

New Ther1-derived SINE Squam3 in Scaled Reptiles

Nikita S Vassetzky (✉ nvas@eimb.ru)

Engelhardt Institute of Molecular Biology, Moscow, Russia <https://orcid.org/0000-0003-4453-4963>

Sergei A Kosushkin

Engelhardt Institute of Molecular Biology

Vitaly I Korchagin

Institute of Gene Biology


Alexey P Ryskov

Institute of Gene Biology

Research

Keywords: SINEs, retrotransposons, Squamata, Reptilia, evolution

DOI: <https://doi.org/10.21203/rs.3.rs-132339/v1>

License:  This work is licensed under a Creative Commons Attribution 4.0 International License. [Read Full License](#)

Abstract

Background

SINEs compose a significant part of animal genomes and are used to study the evolution of taxa. Despite significant advances in SINE studies in vertebrates and higher eukaryotes in general, their own evolution is poorly understood.

Results

We have found and described in detail a new SINE family Squam3 specific for scaled reptiles (Squamata). The subfamilies of this SINE demonstrate different distribution in the genomes of squamates, which together with the data on similar SINEs in the tuatara allowed us to propose a scenario of their evolution in the context of reptilian evolution.

Conclusions

Ancestral SINEs preserved in small numbers in most genomes can give rise to taxon-specific SINE families. Analysis of this aspect of SINEs can shed light on the history and mechanisms of SINE variation in reptilian genomes.

Background

The genomes are invaded by various repetitive elements, the most abundant of which (at least in higher eukaryotes) are Long (LINEs) and Short Interspersed Elements (SINEs). The amplification cycle of these retrotransposons includes the transcription of their genomic copies, reverse transcription of these transcripts, and integration of DNA into the genome. LINEs rely on the transcription by the cellular RNA polymerase II, while the reverse transcription and integration are fulfilled by their own enzymes. SINEs encode no enzymes and employ the cell machinery for their transcription (unlike LINEs, they are transcribed by RNA polymerase III (pol III) similarly to tRNAs) and their partner LINEs for the reverse transcription/integration. Accordingly, SINEs have pol III promoters for transcription and sequences recognized by the enzymes of their partner LINE for reverse transcription/integration.

A typical SINE consists of the head derived from one of the cellular RNA species (tRNA, 7SL RNA, or 5S RNA); the body the terminal part of which is recognized by the partner reverse transcriptase (RT); and the tail, a stretch of simple repeats. There are variations; certain SINEs have no body or their body contains sequences of unknown origin and function (some of them called central domains) are shared between otherwise unrelated SINE families, etc. (Vassetzky & Kramerov, 2013).

LINEs are found in the genomes of all higher eukaryotes. Clearly, SINEs cannot exist without LINEs but not vice versa, there are rare genomes that have LINEs but lack SINEs (e.g., *Saccharomyces* or *Drosophila*). During evolution, LINE (sub)families can become inactive; clearly, partner SINEs also cease to amplify. If another LINE family becomes active in a particular genome, replacing the sequence recognized by its RT can reanimate a SINE (Kramerov & Vassetzky, 2011a). Usually, a genome harbors one or a few SINE families; some of them can be inactive and were amplified in the ancestors. The analysis of SINE variation in different taxa allows us to use them as reliable phylogenetic markers (Suh et al., 2011).

The main lineages of the reptile-bird clade are scaled reptiles (Squamata), tuatara (Rhynchocephalia), turtles (Testudines), crocodiles (Crocodylia), and birds (Aves). Squamata, the largest order of reptiles, include the following major lineages: Serpentes (snakes), Iguania (including iguanids, agamids, chameleons), Anguimorpha, Scincomorpha, Lacertoidea, Gekkota, and Amphisbaenia. Phylogenetic relations among squamate reptiles are highly controversial due to the conflicting signals provided by molecular, morphological, and paleontological data. Together with tuatara, the only extant representative species of order Rhynchocephalia, they form monophyletic superorder Lepidosauria, which is the sister group to Archosauria, the clade that contains crocodiles and birds.

The first reptile SINE was found in 1990 in the Chinese pond turtle (Endoh et al., 1990); currently, we know around ten reptile-specific SINE families (Vassetzky & Kramerov, 2013), and the genomes also harbor degraded copies of ancient SINEs, e.g., AmnSINE (active in the ancestor of amniotes (Nishihara et al., 2006). Another example is Ther1 initially described as a mammalian SINE (MIR) but renamed later (N. Gilbert & Labuda, 2000; Smit & Riggs, 1995). Despite active sequencing of genomes in various species of lizards and snakes, no detailed comparative genomic studies of a SINE family in different taxa at the order level are available.

Recently, we have found a new SINE named Squam3 in the genomes of *Darevskia* and *Anolis* lizards. Further analysis demonstrated their distribution throughout squamates; a similar SINE family was found in the tuatara (Gemmell et al., 2020) but not in other reptiles or birds. However, Squam3 remained unnoticed in almost 40 genomes of squamates.

Below is the analysis of the structure, distribution, and evolution of this Squamata SINE and its relatives.

Results

Squam3 Identification

The consensus sequence of *Darevskia* Squam3 was used to search the genomes of scaled reptiles. It was found in all sequenced genomes (as well as in a variety of GenBank sequences of squamate species whose genomes have not been sequenced; Supplementary Table 1). No Squam3 was found beyond

Squamata (see below). The analysis of their consensus sequences has revealed three major subfamilies that we called Squam3A, Squam3B, and Squam3C.

Squam3 Structure

Squam3 is a typical SINE (Kramerov & Vassetzky, 2011b) composed of the tRNA-derived *head*, the *body* with a central domain and the 3'-terminus matching that of the partner LINE, and the *tail*, a stretch of several simple repeats. The consensus sequences range from 218 to 239 nt (without tail). There is no clear preference for a particular tRNA species (which is not uncommon among SINEs).

The *body* is similar to a fragment of the CORE central domain; the convincing similarity spans over 28 nt (double-overlined in Fig. 1). There is also a convincing similarity with the very 3'-terminus of LINEs of the L2 clade identified in *Darevskia valentini* (data not shown) and a less convincing similarity with L2 LINEs of *Anolis carolinensis* (L2-26_ACar and L2-24_ACar in Repbase).

The tail of Squam3 is largely composed of (TAAA)_n or (CTT)_n; however, certain species have (GTT)_n, (ATT)_n, or poly(A) (Table 1). Squam3 has a very low rate of target site duplications. This is unusual but not exceptional among SINEs and can point to an alternative cleavage pattern in different DNA strands by the partner LINE endonuclease (Kramerov & Vassetzky, 2011b).

Table 1

Squam3 SINE in scaled reptiles. Major subfamilies are described by the proportion and estimated number of full-length copies, the mean sequence similarity, unit. Certain parameters of genome assemblies are given in the left columns.

	Family	Species	Squam3 SINE					Genome assembly				
			subfamily	copies	length w/o tail, nt	similarity	tail	level	N	N ₅₀		
Squamata	Gekkota	Gekkonidae	<i>Gekko japonicus</i>	3A (21%)	54,829	224	60%	(TAAA) _n	□	4%	707,733	
				3B (10%)	26,109	238	75%	(CTT) _n				
				3B3 (69%)	180,151	271	81%	(CTT) _n				
			<i>Paroedura picta</i>	3A (12%)	17,761	221	57%	(TAAA) _n	□	9%	4,106,116	
				3B (19%)	28,122	238	61%	(CTT) _n				
				3B3 (57%)	84,367	267	74%	(CTT) _n				
	Eublepharidae	<i>Eublepharis macularius</i>	3A (50%)	65,672	224	60%	(TAAA) _n	□	2%	663,762		
			3B (50%)	65,672	239	85%	(GTT) _n					
Lacertoidea	Lacertidae	<i>Darevskia valentini</i>	3A (48%)	25,848	218	63%	(TAAA) _n	□	16%	658,539		
			3B (52%)	28,003	238	91%	(CTT) _n					
		<i>Lacerta agilis</i>	3A (16%)	17,446	219	62%	(TAAA) _n	□	0%	86,565,987		
			3B (84%)	91,590	238	92%	(CTT) _n					
		<i>Lacerta bilineata</i>	3A (39%)	24,123	219	63%	(TAAA) _n	□	0%	368,212		
			3B (61%)	37,732	238	75%	(CTT) _n					
		<i>Lacerta viridis</i>	3A (39%)	24,836	219	61%	(TAAA) _n	□	0%	662,519		
			3B (61%)	38,847	238	94%	(CTT) _n					
		<i>Podarcis muralis</i>	3A (35%)	16,967	220	61%	(TAAA) _n	□	0%	92,398,148		
			3B (65%)	31,440	238	88%	(CTT) _n					
		<i>Zootica vivipara</i>	3A (38%)	10,036	220	61%	(TAAA) _n	□	3%	92,810,032		
			3B (62%)	16,374	238	89%	(CTY) _n					
			Teiidae	<i>Salvator merianae</i>	3A (53%)	4,892	221	54%	(TAAA) _n	□	2%	55,382,274
					3B (47%)	4,338	234	85%	(CTT) _n			
Serpentes	Colubridae	<i>Pantherophis guttatus</i>	3C	12,936	226	64%	(TAAA) _n	□	5%	16,790,024		
			3C	12,961	226	63%	(TAAA) _n	□	3%	14,519,768		
		<i>Ptyas mucosa</i>	3C	19,524	226	63%	(TAAA) _n	□	3%	15,963,960		
		<i>Thamnophis elegans</i>	3C	16,934	226	64%	(TAAA) _n	□	0%	440,193		
		<i>Thamnophis sirtalis</i>	3C	12,410	226	63%	(TAAA) _n	□	21%	647,592		

Family	Species	Squam3 SINE					Genome assembly			
		subfamily	copies	length w/o tail, nt	similarity	tail	level	N	N ₅₀	
Elapidae	<i>Thermophis baileyi</i>	3C	15,914	228	65%	(TAAA) _n	□	8%	2,413,955	
	<i>Emydocephalus ijimae</i>	3C	15,782	226	62%	(TAAA) _n	□	0%	18,937	
	<i>Hydrophis cyanocinctus</i>	3C	15,094	226	64%	(TAAA) _n	□	4%	7,437	
	<i>Hydrophis hardwickii</i>	3C	15,782	228	62%	(TAAA) _n	□	4%	5,391	
	<i>Hydrophis melanocephalus</i>	3C	14,271	226	63%	(TAAA) _n	□	11%	59,810	
	<i>Laticauda colubrina</i>	3C	19,118	226	63%	(TAAA) _n	□	13%	3,139,541	
	<i>Laticauda laticaudata</i>	3C	27,835	226	61%	(TAAA) _n	□	0%	39,330	
	<i>Naja naja</i>	3C	10,813	226	64%	(TAAA) _n	□	6%	224,088,900	
	<i>Notechis scutatus</i>	3C	27,122	226	63%	(TAAA) _n	□	5%	5,997,050	
	<i>Ophiophagus hannah</i>	3C	11,613	226	63%	(TAAA) _n	□	13%	241,519	
	<i>Pseudonaja textilis</i>	3C	17,187	226	65%	(TAAA) _n	□	2%	14,685,528	
Pythonidae	<i>Python bivittatus</i>	3A (43%)	9,349	221	58%	(TAAA) _n	□	4%	213,970	
		3C (57%)	12,393	237	75%	(A) _n				
Viperidae	<i>Crotalus horridus</i>	3C	15,006	226	63%	(TAAA) _n	□	12%	23,829	
	<i>Crotalus pyrrhus</i>	3C	15,556	226	64%	(TAAA) _n	□	0%	5,299	
	<i>Crotalus viridis viridis</i>	3C	18,694	226	63%	(TAAA) _n	□	6%	179,897,795	
	<i>Protobothrops flavoviridis</i>	3C	20,667	226	64%	(TAAA) _n	□	3%	467,050	
	<i>Protobothrops mucrosquamatus</i>	3C	20,184	228	64%	(TAAA) _n	□	8%	424,052	
	<i>Vipera berus berus</i>	3C	19,964	226	64%	(TAAA) _n	□	14%	126,452	
Shinisauria	Shinisauridae	<i>Shinisaurus crocodilurus</i>	3A	165,288	225	58%	(TAAA) _n	□	8%	1,469,749

	Family	Species	Squam3 SINE					Genome assembly			
			subfamily	copies	length w/o tail, nt	similarity	tail	level	N	N ₅₀	
	Anguimorpha	Anguinae	<i>Ophisaurus gracilis</i>	3A	35,118	225	71%	(TAAA) _n	□	3%	1,273,270
	Varanoidea	Varanidae	<i>Varanus komodoensis</i>	3A	108,651	229	66%	(TAAA) _n	□	1%	23,831,982
	Iguania	Agamidae	<i>Pogona vitticeps</i>	3A	4,542	221	53%	(TAAA) _n	□	4%	2,290,546
		Dactyloidae	<i>Anolis carolinensis</i>	3A	457	217	54%	(TAA) _n	□	5%	150,641,573
Rhynchocephalia			<i>Sphenodon punctatus</i>	—	—	—	—	—	□	10%	3,052,611
Testudines			<i>Trachemys scripta elegans</i>	—	—	—	—	—	□	1%	147,425,149
Crocodylia			<i>Crocodylus porosus</i>	—	—	—	—	—	□	5%	84,437,661
Aves			<i>Gallus gallus</i>	—	—	—	—	—	□	1%	91,315,245

Squam3 Subfamilies

Genomic copies of SINEs are subject to random mutations; accordingly, single-nucleotide mutations can be used to identify subfamilies only for highly conserved SINEs. We use extended insertions/deletions to distinguish the subfamilies. This allowed us to identify three major Squam3 subfamilies designated as Squam3A, Squam3B, and Squam3C (Fig. 1). Squam3B has a characteristic 11-nt insertion (marked in pink in Fig. 1), and Squam3C has a characteristic 7-nt insertion (marked in blue in Fig. 1). There are also minor differences between Squam3 subfamilies. In addition, there are subsubfamilies, one of these (Squam3B3) has become a major variant in the two Gekkonidae species.

Further analysis of Squam3-related sequences in the tuatara genome has revealed a similar SINE family (tuaMIRa) with a 32-nt insertion (marked in amaranth in Fig. 1). This insertion restores the CORE central domain and makes the element similar to Ther1 (MIR) although with a 12-nt deletion in the LINE-derived region (marked in violet in Fig. 1). Moreover, there is another element (tuaMIRb) with a similar insertion that lacks the ~40-nt region (between the CORE and the LINE-derived region) conserved in other Squam3- and Ther1-related SINEs but has a much longer L2 LINE-derived region due to the 77-nt insertion (marked in mango in Fig. 1). The sequences of these tuatara SINE families were recently reported (Gemmell et al., 2020) but only the relation to MIR (former name of Ther1) and the mean divergence of all Ther1-related sequences were mentioned.

Apart from that, Squam3 subfamilies differ by the tail, which is largely (TAAA)_n in Squam3A/C or (CTT)_n in Squam3B. The mean sequence similarity also differs between subfamilies, it peaks in Squam3B (up to 94%) but is lower in Squam3C (~63%) and Squam3A (54–63%). Figure 2 visualizes the diversity of Squam3 in the available genomes of lizards, snakes, and tuatara. Squam3C in most snake species demonstrates little variation between species, which contrasts the diversity within Squam3A and Squam3B subfamilies. The tuatara SINEs clearly constitute a separate cluster with Ther1.

The number of Squam3 full-length copies varied over a wide range: from ~500 in *Anolis carolinensis* to ~260,000 in *Gekko japonicus*. The mean similarity of Squam3 subfamilies in most species is 60–65% with the notable exceptions of Squam3B (~90%) and Squam3A in Iguania (53%).

Distribution of Squam3 in Reptile Genomes

After establishing the subfamilies of Squam3, their consensus sequences were searched in the same manner in all genomes of squamates and neighboring taxa. Overall, the genomes of 38 squamates, tuatara, turtle (*Trachemys scripta elegans*), crocodile (*Crocodylus porosus*), and bird (*Gallus gallus*) were analyzed. Squam3 was found in all squamates but not in other reptiles or birds (Table 1). Similar SINE families were found in the tuatara (*Sphenodon punctatus*). When this work was in progress, (Gemmell et al., 2020) reported these SINEs, so we use their nomenclature of tuatara SINEs.

The genomes of Gekkota and Lacertoidea (Gekkonidae, Eublepharidae, Lacertidae, and Teiidae families) had both Squam3A and Squam3B subfamilies in similar proportions (although the proportion of Squam3A could be occasionally as low as 12%). Snakes had the Squam3C subfamily except for the python, which had 43% Squam3A. The rest of the squamates (Shinisauridae, Anguinae, Varanidae, Agamidae, and Dactyloidae families) had the Squam3A subfamily alone (Table 1). The analysis of non-genomic NCBI sequences largely confirms this pattern except that a few highly divergent Squam3A sequences were found in three more snake families (Elapidae, Lamprophiidae, and Viperidae) (Supplementary Table 1). We specifically searched for Squam3A in one of the advanced snakes (*Vipera berus*), and our estimate is 330 copies.

The tuatara (Sphenodontidae) has a set of tuaMIR families related to Squam3 and Ther1. Thus, we specifically searched for these sequences in the genomes of Squamata. No tuaMIRb or tuaMIRc were found, while minor tuaMIRa quantities exist in all squamate genomes analyzed ranging from a single full-length copy to ~500 (in *Shinisaurus crocodilurus*) (Supplementary Table 2). All snakes have a single tuaMIRa copy corresponding to the same locus (as judged by very similar flanking regions).

Discussion

One of the most intriguing aspects of SINEs is how they emerge and evolve. This study gives us a unique opportunity to trace this for a single SINE family in a very wide range of taxa.

The Squam3 SINE family was found in scaled reptiles (Squamata) but not in the tuatara (Rhynchocephalia) and further lineages including crocodiles, birds, and turtles. We have found three major subfamilies distinguished by relatively long insertions/deletions (Squam3A, Squam3B, and Squam3C). They also differ by the number of copies and the mean sequence similarity, which points to the age of the SINE family (to be precise, to the time of its amplification) since SINE genomic copies are not subject to selective pressure and gradually accumulate mutations with time.

Evolution of Squam3

Overall, it looks like there was a small pool (a few hundred?) of not very active Squam3A in the genomes of ancestral Squamata. In some lineages (Shinisauridae and Varanidae), Squam3A amplified quite actively without much sequence modifications (to reach ~ 165,000 copies in *Shinisaurus crocodilurus*, this number of Squam3 copies was higher only in the *Gekko japonicus* with a ~ twice larger genome). Squam3A amplification was also active in Anguillidae (~ 35,000 copies in *Ophisaurus gracilis*) but it started relatively recently considering the high mean similarity (71%) of the SINE sequences in this legless lizard. On the contrary, Squam3A gradually declined in Agamidae (~ 4,500 copies and 53% mean similarity in *Pogona vitticeps*). Finally, Squam3A ceased to propagate (and evolve) in Dactyloidae (< 500 copies in *Anolis carolinensis*).

While other Squam3 subfamilies emerged in squamate lineages, Squam3A continued to amplify in Gekkota and Lacertoidea (from ~ 5,000 to ~ 65,000 copies) but not in snakes (except primitive ones, ~ 9,000 in *Python bivittatus*). We could find only ~ 300 copies in *Vipera berus*; individual copies were also found among non-genomic sequences in four other snake families (Supplementary Table 2).

After Squam3A declined in the Gekkota and Lacertoidea, their genomes gave rise to the Squam3B subfamily. It is arguably the youngest Squam3 subfamily. Amazingly, the mean similarity of Squam3B is very high in *Lacerta agilis* (92%) and *L. viridis* (94%) but as low as 75% in *L. bilineata*. This indicates that Squam3B is likely active in *L. viridis* and *L. agilis* but not in *L. bilineata* representing the same genus. In Gekkonidae, the more prolific subfamily Squam3B3 emerged (~ 180,000 copies in *Gekko japonicus*, which is the top number of all Squam3 subfamilies). For some reason, the activity of both Squam3A and Squam3B was low in Teiidae (*Salvator merianae*) but still, Squam3B amplified later than Squam3A.

The Squam3C subfamily is limited to snakes; moreover, it is the only major subfamily in most snakes. It looks like minor Squam3A quantities were present in all squamates but did not propagate in most snakes. Instead, the Squam3C in advanced snakes (Caenophidia) became active slightly later or in the same period of time (the mean Squam3C similarity is 61–65% vs. 51–71% in Squam3A). This pattern is not true for *Python bivittatus* representing more primitive snakes, where the amplification of Squam3A was followed by that of Squam3C (with the mean similarities of 58% and 75%, respectively).

Origin of Squam3

We were very excited to find what is called the missing link by paleontologists in the tuatara. The genome of *Sphenodon punctatus* has three SINE families that are similar to Squam3 in the left ~ 120 nt except the 32-nt deletion in Squam3 relative to two of them (tuaMIRa and tuaMIRb). Thus, a large CORE fragment was deleted in two tuaMIR SINEs. Another tuatara SINE (tuaMIRc) has this deletion and is similar to Squam3 within this region (but differs in the head and LINE-derived regions).

It is plausible that Ther1 (tuaMIRa) that was active in the ancestor of mammals, reptiles, and birds (N. Gilbert & Labuda, 2000) acquired the 32-nt deletion within the CORE domain in the Lepidosauria ancestor, and this precursor SINE gave rise to tuaMIRc in the tuatara and Squam3 in Squamata.

Conclusions

We report a new SINE family Squam3 found in all (38 to the time of analysis) sequenced genomes of scaled reptiles (Squamata). Despite the ever-increasing amount of genomic data for lizards and snakes, this quite prolific SINE was not reported previously. The evolutionary dynamics of SINE families and subfamilies is obscure and linked to the divergence of the genomes. This study is a step forward in understanding how SINEs emerge and decline. We identified and described Squam3 subfamilies and directly compared their structural traits and copy number across a variety of major squamate taxa in comparison with related tuatara SINE families. This study gives an insight into how new SINE families emerge and evolve.

Methods

Most genomic banks were downloaded from NCBI Genomes (<https://www.ncbi.nlm.nih.gov/genome>) except *Anolis carolinensis*, *Podarcis muralis* (Ensembl, <https://www.ensembl.org>), *Ophisaurus gracilis*, *Shinisaurus crocodilurus* (diArk, <https://www.diark.org/diark>), and *Darevskia valentini* (*Darevskia* (ID 327916) - BioProject - NCBI, n.d.). Non-genomic sequences of squamates (excluding the species with sequenced genomes) were also extracted from NCBI (<https://www.ncbi.nlm.nih.gov/taxonomy/advanced>). If no data on the genome size was available in publications or the Animal Genome Size Database (Gregory, 2020), it was calculated as the mean of most close species.

We used custom Perl scripts based on the Smith-Waterman search to find genomic copies of SINEs with at least 65% identity and 90% length overlap with the consensus. After all Squam3 families were identified, the genome bank was successively depleted using their consensus sequences.

Multiple sequence alignments were generated using *MAFFT* (Yamada et al., 2016) and edited by *GeneDoc* (Nicholas & Nicholas, 1997). Subfamilies were identified manually and analyzed in a larger sample if necessary. We considered only ample subfamilies ($\geq 1\%$ of the total number of full-length copies). A

search for tuaMIR SINEs in reptile/bird genomes was carried out by initial identification of all copies with at least 65% similarity to the consensus sequences followed by manual subsampling and realigning of candidate copies possibly containing specific mutations separating them from tuaMIRa sequences. The mean similarity was determined for 100 randomly selected sequences (or all available if less) using the *alifold* program (Eddy S., Rivas Laboratory, Cambridge, 2005). A neighbor-joining tree was constructed using MEGA software with 1000 bootstrap replications and the “partial deletion” option.

Declarations

Ethics approval and consent to participate

Not applicable

Consent for publication

Not applicable

Availability of data and materials

The data generated are available in the manuscript supporting files. The banks of Squam3 SINEs, as well as multiple alignments of random sets of SINE sequences, are available for each species on request.

Competing interests

The authors declare that they have no competing interests.

Funding

This research was funded by the Russian Science Foundation (RSF) Project No. 19-14-00083.

Authors' contributions

NSV and APR, conceptualization; all, genomic data analysis; NSV and SAK, study design and manuscript preparation; APR and VIK, supervision; APR, project administration and funding acquisition. All authors read and approved the final manuscript.

Acknowledgments

We thank Dr. Dmitri Kramerov for critical reading of the manuscript.

References

- Aird, S. D., Arora, J., Barua, A., Qiu, L., Terada, K., & Mikheyev, A. S. (2017). Population genomic analysis of a pitviper reveals microevolutionary forces underlying venom chemistry. *Genome Biology and Evolution*, 9(10), 2640–2649. <https://doi.org/10.1093/gbe/evx199>
- Alföldi, J., Di Palma, F., Grabherr, M., Williams, C., Kong, L., Mauceli, E., Russell, P., Lowe, C. B., Glor, R. E., Jaffe, J. D., Ray, D. A., Boissinot, S., Shedlock, A. M., Botka, C., Castoe, T. A., Colbourne, J. K., Fujita, M. K., Moreno, R. G., Ten Hallers, B. F., ... Lindblad-Toh, K. (2011). The genome of the green anole lizard and a comparative analysis with birds and mammals. *Nature*, 477(7366), 587–591. <https://doi.org/10.1038/nature10390>
- Andrade, P., Pinho, C., de Lanuza, G. P. i., Afonso, S., Brejcha, J., Rubin, C. J., Wallerman, O., Pereira, P., Sabatino, S. J., Bellati, A., Pellitteri-Rosa, D., Bosakova, Z., Bunikis, I., Carretero, M. A., Feiner, N., Marsik, P., Paupério, F., Salvi, D., Soler, L., ... Carneiro, M. (2019). Regulatory changes in pterin and carotenoid genes underlie balanced color polymorphisms in the wall lizard. *Proceedings of the National Academy of Sciences of the United States of America*, 116(12), 5633–5642. <https://doi.org/10.1073/pnas.1820320116>
- Brian Simison, W., Parham, J. F., Papenfuss, T. J., Lam, A. W., & Henderson, J. B. (2020). An Annotated Chromosome-Level Reference Genome of the Red-Eared Slider Turtle (*Trachemys scripta elegans*). *Genome Biology and Evolution*, 12(4), 456–462. <https://doi.org/10.1093/gbe/evaa063>
- Castoe, T. A., De Koning, A. P. J., Hall, K. T., Card, D. C., Schield, D. R., Fujita, M. K., Ruggiero, R. P., Degner, J. F., Daza, J. M., Gu, W., Reyes-Velasco, J., Shaney, K. J., Castoe, J. M., Fox, S. E., Poole, A. W., Polanco, D., Dobry, J., Vandeweghe, M. W., Li, Q., ... Pollock, D. D. (2014). Erratum: The Burmese python genome reveals the molecular basis for extreme adaptation in snakes (Proceedings of the National Academy of Sciences of the United States of America (2013) 110, 51, (20645-20650) DOI: 10.1073/pnas.1314475110). *Proceedings of the National Academy of Sciences of the United States of America*, 111(8), 3194. <https://doi.org/10.1073/pnas.1324133111>
- Crotalus horridus* (ID 16679) - Genome - NCBI. (n.d.). Retrieved December 2, 2020, from https://www.ncbi.nlm.nih.gov/genome/16679?genome_assembly_id=274149
- Crotalus viridis viridis* (ID 71654) - Genome - NCBI. (n.d.). Retrieved December 2, 2020, from https://www.ncbi.nlm.nih.gov/genome/71654?genome_assembly_id=434976
- Darevskia* (ID 327916) - BioProject - NCBI. (n.d.). Retrieved December 9, 2020, from <https://www.ncbi.nlm.nih.gov/bioproject/327916>

- Eddy S. Rivas Laboratory, Cambridge, U. (2005). *SQUID - C function library for sequence analysis*.
- Endoh, H., Nagahashi, S., & Okada, N. (1990). A highly repetitive and transcribable sequence in the tortoise genome is probably a retroposon. *Eur. J. Biochem.*, *189*(1), 25–31. <http://www.ncbi.nlm.nih.gov/pubmed/1691979>
- Gao, J., Li, Q., Wang, Z., Zhou, Y., Martelli, P., Li, F., Xiong, Z., Wang, J., Yang, H., & Zhang, G. (2017). Sequencing, de novo assembling, and annotating the genome of the endangered Chinese crocodile lizard *Shinisaurus crocodilurus*. *GigaScience*, *6*(7), 1–6. <https://doi.org/10.1093/gigascience/gix041>
- Gemmell, N. J., Rutherford, K., Prost, S., Tollis, M., Winter, D., Macey, J. R., Adelson, D. L., Suh, A., Bertozzi, T., Grau, J. H., Organ, C., Gardner, P. P., Muffato, M., Patricio, M., Billis, K., Martin, F. J., Flicek, P., Petersen, B., Kang, L., ... Edmonds, H. (2020). The tuatara genome reveals ancient features of amniote evolution. *Nature*, *584*(7821), 403–409. <https://doi.org/10.1038/s41586-020-2561-9>
- GenomeArk - Lacerta agilis*. (n.d.). Retrieved December 2, 2020, from https://vgp.github.io/genomeark/Lacerta_agilis/
- GenomeArk - Thamnophis elegans*. (n.d.). Retrieved December 2, 2020, from https://vgp.github.io/genomeark/Thamnophis_elegans/
- Georges, A., Li, Q., Lian, J., O'Meally, D., Deakin, J., Wang, Z., Zhang, P., Fujita, M., Patel, H. R., Holleley, C. E., Zhou, Y., Zhang, X., Matsubara, K., Waters, P., Graves, J. A. M., Sarre, S. D., & Zhang, G. (2015). High-coverage sequencing and annotated assembly of the genome of the Australian dragon lizard *Pogona vitticeps*. *GigaScience*, *4*(1). <https://doi.org/10.1186/s13742-015-0085-2>
- Ghosh, A., Johnson, M. G., Osmanski, A. B., Louha, S., Bayona-Vásquez, N. J., Glenn, T. C., Gongora, J., Green, R. E., Isberg, S., Stevens, R. D., & Ray, D. A. (2019). A High-Quality Reference Genome Assembly of the Saltwater Crocodile, *Crocodylus porosus*, Reveals Patterns of Selection in Crocodylidae. *Genome Biology and Evolution*, *12*(1), 3635–3646. <https://doi.org/10.1093/gbe/evz269>
- Gilbert, C., Meik, J. M., Dashevsky, D., Card, D. C., Castoe, T. A., & Schaack, S. (2014). Endogenous hepadnaviruses, bornaviruses and circoviruses in snakes. *Proceedings of the Royal Society B: Biological Sciences*, *281*(1791). <https://doi.org/10.1098/rspb.2014.1122>
- Gilbert, N., & Labuda, D. (2000). Evolutionary inventions and continuity of CORE-SINEs in mammals. *J. Mol. Biol.*, *298*(3), 365–377. http://www.ncbi.nlm.nih.gov/80/entrez/query.fcgi?cmd=Retrieve&db=PubMed&list_uids=10772856&dopt=Abstract
- Gregory, T. R. (2020). *Animal Genome Size Database*. <http://www.genomesize.com>
- Hara, Y., Takeuchi, M., Kageyama, Y., Tatsumi, K., Hibi, M., Kiyonari, H., & Kuraku, S. (2018). Madagascar ground gecko genome analysis characterizes asymmetric fates of duplicated genes. *BMC Biology*, *16*(1), 1–19. <https://doi.org/10.1186/s12915-018-0509-4>
- Hillier, L. W., Miller, W., Birney, E., Warren, W., Hardison, R. C., Ponting, C. P., Bork, P., Burt, D. W., Groenen, M. A. M., Delany, M. E., Dodgson, J. B., Chinwalla, A. T., Cliften, P. F., Clifton, S. W., Delehaunty, K. D., Fronick, C., Fulton, R. S., Graves, T. A., Kremitzki, C., ... Wilson, R. K. (2004). Sequence and comparative analysis of the chicken genome provide unique perspectives on vertebrate evolution. *Nature*, *432*(7018), 695–716. <https://doi.org/10.1038/nature03154>
- Hydrophis cyanocinctus (ID 75161) - Genome - NCBI*. (n.d.). Retrieved December 2, 2020, from https://www.ncbi.nlm.nih.gov/genome/75161?genome_assembly_id=437861
- Hydrophis hardwickii (ID 75162) - Genome - NCBI*. (n.d.). Retrieved December 2, 2020, from https://www.ncbi.nlm.nih.gov/genome/75162?genome_assembly_id=437862
- Kishida, T., Go, Y., Tatsumoto, S., Tatsumi, K., Kuraku, S., & Toda, M. (2019). Loss of olfaction in sea snakes provides new perspectives on the aquatic adaptation of amniotes. *Proceedings of the Royal Society B: Biological Sciences*, *286*(1910). <https://doi.org/10.1098/rspb.2019.1828>
- Kolora, S. R. R., Weigert, A., Saffari, A., Kehr, S., Walter Costa, M. B., Spröer, C., Indrischek, H., Chintalapati, M., Lohse, K., Doose, G., Overmann, J., Bunk, B., Bleidorn, C., Grimm-Seyfarth, A., Henle, K., Nowick, K., Faria, R., Stadler, P. F., & Schlegel, M. (2019). Divergent evolution in the genomes of closely related lacertids, *Lacerta viridis* and *L. bilineata*, and implications for speciation. *GigaScience*, *8*(2), 22. <https://doi.org/10.1093/gigascience/giy160>
- Kramerov, D. A., & Vassetzky, N. S. (2011a). Origin and evolution of SINEs in eukaryotic genomes. *Heredity*, *107*(6), 487–495. http://www.ncbi.nlm.nih.gov/entrez/query.fcgi?cmd=Retrieve&db=PubMed&dopt=Citation&list_uids=21673742
- Kramerov, D. A., & Vassetzky, N. S. (2011b). SINEs. *Wiley Interdiscip. Rev. RNA*, *2*(6), 772–786. <https://doi.org/10.1002/wrna.91>
- Li, J. T., Gao, Y. D., Xie, L., Deng, C., Shi, P., Guan, M. L., Huang, S., Ren, J. L., Wu, D. D., Ding, L., Huang, Z. Y., Nie, H., Humphreys, D. P., Hillis, D. M., Wang, W. Z., & Zhang, Y. P. (2018). Comparative genomic investigation of high-elevation adaptation in ectothermic snakes. *Proceedings of the National Academy of Sciences of the United States of America*, *115*(33), 8406–8411. <https://doi.org/10.1073/pnas.1805348115>
- Lind, A. L., Lai, Y. Y., Mostovoy, Y., Holloway, A. K., Iannucci, A., Mak, A. C. Y., Fondi, M., Orlandini, V., Eckalbar, W. L., Milan, M., Rovatsos, M., Kichigin, I. G., Makunin, A. I., Johnson Pokorná, M., Altmanová, M., Trifonov, V. A., Schijlen, E., Kratochvíl, L., Fani, R., ... Bruneau, B. G. (2019). Genome of the Komodo dragon reveals adaptations in the cardiovascular and chemosensory systems of monitor lizards. *Nature Ecology and Evolution*, *3*(8), 1241–1252. <https://doi.org/10.1038/s41559-019-0945-8>

- Liu, Y., Zhou, Q., Wang, Y., Luo, L., Yang, J., Yang, L., Liu, M., Li, Y., Qian, T., Zheng, Y., Li, M., Li, J., Gu, Y., Han, Z., Xu, M., Wang, Y., Zhu, C., Yu, B., Yang, Y., ... Gu, X. (2015). Gekko japonicus genome reveals evolution of adhesive toe pads and tail regeneration. *Nature Communications*, 6. <https://doi.org/10.1038/ncomms10033>
- Nicholas, K. B., & Nicholas, H. B. J. (1997). *GeneDoc: Analysis and Visualization of Genetic Variation*. <http://www.nrbsc.org/gfx/genedoc/index.html>
- Nishihara, H., Smit, A. F., & Okada, N. (2006). Functional noncoding sequences derived from SINEs in the mammalian genome. *Genome Res.*, 16(7), 864–874. http://www.ncbi.nlm.nih.gov/entrez/query.fcgi?cmd=Retrieve&db=PubMed&dopt=Citation&list_uids=16717141
- Notechis scutatus (ID 14408) - Genome - NCBI*. (n.d.). Retrieved December 2, 2020, from https://www.ncbi.nlm.nih.gov/genome/14408?genome_assembly_id=408294
- Pantherophis obsoletus (ID 88953) - Genome - NCBI*. (n.d.). Retrieved December 2, 2020, from https://www.ncbi.nlm.nih.gov/genome/88953?genome_assembly_id=889057
- Pseudonaja textilis (ID 72610) - Genome - NCBI*. (n.d.). Retrieved December 2, 2020, from https://www.ncbi.nlm.nih.gov/genome/72610?genome_assembly_id=408420
- Ptyas mucosa (ID 44753) - Genome - NCBI*. (n.d.). Retrieved December 2, 2020, from https://www.ncbi.nlm.nih.gov/genome/44753?genome_assembly_id=884075
- Roscito, J. G., Sameith, K., Pippel, M., Francoijs, K. J., Winkler, S., Dahl, A., Papoutsoglou, G., Myers, G., & Hiller, M. (2018). The genome of the tegu lizard *Salvator merianae*: combining Illumina, PacBio, and optical mapping data to generate a highly contiguous assembly. *GigaScience*, 7(12), 1–13. <https://doi.org/10.1093/gigascience/giy141>
- Shibata, H., Chijiwa, T., Oda-Ueda, N., Nakamura, H., Yamaguchi, K., Hattori, S., Matsubara, K., Matsuda, Y., Yamashita, A., Isomoto, A., Mori, K., Tashiro, K., Kuhara, S., Yamasaki, S., Fujie, M., Goto, H., Koyanagi, R., Takeuchi, T., Fukumaki, Y., ... Ogawa, T. (2018). The habu genome reveals accelerated evolution of venom protein genes. *Scientific Reports*, 8(1), 1–11. <https://doi.org/10.1038/s41598-018-28749-4>
- Smit, A. F., & Riggs, A. D. (1995). MIRs are classic, tRNA-derived SINEs that amplified before the mammalian radiation. *Nucleic Acids Res.*, 23(1), 98–102. <http://www.ncbi.nlm.nih.gov/pubmed/7870595>
- Song, B., Cheng, S., Sun, Y., Zhong, X., Jin, J., Guan, R., Murphy, R. W., Che, J., Zhang, Y., & Liu, X. (2015). A genome draft of the legless anguid lizard, *Ophisaurus gracilis*. *GigaScience*, 4(1), 15–17. <https://doi.org/10.1186/s13742-015-0056-7>
- Suh, A., Paus, M., Kieffmann, M., Churakov, G., Franke, F. A., Brosius, J., Kriegs, J. O., & Schmitz, J. (2011). Mesozoic retroposons reveal parrots as the closest living relatives of passerine birds. *Nat. Commun.*, 2, 443. <https://doi.org/10.1038/ncomms1448>
- Suryamohan, K., Krishnankutty, S. P., Guillory, J., Jevit, M., Schröder, M. S., Wu, M., Kuriakose, B., Mathew, O. K., Perumal, R. C., Koludarov, I., Goldstein, L. D., Senger, K., Dixon, M. D., Velayutham, D., Vargas, D., Chaudhuri, S., Muraleedharan, M., Goel, R., Chen, Y. J. J., ... Seshagiri, S. (2020). The Indian cobra reference genome and transcriptome enables comprehensive identification of venom toxins. *Nature Genetics*, 52(1), 106–117. <https://doi.org/10.1038/s41588-019-0559-8>
- Thamnophis sirtalis (ID 16688) - Genome - NCBI*. (n.d.). Retrieved December 2, 2020, from https://www.ncbi.nlm.nih.gov/genome/16688?genome_assembly_id=245767
- Ullate-Agote, A., Milinkovitch, M. C., & Tzika, A. C. (2014). The genome sequence of the corn snake (*Pantherophis guttatus*), a valuable resource for EvoDevo studies in squamates. *International Journal of Developmental Biology*, 58(10–12), 881–888. <https://doi.org/10.1387/ijdb.150060at>
- Vassetzky, N. S., & Kramerov, D. A. (2013). SINEBase: a database and tool for SINE analysis. *Nucleic Acids Research*, 41(Database issue), D83-9. <https://doi.org/10.1093/nar/gks1263>
- Vipera berus berus (ID 14467) - Genome - NCBI*. (n.d.). Retrieved December 2, 2020, from https://www.ncbi.nlm.nih.gov/genome/14467?genome_assembly_id=214193
- Vonk, F. J., Casewell, N. R., Henkel, C. V., Heimberg, A. M., Jansen, H. J., McCleary, R. J. R., Kerckamp, H. M. E., Vos, R. A., Guerreiro, I., Calvete, J. J., Wüster, W., Woods, A. E., Logan, J. M., Harrison, R. A., Castoe, T. A., De Koning, J., Pollock, D. D., Yandell, M., Calderon, D., ... Richardson, M. K. (2013). The king cobra genome reveals dynamic gene evolution and adaptation in the snake venom system. *Proceedings of the National Academy of Sciences of the United States of America*, 110(51), 20651–20656. <https://doi.org/10.1073/pnas.1314702110>
- Xiong, Z., Li, F., Li, Q., Zhou, L., Gamble, T., Zheng, J., Kui, L., Li, C., Li, S., Yang, H., & Zhang, G. (2016). Draft genome of the leopard gecko, *Eublepharis macularius*. *GigaScience*, 5(1). <https://doi.org/10.1186/s13742-016-0151-4>
- Yamada, K. D., Tomii, K., & Katoh, K. (2016). Application of the MAFFT sequence alignment program to large data - Reexamination of the usefulness of chained guide trees. *Bioinformatics*, 32(21), 3246–3251. <https://doi.org/10.1093/bioinformatics/btw412>

Yurchenko, A. A., Recknagel, H., & Elmer, K. R. (2020). Chromosome-Level Assembly of the Common Lizard (*Zootoca vivipara*) Genome. *Genome Biology and Evolution*, 12(11), 1953–1960. <https://doi.org/10.1093/gbe/evaa161>