthologous genes of *G. japonicus* with *A. carolinensis* and *G. japonicus* with *G. gallus*. The chromosome sequences and orthologous genes of each of these three species were used as input to Symap version 5.0.6[92] to identify HSBs of *G. japonicus* with *A. carolinensis* and *G. japonicus* with *G. gallus*, and the syntenic regions were visualized with Circos version 0.69.[91] The EBR is defined as the interval between two adjacent HSBs on the same chromosome, i.e., the end position of the preceding HSB plus 1 bp to the start position of the following HSB minus 1 bp. GO enrichment was performed with clusterProfiler version 4.0[93] in R version 4.1.0.

The Wilcoxon rank-sum test was performed in R version 4.1.0 to test whether the gene density, GC content and repeat content of the EBR of each chromosome were significantly different from those of the whole chromosome. Linear regression analysis was performed in Python version 3.8.1 using the module statsmodel and visualized with the seaborn module. A t-test with unequal variances was used to test whether each repetitive element in the EBRs was significantly higher than the same repetitive element in the rest of the genome, calculated based on 10 kbp windows, performed in R version 4.1.0.

Phylogenetic and gene family analysis

A total of 22 species genomes were selected for gene family and phylogenetic analyses, including four species of snakes (*O. hannah*, *Protobothrops mucrosquamatus*, *P. bivittatus* and *Pantherophis guttatus*), and five species of lizards (*A. carolinensis*, *P. muralis*, *P. vitticeps*, *G. japonicus* (this study), *G. japonicus* (previous version)), three species of birds (*G. Gallus*, *Meleagris gallopavo*, *Taeniopygia guttata*), four species of crocodiles (*Alligator misissippiensis*, *A. sinensis*, *Crocodylus porosus*, and *Gavialis gangeticus*), three species of turtles (*Chelonia mydas*, *Chrysemys picta bellii*, and *Perodiscus sinensis*), two species of mammals (*Homo sapiens* and *M. musculus*) and an amphibian (*Xenopus tropicalis*).

Gene family clustering is an essential step in phylogenetic analysis. Protein sequences of various species were clustered based on sequence similarity by OrthoMcl version 1.4 (e-value of BLASTP is 1e-5, and alignments with an identity of about 30% or coverage <50% were dropped).[94] The single-copy orthologous genes were filtered, and only the genes with an aa length ≥100 were retained. Muscle version 3.8.31[95] was used to perform multiple sequence alignment for each single-copy orthologous gene set. The alignments were concatenated and converted into a super-gene alignment in Phylip format.

RAXML version 8[96] was used to construct the phylogenetic tree based on the maximum likelihood method. Using the phylogenetic tree constructed, combined with TimeTree[97] and literature to obtain time correction points, the MCMCTree program in PAML version 4.7[39] and R8s version 1.50[98] was used to estimate the divergence time. CAFE version 4.2.1[99] was used to estimate the expansion and contraction events of gene families by employing a random birth and death model to find gene gain and loss across the phylogenetic tree. According to the results of gene family clustering, for each single-copy orthologous gene set, the branch-site model in PAML version 4.7[39] was used to detect whether genes were under positive selection in *G. japonicus*.

This model helped test positive selection on specific branches. It assumed four classes of sites, the first class of sites 0 was assumed to have $0 < \omega_0 < 1$ in all branches, the second class of sites 1 is assumed to have $\omega_1 = 1$ in all branches, the third class of sites 2a is assumed to have $\omega_{2a} = \omega_{2b} \geq 1$ in the foreground and $0 < \omega_{2a} = \omega_0 < 1$ in the background, and the fourth class of sites 2b is assumed to have $\omega_{2b} = \omega_{2a} \geq 1$ in the foreground and $\omega_{2b} = \omega_1 = 1$ in the background. It only allows positive selections in the foreground, but may result in false positives when positive selection is present in the background.[121,122] Go/KEGG enrichment analysis was performed using Fisher's exact test for the expanded and contracted gene families, specific gene families and positively selected genes in *G. japonicus*, respectively. The p-value was then corrected based on the false discovery rate (FDR) method to obtain q-value, and results less than 0.05 were retained.

T2R analysis

Seven sauropsids, including *A. carolinensis*, *A. sinensis*, *G. gallus*, *G. japonicus*, *P. bivittatus*, *P. sinensis*, and *S. punctatus*, were used across all extant taxa as representative small datasets for gene identification to compare and complement previous gene identification results, and for evolutionary analysis.

The T2R protein sequences of western clawed frog, human, mouse and chicken were downloaded from NCBI and used as reference sequences to identify the T2R sequences of other species using GeMoMa version 1.7.1[100] and the TBLASTN method with the threshold set as 1e-10. Next, the transmembrane regions of the candidate sequences were predicted using TMHMM version 2.0.[101] Only sequences with transmembrane number greater than or equal to 7 were retained. We also removed sequences with codon number lower than or equal to 270 to ensure the high quality and structural integrity of the sequences. Finally, we obtained a high-quality T2R gene set containing only intact genes, beginning with a start codon and ending with a stop codon. We then named the genes with species name acronyms and numbers for subsequent analysis.

We used MACSE version 2.05[102] to align the T2R coding DNA sequences and manually removed low-quality regions in Mega-X version 10.0.5.[103] The phylogenetic tree of the T2R proteins alignment for seven species was constructed using RAxML-NG v. 0.9.0[104] based on the best model JTT+I+G4+F selected by ModelTest-NG version 0.1.7.[105] The subtree of *G. japonicus* was extract using ETE 3.[106] All trees were edited in Adobe Illustrator CC 2019, iTOL[107] and R package ggtree.[108] M3 vs. M0, M2a vs. M1a, and M8 vs. M7/M8a in PAML version 4.7[39] and FEL[41] and FUBAR[42] in HyPhy version 2.5.2[40] were performed to identify positively selected sites in T2R genes. The aBSREL[123] in HyPhy version 2.5.2[40] was used to find some branches in the T2R tree of seven species and the T2R tree of *G. japonicus* under positive selection. Multiple sequence alignments of the T2R gene were divided into ER, TM, and IR regions regarding a chicken T2R protein (XP_004938201.1 or Q2AB83) in UniProt.[124] YN00 in PAML version 4.7[39] was used to calculate pairwise ω values in ER, TM and IR regions. All statistical tests were performed in R version 4.1.0.

CYP analysis

Functional CYP gene sequences of human were obtained from the Cytochrome P450 Homepage[125] then we used GeMoMa version 1.7.1[100] with TBLASTN method to search the genomes of 13 sauropsids including two crocodiles (*A. misissippiensis* and *A. sinensis*), two lizards (*A. carolinensis* and *P. vitticeps*), two geckos (*G. japonicus* (this study) and *P. picta*), two snakes (*V. berus* and *O. hannah*), three turtles

(*C. mydas*, *C. picta bellii*, and *P. sinensis*), tuatara (*S. punctatus*) and one bird (*G. gallus*), using the CYP protein sequences of human as references. We filtered the protein sequences with lengths of less than 350 aa. Next, we aligned them with *Drosophila melanogaster* CYP301a1 using MAFFT version 7.475[109] (strategy L-INS-i). We used trimAl version 1.4.rev15[126] (gappyout) to trim the low-quality regions in the protein alignment. RAxML-NG v. 0.9.0[104] was used to build an ML tree using the best model JTT+G4 found by ModelTest-NG version 0.1.7[105] and 1000 repeated samples to test the confidence of each branch, and *D. melanogaster* CYP301a1 as the root of the tree.

MIPhy was used to identify clades and assess their instability scores. Then, PAML version 4.7[39] and HyPhy version 2.5.2[40] were used to perform selection pressure analysis for each clade. The M0 model in PAML was used to calculate the average ω for each clade, and M2a vs. M1a and M8 vs. M7 in PAML were used to test whether there are positively selected sites in each clade. The branch-site model in PAML was used to test whether branches of *G. japonicus* were under positive selection, and the aBSREL model in HyPhy was used to find positively selected branches in each clade. NOTUNG version 3.0.26[110] was used to identify gene duplication and loss events in sauropsid CYPs. PICs were calculated in R package ape[111] and all other statistical tests were run in R version 4.1.0.

CBP analysis

The identified CBP gene family sequences of *G. gecko*[44] and *A. carolinensis*[45] were used as a reference to search the third-generation genome sequences of *G. japonicus* and were used as a database to extract the CBP gene family of *G. japonicus* by BLASTP, setting the e-value as 10^{-5}. Then, the genomes and annotated data of the other selected species were download. They were *P. picta*, *A. carolinensis*, *P. muralis*, *Z. vivipara*, *L. agilis*, *P. vitticeps*, *P. bivittatus*, *O. hannah*, *Thamnophis elegans*, *P. sinensis*, *C. picta*, *C. mydas*, *A. sinensis*, *A. mississippiensis*, *C. porosus*, *X. tropicalis*, *M. musculus*, *E. macularius* and *S. crocodilurus*. The same method was performed to obtain the CBP gene family in each species mentioned above. Sixty-six aa sequences of the CBP gene family members of *G. japonicus* were obtained.

The physicochemical properties and components of CBP aa sequences obtained from *G. japonicus* were analyzed using online website ProtParam (https://web.expasy.org/protparam/) to calculate the basic information of CBP aa sequences such as the molecular weight (MW) and isoelectric point (PI). The aa composition ratio of each sequence was calculated using the local software Mega-X version 10.0.5.[103] Each sequence motif was analyzed using the MEME online page (http://meme-suite.org/). Then, the conservative domain characteristics of the sequences were analyzed using the NCBI online tool (https://www.ncbi.nlm.nih.gov/cdd/). The local software TBtools version 1.074[112] was used to visually analyze the gene structure and the distribution of motif and domain in each sequence of the acquired gene family members. R packages ggplot2 and gggenes were used to plot and analyze the position, arrangement and orientation of gene family members in the chromosomes.

Phylogenetic analysis was performed among the CBP gene family members of *G. japonicus*. MAFFT version 7.475[109] (strategy L-INS-i) was used for multi-sequence alignment. ModelTest-NG version 0.1.7[105] was used to calculate the optimal model for constructing the maximum likelihood (ML) tree. RAxML-NG v. 0.9.0[104] was used to build an ML tree based on the best model WAG+G4+F, and 1000 repeated samples were used to test the confidence of each branch. ML trees were constructed for all selected reptile species in the same way to explore the evolutionary origin of the CBP gene family of *G. japonicus*.

Regional synteny analysis was performed on the number and distribution of CBPs to compare these sequences among the selected species. Sequences and gene annotation files inside the flanking genes of the CBP sequences were intersected using BEDTools version 2.29.2.[90] Synteny analysis of the CBP gene families between the two species was performed by MCScanX.[89] TimeTree[97] was used to construct the time tree between the selected species, which was used as the basis for drawing the sequence of the gene family arrangement map of the species.

TRP analysis

TRP genes of representative sauropsids, two mammals and one amphibians were obtained from NCBI, including one frog (*X. tropicalis*), human (*H. sapiens*), mouse (*M. musculus*), two birds (*G. gallus* and *T. guttata*), one lizard (*A. carolinensis*), one gecko (*G. japonicus*, previous version), one snake (*P. bivittatus*), three turtles (*C. mydas*, *C. picta* and *P. sinensis*) and four alligators (*A. mississippiensis*, *A. sinensis*, *C. porosus* and *G. gangeticus*).

Then, we used GeMoMa version 1.7.1[100] with the TBLASTN method to search the genomes of two geckos (*G. japonicus* (this study) and *P. picta*), two snakes (*V. berus* and *O. hannah*) and tuatara (*S. punctatus*) using the TRP protein sequences of *X. tropicalis*, *H. sapiens*, *G. gallus*, *A. carolinensis* and *G. japonicus* (previsou version) as references. TMHMM version 2.0[101] was used to predict TM topology, sequences that did not have at least three predicted TM segments were removed (TRP commonly has six transmembrane regions, here we chose a low threshold to retain some partial and intact sequences). The TRP protein sequences of 19 species and the VDAC protein sequences of *X. tropicalis* as an outgroup were aligned using MAFFT version 7.475[109] (strategy L-INS-i). The aligned sequences were trimmed using trimAl version 1.4.rev15[126] (gappyout). The ML tree was constructed using RAxML-NG version 0.9.0[104] (1,000 bootstraps) with the best-fit model JTT+G4+F selected by ModelTest-NG version 0.1.7,[105] and rooted with the VDAC protein sequences of *X. tropicalis*.

The predicted genes were named by their phylogenetic relationships and BLASTP results. We visualized and edited our tree in Adobe Illustrator CC 2019, iTOL[107] and R package ggtree.[108] Single TRP subfamily members were aligned using the $Prank_{+F}$[127] algorithm in GUIDANCE2,[113] and unreliable codons (with scores <0.9) were replaced with "NNN" based on the residue filtering principle. Four pairs of site models M3 vs. M0, M2a vs. M1a, and M8 vs. M7/M8a in PAML version 4.7[39] were performed to identify positively selected sites and calculate average ω values for each TRP gene. CmC and CmD model in PAML version 4.7[39] are well suited to test for divergent selection, and we used them to test for shifts in selection pressures. CmC assumes that throughout the phylogeny one class of sites has a ω value greater

than 0 and less than 1, another class of sites has a ω value equal to 1. In two or more partitions one class of sites is free to evolve differently with unconstrained ω values. CmD is similar to CmC, but all three classes of sites are unconstrained.

We also used the branch-site and branch model in in PAML version 4.7[39] to test for positive and relaxed selection. The branch model helps test relaxed selection. It can calculate average ω values between different partitions. All models were run at least three times with different initial values of κ and ω (except for some null models with fixed values) to ensure the robustness of the results. FDR control was performed with the R package q-value based on the method of Storey et al.[114] We also performed Bonferroni correction with similar results in R version 4.1.0. YN00 in PAML version 4.7[39] was used to calculate the pairwise ω values of 1:1 orthologous TRP genes between *G. japonicus* and *A. carolinensis*. All other statistical tests were run in R version 4.1.0.

QUANTIFICATION AND STATISTICAL ANALYSIS

Sampling, sequencing and data analysis

One healthy adult female gecko (from a group of 19 geckos captured in the wild) was chosen for further DNA extraction, RNA extraction, Hi-C cell crosslinking experiment, Illumina Hiseq sequencing, PacBio SMRT sequencing, and Hi-C analysis. All statistical analyses were performed in R version 4.1.0, which are described in the method details. No plausible methods were used to determine whether the data met assumptions of the statistical approach.

K-mer analysis

K-mer i.e., iteratively select sequences of length K bases from a continuous sequence, if the length of each sequence is L, then L-K+1 K-mer can be obtained. we usually take K = 17 for the analysis. According to the Lander_waterman algorithm, the genome size (G) satisfies the following equation:

$$C_{base} = \frac{C_{k-mer} \times L}{L - K+1}$$

$$G = \frac{n_{k-mer}}{C_{k-mer}} = \frac{n_{base}}{C_{base}}$$

n_{base} and $n_{k\text{-}mer}$ are the total number of bases and the number of K-mer, C_{base} and $C_{k\text{-}mer}$ are the desired depth of covered bases and the desired depth of covered K-mer.

In addition, the depth distribution of K-mer obeys a Poisson distribution:

$$P_{Kspecies}(x) = \frac{C^x}{x!}e^{-c}$$

In the formula, c is the depth of K-mer corresponding to the peak and is the estimated value of K-mer depth. Based on the above assumptions, K-mers were taken base by base from all sequenced sequences, K-mer frequency distribution was plotted, and K-mer depth distribution curve and depth product curve were calculated and obtained. The genome size was estimated by obtaining an estimate of the K-mer depth, c, based on the curve.