

Differential Conservation and Loss of Chicken Repeat 1 (CR1) Retrotransposons in Squamates Reveal Lineage-Specific Genome Dynamics Across Reptiles

Simone M. Gable ¹, Nicholas A. Bushroee¹, Jasmine M. Mendez¹, Adam Wilson¹, Brendan J. Pinto ^{2,3}, Tony Gamble ^{3,4,5}, Marc Tollis ^{1,*}

¹School of Informatics, Computing, and Cyber Systems, Northern Arizona University, Flagstaff, AZ, USA

²Center for Evolution and Medicine, Arizona State University, Tempe, AZ, USA

³Department of Zoology, Milwaukee Public Museum, Milwaukee, WI, USA

⁴Department of Biological Sciences, Marquette University, Milwaukee, WI, USA

⁵Bell Museum of Natural History, University of Minnesota, St. Paul, MN, USA

*Corresponding author: E-mail: marc.tollis@nau.edu.

Accepted: July 11, 2024

Abstract

Transposable elements (TEs) are repetitive DNA sequences which create mutations and generate genetic diversity across the tree of life. In amniote vertebrates, TEs have been mainly studied in mammals and birds, whose genomes generally display low TE diversity. Squamates (Order Squamata; including ~11,000 extant species of lizards and snakes) show as much variation in TE abundance and activity as they do in species and phenotypes. Despite this high TE activity, squamate genomes are remarkably uniform in size. We hypothesize that novel, lineage-specific genome dynamics have evolved over the course of squamate evolution. To understand the interplay between TEs and host genomes, we analyzed the evolutionary history of the chicken repeat 1 (CR1) retrotransposon, a TE family found in most tetrapod genomes which is the dominant TE in most reptiles. We compared 113 squamate genomes to the genomes of turtles, crocodylians, and birds and used ancestral state reconstruction to identify shifts in the rate of CR1 copy number evolution across reptiles. We analyzed the repeat landscapes of CR1 in squamate genomes and determined that shifts in the rate of CR1 copy number evolution are associated with lineage-specific variation in CR1 activity. We then used phylogenetic reconstruction of CR1 subfamilies across amniotes to reveal both recent and ancient CR1 subclades across the squamate tree of life. The patterns of CR1 evolution in squamates contrast other amniotes, suggesting key differences in how TEs interact with different host genomes and at different points across evolutionary history.

Key words: reptiles, squamates, transposable elements, retrotransposons.

© The Author(s) 2024. Published by Oxford University Press on behalf of Society for Molecular Biology and Evolution.

This is an Open Access article distributed under the terms of the Creative Commons Attribution-NonCommercial License (<https://creativecommons.org/licenses/by-nc/4.0/>), which permits non-commercial re-use, distribution, and reproduction in any medium, provided the original work is properly cited. For commercial re-use, please contact reprints@oup.com for reprints and translation rights for reprints. All other permissions can be obtained through our RightsLink service via the Permissions link on the article page on our site—for further information please contact journals.permissions@oup.com.

Significance

The chicken repeat 1 (CR1) retrotransposon is the most common transposable element (TE) found in the genomes of birds and reptiles, yet the origins and diversity of CR1 in squamates (i.e. lizards and snakes) have been understudied. We found that CR1 evolution in squamates is characterized by uniquely rapid shifts in genomic copy number driven by large variations in CR1 activity. We also found that some squamate genomes have retained extremely ancient CR1 subfamilies, contrasted by the loss of these CR1 subfamilies in other squamates, and that purifying selection plays a key role in the distribution of CR1 insertions across the squamate genome. Our study demonstrates that squamates are a promising model for studying the evolutionary dynamics of TEs in vertebrate genomes.

Introduction

Transposable elements (TEs) are repetitive DNA sequences that are important contributors to genome organization and diversity in eukaryotes. TEs create mutations by mobilizing throughout the genome via transposition, generating genetic diversity that evolutionary forces can act on, in some cases forming the basis of new phenotypes (Kidwell and Lisch 2001). TEs also constitute between 10% and 80% of some vertebrate genomes (Jang et al. 2019), including at least 60% of the human genome (de Koning et al. 2011), and high TE abundance is linked to increased genome size (Kapusta et al. 2017; Blommaert 2020). Comparative genomic studies of TE dynamics across species shed light on the origins and maintenance of a near ubiquitous form of biological variation.

Across the genomes of mammals, birds, and reptiles (i.e. amniotes), the chicken repeat 1 (CR1) autonomous non-terminal repeat (non-LTR) retrotransposon is one of the most abundant and active TE families (Chalopin et al. 2015; Galbraith et al. 2021). First described from the chicken genome (*Gallus gallus*; Stumph et al. 1984; International Chicken Genome Sequencing Consortium 2004), CR1 elements are similar to other vertebrate non-LTR retrotransposon families, such as L1 and L2, and are equipped with the machinery to “copy and paste” themselves into new genomic locations by retrotransposition using their own encoded proteins, including a reverse transcriptase (RT) (Ichiyanagi and Okada 2008). While CR1 is found only scarcely across mammalian genomes, which are instead dominated by L1 retrotransposons and non-autonomous retrotransposons (i.e. short interspersed nuclear elements) (Chalopin et al. 2015), CR1 is the dominant repetitive element in the genomes of birds and reptiles (i.e. sauropsids), comprising ~4% of most avian genomes and up to 18% of some reptile genomes (Shedlock 2006; Shaffer et al. 2013; Green et al. 2014; Chalopin et al. 2015; Pasquesi et al. 2018; Gable et al. 2023).

While CR1 sequences have been found in the genomes of species from all amniote groups, the abundance and evolutionary origins of CR1 within each of the amniote clades vary widely (Shedlock 2006; Suh et al. 2015). In a given amniote genome, CR1 elements can be further divided into subfamilies, each with unique phylogenetic histories

indicative of a pattern of diversification that occurred during the evolution of their host genomes (Suh et al. 2015). The genomes of birds, which generally lack TE abundance and diversity overall (Kapusta and Suh 2017; Sotero-Caio et al. 2017), contain relatively few CR1 elements, derived from sets of avian-specific CR1 subfamilies (Suh et al. 2015). Compared to birds, other archosaurian-line sauropsids such as turtles and crocodylians have many more CR1 copies in their genomes (Chalopin et al. 2015; Sotero-Caio et al. 2017). CR1 elements in turtle and crocodylian genomes are also derived from a far more diverse set of CR1 subfamilies than in bird genomes (Suh et al. 2015), suggesting an ancient origin of CR1 retrotransposons that stems from the genome of the ancestral amniote, followed by the subsequent loss of most CR1 diversity in birds (Suh et al. 2015; Galbraith et al. 2021).

Outside turtles and archosaurs, the remaining living reptiles are the lepidosaurs, which include ~11,000 extant species of squamates (Order Squamata; i.e. lizards and snakes) and a single species of tuatara (*Sphenodon punctatus*, which constitutes the outgroup to squamates; Uetz and Hošek 2023). The genome of the tuatara contains CR1 elements which cluster phylogenetically with the more early branching CR1 subfamilies found in turtles, archosaurs, and mammals, suggesting ancient features of amniote genome evolution (Gemmell et al. 2020). Phylogenetic reconstructions of CR1 sequences in the genome of the green anole lizard (*Anolis carolinensis*) revealed the coexistence of numerous CR1 subfamilies (Novick et al. 2009), and CR1 subfamilies from anole and Burmese python (*Python bivittatus*) genomes were found to be nested deep within the amniote CR1 phylogeny (Suh et al. 2015). This suggests that the most ancient CR1 subfamilies otherwise shared across most amniotes have been lost in squamate genomes, as they have in birds. However, a lack of genome assemblies from representatives of many squamate lineages have only been made available very recently (Card et al. 2023; Gable et al. 2023; Pinto et al. 2023), limiting our ability to accurately model ancestral states in reptile genome evolution (Shedlock 2006).

Here, we analyzed the evolutionary history of the CR1 retrotransposon across amniote evolution using a wide taxonomic sampling of whole-genome assemblies from

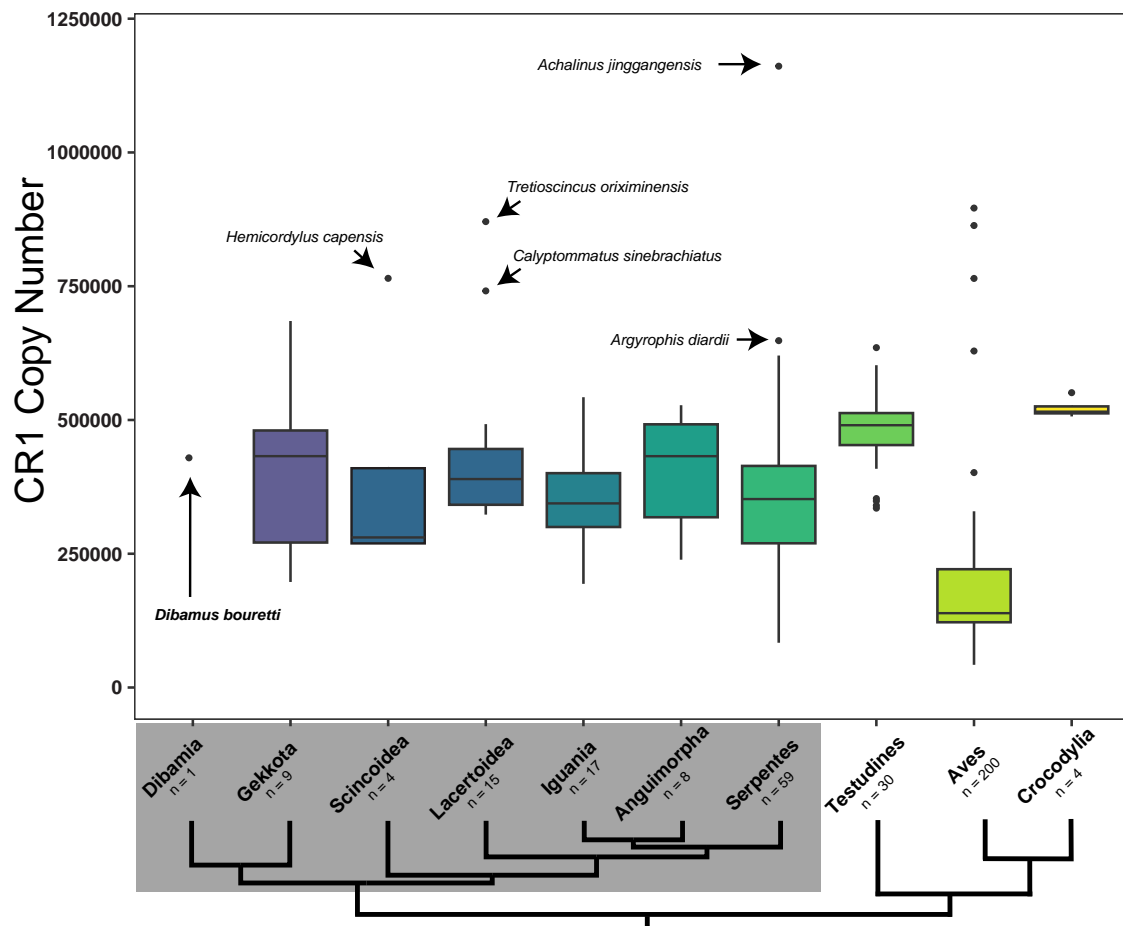


Fig. 1. Variation in CR1 element copy number in 113 genomes from all seven major squamate clades, in comparison with birds and other reptiles. The phylogenetic relationships and sample sizes for the seven major squamate clades (within gray box: Dibamia, Gekkota, Scincoidea, Lacertoidea, Iguania, Anguimorpha, Serpentes), turtles (Testudines), birds (Aves), and crocodylians (Crocodylia) are given. Vertical bars represent median; small black circles represent outliers.

345 species of reptiles and birds, including 113 squamates. Our study focused on questions about the abundance, activity, and diversity of CR1 retrotransposons in the genomes of squamates and other amniotes. We identified contrasting patterns of loss and retention of CR1 throughout amniote evolution, particularly in squamates, signaling important differences in TE evolutionary dynamics across the vertebrate classes. We also demonstrate that the diversification of squamates, which spans a ~200 million year fossil history and ~11,000 extant species (Brownstein et al. 2023; Uetz and Hošek 2023), was accompanied by rapid shifts in genome evolution, multiple episodes of CR1 expansion, as well as loss in CR1 copy number—a dynamic history of TE evolution within the host lineage that differs greatly from birds and other reptiles. We suggest that squamates are a promising model for TE evolution, with the potential to shed light on key biological questions about the origins of genome size and structural variation in vertebrates.

Results

Variation in CR1 Copy Number Across Squamates Compared to Other Reptiles

We identified and annotated repeats in genome assemblies for 113 squamate species representing seven major clades (Iguania, $n = 17$; Anguimorpha, $n = 8$; Serpentes, $n = 59$; Lacertoidea, $n = 15$; Scincoidea, $n = 4$; Gekkota, $n = 9$; Dibamia, $n = 1$; Fig. 1), 30 turtle species representing all extant families of cryptodires and pleurodires (Shaffer et al. 2017; Gable et al. 2022), and 204 archosaur species including 200 birds and four crocodylians using de novo repeat detection, followed by classification based on repeat databases (Materials and Methods; [supplementary table S1, Supplementary Material](#) online). The full results of all repeat types detected in all analyzed genomes are available in [supplementary data, Supplementary Material](#) online, and Zenodo at <https://doi.org/10.5281/zenodo.10521011> (see Data Availability). The CR1 family was the dominant

element on average across all reptile genomes, including squamates. Our analyses calculated 10.2 Gbp (10,236,613,375 bp) of CR1 sequences across 113 squamate genomes after filtering for nested elements, with CR1 accounting for ~5.2% of the average squamate genome (supplementary table S1, Supplementary Material online), and non-LTR retrotransposons in general occupying ~15.9% of the average squamate genome.

The amount of CR1 insertions in the genome varied dramatically between and within squamate clades (Fig. 1). Genomes of species classified in Lacertoidea ($n = 15$) and Serpentes ($n = 59$) showed the most variation in CR1 copy number with ranges of 547,627 inserts and 1,077,559 inserts, respectively. These large ranges are associated with the outliers in these groups, although Serpentes had a range of 536,785 even when excluding outliers. Serpentes included two outliers: *Achalinus jinggangensis* (family Xenodermidae; Zong and Ma 1983) with 1,161,213 CR1 copies accounting for over 18% of the genome, the most of any genome in our dataset, and the blind snake *Argyrophis diardii* (family Typhlopidae; Scolecophidia; Pyron and Wallach 2014) with 648,249 insertions. The outliers in Lacertoidea were the only two species from the lizard family Gymnophthalmidae, *Calyptommatus sinebrachiatus* and *Tretioscincus oriximinensis*. The outlier in Scincoidea, *Hemicordylus capensis*, is also the only representative of its family (Cordylidae) and therefore may not represent a true outlier due to the small sample size for Scincoidea ($n = 4$).

CR1 copy number varied much more extensively across squamates compared to turtles and archosaurs (supplementary fig. S1, Supplementary Material online). Both the maximum and minimum CR1 copy numbers for squamates were found in snakes, ranging from 1,161,213 in *A. jinggangensis* to 83,654 in *Simalia boeleni* (family Pythonidae; Reynolds et al. 2014). The mean CR1 copy number in squamate genomes was 373,541.4, with a median of 361,430. In contrast, CR1 copy number in turtle genomes ranged 300,038 inserts, from 635,205 (*Pelodiscus sinensis*, family Trionychidae, Cryptodira; Stuckas and Fritz 2011) to 335,167 inserts in *Dermatemys mawii* (family Dermatemydidae, Cryptodira; González-Porter et al. 2013), with a mean of 482,047 and a median of 490,220. CR1 copy number in crocodylians showed the smallest range of 44,448 inserts, with a mean of 527,897.2 and a median of 516,830.

CR1 copy number in avian genomes was the lowest across the sauropsid groups analyzed with a mean of 176,281 inserts and a median of 138,351, reflecting their small genome sizes and overall low TE abundance. However, there were five outlier bird species with much higher CR1 copy numbers, including the genomes of woodpeckers and barbets (Manthey et al. 2018). These outlier species contained an average CR1 copy number of

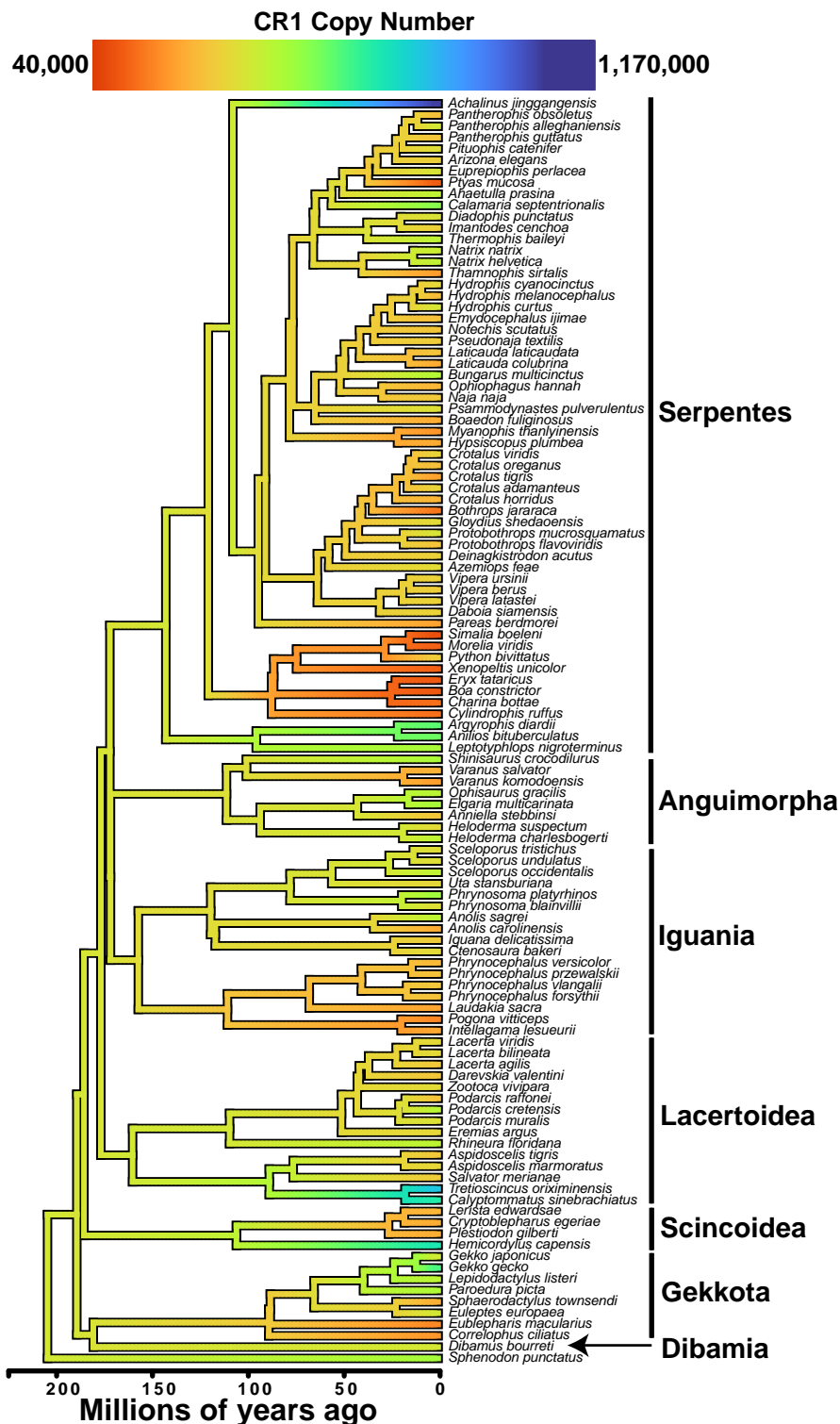
710,915.8, whereas an average CR1 copy number in birds excluding outliers was 162,573. The maximum CR1 copy number in birds was in *Eubucco bourcierii* (Piciformes, family Capitonidae) with 896,033 inserts, over 200,000 inserts more than the maximum CR1 copy number in turtle genomes. The minimum CR1 copy number in birds was in *Parabuteo unicinctus* (Accipitriformes, subfamily Buteoninae) with 42,426 copies.

Highly Heterogenous Rates of CR1 Copy Number Evolution in Squamates Compared to Other Reptiles

We estimated the rate of CR1 copy number evolution in terms of number of inserts per million years for each branch of the species-level phylogenies for squamates, turtles, and archosaurs (including birds and crocodylians both together and separately) using ancestral state reconstruction (Fig. 2; supplementary figs. S2 and S3, Supplementary Material online; also see Materials and Methods). The distributions of evolutionary rates for CR1 copy number from the squamate, archosaur, and turtle phylogenies, respectively, were all roughly bell-shaped around their means (Fig. 3); however, there were significant differences between all of them ($P < 0.005$, $D > 0.25$, nonparametric asymptotic two-sample Kolmogorov–Smirnov test).

The estimated rates of CR1 copy number evolution for turtles were normally distributed ($P = 0.61$, Shapiro–Wilk normality test), while estimated rates of CR1 copy number evolution for squamates, archosaurs, and the avian-only analysis were nonnormal ($P < 0.0001$, Shapiro–Wilk normality test), consistent with residual analyses (supplementary fig. S4, Supplementary Material online). The variances of each distribution were also heterogenous ($P = 0.0136$, Levene's test). Because the estimated rate of CR1 copy number evolution on a given branch was either positive (indicating overall gain from the ancestral state) or negative (indicating overall loss from the ancestral state), we alternatively log-transformed the data using $\log(1 + \text{CR1 rate for the branch} - \text{the minimum CR1 rate})$. The Shapiro–Wilk normality test results for transformed data were as follows: turtles, $W = 0.36094$, $P = 1.362e-14$; squamates, $W = 0.11339$, $P < 0.001$; archosaurs and avian-only, $W = 0.23974$, $P < 0.001$. The distribution of the estimated rates of CR1 copy number evolution for the squamate phylogeny was also significantly different from those of other reptiles based on transformed data ($P < 0.0001$, $D = 0.99548$, nonparametric asymptotic two-sample Kolmogorov–Smirnov test).

The estimated rate of CR1 copy number evolution for squamates was on average negative, ranging from $-34,679.7$ copies per million years (copies/MY) to 39,034.5 copies/MY, with a mean of -19.1 and a median of -156.2 . Meanwhile, the estimated rate of CR1 copy number evolution for turtles was overall slightly positive,



Downloaded from https://academic.oup.com/gbe/article/16/8/evae157/7717631 by Law Library user on 24 September 2024

Fig. 2. Species tree for 111 squamates representing seven major clades with ancestral state estimation for CR1 copy number mapped onto branches. Species tree was based on 2,999 nucleotide sequences of high-confidence single-copy orthologs (see Methods and Materials). The units for branch length are in terms of millions of years. Outgroups other than *Sphenodon* (*Homo sapiens*, *G. gallus*, *A. mississippiensis*, *C. picta*) not shown. CR1 insert number was modeled as a continuous trait for all genomes with annotated repeats using fastAnc in the package phytools (Revell 2012).

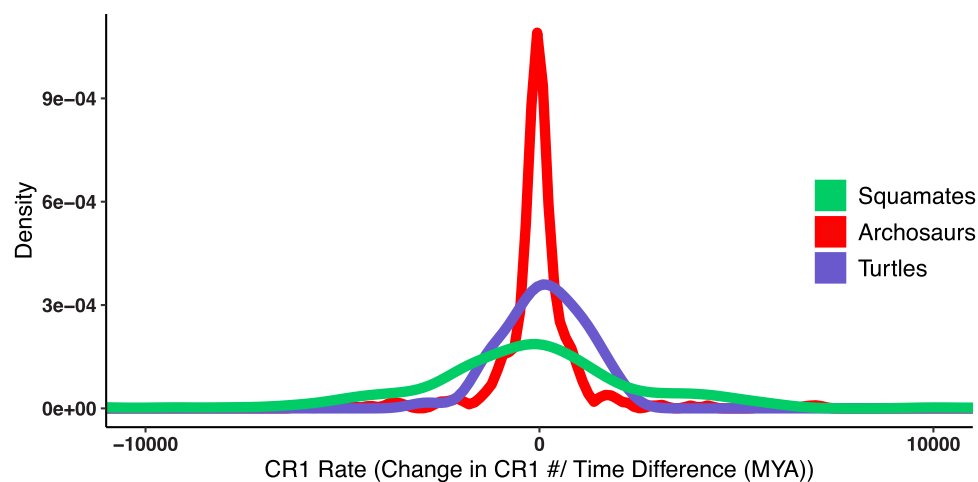


Fig. 3. Estimated rates of CR1 copy number evolution are significantly different and more varied in squamates compared to other reptiles. Comparisons of these distributions were made using the Kolmogorov–Smirnov test ($P < 0.0001$). Overlaid density plot of normalized archosaur, squamate, and turtle CR1 rates. Turtle CR1 rates range from $-2,877.3$ to $2,031.0$ with a mean of 128.8 and a median of 146.8 , while squamate CR1 rates range from $-34,679.7$ to $39,034.5$ with a mean of -19.1 and a median of -156.2 . Archosaur CR1 rates average -22.0 with a median of -36.4 , and avian-only rates average -29.8 with a median of -38.6 .

ranging from $-2,877.3$ copies/MY to $2,031.0$ copies/MY, with a mean of 128.8 and a median of 146.8 . Among archosaurs, the estimated rate of CR1 copy number evolution for birds was negative with a mean of -29.8 copies/MY and a median of -38.6 ; meanwhile, like turtles, crocodylians showed an average net increase in CR1 copy number with a mean of 276.9 copies/MY and a median of 417.6 . Estimated rates of CR1 copy number evolution in archosaurs ranged from $-6,062.4$ copies/MY to $7,102.1$ copies/MY, which was less than one-fifth the range for squamates.

Episodic Bursts of CR1 Activity Across Squamate Evolutionary History

In addition to overall variation in CR1 copy number, as well as the estimated rate of CR1 copy number evolution being highly varied between squamates, we also found that CR1 activity was highly variable across squamate species and clades (Fig. 4). We used repeat landscapes of each analyzed genome to model CR1 activity over time, based on the Kimura 2-parameter (K2P) divergence of each CR1 insert from their family consensus sequence as a proxy for element age (see Materials and Methods). Among the repeat landscapes of CR1 divergence from representatives of ten squamate clades (Fig. 4), the scolecophidian snake *A. diardii* showed the largest proportion of the genome consisting of CR1 inserts at $<10\%$ divergence from consensus. In contrast, some squamate genomes contained relatively low amounts of CR1 insertions in more recent bins of K2P divergence. The *Boa constrictor* landscape revealed limited CR1 activity across all divergence bins, matching the

pattern seen in other henophidian snake genomes which generally lack CR1 copies compared to other squamates (supplementary table S1, Supplementary Material online). These results are consistent with the relative losses in CR1 copy number in henophidian snakes that we detected using ancestral state reconstruction (Fig. 2).

To test for significant differences in CR1 activity over time across squamate genomes, we used pairwise Wilcoxon's rank-sum tests to compare all the distributions of CR1 K2P divergence calculated for 113 squamate genomes (see Materials and Methods). The homalapsid snake *Myanophis thanlyinensis* (Köhler et al. 2021) had the most significant pairwise comparisons at 5% and 10% K2P divergence (i.e. ~ 62.5 and ~ 125 MYA, respectively, see Discussion), and the second most significant pairwise comparisons at 15% K2P divergence (supplementary table S1, Supplementary Material online). Most species with the highest amount of significant pairwise comparisons were henophidian snakes (e.g. boas and pythons), reflecting the loss of CR1 copy number that is shared by these species (Fig. 2). CR1 landscapes of squamate species with some of the highest CR1 copy numbers included the xenodermid snake *A. jinggangensis* and the gymnophthalmid lizards *T. oriximinensis* and *C. sinebrachiatus*, which were significantly different from most squamate landscapes across all divergence ranges. However, the cordylid *H. capensis* (Scincoidea: Cordylidae), an outlier with a relatively higher number of insertions indicative of a large CR1 expansion (Fig. 1), had many more significantly different comparisons at the 10% and 15% levels of CR1 K2P divergence than the 5% level (supplementary table S1, Supplementary Material online). This suggests some

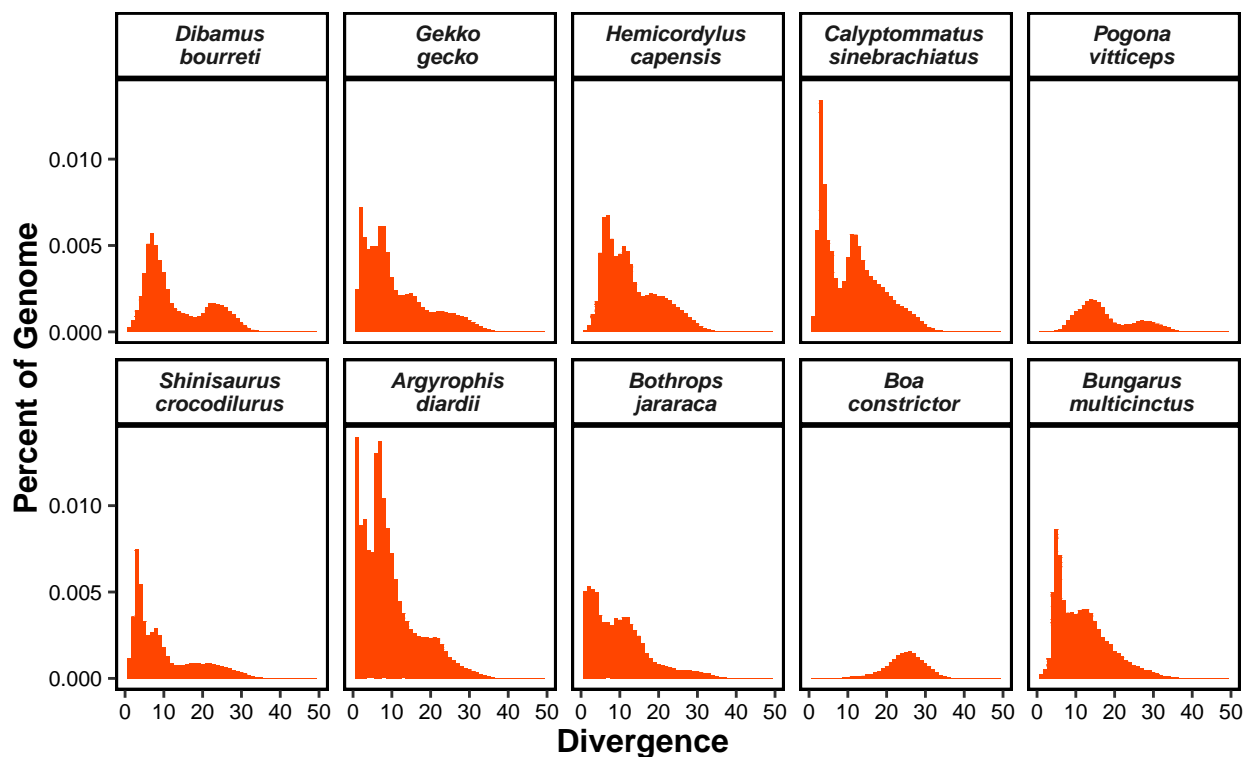


Fig. 4. CR1 repeat landscapes for 10 squamate species representative of major clades. Representatives were chosen based on primarily number of significant Wilcoxon's pairwise comparisons ($P < 0.05$) at 5% divergence and secondarily assembly quality (e.g. N50, % gaps, completeness). Represented clades: Dibamia: *D. bourreti*, Gekkota: *G. gecko*, Scincoidea: *H. capensis*, Lacertoidea: *C. sinebrachiatus*, Iguania: *Pogona vitticeps*, Anguimorpha: *Shinisaurus crocodilurus*; Serpentes (Scolophorida: *A. diardii*; Viperidae: *Bothrops jararaca*; Henophidia: *B. constrictor*; Elapidae: *Bungarus multicinctus*).

expansions of CR1 in cordylids may be older than in other squamates (~187.5 MYA, see Discussion); however, more sampling from cordylids and other scincoids is required to more accurately determine the age of this expansion.

CR1 Insertions Are Mostly Truncated and Are Found Far from Exons in Squamate Genomes

To understand how CR1 retrotransposons are distributed across squamate genomes, including their interactions with protein-coding genes, we analyzed the genomic locations of CR1 inserts, their lengths, and their proximity to gene (exonic) annotations in 30 squamate genome assemblies (i.e. "exon" feature annotations; Table 1; supplementary fig. S5, Supplementary Material online; see Materials and Methods). We found that in all squamate genomes the vast majority (97%) of CR1 inserts are truncated rather than full length. The mean distance of CR1 inserts to exons in base pairs (49,144 bp) was greater than the mean intergenic distance overall in squamate genomes (41,330 bp) (Wilcoxon's rank-sum test, $z = -2.8909$, $P = 0.00386$). The mean distance of CR1 inserts upstream of exons (52,227 bp) was greater than the mean distance of CR1 inserts downstream of exons (47,568 bp). We found a nonsignificant moderate positive correlation between

average distance in base pairs of CR1 inserts to exons and the average CR1 insert length in base pairs (phylogenetic generalized least squares [PGLS]: TE length ~ TE-exon distance, $P = 0.0524$, coefficient = 0.0026, lambda = 0.938742, Akaike information criterion [AIC] = 364.6591). When we added average intergenic distance as a response variable (i.e. TE length * average intergenic distance ~ TE-exon distance), we found a significant positive correlation between TE length and average intergenic distance to TE-exon distance (PGLS, $P < 0.0001$, coefficient = 498, lambda = 1.060542, AIC = 1,012.571). In the analyzed squamate genomes, we also found an average of 316 CR1 inserts which overlapped exons for a mean length of overlap of 239 bp. We did not detect a significant difference between average GC content 1 kb upstream or downstream of CR1 inserts (43%) and average total genomic GC content (42%) in the squamate genomes.

Conservation and Loss of CR1 Subfamilies Across Reptile Evolution

In order to model the evolutionary history of full-length, active CR1 sequences that were producing new copies throughout the history of amniotes, we extracted, aligned, and filtered CR1 open reading frame 2 (ORF2) RT sequences

Table 1 CR1 insertion length, distance to exons, and overlap with exons data for 29 squamate genome assemblies

Major clade	Family	Species	Mean CR1 length (bp)	No. of full-length CR1s	No. of truncated CR1s ^a	Mean CR1–exon distance (absolute value) (bp)	Mean intergenic distance (bp)	No. of overlapping CR1s–exons	
Gekkota	Eublepharidae	<i>Eublepharis macularius</i>	1,059	2,063	37,126	99,118	58,680	190	
	Sphaerodactylidae	<i>E. europaea</i>	842	894	48,871	53,794	53,086	103	
		<i>Sphaerodactylus townsendi</i>	989	3,130	50,002	50,184	48,630	281	
Scincoidea	Gekkonidae	<i>Gekko japonicus</i>	903	1,805	67,648	48,602	43,394	153	
	Cordylidae	<i>H. capensis</i>	896	2,360	77,170	30,385	48,305	1,768	
	Lacertidae	<i>Lacerta agilis</i>	736	87	17,142	43,853	33,777	126	
		<i>Lacerta bilineata</i>	833	502	19,003	26,491	22,091	81	
Serpentes	Colubridae	<i>Lacerta viridis</i>	805	367	20,723	36,909	25,969	65	
		<i>Podarcis cretensis</i>	929	1,367	24,939	41,934	35,101	208	
		<i>P. raffonei</i>	992	1,743	24,141	39,256	35,028	260	
		<i>Ahaetulla prasina</i>	865	1,664	79,851	70,043	49,209	744	
		<i>Pantherophis alleghaniensis</i>	730	57	36,399	40,117	37,592	256	
		<i>Pantherophis guttatus</i>	736	79	32,585	38,855	36,644	216	
Elapidae		<i>Pantherophis obsoletus</i>	729	0	32,775	40,052	37,091	181	
		<i>Pituophis catenifer</i>	762	244	36,440	58,946	48,621	402	
		<i>Notechis scutatus</i>	1,066	3,797	53,031	53,753	37,403	252	
		<i>Pseudonaja textilis</i>	953	2,181	48,730	54,312	39,363	293	
		<i>Python bivittatus</i>	758	15	13,589	18,901	17,756	143	
		<i>Thamnophis elegans</i>	703	322	48,938	63,500	48,155	303	
		<i>Crotalus adamanteus</i>	977	2,788	62,564	22,951	23,369	477	
		<i>Crotalus oreganus</i>	906	1,326	48,528	51,874	42,640	405	
		<i>Crotalus tigris</i>	977	1,973	44,322	39,623	31,008	363	
		<i>Gloydius shedaoensis</i>	919	1,911	66,821	48,825	37,989	439	
		<i>Protobothrops mucrosquamatus</i>	809	530	52,118	34,806	27,115	91	
	Anguimorpha	Helodermatidae	<i>Heloderma suspectum</i>	1,153	5,960	73,910	61,733	84,786	115
		Varanidae	<i>Varanus komodoensis</i>	919	1,474	56,387	52,828	41,144	476
<i>Varanus salvator</i>			1,082	4,329	56,221	78,707	52,414	432	
Iguania	Dactyloidae	<i>Anolis carolinensis</i>	1,010	1,756	31,640	33,444	45,471	434	
	Agamidae	<i>Laudakia sacra</i>	1,351	6,516	40,026	80,867	61,179	153	
		<i>Pogona vitticeps</i>	802	54	18,032	59,646	36,883	71	
Average		906	1,710	43,989	49,144	41,330	316		

^aOnly CR1 inserts >500 bp were considered.

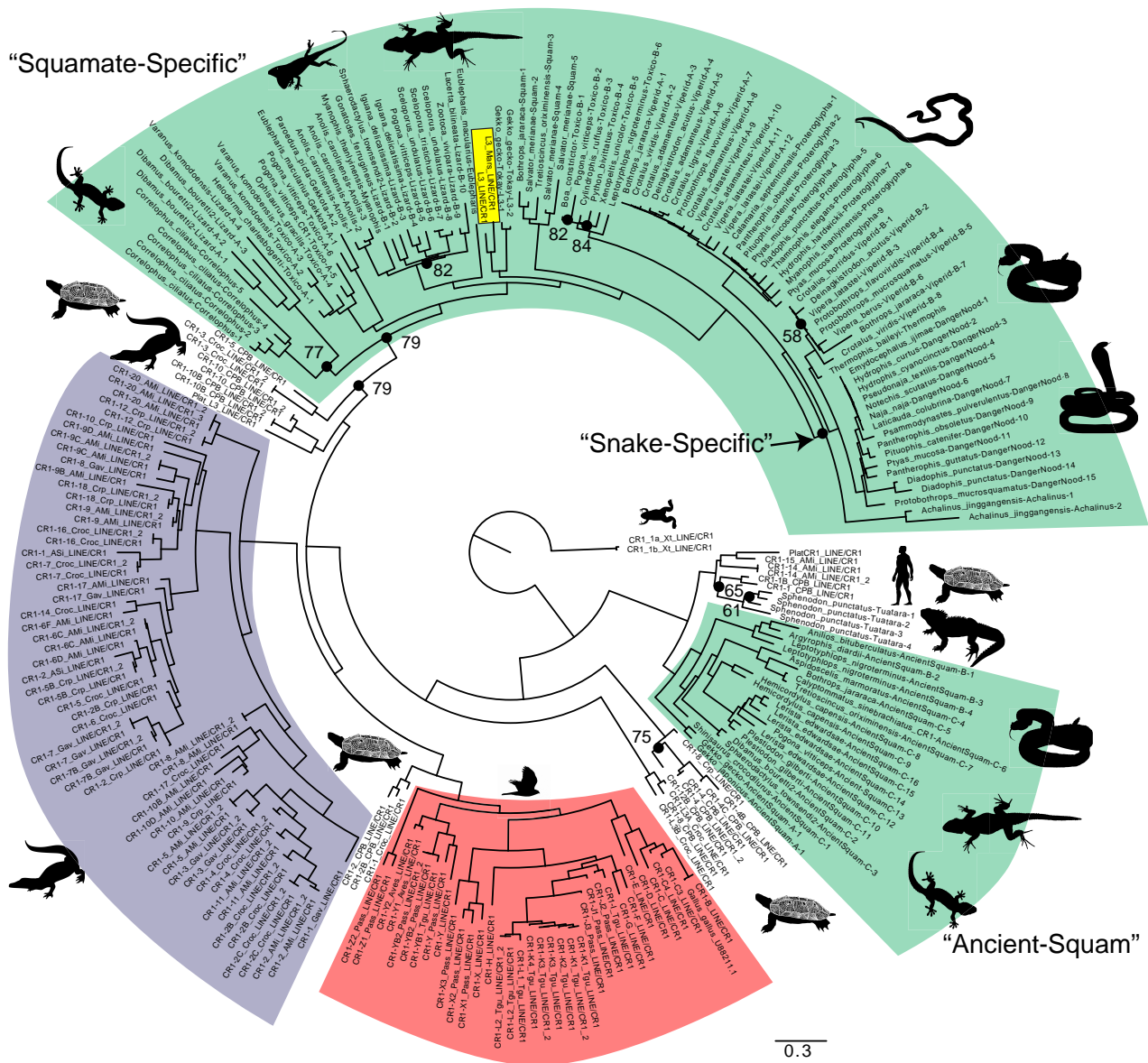


Fig. 5. Maximum-likelihood phylogeny of 244 amniote CR1 subfamily consensus sequences with new squamate CR1 subfamilies labeled. The tree file in Newick format for this phylogeny is given in [supplementary data, Supplementary Material](#) online. Branch lengths are in terms of substitutions per site. Shaded areas and animal silhouettes indicate the taxonomy of the genomes where the CR1 consensus sequences were found (i.e. several types of lizards and snakes, tuatara, turtles, crocodylians, birds, mammals, and the outgroup amphibian *Xenopus*). Squamate subfamilies were named based on the clade most inclusive of the species represented in each CR1 subfamily. We identified 114 total squamate CR1 subfamilies from 63 squamate species, along with 4 total tuatara CR1 subfamilies. There were 36 total avian subfamilies, 69 crocodylian subfamilies, 15 turtle subfamilies, and 3 mammal subfamilies (not colored). Within the squamate-specific CR1 clade, we have highlighted L3 sequences, a subtype of CR1 found in mammals (Gentles et al. 2007), that clustered with consensus sequences from a gecko. Bootstrap values <90% are shown. RepBase abbreviations: “Plat” = Platypus, “Ami” = *Alligator mississippiensis*, “CPB” = *Chrysemys picta bellii*, “Croc” = Crocodylian, “Crp” = *Crocodylus porosus*. Animal silhouettes from [www.phylopic.org](#) under the public domain.

from each squamate genome and used maximum-likelihood phylogenetic inference to cluster the sequences into subfamilies. We incorporated consensus sequences of 118 CR1 subfamilies identified in lepidosaurs with 106 additional CR1 consensus sequences from amniotes and an outgroup amphibian (*Xenopus tropicalis*) from RepBase (Bao et al. 2015) also using maximum-likelihood phylogenetic reconstruction

([supplementary table S1, Supplementary Material](#) online). CR1 subfamily consensus sequences from RepBase consisted of 36 avian subfamilies, 69 crocodylian subfamilies, 15 turtle subfamilies, and 3 mammal subfamilies.

Our phylogenetic analysis revealed evidence of multiple lineages of CR1 inhabiting the genomes of amniotes (Fig. 5). Many CR1 subfamilies recovered from crocodylian,

mammalian, turtle, and lepidosaurian genomes clustered together in a pattern that was inconsistent with the host phylogenetic relationships. Crocodylian CR1 subfamilies were found in numerous places along the amniote CR1 phylogeny, including 60 monophyletic CR1 subfamilies we recovered only from crocodylian genomes (Fig. 5, light purple). CR1 consensus sequences from mammalian and turtle genomes were scattered across the CR1 phylogeny, while our analysis of the tuatara genome revealed several tuatara-specific CR1 subfamilies that clustered with CR1 sequences from mammalian, crocodylian, and turtle genomes. All avian CR1 subfamilies were grouped together (Fig. 5, light red).

Our analysis enabled a deeper look into CR1 evolution in squamates that has not been possible until recently. The 114 CR1 subfamilies we identified in squamates could be classified into two main clades (Fig. 5, light green). The first is an ancient CR1 clade (“Ancient-Squam,” 100% bootstrap support) comprised of consensus sequences we collected from 16 squamate genomes that forms the sister taxon to the clade of CR1 subfamilies collected from the tuatara, turtle, crocodylian, and mammalian genomes, which together formed an outgroup to all other amniote CR1 consensus sequences. The Ancient-Squam CR1 clade contained consensus sequences recovered from at least one species from each of the seven major squamate clades (Iguania, Anguimorpha, Serpentes, Gekkota, Dibamia, Lacertoidea, Scincoidea).

The remaining CR1 consensus sequences from squamate genomes formed a large clade of CR1 subfamilies that was unique to squamates, including sequences recovered from species across the squamate phylogeny (“squamate-specific,” 79% bootstrap support). Within this group, several CR1 subfamilies found in squamates clustered in patterns that were consistent with host phylogenetic relationships (see Discussion). These included CR1 subfamilies Viperid-A and Viperid-B that were unique to viperid snakes. Lizard-A sequences found in the dibamid (*Dibamus bourreti*) genome clustered with CR1 retrieved from several gecko genomes, which formed the outgroup to all other squamate-specific CR1 subfamilies.

There were several areas of the tree showing high-support grouping of subfamilies from species with large phylogenetic distances (e.g. Squam, Toxico-A, Toxico-B). While CR1 subfamilies from the Ancient-Squam clades were found only in squamate genomes, the squamate-specific clade contained one exception: an L3 subfamily from a mammalian host (L3_Mars_LINE/CR1 and L3_LINE/CR1) that was sister to the Tokay-L3 subfamily found in a gecko genome (*Gekko gecko*) with 100% branch support (Fig. 5, yellow box).

Discussion

Our study has revealed contrasting patterns of diversification, activity, conservation, and loss of the CR1

retrotransposon family throughout amniote and specifically squamate evolution. We found that (i) CR1 copy number varies significantly across the genomes of squamates and that this variation is much more extensive than in other reptile groups; (ii) there has also been much more extensive variation in CR1 rates of evolution and CR1 retrotransposition activity across squamates compared to other reptiles; and (iii) squamate genomes contain a diversity of shared and derived CR1 subfamilies indicative of both the persistence and evolutionary loss of several ancient CR1 lineages over amniote evolution. We also found that CR1 elements are mostly truncated in sequence and tend to be far upstream or downstream of exons in squamate genomes. These results point to a dynamic genomic environment in squamates that stands out in comparison with other sauropsids, which may shed light on the evolutionary forces that govern genome size and structure in vertebrates.

Throughout >300 million years of amniote evolution, the number of CR1 retrotransposon copies inhabiting and diversifying within amniote genomes has varied widely, reflecting different modes of genome evolution among mammals, birds, and reptiles. Our estimated rates of CR1 evolution support a “slow and steady” mode of genomic evolution for turtles and crocodylians, consistent with molecular clock-based analyses based on 4-fold degenerate sites (Shaffer et al. 2013; Green et al. 2014; Tollis et al. 2017; Gemmell et al. 2020). Meanwhile, avian genomic evolution, which parallels squamate evolution in terms of species diversity (i.e. both clades have ~11,000 extant species), features a much smaller range of CR1 copy numbers than squamates, with few outliers (Manthey et al. 2018, Galbraith et al. 2021). The much wider variation in CR1 copy number and the distribution of CR1 evolutionary rates in squamates suggest many rapid shifts in genome evolution in this group of reptiles.

Most of the squamate genomes that we analyzed contained a relatively high proportion of CR1 insertions between 0% and 10% K2P divergence from consensus (Fig. 4; supplementary fig. S6, Supplementary Material online), suggesting relatively recent CR1 activity. Applying a conservative genome-wide substitution rate of 0.0008 substitutions per site per million years for squamates (Perry et al. 2018; Finger et al. 2022), we estimate this range of CR1 divergence reflects CR1 activity within the last 125 million years of squamate evolution, which is younger than the divergence of the seven major squamate clades analyzed here (Fig. 2; Gamble et al. 2015; Pyron 2017; Burbrink et al. 2020; Simões et al. 2020; Brownstein et al. 2023). These results suggest that differences in the success of CR1 retrotransposons across squamates may be due to lineage-specific differences in CR1 activity over relatively short timespans.

While our analyses support the hypothesis that the common ancestor of amniotes hosted multiple subfamilies of

CR1 in its genome, followed by multiple losses during amniote evolution (Suh et al. 2015), we have expanded into squamate taxonomic diversity to reveal a more dynamic CR1 evolutionary history than has been previously suggested. For instance, the Ancient-Squam CR1 clade contained consensus sequences from at least one species from each of seven major squamate clades (Iguania, Anguimorpha, Serpentes, Gekkota, Dibamia, Lacertoidea, Scincoidea), suggesting that this CR1 subfamily originated in the genome of the common ancestor of all extant squamates and has persisted for as much as 200 million years (Brownstein et al. 2023). This conservation of the Ancient-Squam CR1 clade in the genomes of some species is contrasted by its apparent evolutionary loss in the genomes of squamates where we did not find Ancient-Squam CR1 sequences.

Certain CR1 subfamilies in the squamate-specific clade phylogenetically clustered in patterns that are reminiscent of the host phylogeny, supporting a model of CR1 vertical transmission between ancestor and descendent genomes. For instance, Lizard-A sequences found in the dibamid (*D. bourreti*) genome clustered with CR1 subfamilies retrieved from several gecko genomes, which in our analyses formed the outgroup to all other squamate-specific CR1 subfamilies (with 79% bootstrap support). This result mirrors most species trees for squamates published in the molecular era, which place geckos and/or dibamids at the root of the squamate phylogeny (Pyron et al. 2013; Streicher and Wiens 2017; Burbrink et al. 2020; Singhal et al. 2021). We also recovered a large snake-specific CR1 clade with relatively short branch lengths throughout, which possibly coincided with the snake radiation that started ~150 million years ago (Fig. 2; Caldwell et al. 2015). The shortest branch lengths in the squamate CR1 phylogeny were all in venomous snake CR1 subfamilies, perhaps reflecting the rapid diversification of those snake families during the late Cretaceous and Paleogene periods (Pyron and Burbrink 2012; Lee et al. 2016). In contrast, some species of squamates contained groups of CR1 subfamilies that were reciprocally monophyletic (i.e. found only in that species), such as *Correlophus ciliatus*, *D. bourreti*, and *A. carolinensis*, indicating more recent diversification of CR1 subfamilies unique to these species' evolutionary histories.

Our reconstruction of CR1 evolution relies on the assumption that CR1 is vertically transmitted between ancestor and descendant genomes through the germline, rather than horizontally transferred between very distantly related host taxa. Since phylogenetic reconstructions of non-LTR retrotransposons generally follow their host phylogenies (Malik et al. 1999; Kordiš et al. 2006; Waters et al. 2007; Boissinot and Sookdeo 2016), this model of vertical transmission (with occasional discordance due to incomplete lineage sorting, introgression, and loss via deletion) is reasonable most of the time. However, horizontal transfer

of non-LTR retrotransposons has been demonstrated in reptiles, as in the transfer of Bov-B elements between boid snakes and a ruminant mammalian ancestor (Ivancevic et al. 2018), as well as between squamates and potentially several ectoparasitic hosts (Pasquesi et al. 2018). In particular, horizontal transfer of CR1 has occurred between *Maculinea* butterflies and *Bombyx mori* silk moths (Novikova et al. 2007). In general, these instances of horizontal transfer can be detected when the element phylogeny is inconsistent with the host species phylogeny. In our phylogenetic analysis of CR1 consensus sequences, the CR1/L3 subfamily identified in the Tokay gecko (*G. gecko*) clustered with mammalian L3 deep within the otherwise squamate-specific CR1 clade (Fig. 5). Previously, it was suggested that this pattern of a mammalian L3 grouping with lepidosaurian CR1 sequences supported a model of shared ancestry followed by loss in mammals and other amniotes (Suh et al. 2015). After expanding the taxonomic sampling of squamates for our study, we suggest a more parsimonious explanation may be horizontal transfer between a squamate and a mammalian host.

Variation in dynamics of TE movement has been identified as a main factor in determining eukaryotic genome size (Gregory 2002; Blommaert 2020). Comparative research in birds and mammals support an "accordion" model whereby DNA loss counteracts TE expansion (via transposition) and results in stable genome sizes over evolutionary timescales. Across squamates, genome size is tightly constrained (Pasquesi et al. 2018; Armstrong et al. 2020; Gable et al. 2023), showing little variation across large phylogenetic distances. Our results suggest that while CR1 has been highly active over squamate evolutionary history, there is also evidence for the loss of CR1 subfamilies in the genomes of some squamate lineages, supporting an accordion model for squamates. For instance, we estimated that the net rate of CR1 copy number evolution for the entire squamate phylogeny is negative. However, not much is known about patterns of DNA deletion in squamates. Future studies should incorporate measures of TE copy number expansion, genomic content, degree of truncation or divergence, and key changes in squamate life history to account for different dynamics acting between different taxonomic units, similar to recent studies on mammalian TE diversity (Osmanski et al. 2023). The role of deletions in maintaining genome size, whether by nonallelic recombination resulting from the accumulation of TEs such as CR1, or via smaller deletions at the base pair level, will shed light on the mechanisms that control genome size variation in amniotes.

CR1 has been highly active in squamates and has produced large copy numbers in squamate genomes; however, similar to initial studies of the green anole genome (Novick et al. 2009; Alföldi et al. 2011), we found that on average

only ~3% of CR1 inserts were full length (and therefore potentially active) across squamate genomes. This may be due to: (i) the retrotransposition process being mistake-prone and failing to produce many full-length CR1 insertions; (ii) shorter elements being the result of DNA deletions; or (iii) selection which constrains the number of full-length elements in the genome. While the other two choices are possibilities, we think purifying selection (the third choice) plays a large role in squamate TE dynamics. In natural populations of green anoles, full-length non-LTR retrotransposon insertions are found at significantly lower allele frequencies than truncated ones (Tollis and Boissinot 2013; Ruggiero et al. 2017; Bourgeois et al. 2020), suggesting selection acts against TEs according to their length, possibly due to the deleterious effects of ectopic recombination (Petrov et al. 2003; Boissinot et al. 2006). We found that across squamate genomes CR1 insertions were relatively far from exons, suggesting selection against TE insertions in genic regions. Truncated CR1 elements may be slightly more tolerated near genes, as well as genome-wide, due to weaker purifying selection, perhaps because they are more likely to become fixed in natural populations of lizards (Tollis and Boissinot 2013; Ruggiero et al. 2017; Bourgeois et al. 2020). This would explain why CR1 elements persist in squamate genomes despite purifying selection widely acting against TE accumulation.

The mutational effect of TE activity can be deleterious but also sometimes beneficial to the host genome, resulting in new phenotypes and positive selection (Pastuzyn et al. 2018; Rishishwar et al. 2018; Schrader and Schmitz 2019). We found that CR1 insertions were much closer to genes on average in the genome of the scinoid *Hemicordylus*, which also contained the highest number of CR1 insertions overlapping exons (1,768). Understanding the contribution of TEs to exon sequences will be important for investigations of the evolution and genetic regulation underlying fascinating squamate phenotypes, including leglessness (Roscito et al. 2022). The variation in squamate CR1 evolution uncovered in this study may also explain the unmatched abundance of microsatellites in squamate genomes (Adams et al. 2016), particularly in snakes (Pasquesi et al. 2018) where CR1 elements often contain microsatellites on their 3' ends (Castoe et al. 2011). Understanding the role of CR1 activity in microsatellite expansion, and protein-coding and regulatory regions in general, will shed light on CR1's potential to be a mechanism of phenotypic evolution in squamates and other vertebrates.

TE evolution directly impacts genome architecture and drives genetic diversity across eukaryotes. Using expanded genomic datasets for reptiles, we uncovered evidence of evolutionary gain and loss of CR1 retrotransposon diversity throughout amniote genome evolution, identifying unique patterns of CR1 diversity and activity in squamate genomes. With diverse TE profiles and evidence of recent TE activity,

the genomic ecosystem of squamates appears less constrained than all other amniotes. Our study underscores squamates as a promising research model for not only TE activity and diversity, but also outstanding questions in the evolutionary dynamics of genome size, structure, and function in vertebrates.

Materials and Methods

Genomic Data

We downloaded whole-genome assemblies for 113 squamate species and the tuatara (*S. punctatus*, accession: GCA_003113815.1) from the National Center for Biotechnology Information (NCBI) (Sayers et al. 2022), Genome Warehouse (Chen et al. 2021), DNA Zoo (Dudchenko et al. 2017), and other open-source data repositories from individual publications (see [supplementary table S1, Supplementary Material](#) online). For species with multiple assemblies available, the representative genome was chosen based on highest contig N50 (following Bradnam et al. 2013). In total, we collected 114 lepidosaur genomes. We also downloaded whole-genome assemblies for 30 turtle, 200 avian, and 4 crocodylian species from the NCBI (see [supplementary table S1, Supplementary Material](#) online).

De Novo Repeat Annotation

We identified repetitive elements in each genome assembly using the de novo repeat finder RepeatModeler2.0 (Flynn et al. 2020) to build species-specific repeat libraries with search engine option "ncbi." We then used RepeatMasker4.1 (Smit et al. 2013) to query each species-specific repeat library against the appropriate reference genome, generating repeat annotations for each assembly using the options "-xsmall," "-no_is," and "-a" to generate a *.align alignment output file for downstream analyses. To remove nested elements from TE annotations, a Perl script from Kapusta et al. (2017) (parseRM.pl, available at <https://github.com/4ureliek/Parsing-RepeatMasker-Outputs>) was used to parse *.align files. We calculated K2P corrected percentages of divergence from TE family consensus sequences accounting for CpG sites using calcDivergince.pl in RepeatMasker. All CR1 landscapes were generated using a custom R script. We used pairwise Wilcoxon's rank-sum tests in R to compare the CR1 landscapes of 113 squamate species. We focused on the total proportion of the genome consisting of CR1 insertions in bins of 5%, 10%, and 15% K2P divergence to capture recent activity and used a false-discovery rate *P*-value adjustment for multiple comparisons (p.adjust.method = "BH") (Benjamini and Hochberg 1995).

Squamate, Turtle & Archosaur Species Tree Inference

To infer ancestral states across reptiles, we generated a species tree using markers from whole-genome sequence for

111 squamates, with the tuatara, the painted turtle (*Chrysemys picta*), the American alligator (*Alligator mississippiensis*), the chicken (*G. gallus*), and human for outgroups (supplementary table S1, Supplementary Material online). Two squamate genomes were excluded from species tree inference due to low genome completeness where high-quality taxonomic representatives of those lineages were available (*Gonatodes ferrugineus* and *Thamnophis elegans*). We extracted nucleotide sequences of 7,453 high-confidence single-copy sauropsid orthologs using Benchmarking Universal Single-Copy Orthologs (BUSCO) v. 5.4.4 (OrthoDB v10, Manni et al. 2021). We used MAFFT v7.505 (Katoh and Standley 2013) to generate alignments for each ortholog and TrimAL v1.4.rev15 (Capella-Gutierrez et al. 2009) to impose a gap threshold set to remove all positions in alignments with gaps in 30% or more of sequences (-gt 0.7). We generated summary statistics for all alignments with AMAS v3.04 (Borowiec 2016) and filtered for alignments with <15% gaps, >1,500 bp in length, <15,000 bp in length, and a minimum of 80% taxa representation.

After filtering, we analyzed 2,999 orthologs totaling 8,625,493 sites in IQ-Tree v2.2.0.4 (Minh et al. 2020) to infer gene trees using maximum likelihood with ModelFinder implemented to determine the best-fit model for each partition (see Data Availability). The resulting gene trees were used to infer a coalescent-consistent multilocus species tree in ASTRAL-III (Zhang et al. 2018), with branch support measured in local posterior probabilities (Sayyari and Mirarab 2016). The species tree was rooted using the *H. sapiens* outgroup. All nodes were fully supported based on local posterior probabilities, aside from three nodes known to be areas of phylogenetic conflict within squamates (supplementary fig. S7, Supplementary Material online): Dibamia–Gekkota (66%, Burbrink et al. 2020; Singhal et al. 2021), the Iguania–Anguimorpha (91%, Losos et al. 2012; Simões et al. 2018), and the Pseudaspidae–Elapidae (43%, Zaher et al. 2019).

The turtle coalescent-based phylogeny was inferred with the same methods described above, with genomes of 30 turtle species and five outgroups (the tuatara, the green anole [*Anolis carolinensis*], the American alligator [*A. mississippiensis*], the chicken [*G. gallus*], and human; supplementary table S1, Supplementary Material online). We extracted nucleotide sequences of 7,448 high-confidence single-copy sauropsid orthologs, and after filtering trimmed alignments based on criteria described above, a total of 3,342 orthologs with 9,716,540 sites were used to infer gene trees and the resulting coalescent-based species tree (rooted with *H. sapiens*). The coalescent-based species tree for turtles fully recovered the Lepidosauria, Archelosauria, Pleurodira, and Cryptodira clades (Shaffer et al. 2017; Gable et al. 2022). All nodes were fully supported based on local posterior probabilities (supplementary fig. S8, Supplementary Material online).

We inferred the archosaur coalescent-based phylogeny with the same methods described above, using 200 bird species and four crocodylian species, as well as four outgroups (the tuatara, the green anole, the painted turtle, and human). We extracted nucleotide sequences of 8,311 high-confidence single-copy orthologs using BUSCO v. 5.4.4 (OrthoDB v10 avian database). We used a total of 2,981 filtered orthologs with 8,417,077 sites. Most branches in the archosaur tree received full branch support, apart from order-level relationships in Neoavian birds (supplementary fig. S9, Supplementary Material online), which are known areas of uncertainty in avian phylogenetics (Prum et al. 2015; Kuhl et al. 2021). The few areas of low branch support did not impact the estimates of CR1 evolutionary rate variation, given that bird genomes largely lack variation in repeat content (Kapusta and Suh 2017, supplementary table S1, Supplementary Material online).

Divergence Time Estimation

We estimated divergence times based on the coalescent-based squamate, turtle, and archosaur phylogenies separately using the penalized-likelihood approach implemented in treePL (Smith and O’Meara 2012). We used treePL with minimum and maximum fossil calibrations for five nodes (Amniota: minimum 318 millions of years ago or My, maximum 332.9 My; Sauropsida: minimum 255.9 My, maximum 295.9 My; Archosauria: minimum 247.1 My, maximum 260.2 My; Lepidosauria: minimum 238 My, maximum 252.7 My; Squamata: minimum 168.9 My, maximum 209.5 My) from Benton et al. 2015 (timetree for squamates shown in supplementary fig. S10, Supplementary Material online; for others see Data Availability). An optimal smoothing parameter was chosen based on cross-validation testing, which sequentially removes terminal taxa to produce an estimate of branches from remaining data given an optimal smoothing parameter, which is found from the fit of the true branch and the pruned branch (Sanderson 2002). The *randomcv* option was used to run randomly sampled cross-validation analysis, and the *prime* option was initially run to find the best values for *opt*, *optad*, and *optcvad* optimizers. We also used the *thorough* option to ensure the run iterated until convergence.

Ancestral State Estimation Using CR1 Copy Numbers

To investigate differences in reptile CR1 evolution, we modeled CR1 copy number as a continuous trait to trace the evolution of this element across reptile diversification. We calculated CR1 copy number for each species using RepeatMasker annotation files to generate a data frame of trait values and used our species tree to inform taxa relationships. Any duplicate CR1 sequences in repeat annotations were removed using seqkit “rmdup” command. We

first inferred ancestral states of CR1 insert copy numbers on each branch of the species-level phylogenies for squamates, turtles, and archosaurs (birds and crocodylians both together and separately) with ancestral state reconstruction using fastAnc in the phytools R package (Revell 2012) with the species trees and CR1 trait file as input.

Within each phylogeny, we estimated the rate of CR1 copy number evolution for each branch by calculating the estimated CR1 copy number on the younger node minus the estimated CR1 copy number on the older node, divided by the length of the branch in millions of years (i.e. Δ CR1 copy number/ Δ Time [My]). Because the fastAnc method assumes homogenous rates across a phylogeny, we tested for evidence of rate heterogeneity in CR1 evolution using the Savage–Dickey metric implemented in the evrates R package (Martin et al. 2023) and found no significant evidence of heterogeneity. We set variance and confidence intervals to “TRUE” (i.e. vars = TRUE, CI = TRUE) to measure variability in estimates.

Phylogenomic Estimation of Squamate CR1—Within-Species Analyses

To reveal the phylogenetic structure of CR1 elements found in squamate genomes, we inferred CR1 phylogenies using full-length, active CR1 elements from the genomes of 78 squamate species representing 25 families and all 7 major clades (supplementary table S1, Supplementary Material online; Burbrink et al. 2020; Singhal et al. 2021). Our full dataset of 111 squamate genomes was filtered for taxonomic representation primarily and assembly contiguity (i.e. N50) secondarily to reduce computational burden in manual curation of consensus sequences. In order to extract CR1 sequences from each assembly, we first converted the RepeatMasker *.out output files to standard BED annotation files via the “rmsk2bed” command in “bedops” (Neph et al. 2012) with option “—keep-header.” Family-specific BED files of full-length CR1 elements were generated using bash scripting, with a minimum cut-off value of 2,000 bp used to initially filter for at least 60% complete ORF2 region (see supplementary table S1, Supplementary Material online for cut-offs used for each assembly) and maximum 6,000 bp cut-off to remove artifactual sequencing errors (e.g. chimeric sequences). The CR1 ORF2 is known to be highly conserved across species due to the presence of an RT region, so the ORF2 is commonly isolated for comparative studies (Novick et al. 2009). We then extracted CR1 sequences for each species using bedtools getfasta with bedtoolsv2.31.1 (Quinlan and Hall 2010) using the CR1 BED file and respective assembly as input.

We aligned the extracted CR1 nucleotide sequences for each species to a reference sequence of the chicken (*G. gallus*, GenBank accession: U88211.1, accessed

September 2021) complete consensus CR1 ORF2 using MAFFT (Katoh and Standley 2013). We used the ORF Finder tool in Geneious as an extra step to verify the boundaries of ORF2 and manually trimmed the CR1 sequences for each squamate species to the boundaries of the reference ORF2, followed by iterative realignment. After manual inspection, any individual sequences forcing large gaps (i.e. >500 bp) were removed from alignments.

We inferred species-specific CR1 phylogenies in IQ-Tree2 (Minh et al. 2020) with MFP and 1,000 bootstrap replicates (“-m MFP -B 1000”). The resulting 63 CR1 phylogenies were rooted using *G. gallus* as the outgroup and visualized in FigTree (Rambaut 2010). Subclades were manually annotated based on cladistic groupings with the most recent common ancestor of each subclade with 100% parametric bootstrap. Sequences belonging to each subclade were then extracted from CR1 alignments using “seqtk” (Li 2013) “subseq” to make concatenated multi-FASTA files for subclades from each species.

To verify that our CR1 sequences contained operable ORFs, we screened all sequences for the presence of a RT and ORF2 domain. We used usearch11.0.667 (Edgar 2010) with options “-fastx_findorfs -orfstyle 7 -mincodons 16” to retrieve the ORF2 domain. We used hmsearch (Finn et al. 2011) to generate a domain table for the RT domain against the Pfamv.28.0 database (Punta et al. 2012) as of May 2015 (includes 16,230 families) and collected all peptides annotated with RT. We then indexed the amino acid file via “esl-sfetch” and generated a multi-FASTA file of identified RT hits using the index and RT domain table as input via “esl-sfetch.” Finally, we used faSomeRecords from UCSC Executables (Nassar et al. 2023) to generate new multi-FASTA files with sequences verified for presence of intact ORFs and RTs. Filtered, trimmed sequences were realigned using MAFFT, and 50%-majority rule consensus sequences were inferred for each subfamily, resulting in a total of 118 lepidosaur CR1 subfamilies from 63 species. Sequences from 15 squamate species were filtered out for lack of intact ORF2 and RT domain (e.g. *Boaedon fuliginosus*, *Eryx tataricus*, see supplementary table S1, Supplementary Material online).

Phylogenomic Estimation of Squamate CR1—Between-Species Analyses

To put CR1 evolution of squamate genomes in the context of amniote genome evolution as a whole, all available amniote CR1 consensus sequences were downloaded from RepBase v24 and aligned in Geneious Prime via MAFFT. The alignment was trimmed to target the ORF2 region and any sequences <1,500 bp in length after trimming were filtered out. The amniote CR1 alignment was then filtered for intact ORF2 and RT domains using the scripts described above. A maximum-likelihood phylogeny was

inferred using IQ-Tree2 with 1,000 nonparametric bootstrap replicates and the best-fit model determined by ModelFinder (Kalyaanamoorthy et al. 2017) (options “-m MFP-B 1000”) with all newly identified squamate CR1 sub-families and amniote CR1 consensus sequences from RepBase.

CR1 Genomic Distribution Analyses

To uncover the genomic distribution of squamate CR1 insertions, we downloaded protein sets for 20 squamate species available from RefSeq (O’Leary et al. 2016), Genome Warehouse, and Zenodo (last accessed December 2023). For the remaining 91 species without publicly available gene annotations, we generated annotation files using LiftOffv1.6.3 (Shumate and Salzberg 2021) with the annotation of each species’ closest relative (according to our species tree; see [supplementary fig. S5, Supplementary Material](#) online). To ensure annotation quality, we generated proteome files using both LiftOff and original annotations via AGAT perl script `agat_sp_extract_sequences.pl` (option -p for amino acid sequences) (Dainat 2023) and measured proteome completeness via BUSCO v. 5.4.7 (OrthoDB v10 sauropsida database) in protein mode. We filtered all squamate proteomes by percent of complete BUSCOs > 80%, resulting in 30 accepted squamate genome annotations.

To understand the proximity of CR1 insertions to protein-coding regions, we combined our de novo repeat annotations with protein annotations for the 30 squamate species. We first parsed our protein annotations to isolate only exons and the largest isoform of each using custom bash scripts (see Data Availability). We then used *bedtools closest* to measure the distance of each squamate CR1 element to the nearest exon (options -t all -D b), using the -D option to calculate distance with respect to upstream/downstream of exons. We also used *bedtools window* with *bedtools overlap* to identify CR1 elements overlapping exons (default settings).

For the purposes of investigating the CR1 genomic environment, we filtered CR1–exon distance results by CR1 length, with a minimum length of 500 bp and a maximum length of 5,000 bp to avoid falsely identified repeats due to artifactual errors. To delineate between full-length and truncated CR1s, we separated results as follows: full-length CR1s > 2,500 bp (over 50% total length) and truncated CR1s 500 to 2,500 bp (at least 10% total length). We used a PGLS model to determine the impact of CR1–exon distances on CR1 length (generated using “nlme” (Pinheiro et al.) and “ape” (Paradis et al. 2004) R packages). We used Pagel’s lambda correlation structure with an initial lambda of 0.5 and a maximum-likelihood model fitting method. Using our species tree inferred for ancestral state reconstruction, we pruned to match our CR1–exon dataset,

dropping one species absent from our tree, resulting in 29 species. To account for variable intergenic distances across squamate species, we also calculated intergenic distances for 29 squamate genomes using custom Python scripts and generated a PGLS model with both CR1 length and intergenic distance as response variables with the same settings.

Lastly, we calculated GC content for 1 kb flanking regions downstream and upstream of all CR1 inserts, filtered by length. We used *seqkit subseq* with flags “-u” and “-d” along with “-f” and our de novo repeat annotations to generate flanking regions, followed by *seqkit fx2tab* with option “-gc” to calculate GC% for each flanking region.

Supplementary Material

[Supplementary material](#) is available at *Genome Biology and Evolution* online.

Acknowledgments

The authors would like to acknowledge the Office of Undergraduate Research and Creative Activity and the Monsoon High Performance Computing Cluster at Northern Arizona University. This work was supported by the Division of Environmental Biology (National Science Foundation) under grant DEB 2323124 to M.T., a State of Arizona Technology Research Initiative Fund (TRIF) Faculty Support Grant awarded to M.T., and in part through the National Institutes of Health U54 CA217376. Unpublished genome assemblies for *Uta stansburiana* and *Intellagama lesueurii* are used with permission from the DNA Zoo Consortium ([dnazoo.org](#)). The authors would like to acknowledge the Vertebrate Genomes Project for making genome assemblies available for *Euleptes europaea*, *Lacerta agilis*, *Podarcis raffonei*, *Rhineura floridana*, *Thamnophis elegans*, *Vipera latastei*, *Gopherus evegoodei*, *Chelonia mydas*, *Emys orbicularis*, *Gopherus flavomarginatus*, and *Malaclemys terrapin*. The authors thank Emma Reich, Panth Patel, and Vahid Fard for feedback on this manuscript.

Data Availability

Whole-genome files, alignments, tree files, and repeat annotations have been made publicly available at NCBI (PRJNA1073866; PRJNA1073867) and Zenodo at <https://doi.org/10.5281/zenodo.10521011>.

Literature Cited

Adams RH, Blackmon H, Reyes-Velasco J, Schield DR, Card DC, Andrew AL, Waynewood N, Castoe TA. Microsatellite landscape evolutionary dynamics across 450 million years of vertebrate genome evolution. *Genome*. 2016;59(5):295–310. <https://doi.org/10.1139/gen-2015-0124>.

- Alföldi J, Di Palma F, Grabherr M, Williams C, Kong L, Mauceli E, Russell P, Lowe CB, Glor RE, Jaffe JD, et al. The genome of the green anole lizard and a comparative analysis with birds and mammals. *Nature*. 2011;477(7366):587–591. <https://doi.org/10.1038/nature10390>.
- Armstrong J, Hickey G, Diekhans M, Fiddes IT, Novak AM, Deran A, Fang Q, Xie D, Feng S, Stiller J, et al. Progressive Cactus is a multiple-genome aligner for the thousand-genome era. *Nature*. 2020;587(7833):246–251. <https://doi.org/10.1038/s41586-020-2871-y>.
- Bao W, Kojima KK, Kohany O. Repbase Update, a database of repetitive elements in eukaryotic genomes. *Mob DNA*. 2015;6:11. <https://doi.org/10.1186/s13100-015-0041-9>.
- Benjamini Y, Hochberg Y. Controlling the false discovery rate: a practical and powerful approach to multiple testing. *J R Stat Soc Ser B Methodol*. 1995;57(1):289–300. <https://doi.org/10.1111/j.2517-6161.1995.tb02031.x>.
- Benton MJ, Donoghue PCJ, Vinther J, Asher RJ, Friedman M, Near TJ. Constraints on the timescale of animal evolutionary history. *Palaeontol Electron*. 2015;18(1):1–106. <https://doi.org/10.26879/424>.
- Blommaert J. Genome size evolution: towards new model systems for old questions. *Proc R Soc B Biol Sci*. 2020;287(1933):20201441. <https://doi.org/10.1098/rspb.2020.1441>.
- Boissinot S, Davis J, Entezam A, Petrov D, Furano AV. Fitness cost of LINE-1 (L1) activity in humans. *Proc Natl Acad Sci U S A*. 2006;103(25):9590–9594. <https://doi.org/10.1073/pnas.0603334103>.
- Boissinot S, Sookdeo A. The evolution of line-1 in vertebrates. *Genome Biol Evol*. 2016;8(12):3485–3507. <https://doi.org/10.1093/gbe/evw247>.
- Borowiec ML. AMAS: a fast tool for alignment manipulation and computing of summary statistics. *PeerJ*. 2016;4:e1660. <https://doi.org/10.7717/peerj.1660>.
- Bourgeois Y, Ruggiero RP, Hariyani I, Boissinot S. Disentangling the determinants of transposable elements dynamics in vertebrate genomes using empirical evidences and simulations. *PLoS Genet*. 2020;16(10):e1009082. <https://doi.org/10.1371/journal.pgen.1009082>.
- Bradnam KR, Fass JN, Alexandrov A, Baranay P, Bechner M, Birol I, Boisvert S, Chapman JA, Chapuis G, Chikhi R, et al. Assemblathon 2: evaluating de novo methods of genome assembly in three vertebrate species. *GigaScience*. 2013;2(1):10. <https://doi.org/10.1186/2047-217X-2-10>.
- Brownstein CD, Simões TR, Caldwell MW, Lee MSY, Meyer DL, Scarpetta SG. The affinities of the Late Triassic *Cryptovaranoidea* and the age of crown squamates. *R Soc Open Sci*. 2023;10(10):230968. <https://doi.org/10.1098/rsos.230968>.
- Burbrink FT, Grazziotin FG, Pyron RA, Cundall D, Donnellan S, Irish F, Keogh JS, Kraus F, Murphy RW, Noonan B, et al. Interrogating genomic-scale data for Squamata (lizards, snakes, and amphisbaenians) shows no support for key traditional morphological relationships. *Syst Biol*. 2020;69(3):502–520. <https://doi.org/10.1093/sysbio/syz062>.
- Caldwell MW, Nydam RL, Palci A, Apesteguía S. The oldest known snakes from the Middle Jurassic-Lower Cretaceous provide insights on snake evolution. *Nat Commun*. 2015;6:5996. <https://doi.org/10.1038/ncomms6996>.
- Capella-Gutierrez S, Silla-Martinez JM, Gabaldon T. trimAl: a tool for automated alignment trimming in large-scale phylogenomic analyses. *Bioinformatics*. 2009;25(15):1972–1973. <https://doi.org/10.1093/bioinformatics/btp348>.
- Card DC, Jennings WB, Edwards SV. Genome evolution and the future of phylogenomics of nonavian reptiles. *Animals*. 2023;13(3):471. <https://doi.org/10.3390/ani13030471>.
- Castoe TA, Hall KT, Guibotsy Mboulas ML, Gu W, de Koning APJ, Fox SE, Poole AW, Vemulapalli V, Daza JM, Mockler T, et al. Discovery of highly divergent repeat landscapes in snake genomes using high-throughput sequencing. *Genome Biol Evol*. 2011;3:641–653. <https://doi.org/10.1093/gbe/evr043>.
- Chalopin D, Naville M, Plard F, Galiana D, Volff J-N. Comparative analysis of transposable elements highlights mobilome diversity and evolution in vertebrates. *Genome Biol Evol*. 2015;7(2):567–580. <https://doi.org/10.1093/gbe/ew005>.
- Chen M, Ma Y, Wu S, Zheng X, Kang H, Sang J, Xu X, Hao L, Li Z, Gong Z, et al. Genome warehouse: a public repository housing genome-scale data. *Genomics Proteomics Bioinformatics*. 2021;19(4):584–589. <https://doi.org/10.1016/j.gpb.2021.04.001>.
- Dainat J. NBISweden/AGAT: AGAT-v1.2.0; 2023 [accessed 2024 Feb 1]. <https://zenodo.org/record/3552717>.
- de Koning APJ, Gu W, Castoe TA, Batzer MA, Pollock DD. Repetitive elements may comprise over two-thirds of the human genome. *PLoS Genet*. 2011;7(12):e1002384. <https://doi.org/10.1371/journal.pgen.1002384>.
- Dudchenko O, Batra SS, Omer AD, Nyquist SK, Hoeger M, Durand NC, Shamim MS, Machol I, Lander ES, Aiden AP, et al. De novo assembly of the *Aedes aegypti* genome using Hi-C yields chromosome-length scaffolds. *Science*. 2017;356(6333):92–95. <https://doi.org/10.1126/science.aal3327>.
- Edgar RC. Search and clustering orders of magnitude faster than BLAST. *Bioinformatics*. 2010;26(19):2460–2461. <https://doi.org/10.1093/bioinformatics/btq461>.
- Finger N, Farleigh K, Bracken JT, Leaché AD, François O, Yang Z, Flouri T, Charran T, Jezkova T, Williams DA, et al. Genome-scale data reveal deep lineage divergence and a complex demographic history in the Texas Horned Lizard (*Phrynosoma cornutum*) throughout the Southwestern and Central United States. *Genome Biol Evol*. 2022;14(1):evab260. <https://doi.org/10.1093/gbe/evab260>.
- Finn RD, Clements J, Eddy SR. HMMER web server: interactive sequence similarity searching. *Nucleic Acids Res*. 2011;39:W29–W37. <https://doi.org/10.1093/nar/gkr367>.
- Flynn JM, Hubble Robert, Goubert Clément, Rosen Jeb, Clark Andrew G, Feschotte Cédric, Smit Arian F. RepeatModeler2 for automated genomic discovery of transposable element families. *Proc Natl Acad Sci U S A*. 2020;117(17):9451–9457. <https://doi.org/10.1073/pnas.1921046117>.
- Gable SM, Byars MI, Literman R, Tollis M. A genomic perspective on the evolutionary diversification of turtles. *Syst Biol*. 2022;71(6):1331–1347. <https://doi.org/10.1093/sysbio/syab019>.
- Gable SM, Mendez JM, Bushroo NA, Wilson A, Byars MI, Tollis M. The state of squamate genomics: past, present, and future of genome research in the most speciose terrestrial vertebrate order. *Genes (Basel)*. 2023;14(7):1387. <https://doi.org/10.3390/genes14071387>.
- Galbraith JD, Kortschak RD, Suh A, Adelson DL. Genome stability is in the eye of the beholder: cR1 retrotransposon activity varies significantly across avian diversity. *Genome Biol Evol*. 2021;13(12):evab259. <https://doi.org/10.1093/gbe/evab259>.
- Gamble T, Greenbaum E, Jackman TR, Bauer AM. Into the light: diurnality has evolved multiple times in geckos. *Biol J Linn Soc*. 2015;115(4):896–910. <https://doi.org/10.1111/bij.12536>.
- Gemmell NJ, Rutherford K, Prost S, Tollis M, Winter D, Macey JR, Adelson DL, Suh A, Bertozzi T, Grau JH, et al. The tuatara genome reveals ancient features of amniote evolution. *Nature*. 2020;584(7821):403–409. <https://doi.org/10.1038/s41586-020-2561-9>.
- Gentles AJ, Wakefield MJ, Kohany O, Gu W, Batzer MA, Pollock DD, Jurka J. Evolutionary dynamics of transposable elements in the

- short-tailed opossum *Monodelphis domestica*. *Genome Res.* 2007;17(7):992–1004. <https://doi.org/10.1101/gr.6070707>.
- González-Porter GP, Maldonado JE, Flores-Villela O, Vogt RC, Janke A, Fleischer RC, Hailer F. Cryptic population structuring and the role of the Isthmus of Tehuantepec as a gene flow barrier in the critically endangered Central American River Turtle. *PLoS One.* 2013;8(9):e71668. <https://doi.org/10.1371/journal.pone.0071668>.
- Green RE, Braun EL, Armstrong J, Earl D, Nguyen N, Hickey G, Vandeweghe MW, St John JA, Capella-Gutiérrez S, Castoe TA, et al. Three crocodylian genomes reveal ancestral patterns of evolution among archosaurs. *Science.* 2014;346(6215):1254449. <https://doi.org/10.1126/science.1254449>.
- Gregory TR. Genome size and developmental complexity. *Genetica.* 2002;115(1):131–146. <https://doi.org/10.1023/A:1016032400147>.
- Ichiyanagi KI, Okada N. Mobility pathways for vertebrate L1, L2, CR1, and RTE clade retrotransposons. *Mol Biol Evol.* 2008;25(6):1148–1157. <https://doi.org/10.1093/molbev/msn061>.
- International Chicken Genome Sequencing Consortium. Sequence and comparative analysis of the chicken genome provide unique perspectives on vertebrate evolution. *Nature.* 2004;432(7018):695–716. <https://doi.org/10.1038/nature03154>.
- Ivancevic AM, Kortschak RD, Bertozzi T, Adelson DL. Horizontal transfer of BovB and L1 retrotransposons in eukaryotes. *Genome Biol.* 2018;19(1):85. <https://doi.org/10.1186/s13059-018-1456-7>.
- Jang HS, Shah NM, Du AY, Dailey ZZ, Pehrsson EC, Godoy PM, Zhang D, Li D, Xing X, Kim S, et al. Transposable elements drive widespread expression of oncogenes in human cancers. *Nat Genet.* 2019;51(4):611–617. <https://doi.org/10.1038/s41588-019-0373-3>.
- Kalyanamoorthy S, Minh BQ, Wong TKF, Von Haeseler A, Jermini LS. ModelFinder: fast model selection for accurate phylogenetic estimates. *Nat Methods.* 2017;14(6):587–589. <https://doi.org/10.1038/nmeth.4285>.
- Kapusta A, Suh A. Evolution of bird genomes—a transposon’s-eye view. *Ann N Y Acad Sci.* 2017;1389(1):164–185. <https://doi.org/10.1111/nyas.13295>.
- Kapusta A, Suh A, Feschotte C. Dynamics of genome size evolution in birds and mammals. *Proc. Natl. Acad. Sci U S A.* 2017;114(8):E1460–E1469. <https://doi.org/10.1073/pnas.1616702114>.
- Katoh K, Standley DM. MAFFT multiple sequence alignment software version 7: improvements in performance and usability. *Mol Biol Evol.* 2013;30(4):772–780. <https://doi.org/10.1093/molbev/mst010>.
- Kidwell MG, Lisch DR. Perspective: transposable elements, parasitic DNA, and genome evolution. *Evolution.* 2001;55(1):1–24. <https://doi.org/10.1111/j.0014-3820.2001.tb01268.x>.
- Köhler G, Khaing KPP, Than NL, Baranski D, Schell T, Greve C, Janke A, Pauls SU. A new genus and species of mud snake from Myanmar (Reptilia, Squamata, Homalopsidae). *Zootaxa.* 2021;4915(3):301–325. <https://doi.org/10.11646/zootaxa.4915.3.1>.
- Kordiš D, Lovšin N, Gubenšek F. Phylogenomic analysis of the L1 retrotransposons in deuterostomia. *Syst Biol.* 2006;55(6):886–901. <https://doi.org/10.1080/10635150601052637>.
- Kuhl H, Frankl-Vilches C, Bakker A, Mayr G, Nikolaus G, Boerno ST, Klages S, Timmermann B, Gahr M. An unbiased molecular approach using 3'-UTRs resolves the avian family-level tree of life. *Mol Biol Evol.* 2021;38(1):108–127. <https://doi.org/10.1093/molbev/msaa191>.
- Lee MSY, Sanders KL, King B, Palci A. Diversification rates and phenotypic evolution in venomous snakes (Elapidae). *R Soc Open Sci.* 2016;3(1):150277. <https://doi.org/10.1098/rsos.150277>.
- Li H. seqtk: a fast and lightweight tool for processing FASTA or FASTQ sequences; 2013 [accessed 2024 Feb 1]. <https://github.com/lh3/seqtk>.
- Losos JB, Hillis DM, Greene HW. Who speaks with a forked tongue? *Science.* 2012;338(6113):1428–1429. <https://doi.org/10.1126/science.1232455>.
- Malik HS, Burke WD, Eickbush TH. The age and evolution of non-LTR retrotransposable elements. *Mol Biol Evol.* 1999;16(6):793–805. <https://doi.org/10.1093/oxfordjournals.molbev.a026164>.
- Manni M, Berkeley MR, Seppely M, Zdobnov EM. BUSCO: Assessing genomic data quality and beyond. *Curr Protoc.* 2021:e323(1). <https://doi.org/10.1002/cpz1.323>.
- Manthey JD, Moyle RG, Boissinot S. Multiple and independent phases of transposable element amplification in the genomes of piciformes (woodpeckers and allies). *Genome Biol Evol.* 2018;10(6):1445–1456. <https://doi.org/10.1093/gbe/evy105>.
- Martin BS, Bradburd GS, Harmon LJ, Weber MG. Modeling the evolution of rates of continuous trait evolution. *Syst Biol.* 2023;72(3):590–605. <https://doi.org/10.1093/sysbio/syaa068>.
- Minh BQ, Schmidt HA, Chernomor O, Schrempf D, Woodhams MD, von Haeseler A, Lanfear R. IQ-TREE 2: new models and efficient methods for phylogenetic inference in the genomic era. *Mol Biol Evol.* 2020;37(5):1530–1534. <https://doi.org/10.1093/molbev/msaa015>.
- Nassar LR, Barber GP, Benet-Pagès A, Casper J, Clawson H, Diekhans M, Fischer C, Gonzalez JN, Hinrichs AS, Lee BT, et al. The UCSC Genome Browser database: 2023 update. *Nucleic Acids Res.* 2023;51(D1):D1188–D1195. <https://doi.org/10.1093/nar/gkac1072>.
- Neph S, Kuehn MS, Reynolds AP, Haugen E, Thurman RE, Johnson AK, Rynes E, Maurano MT, Vierstra J, Thomas S, et al. BEDOPS: high-performance genomic feature operations. *Bioinformatics.* 2012;28(14):1919–1920. <https://doi.org/10.1093/bioinformatics/bts277>.
- Novick PA, Basta H, Floumanhaft M, McClure MA, Boissinot S. The evolutionary dynamics of autonomous non-LTR retrotransposons in the lizard *Anolis carolinensis* shows more similarity to fish than mammals. *Mol Biol Evol.* 2009;26(8):1811–1822. <https://doi.org/10.1093/molbev/msp090>.
- Novikova O, Śliwińska E, Fet V, Settele J, Blinov A, Woyciechowski M. CR1 clade of non-LTR retrotransposons from *Maculinea* butterflies (Lepidoptera: Lycaenidae): evidence for recent horizontal transmission. *BMC Evol. Biol.* 2007;7:93. <https://doi.org/10.1186/1471-2148-7-93>.
- O’Leary NA, Wright MW, Brister JR, Ciufu S, Haddad D, McVeigh R, Rajput B, Robbertse B, Smith-White B, Ako-Adjei D, et al. Reference sequence (RefSeq) database at NCBI: current status, taxonomic expansion, and functional annotation. *Nucleic Acids Res.* 2016;44(D1):D733–D745. <https://doi.org/10.1093/nar/gkv1189>.
- Osmanski AB, Paulat NS, Korstian J, Grimshaw JR, Halsey M, Sullivan KAM, Moreno-Santillán DD, Crookshanks C, Roberts J, Garcia C, et al. Insights into mammalian TE diversity through the curation of 248 genome assemblies. *Science.* 2023;380(6643):eabn1430. <https://doi.org/10.1126/science.abn1430>.
- Paradis E, Claude J, Strimmer K. APE: analyses of phylogenetics and evolution in R language. *Bioinformatics.* 2004;20(2):289–290. <https://doi.org/10.1093/bioinformatics/btg412>.
- Pasquesi GIM, Adams RH, Card DC, Schield DR, Corbin AB, Perry BW, Reyes-Velasco J, Ruggiero RP, Vandeweghe MW, Shortt JA. Squamate reptiles challenge paradigms of genomic repeat element evolution set by birds and mammals. *Nat Commun.* 2018;9(1):2774. <https://doi.org/10.1038/s41467-018-05279-1>.
- Pastuzyn ED, Day CE, Kearns RB, Kyrke-Smith M, Taibi AV, McCormick J, Yoder N, Belnap DM, Erlendsson S, Morado DR, et al. The neuronal gene arc encodes a repurposed retrotransposon gag protein that mediates intercellular RNA transfer. *Cell.*

- 2018;172(1-2):275–288.e18. <https://doi.org/10.1016/j.cell.2017.12.024>.
- Perry BW, Card DC, McGlothlin JW, Pasquesi GIM, Adams RH, Schield DR, Hales NR, Corbin AB, Demuth JP, Hoffmann FG, et al. Molecular adaptations for sensing and securing prey and insight into amniote genome diversity from the garter snake genome. *Genome Biol Evol.* 2018;10(8):2110–2129. <https://doi.org/10.1093/gbe/evy157>.
- Petrov DA, Aminetzach YT, Davis JC, Bensasson D, Hirsh AE. Size matters: non-LTR retrotransposable elements and ectopic recombination in *Drosophila*. *Mol Biol Evol.* 2003;20(6):880–892. <https://doi.org/10.1093/molbev/msg102>.
- Pinheiro J, Bates D, R Core Team. nlme: linear and nonlinear mixed effects models. [accessed 2024 Feb 1]. <https://cran.r-project.org/package=nlme>.
- Pinto BJ, Gamble T, Smith CH, Wilson MA. A lizard is never late: squamate genomics as a recent catalyst for understanding sex chromosome and microchromosome evolution. *J Hered.* 2023;114(5):445–458. <https://doi.org/10.1093/jhered/esad023>.
- Prum RO, Berv JS, Dornburg A, Field DJ, Townsend JP, Lemmon EM, Lemmon AR. A comprehensive phylogeny of birds (*Aves*) using targeted next-generation DNA sequencing. *Nature.* 2015;526(7574):569–573. <https://doi.org/10.1038/nature15697>.
- Punta M, Coghill PC, Eberhardt RY, Mistry J, Tate J, Boursnell C, Pang N, Forslund K, Ceric G, Clements J, et al. The Pfam protein families database. *Nucleic Acids Res.* 2012;40(D1):D290–D301. <https://doi.org/10.1093/nar/gkr1065>.
- Pyron RA. Novel approaches for phylogenetic inference from morphological data and total-evidence dating in squamate reptiles (lizards, snakes, and amphisbaenians). *Syst Biol.* 2017;66(1):38–56. <https://doi.org/10.1093/sysbio/syw068>.
- Pyron RA, Burbrink FT. Extinction, ecological opportunity, and the origins of global snake diversity. *Evolution.* 2012;66(1):163–178. <https://doi.org/10.1111/j.1558-5646.2011.01437.x>.
- Pyron RA, Burbrink FT, Wiens JJ. A phylogeny and revised classification of Squamata, including 4161 species of lizards and snakes. *BMC Evol Biol.* 2013;13(1):93. <https://doi.org/10.1186/1471-2148-13-93>.
- Pyron RA, Wallach V. Systematics of the blindsnakes (Serpentes: Scolecophidia: Typhlopoidea) based on molecular and morphological evidence. *Zootaxa.* 2014;3829:1–81. <https://doi.org/10.11646/zootaxa.3829.1.1>.
- Quinlan AR, Hall IM. BEDTools: a flexible suite of utilities for comparing genomic features. *Bioinformatics.* 2010;26(6):841–842. <https://doi.org/10.1093/bioinformatics/btq033>.
- Rambaut A. FigTree v1.3.1. Edinburgh: Institute of Evolutionary Biology, University of Edinburgh; 2010 [accessed 2024 Feb 1]. <http://tree.bio.ed.ac.uk/software/figtree/>.
- Revell LJ. Phytools: an R package for phylogenetic comparative biology (and other things). *Methods Ecol. Evol.* 2012;3(2):217–223. <https://doi.org/10.1111/j.2041-210X.2011.00169.x>.
- Reynolds RG, Niemiller ML, Revell LJ. Toward a Tree-of-Life for the boas and pythons: multilocus species-level phylogeny with unprecedented taxon sampling. *Mol Phylogenet Evol.* 2014;71:201–213. <https://doi.org/10.1016/j.ympev.2013.11.011>.
- Rishishwar L, Wang L, Wang J, Yi SV, Lachance J, Jordan IK. Evidence for positive selection on recent human transposable element insertions. *Gene.* 2018;675:69–79. <https://doi.org/10.1016/j.gene.2018.06.077>.
- Roscito JG, Sameith K, Kirilenko BM, Hecker N, Winkler S, Dahl A, Rodrigues MT, Hiller M. Convergent and lineage-specific genomic differences in limb regulatory elements in limbless reptile lineages. *Cell Rep.* 2022;38(3):110280. <https://doi.org/10.1016/j.celrep.2021.110280>.
- Ruggiero RP, Bourgeois Y, Boissinot S. LINE insertion polymorphisms are abundant but at low frequencies across populations of *Anolis carolinensis*. *Front Genet.* 2017;8:44. <https://doi.org/10.3389/fgene.2017.00044>.
- Sanderson MJ. Estimating absolute rates of molecular evolution and divergence times: a penalized likelihood approach. *Mol Biol Evol.* 2002;19(1):101–109. <https://doi.org/10.1093/oxfordjournals.molbev.a003974>.
- Sayers EW, Bolton EE, Brister JR, Canese K, Chan J, Comeau DC, Connor R, Funk K, Kelly C, Kim S, et al. Database resources of the national center for biotechnology information. *Nucleic Acids Res.* 2022;50(D1):D20–D26. <https://doi.org/10.1093/nar/gkab1112>.
- Sayyari E, Mirarab S. Fast coalescent-based computation of local branch support from quartet frequencies. *Mol Biol Evol.* 2016;33(7):1654–1668. <https://doi.org/10.1093/molbev/msw079>.
- Schrader L, Schmitz J. The impact of transposable elements in adaptive evolution. *Mol Ecol.* 2019;28(6):1537–1549. <https://doi.org/10.1111/mec.14794>.
- Shaffer HB, McCartney-Melstad E, Near TJ, Mount GG, Spinks PQ. Phylogenomic analyses of 539 highly informative loci dates a fully resolved time tree for the major clades of living turtles (Testudines). *Mol Phylogenet Evol.* 2017;115:7–15. <https://doi.org/10.1016/j.ympev.2017.07.006>.
- Shaffer HB, Minx P, Warren DE, Shedlock AM, Thomson RC, Valenzuela N, Abramyan J, Amemiya CT, Badenhorst D, Biggar KK, et al. The western painted turtle genome, a model for the evolution of extreme physiological adaptations in a slowly evolving lineage. *Genome Biol.* 2013;14(3):R28. <https://doi.org/10.1186/gb-2013-14-3-r28>.
- Shedlock AM. Phylogenomic investigation of CR1 LINE diversity in reptiles. *Syst Biol.* 2006;55(6):902–911. <https://doi.org/10.1080/10635150601091924>.
- Shumate A, Salzberg SL. Liftoff: accurate mapping of gene annotations. *Bioinformatics.* 2021;37(12):1639–1643. <https://doi.org/10.1093/bioinformatics/btaa1016>.
- Simões TR, Caldwell MW, Talanda M, Bernardi M, Palci A, Vernygora O, Bernardini F, Mancini L, Nydam RL. The origin of squamates revealed by a Middle Triassic lizard from the Italian Alps. *Nature.* 2018;557(7707):706–709. <https://doi.org/10.1038/s41586-018-0093-3>.
- Simões TR, Vernygora O, Caldwell MW, Pierce SE. Megaevolutionary dynamics and the timing of evolutionary innovation in reptiles. *Nat Commun.* 2020;11(1):3322. <https://doi.org/10.1038/s41467-020-17190-9>.
- Singhal S, Colston TJ, Grundler MR, Smith SA, Costa GC, Colli GR, Moritz C, Pyron RA, Rabosky DL. Congruence and conflict in the higher-level phylogenetics of squamate reptiles: an expanded phylogenomic perspective. *Syst Biol.* 2021;70(3):542–557. <https://doi.org/10.1093/sysbio/syaa054>.
- Smit A, Hubble R, Green P. RepeatMasker Open-4.0; 2013 [accessed 2024 Feb 1]. <http://www.repeatmasker.org>.
- Smith SA, O'Meara BC. treePL: divergence time estimation using penalized likelihood for large phylogenies. *Bioinformatics.* 2012;28(20):2689–2690. <https://doi.org/10.1093/bioinformatics/bts492>.
- Sotero-Caio CG, Platt RN II, Suh A, Ray DA. Evolution and diversity of transposable elements in vertebrate genomes. *Genome Biol Evol.* 2017;9(1):161–177. <https://doi.org/10.1093/gbe/evw264>.
- Streicher JW, Wiens JJ. Phylogenomic analyses of more than 4000 nuclear loci resolve the origin of snakes among lizard families. *Biol Lett.* 2017;13(9):20170393. <https://doi.org/10.1098/rsbl.2017.0393>.
- Stuckas H, Fritz U. Identity of *Pelodiscus sinensis* revealed by DNA sequences of an approximately 180-year-old type specimen and a

- taxonomic reappraisal of *Pelodiscus* species (Testudines: Trionychidae). *J. Zool. Syst. Evol. Res.* 2011;49(4):335–339. <https://doi.org/10.1111/j.1439-0469.2011.00632.x>.
- Stumph WE, Hodgson CP, Tsai MJ, O'Malley BW. Genomic structure and possible retroviral origin of the chicken CR1 repetitive DNA sequence family. *Proc Natl Acad Sci U S A.* 1984;81(21):6667–6671. <https://doi.org/10.1073/pnas.81.21.6667>.
- Suh A, Churakov G, Ramakodi MP, Platt RN, Jurka J, Kojima KK, Caballero J, Smit AF, Vliet KA, Hoffmann FG, et al. Multiple lineages of ancient CR1 retroposons shaped the early genome evolution of amniotes. *Genome Biol Evol.* 2015;7(1):205–217. <https://doi.org/10.1093/gbe/evu256>.
- Tollis M, Boissinot S. Lizards and LINEs: selection and demography affect the fate of L1 retrotransposons in the genome of the green anole (*Anolis carolinensis*). *Genome Biol. Evol.* 2013;5(9):1754–1768. <https://doi.org/10.1093/gbe/evt133>.
- Tollis M, DeNardo DF, Cornelius JA, Dolby GA, Edwards T, Henen BT, Karl AE, Murphy RW, Kusumi K. The Agassiz's desert tortoise genome provides a resource for the conservation of a threatened species. *PLoS One.* 2017;12(5):e0177708. <https://doi.org/10.1371/journal.pone.0177708>.
- Uetz P, Hošek J. The reptile database; 2023 [accessed 2024 Feb 1]. <http://www.reptile-database.org/>.
- Waters PD, Dobigny G, Waddell PJ, Robinson TJ. Evolutionary history of LINE-1 in the major clades of placental mammals. *PLoS One.* 2007;2(1):e158. <https://doi.org/10.1371/journal.pone.0000158>.
- Zaher H, Murphy RW, Arredondo JC, Graboski R, Machado-Filho PR, Mahlow K, Montingelli GG, Quadros AB, Orlov NL, Wilkinson M, et al. Large-scale molecular phylogeny, morphology, divergence-time estimation, and the fossil record of advanced caenophidian snakes (Squamata: Serpentes). *PLoS One.* 2019;14(5):e0216148. <https://doi.org/10.1371/journal.pone.0216148>.
- Zhang C, Rabiee M, Sayyari E, Mirarab S. ASTRAL-III: polynomial time species tree reconstruction from partially resolved gene trees. *BMC Bioinformatics.* 2018;19(Suppl 6):153. <https://doi.org/10.1186/s12859-018-2129-y>.
- Zong Y, Ma J. A new species of the genus *Achalinopsis* from Jiangxi and the restoration of this genus. *Acta Herpetol Sin.* 1983;2(2):61–63.

Associate editor: Rachel Mueller