Horizontal transfer and southern migration: the tale of Hydrophiinae's marine journey.

James D. Galbraith[1], Alastair J. Ludington[1], Richard J. Edwards[2], Kate L. Sanders[1], Alexander Suh*[3,4], David L. Adelson*[1]

1) School of Biological Sciences, University of Adelaide, Adelaide, SA 5005, Australia

2) School of Biotechnology and Biomolecular Sciences, University of New South Wales, Sydney, NSW 2052, Australia

3) School of Biological Sciences, University of East Anglia, Norwich Research Park, NR4 7TU, Norwich, United Kingdom

4) Department of Organismal Biology - Systematic Biology, Evolutionary Biology Centre, Uppsala University, SE-752 36 Uppsala, Sweden

* David L. Adelson and Alexander Suh corresponding authors and contributed equally to this work.

Email: david.adelson@adelaide.edu.au

James: https://orcid.org/0000-0002-1871-2108

Alastair: https://orcid.org/0000-0003-3994-6023

Alexander: https://orcid.org/0000-0002-8979-9992

Kate: https://orcid.org/0000-0002-9581-268X

David: https://orcid.org/0000-0003-2404-5636

Classification

Biological Sciences: Evolution

Keywords

transposable element, comparative genomics, Serpentes

Author Contributions

J.D.G., A.S and D.L.A. designed research; D.L.A. and A.S supervised research; R.J.E. provided olive sea snake genome assembly; J.D.G. and A.L. performed research; and J.D.G., A.S. and D.L.A. wrote the paper with input from A.J.L. and K.L.S.

Abstract

Transposable elements (TEs), also known as jumping genes, are sequences able to move or copy themselves within a genome. As TEs move throughout genomes they can be exapted as coding and regulatory elements, or can promote genetic rearrangement. In so doing TEs act as a source of genetic novelty, hence understanding TE evolution within lineages is key in understanding adaptation to their environment. Studies into the TE content of lineages of mammals such as bats have uncovered horizontal transposon transfer (HTT) into these lineages, with squamates often also containing the same TEs. Despite the repeated finding of HTT into squamates, little comparative research has examined the evolution of TEs within squamates. The few broad scale studies in Squamata which have been conducted found both the diversity and total number of TEs differs significantly across the entire order. Here we examine a diverse family of Australo-Melanesian snakes (Hydrophiinae) to examine if this pattern of variable TE content and activity holds true on a smaller scale. Hydrophiinae diverged from Asian elapids ~15-25 Mya and have since rapidly diversified into 6 amphibious, ~60 marine and ~100 terrestrial species which fill a broad range of ecological niches. We find TE diversity and expansion differs between hydrophiines and their Asian relatives and identify multiple HTTs into Hydrophiinae, including three transferred into the ancestral hydrophiine likely from marine species. These HTT events provide the first tangible evidence that Hydrophiinae reached Australia from Asia via a marine route.

Introduction

Elapids are a diverse group of venomous snakes found across Africa, Asia, the Americas and Australia. Following their divergence from Asian elapids ~30 Mya, the Australo-Melanesian elapids (Hydrophiinae) have rapidly diversified into more than 160 species including ~100 terrestrial snakes, ~60 fully marine sea snakes, and 6 amphibious sea kraits [1]. Both the terrestrial and fully marine hydrophiines have adapted to a wide range of habitats and niches. Terrestrial Hydrophiinae are found across Australia, for example the eastern brown snake *Pseudonaja textilis* in open habitats, the tiger snake (*Notechis scutatus*) in subtropical and temperate habitats, and the inland taipan (*Oxyuranus microlepidotus*) to inland arid habitats [2]. Since transitioning to a marine habitat, many sea snakes have specialised to feed on a single prey such as fish eggs, catfish, eels or burrowing gobies, while others such as *Aipysurus laevis* are generalists [3,4]. Sea kraits (*Laticauda*) are amphibious and have specialised to hunt various fish including eels and anguilliform-like fish at sea, while digesting prey, mating and shedding on land [5]. Since transitioning to marine environments, both sea snakes and sea kraits have been the recipients of multiple independent horizontal

transposon transfer (HTT) events, which may have had adaptive potential [6,7].

Transposable elements (TEs) are mobile genetic elements that can move or copy themselves across the genome, and account for a large portion of most vertebrate genomes [8,9]. Though often given short shrift in genome analyses, TEs are important agents of genome evolution and generate genomic diversity [10,11]. For example, the envelope gene of endogenous retroviruses was exapted by both mammals and viviparous lizards to function in placental development [12]. In addition, unequal crossing over caused by CR1 retrotransposons led to the duplication, and hence diversification, of PLA_2 venom genes in pit vipers [13]. Transposable elements (TEs) are classified into one of two major classes based on their structure and replication method [14]. DNA transposons (Class II) proliferate through a “cut and paste” method, possess terminal inverted repeats and are further split based on the transposase sequence used in replication. Retrotransposons (Class I) are split into LTR retrotransposons and non-LTR retrotransposons, which proliferate through ``copy and paste” methods. Both subclasses of retrotransposons are split into numerous superfamilies based on both coding and structural features [15–17].Within the diverse lineages of higher vertebrates, the evolution of TEs is well described in eutherian mammals and birds. The total repetitive content of both bird and mammal genomes is consistently at 7-10% and 30-50% respectively. Similarly, most lineages of both birds and eutherian mammals are dominated by a single superfamily of non-LTR retrotransposons (CR1s and L1s respectively) and a single superfamily of LTR retrotransposons (endogenous retroviruses in both) [8,18]. Some lineages of birds and mammals contain horizontally transferred retrotransposons which have variably been successful (AviRTE and RTE-BovB respectively) [19,20].

In stark contrast to mammals and birds, squamates have highly variable mobilomes, both in terms of the diversity of their TE superfamilies and the level activity of said superfamilies within each genome [21]. While these broad comparisons have found significant variation in TEs between distant squamate lineages, none have examined how TEs have evolved within a single family of squamates. The one in depth study into the mobilome of snakes found the Burmese python genome is approximately ~21% TE and appears to have low TE expansion, while that of a pit viper is ~45% TE due to the expansion of numerous TE superfamilies and microsatellites since their divergence ~90 Mya [22,23] . Unfortunately it is unclear whether similar expansions have occurred within other lineages of venomous snakes. Here we examine the TE landscape of the family Hydrophiinae, and in doing so discover horizontal transfer events into the ancestral hydrophiine, sea kraits and sea snakes.

Methods

Ab initio TE annotation of the elapid genomes

We used RepeatModeler2 [24] to perform *ab initio* TE annotation of the genome assemblies of 4 hydrophiines (*Aipysurus laevis*, *Notechis scutatus*, *Pseudonaja textilis* and *Laticauda colubrina*) and 2 Asian elapids (*Naja naja* and *Ophiophagus hannah*). We manually curated the subfamilies of TEs identified by RepeatModeler (rm-families) to ensure they encompassed the full TE, were properly classified and that each species' library was non-redundant.

We first purged redundant rm-families from each species library based on pairwise identity to and coverage by other rm-families within the library. Using BLAST [25] we calculated the similarity between all rm-families. Any rm-family with over 75% of its length aligning to a larger rm-family at 90% pairwise identity or higher was removed from the library. We then searched for each non-redundant rm-family within their source genome with BLASTN (-task dc-megablast) and selected the best 30 hits based on bitscore. In order to ensure we could retrieve full length TE insertions, we extended the flanks of each hit by 4000 bp. Using BLASTN (-task dc-megablast) we pairwise aligned each of the 30 extended sequences to others, trimming trailing portions of flanks which did not align to flanks of the other 29 sequences. Following this, we constructed a multiple sequence alignment (MSA) of the 30 trimmed sequences with MAFFT [26] (--localpair). Finally we trimmed each MSA at the TE target site duplications (TSDs) and constructed a consensus from the multiple sequence alignments using Geneious Prime 2021.1.1 (www.geneious.com) which we henceforth refer to as a mc-subfamily (manually curated subfamily).

To classify the mc-subfamilies we searched for intact protein domains in the consensus sequences using RPSBLAST [27] and the CDD library [28] and identified homology to previously described TEs in Repbase using CENSOR online [29] . Using this data in conjunction with the classification set out in Wicker (2007) [14], we classified previously unclassified mc-subfamilies where possible and corrected the classification of mc-subfamilies where necessary. Where possible we used the criteria of Feschotte and Pritham (2007) [17] to identify unclassified DNA transposons using TSDs and terminal inverted repeats. Finally, we removed any genes from the mc-subfamily libraries based on searches using online NCBI BLASTN and BLASTX searches against the nt/nr and UniProt libraries respectively [30,31]. Any mc-subfamilies unable to be classified were labelled as "Unknown".

TE annotation of the elapid genomes

We constructed a custom library for TE annotation of the elapid genome assemblies by combining the mc-subfamilies from the six assemblies with previously described lepidosaur TEs identified using RepeatMasker's "queryRepeatDatabase.pl" utility. Using RepeatMasker, we generated repeat annotations of all six elapid genome assemblies.

Estimating ancestral TE similarity

To estimate the sequence conservation of ancestral TEs, and hence categorise recently expanding TEs as either ancestral or horizontally transferred, we identified orthologous TE insertions and their flanks present in both the *Notechis scutatus* and *Naja naja* genome assemblies. From the *Notechis* repeat annotation, we took a random sample of 5000 TEs over 500 bp in length and extended each flank by 1000 bp. Using BLASTN (-task dc-megablast) we searched for the TEs and their flanks in the *Naja* assembly and selected all hits containing at least 250 bp of both the TE and the flank. Sequences with more than one hit containing flanks were treated as potential segmental duplications. We also removed any potential segmental duplications from the results. We then used the orthologous sequences to estimate the expected range in similarity between TEs present in the most recent common ancestor of Australian and Asian elapids. Based on this information, TEs with 95% or higher pairwise identity to the mc-subfamily used to identify them were treated as likely inserted in hydrophiine genomes since their divergence from Asian elapids. In addition, mc-subfamilies which we had identified as recently expanding in hydrophiines but were not found at 80% or higher pairwise identity in other serpentine genomes, were identified as candidates for horizontal transfer.

Identifying recent TE expansions

In each of the four hydrophiines, using the RepeatMasker output we identified mc-subfamilies comprising at least 100 total kbp having 95% or higher pairwise identity to the mc-subfamily. We treated these mc-subfamilies as having expanded since Hydrophiinae's divergence from Asian elapids. We reduced any redundancy between recently expanding mc-subfamilies by clustering Using CD-HIT-EST (-c 0.95 -n 6 -d 50) [32]. Using BWA [33], we mapped raw transcriptome reads of eye tissue taken from each of the hydrophiines [34] for back to these mc-subfamilies. Retrotransposons with RNA-seq reads mapping across their whole length and DNA transposons with RNA-seq reads mapping to their coding regions were treated as expressed and therefore currently expanding.

Continued expansion or horizontal transfer

Using BLASTN (-task dc-megablast), we searched for homologs of recently expanding mc-subfamilies in a

range of snake genomes including Asian elapids, colubrids, vipers and a python. We classified mc-subfamilies having copies of 80% or higher pairwise identity to the query sequence in other snakes as ancestral. All hydrophiine mc-subfamilies we were unable to find in other snakes were treated as candidates for horizontal transfer. We searched for the horizontal transfer candidates in approximately 600 additional metazoan genomes using BLASTN (-task dc-megablast). We classified all mc-subfamilies present in non-serpentine genomes at 80% or higher pairwise identity and absent from other serpentine genomes at 80% or higher pairwise identity as horizontally transferred into hydrophiines.

Results and Discussion

Genome quality affects repeat annotation

Previous studies have highlighted the importance of genome assembly quality in repeat annotation, with higher sequencing depths and long read technologies critical for resolving TEs [35,36]. Our repeat analysis reveals significant variation in total TE content between genome assemblies (Table 1, Figure 1), however some of this variation is likely due to large differences in assembly quality rather than differential TE expansions or contractions in certain lineages. Most notably, the TE content of the *Ophiophagus* assembly is significantly lower than that of that of the other species (~36% compared to ~46%). The TE content of the *Aipysurus* assembly is also notably lower, however to a lesser extent (41% compared to ~46%). The *Naja*, *Laticauda*, *Notechis*, and *Pseudonaja* assemblies are much higher quality assemblies than the *Ophiophagus* and *Aipysurus* assemblies, having longer contigs and scaffolds (SI Table 1). This discrepancy is because the *Ophiophagus* and *Aipysurus* genomes are both assembled solely from short read data with a low sequencing depth (28x and 30x respectively). In stark contrast the *Naja* genome was assembled from a combination of long read (PacBio and Oxford Nanopore) and short read (Illumina) data, scaffolded using Chicago and further improved using Hi-C and optical mapping (Bionano) technologies. In the middle ground, the *Laticauda*, *Notechis* and *Pseudonaja* assemblies utilized a combination of 10X Chromium linked read and short read technologies. Many of the recently expanded TEs in the *Ophiophagus* and *Aipysurus* genomes likely collapsed during assembly because of their very high sequence similarity. Therefore, the apparent lack of recent activity in *Ophiophagus* and *Aipysurus* is a likely artefact of assembly quality. As the total TE content annotated in the *Naja*, *Laticauda*, *Notechis* and *Pseudonaja* is comparable at 46-48% of the genome and the four genomes are of comparable quality, the majority of the following analyses focuses on these four species.

	Asian elapids		Hydrophiinae			
	Naja naja	***Ophiophagus hannah***	***Laticauda colubrina***	***Pseudonaja textilis***	***Notechis scutatus***	***Aipysurus laevis***
Class I (Retroelements)	30.58	19.73	23	27.81	27.29	25.91
Penelope	1.72	1.37	2.08	2.06	1.98	1.64
LINE/CR1	4.37	4.12	4.12	5.54	6.28	4.67
LINE/L1	3.24	2.21	2.94	4.04	3.4	2.72
LINE/L2	7.14	3.39	1.05	1.41	1.4	1.27
LINE/Rex-Babar	1.17	1.11	1.08	1.44	1.42	1.26
LINE/RTE	1.4	1.52	1.07	1.5	1.43	1.25
LINE/Other	0.62	0.69	0.54	0.66	0.64	0.59
SINE	0.37	0.43	0.3	0.4	0.39	0.36
LTR/Copia	0.71	0.47	1.02	1.27	1.65	0.89
LTR/DIRS	0.81	0.61	0.86	1.04	1.19	1.57
LTR/ERV	0.6	0.42	1.23	1.38	1.69	1.16
LTR/Gypsy	7.01	1.97	5.6	5.61	4.38	7.22
LTR/Other	1.42	1.42	1.11	1.46	1.44	1.31
Class II (DNA Transposons)	7.81	7.45	18.08	9.02	9.2	7.95
DNA/hAT	4.25	4.13	3.78	5.09	5.13	4.47
DNA/PIF-Harbinger	0.02	0.03	11.19	0.02	0.02	0.02
DNA/Tc1-Mariner	3.13	2.88	2.79	3.47	3.63	3.08
DNA/Other	0.22	0.24	0.17	0.24	0.22	0.2
RC/Helitron	0.19	0.17	0.15	0.2	0.2	0.18
Unknown	7.93	8.57	6.46	9.18	9.26	7.89
Total interspersed repeats	46.32	35.75	47.54	46.01	45.75	41.75
Satellite	0.17	0.18	0.13	0.17	0.17	0.18
Simple repeat	1.31	1.37	1.28	1.66	1.67	1.28

Table 1: **Repeat composition of hydrophiine and Asian elapid genome assemblies.** Variation in assembly repeat content varies both within hydrophiines and between hydrophiines and Asian elapids. Genome assemblies annotated with RepeatMasker [37] and a custom library of curated RepeatModeler2 libraries [24] and previously described lepidosaur TEs from the Repbase RepeatMasker library [38].

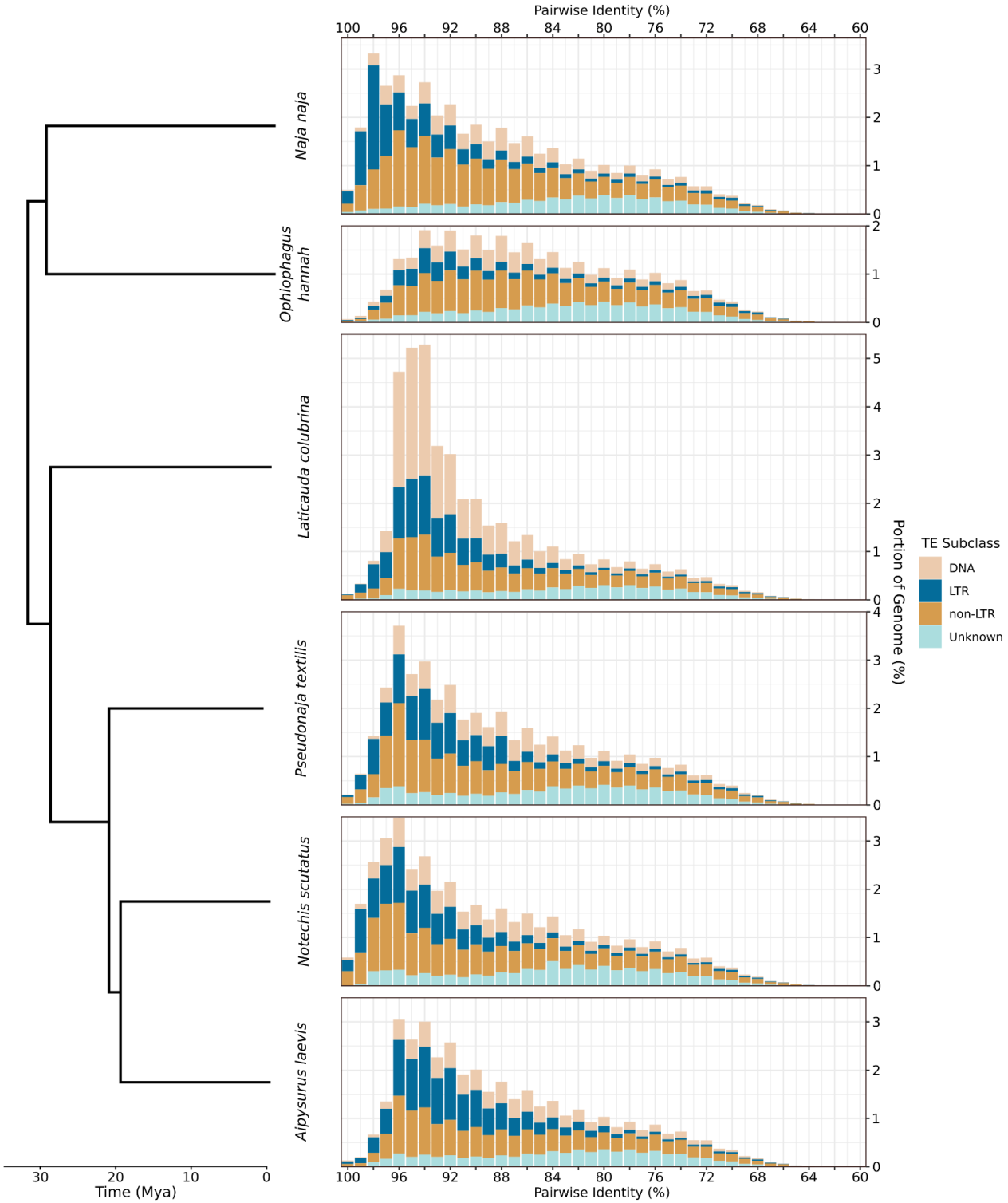

Figure 1. **Overall TE divergence profile of four hydrophiines and two Asian elapids.** Ancestral TE expansion is similar across hydrophiines and Asian elapids while recent expansion varies between species. Due to much lower genome assembly quality resulting in collapsed TEs, little recent expansion in the *Aipysurus laevis* and *Ophiophagus hannah* genomes was detected. TEs were identified using RepeatMasker [37] and a custom repeat library (see methods).

Recent insertions vs ancestral insertions

Recent TE insertions are likely to have diverged only slightly from the sequences RepeatMasker used to identify them, while ancestral insertions will likely be highly divergent. Based on this assumption, we discerned between recent and ancestral insertions using the pairwise identity of TE insertions to the mc-subfamily used to identify them. To estimate the expected divergence of ancestral TE insertions from consensus sequences compared to new insertions we searched for orthologues of 5000 randomly selected *Notechis* TE insertions and their flanks in the *Naja* assembly (Figure 2). From the 5000 TEs we were able to identify 2192 orthologues in *Naja naja*. As expected the median pairwise % identity of the ancestral TEs to the curated consensus sequences was notably lower compared to that of all TEs, at 90.5% compared to 93.6%. Similarly the 95% quantile was lower, at 96.2% compared to 98.4%. Based on these results we treated TEs with a percent identity of 96 % or higher to consensus sequences as having likely been inserted since divergence from Asian elapids.

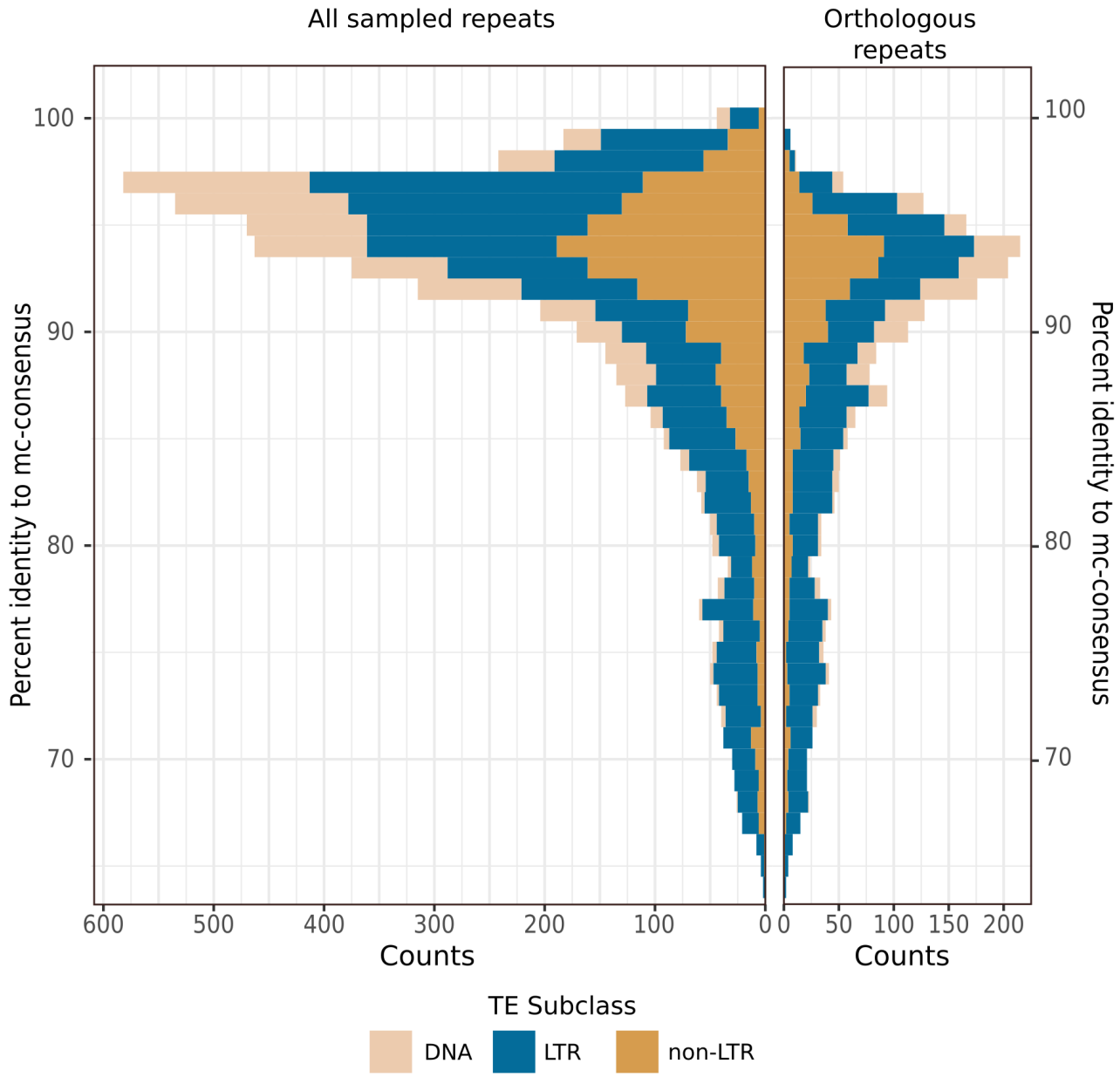

Figure 2: **The similarity of 5000 randomly selected TE insertions in *Notechis scutatus* to the consensus sequence used to identify them compared to that of the subset of 2192 having orthologues in *Naja naja*.** The similarity of ancestral insertions from mc-consensuses used to identify them was notably lower than that of TEs likely inserted since the species diverged. TEs were initially identified in *Notechis* using RepeatMasker [37]. The presence of orthologues in *Naja* was determined using BLASTN (-task dc-megablast) [25].

Recent expansion of specific superfamilies

By comparing TE divergence profiles of the various assemblies, we can gain an overall picture of how TE superfamilies have expanded since the split of Hydrophiinae from Asian elapids (Figures 3-5). Large

expansions of Gypsy retrotransposons are apparent in both the *Naja* and hydrophiine assemblies, however Copia, DIRS and ERVs appear inactive in *Naja* while expanding in hydrophiines. The divergence profile of DNA transposons suggests *Tc1-Mariner* and *hAT* transposons to have either been expanding at a similar rate in all species and/or result from ancestral expansion, with the exception of the explosive expansion of *PIF-Harbinger* transposons in *Laticauda* (see [7]). The greatest variation was seen within non-LTR retrotransposons, with L2s highly active in *Naja* yet completely inactive in hydrophiines. Instead, multiple other LINE superfamilies expanded, in particular CR1s and L1s.. This difference in TE expansion between snake lineages is similar to what was reported by Castoe et. al (2011) [23] , Yin et. al (2016) [39] and Pasquesi et. al (2018) [21], except here we see variation within a family of snakes, not just between families of snakes.

Without highly contiguous assemblies of all species it is difficult to rigorously identify recent or ongoing TE expansions. However, by using transcription as a proxy for transposition we identified currently expressed TE families in present day species as candidates for being active and potentially expanding. To achieve this, we first identified TE subfamilies in each species with over 100 kbp of copies with over 95% pairwise identity to the consensus sequences used to identify them; treating these subfamilies as potentially expanding. By mapping raw transcriptome reads back to these consensuses, we were able to identify expressed TE subfamilies. In all four species, diverse TEs were expressed including subfamilies of Copia, ERV, DIRS, Gypsy, Penelope, CR1, L1s, *Rex1*, RTE, *hAT* and *Tc1-Mariner*.

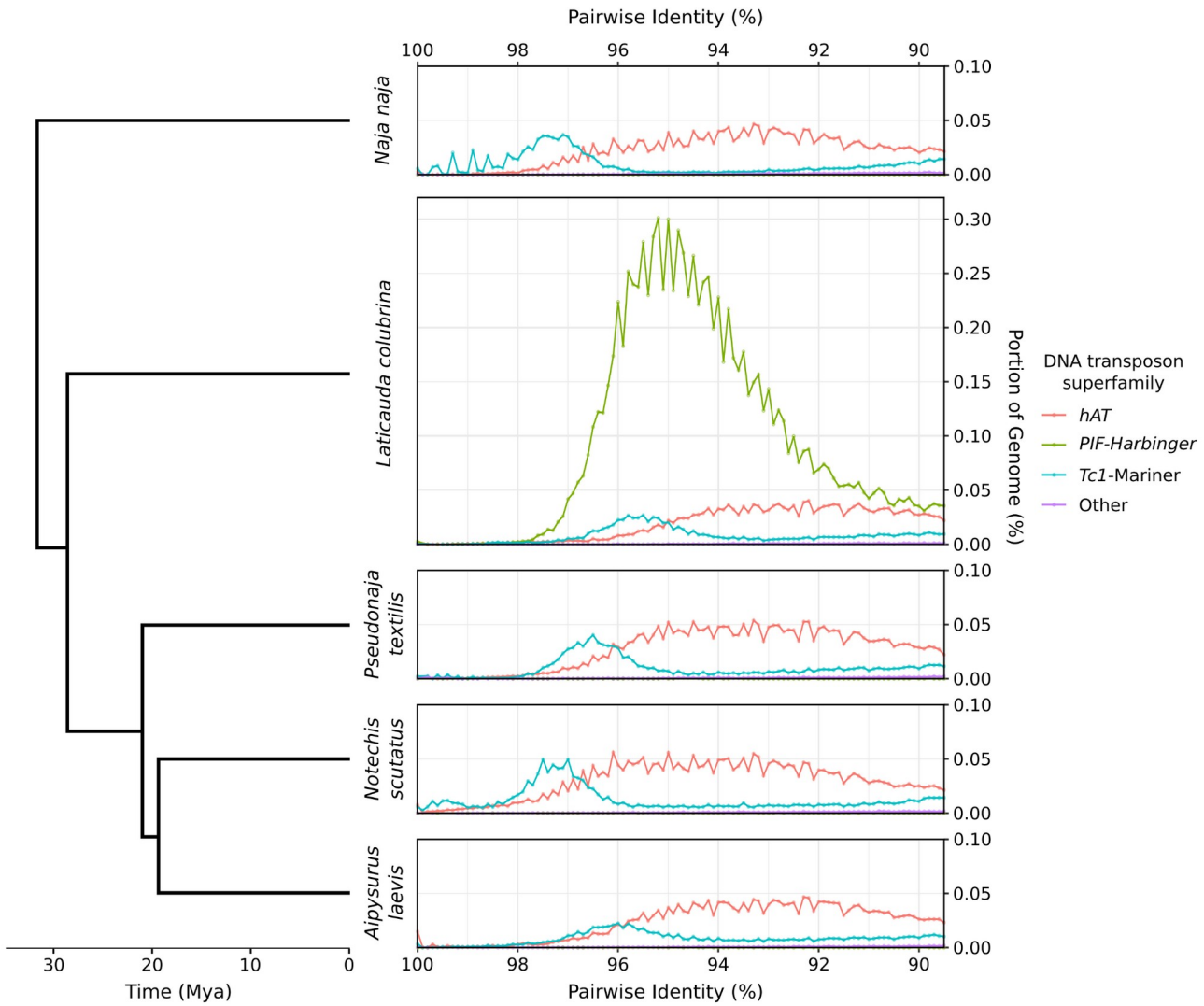

Figure 3 - **Recent DNA transposon divergence profile of four hydrophiines and the Indian cobra.** Across elapids, *hAT* transposons appear inactive, *Tc1-Mariner* transposons appear to have expanded in multiple lineages and *PIF-Harbinger* transposons appear to have expanded rapidly in *Laticauda* following horizontal transfer from a marine species. TEs were identified using RepeatMasker [37] and a custom repeat library (see methods).

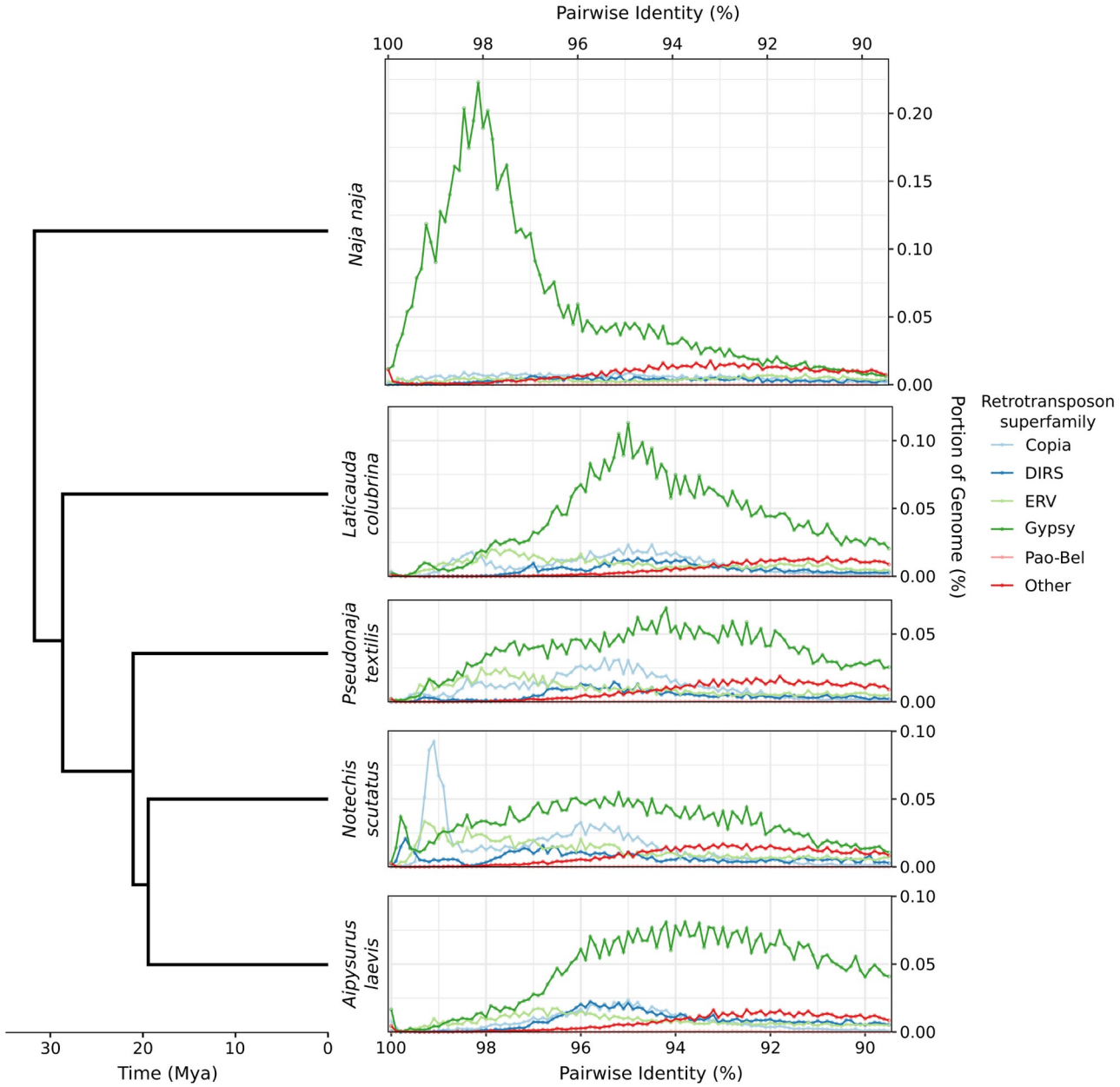

Figure 4 - **Recent LTR retrotransposon divergence profile of four hydrophiines and the Indian cobra.** Gypsy elements are the dominant superfamily in all five species of elapids; ERVs, Copia and DIRS elements have expanded in all Hydrophiinae but have been near inactive in the cobra outgroup. TEs were identified using RepeatMasker [37] and a custom repeat library (see methods).

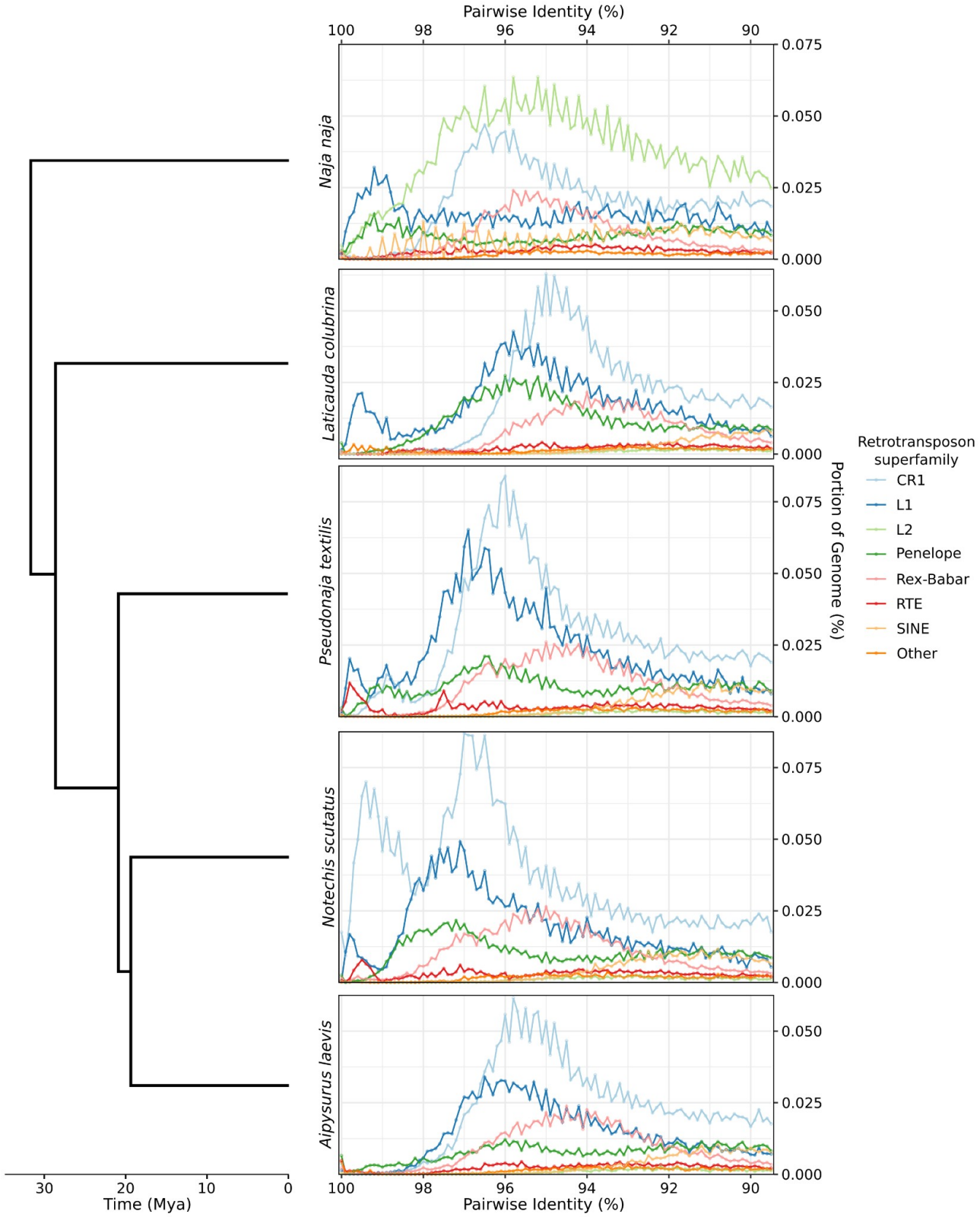

Figure 5 - **Recent non-LTR retrotransposon divergence profile of four hydrophiines and the Indian cobra.** A wide diversity of non-LTR retrotransposons have expanded in all five genomes, with CR1s and L1s being most active in hydrophiines and L2s most active in the cobra outgroup. TEs were identified using RepeatMasker [37] and a custom repeat library (see methods).

Continued expansion or horizontal transfer

The TE subfamilies which we have identified as recently expanded within Hydrophiinae could be ancestral and continuously expanding since diverging from Asian elapids or have been horizontally transferred from long diverged species. Differentiating between ancestral and horizontally transferred TEs is difficult and must meet strict conditions. Horizontally transferred sequences are defined as having a patchy phylogenetic distribution and higher similarity to sequences in another species than would be expected based on divergence time. To identify any TEs which may have been horizontally transferred into Hydrophiinae we conservatively estimated the expected minimum similarity of TEs present in both hydrophiines and Asian elapids using the 2192 orthologous sequences identified in *Notechis* and *Naja* to be 80% (Figure 6). Based on this, any vertically inherited TE subfamily classified as recently expanding in hydrophiines will likely have copies of 80% or higher similarity present in Asian elapids.

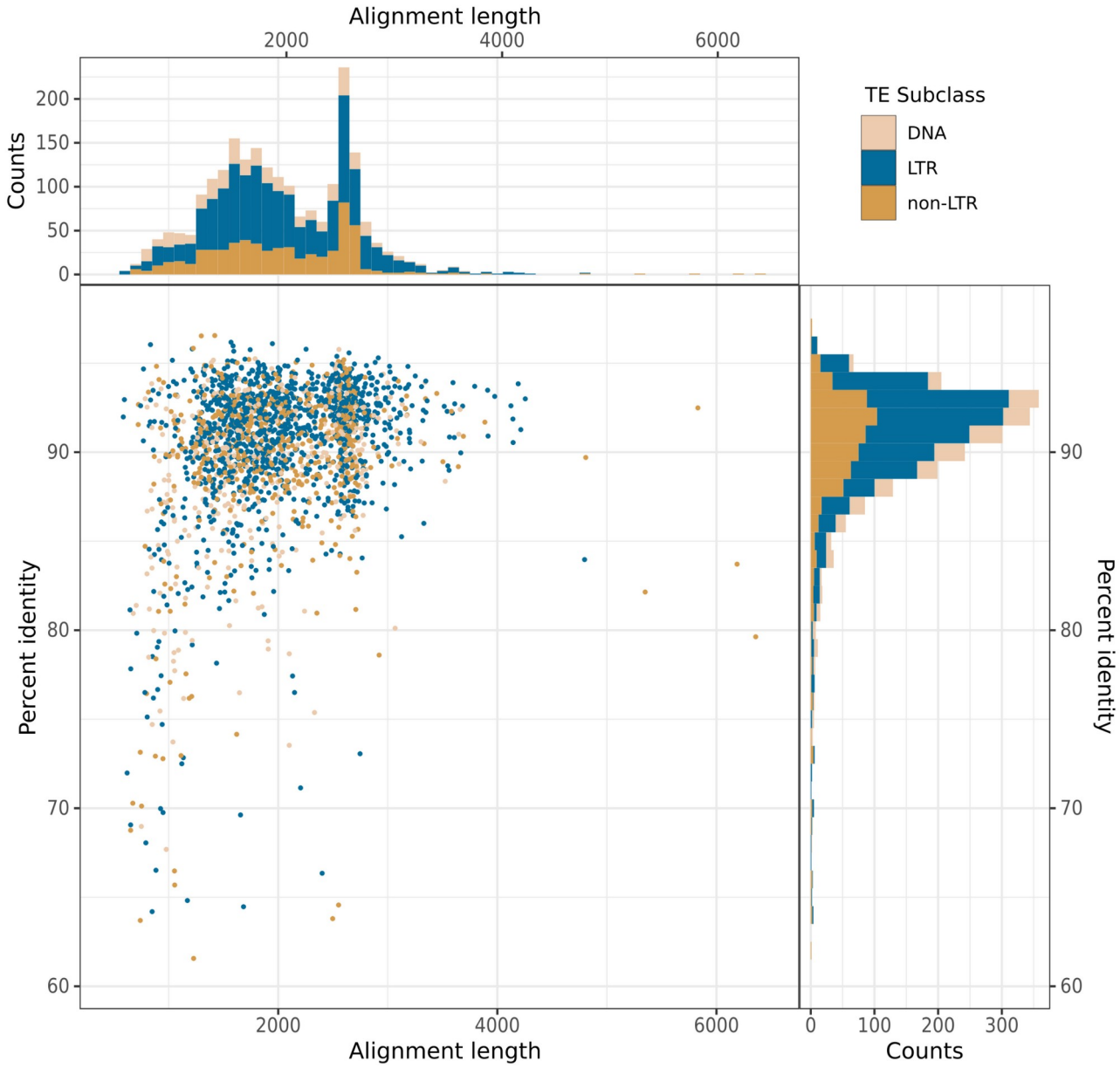

Figure 6. **Similarity (% pairwise identity) of orthologous TEs in *Notechis scutatus* and *Naja naja* genomes.** TE initially identified in *Notechis scutatus* using RepeatMasker, orthologues identified in and pairwise identity calculated for *Naja naja* using BLASTN (-task dc-megablast) [25].

To determine whether any recently expanding TE subfamilies were horizontally transferred into hydrophiines following their divergence from Asian elapids, we searched for them in the genomes of *Naja*, *Ophiophagus* and an additional 8 non-elapid snakes. Some recently expanding subfamilies absent from *Naja* and *Ophiophagus* were present in non-elapid snakes at 80% or higher identity. To be conservative we treated these TEs as ancestral, likely being lost from Asian elapids. The remaining TE subfamilies, those present in hydrophiines but absent from other snakes, were treated as horizontal transfer candidates. To confirm these candidate TEs were horizontally transferred into hydrophiines we searched for them in over 600 metazoan

genomes. This search revealed at least eleven autonomous TEs present in non-serpentine genomes at 80% or higher identity and are therefore likely to have been horizontally transferred into hydrophiines. Of these eleven, three were transferred into the ancestral hydrophiine, five into sea kraits, one into sea snakes and one into the common ancestor of terrestrial hydrophiines and sea snakes (Figure 7).

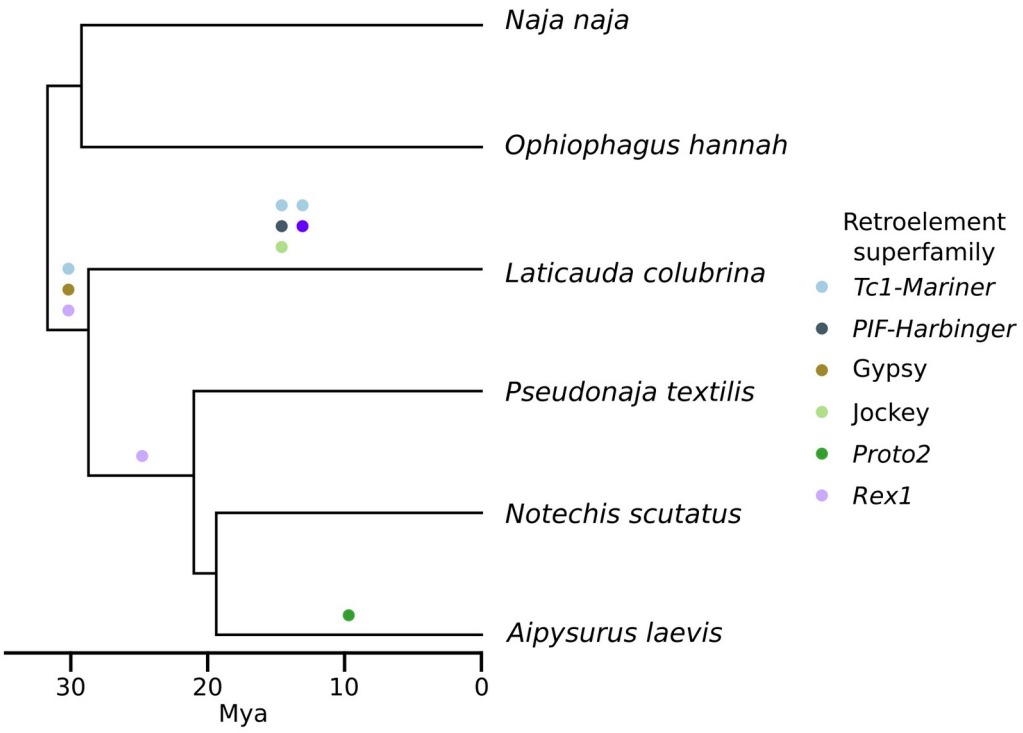

Figure 7: **Horizontal transfer of TEs into hydrophiines since their divergence from Asian elapids.** At least eleven autonomous TEs have been horizontally transferred into hydrophiines, with most likely from marine organisms. We have previously described the horizontal transfer of the *Proto2* to *Aipysurus laevis* in Galbraith et al. (2020) [6] and the *PIF-Harbinger* to *Laticauda* in Galbraith et al. (2021) [7]

We have previously described 2 of the 11 HT events in detail, that of *Proto2-Snek* to *Aipysurus* and *Harbinger-Snek* to*Laticauda*, both of which were likely transferred from a marine species (see [6,7]). Three of the four newly identified HT events identified in *Laticauda* were probably also from an aquatic species, because similar sequences are only found in marine or amphibious species. Therefore, the transfer of these elements likely occurred following the transition of each group to a marine habitat. The exception is a *Tc1-Mariner* which is most similar to sequences identified in hemipterans, a beetle and a spider, however as *Laticauda* is amphibious this is perhaps not surprising.

The *Rex1* transferred to the common ancestor of terrestrial hydrophiines and sea snakes was only identified in the central bearded dragon (*Pogona vitticeps*), an agamid lizard native to the inland woodlands and

shrublands of eastern and central Australia [40]. As this TE is restricted to another species of Australian squamate, this HTT appears to have occurred after hydrophiines reached Australia.

The most interesting of the horizontally transferred TEs are the *Tc1-Mariner*, Gypsy and *Rex1* which were horizontally transferred into the ancestral hydrophiine following its divergence from Asian elapids. Those three are most similar to sequences identified in marine species, either fish or tunicates. Marine elapids (sea kraits and sea snakes) and terrestrial Australian elapids were originally considered two distinct lineages [41–43], however recent adoption of molecular phylogenomics has resolved Hydrophiinae as a single lineage, with sea kraits as a deep-branch and sea snakes nested within terrestrial Australian snakes [1,44,45]. Fossil evidence combined with an understanding of plate tectonics has revealed Hydrophiinae, like many other lineages of Australian reptiles, likely colonised Australia via hopping between islands formed in the Late Oligocene-Early Miocene by the collision of the Australian and Eurasian plates [46–50]. Alternatively, it has also been proposed the common ancestor of Hydrophiinae may have been a semi-marine “proto-*Laticauda*”, which colonised Australia in the Late Oligocene directly from Asia [51]. The horizontal transfer of three TEs into the ancestral hydrophiine likely from a marine organism provides tangible support for the hypothesis that the ancestral hydrophiine was a semi-marine or marine snake.

Conclusion

In our survey of elapid genomes, we have found that TE diversity and their level of expansion varies significantly within a single family of squamates, similar to the variation previously seen across all squamates or within long diverged snakes. This diversity and variation is much greater than what has been reported for mammals and birds. Our finding of HTT into lineages of hydrophiine exposed to novel environments indicates that environment may play a large role in HTT through exposure to new TEs. Additionally, the HTT of three TEs found solely in marine organisms into the ancestral hydrophiine provides evidence that terrestrial Australian elapids are derived from a marine or amphibious ancestor.

As long read genome sequencing becomes feasible for more species, genome assembly quality will continue to increase and multiple genomes of non-model species will be able to be sequenced. Using these higher quality genomes, we will be able to better understand HTT and the role TEs play in adaptive evolution. Due to their rapid adaptation to a wide range of environments and multiple HTT events into different lineages, Hydrophiinae provide the ideal system for such studies.

Acknowledgements

We thank James Nankivell and Vicki Thomson for their discussions on Hydrophiinae's phylogeny and dispersal to Australia.

Supplementary Information

SI Table 1

Genome assembly statistics. Repeat composition was calculated using RepeatMasker [37] and a custom library of curated RepeatModeler libraries and previously described lepidosaur TEs from the Repbase RepeatMasker library [38].

SI Table 2

A list of the metazoan genome assemblies searched for HTT. All genomes were downloaded from NCBI RefSeq and GenBank [52,53].

SI Table 3

Horizontally transferred TEs and the number of hits in species they were identified in. TEs were identified using BLASTN (-task dc-megablast) [25].

References

1. Sanders KL, Lee MSY, Leys R, Foster R, Keogh JS. 2008 Molecular phylogeny and divergence dates for Australasian elapids and sea snakes (hydrophiinae): evidence from seven genes for rapid evolutionary radiations. *J. Evol. Biol.* **21**, 682–695. (doi:10.1111/j.1420-9101.2008.01525.x)

2. Mirtschin P, Rasmussen A, Weinstein S. 2017 *Australia's dangerous snakes : identification, biology and envenoming*. Clayton South, VIC, Australia: CSIRO Publishing. (doi:10.1071/9780643106741)

3. Glodek GS, Voris HK. 1982 Marine Snake Diets: Prey Composition, Diversity and Overlap. *Copeia* **1982**, 661–666. (doi:10.2307/1444667)

4. Voris HK, Voris HH. 1983 Feeding strategies in marine snakes: an analysis of evolutionary, morphological, behavioral and ecological relationships. *Am. Zool.* **23**, 411–425.

5. Brischoux F, Bonnet X. 2009 Life history of sea kraits in New Caledonia. *Zoologia Neocaledonica* **7**, 37–51.

6. Galbraith JD, Ludington AJ, Suh A. 2020 New Environment, New Invaders—Repeated Horizontal Transfer of LINEs to Sea Snakes. *Genome Biol. Evol.*

7. Galbraith JD, Ludington AJ, Sanders KL, Suh A, Adelson DL. 2021 Horizontal transfer and subsequent explosive expansion of a DNA transposon in sea kraits (Laticauda). *bioRxiv*. , 2021.06.13.448261. (doi:10.1101/2021.06.13.448261)

8. Chalopin D, Naville M, Plard F, Galiana D, Volff J-N. 2015 Comparative analysis of transposable elements highlights mobilome diversity and evolution in vertebrates. *Genome Biol. Evol.* **7**, 567–580. (doi:10.1093/gbe/evv005)

9. Sotero-Caio CG, Platt RN 2nd, Suh A, Ray DA. 2017 Evolution and Diversity of Transposable Elements in Vertebrate Genomes. *Genome Biol. Evol.* **9**, 161–177. (doi:10.1093/gbe/evw264)

10. Volff J-N. 2006 Turning junk into gold: domestication of transposable elements and the creation of new genes in eukaryotes. *Bioessays* **28**, 913–922. (doi:10.1002/bies.20452)

11. Kazazian HH Jr. 2004 Mobile elements: drivers of genome evolution. *Science* **303**, 1626–1632. (doi:10.1126/science.1089670)

12. Cornelis G *et al.* 2017 An endogenous retroviral envelope syncytin and its cognate receptor identified in the viviparous placental Mabuya lizard. *Proc. Natl. Acad. Sci. U. S. A.* **114**, E10991–E11000. (doi:10.1073/pnas.1714590114)

13. Ikeda N, Chijiwa T, Matsubara K, Oda-Ueda N, Hattori S, Matsuda Y, Ohno M. 2010 Unique structural characteristics and evolution of a cluster of venom phospholipase A2 isozyme genes of Protobothrops flavoviridis snake. *Gene* **461**, 15–25. (doi:10.1016/j.gene.2010.04.001)

14. Wicker T *et al.* 2007 A unified classification system for eukaryotic transposable elements. *Nat. Rev. Genet.* **8**, 973–982. (doi:10.1038/nrg2165)

15. Jurka J, Kapitonov VV, Kohany O, Jurka MV. 2007 Repetitive sequences in complex genomes: structure and evolution. *Annu. Rev. Genomics Hum. Genet.* **8**, 241–259. (doi:10.1146/annurev.genom.8.080706.092416)

16. Kapitonov VV, Tempel S, Jurka J. 2009 Simple and fast classification of non-LTR retrotransposons based on phylogeny of their RT domain protein sequences. *Gene* **448**, 207–213. (doi:10.1016/j.gene.2009.07.019)

17. Feschotte C, Pritham EJ. 2007 DNA transposons and the evolution of eukaryotic genomes. *Annu. Rev. Genet.* **41**, 331–368. (doi:10.1146/annurev.genet.40.110405.090448)

18. Kapusta A, Suh A, Feschotte C. 2017 Dynamics of genome size evolution in birds and mammals. *Proc. Natl. Acad. Sci. U. S. A.* **114**, E1460–E1469. (doi:10.1073/pnas.1616702114)

19. Ivancevic AM, Kortschak RD, Bertozzi T, Adelson DL. 2018 Horizontal transfer of BovB and L1 retrotransposons in eukaryotes. *Genome Biol.* **19**, 85. (doi:10.1186/s13059-018-1456-7)

20. Suh A, Paus M, Kiefmann M, Churakov G, Franke FA, Brosius J, Kriegs JO, Schmitz J. 2011 Mesozoic retroposons reveal parrots as the closest living relatives of passerine birds. *Nat. Commun.* **2**, 443. (doi:10.1038/ncomms1448)

21. Pasquesi GIM *et al.* 2018 Squamate reptiles challenge paradigms of genomic repeat element evolution set by birds and mammals. *Nat. Commun.* **9**, 2774. (doi:10.1038/s41467-018-05279-1)

22. Kumar S, Stecher G, Suleski M, Hedges SB. 2017 TimeTree: A Resource for Timelines, Timetrees, and Divergence Times. *Mol. Biol. Evol.* **34**, 1812–1819. (doi:10.1093/molbev/msx116)

23. Castoe TA *et al.* 2011 Discovery of highly divergent repeat landscapes in snake genomes using high-throughput sequencing. *Genome Biol. Evol.* **3**, 641–653. (doi:10.1093/gbe/evr043)

24. Flynn JM, Hubley R, Goubert C, Rosen J, Clark AG, Feschotte C, Smit AF. 2020 RepeatModeler2 for automated genomic discovery of transposable element families. *Proc. Natl. Acad. Sci. U. S. A.* **117**, 9451–9457. (doi:10.1073/pnas.1921046117)

25. Zhang Z, Schwartz S, Wagner L, Miller W. 2000 A greedy algorithm for aligning DNA sequences. *J. Comput. Biol.* **7**, 203–214. (doi:10.1089/10665270050081478)

26. Katoh K, Standley DM. 2013 MAFFT multiple sequence alignment software version 7: improvements in performance and usability. *Mol. Biol. Evol.* **30**, 772–780. (doi:10.1093/molbev/mst010)

27. Altschul SF, Madden TL, Schäffer AA, Zhang J, Zhang Z, Miller W, Lipman DJ. 1997 Gapped BLAST

and PSI-BLAST: a new generation of protein database search programs. *Nucleic Acids Res.* **25**, 3389–3402. (doi:10.1093/nar/25.17.3389)

28. Marchler-Bauer A *et al.* 2017 CDD/SPARCLE: functional classification of proteins via subfamily domain architectures. *Nucleic Acids Res.* **45**, D200–D203. (doi:10.1093/nar/gkw1129)

29. Kohany O, Gentles AJ, Hankus L, Jurka J. 2006 Annotation, submission and screening of repetitive elements in Repbase: RepbaseSubmitter and Censor. *BMC Bioinformatics* **7**, 474. (doi:10.1186/1471-2105-7-474)

30. NCBI Resource Coordinators. 2018 Database resources of the national center for biotechnology information. *Nucleic Acids Res.* **46**, D8–D13.

31. The UniProt Consortium. 2021 UniProt: The universal protein knowledgebase in 2021. *Nucleic Acids Res.* **49**, D480–D489.

32. Huang Y, Niu B, Gao Y, Fu L, Li W. 2010 CD-HIT Suite: a web server for clustering and comparing biological sequences. *Bioinformatics* **26**, 680–682. (doi:10.1093/bioinformatics/btq003)

33. Li H. 2013 Aligning sequence reads, clone sequences and assembly contigs with BWA-MEM. *arXiv [q-bio.GN]*.

34. Simões BF *et al.* 2020 Spectral Diversification and Trans-Species Allelic Polymorphism during the Land-to-Sea Transition in Snakes. *Curr. Biol.* **30**, 2608–2615.e4. (doi:10.1016/j.cub.2020.04.061)

35. Rhie A *et al.* 2021 Towards complete and error-free genome assemblies of all vertebrate species. *Nature* **592**, 737–746. (doi:10.1038/s41586-021-03451-0)

36. Peona V *et al.* 2021 Identifying the causes and consequences of assembly gaps using a multiplatform genome assembly of a bird-of-paradise. *Mol. Ecol. Resour.* **21**, 263–286. (doi:10.1111/1755-0998.13252)

37. Smit AF. 2004 Repeat-Masker Open-3.0. *http://www.repeatmasker.org*

38. Bao W, Kojima KK, Kohany O. 2015 Repbase Update, a database of repetitive elements in eukaryotic genomes. *Mob. DNA* **6**, 11. (doi:10.1186/s13100-015-0041-9)

39. Yin W *et al.* 2016 Evolutionary trajectories of snake genes and genomes revealed by comparative analyses of five-pacer viper. *Nat. Commun.* **7**, 13107. (doi:10.1038/ncomms13107)

40. Melville J, Wilson SK. 2019 *Dragon Lizards of Australia: Evolution, Ecology and a Comprehensive Field Guide*. Museums Victoria Publishing. See https://play.google.com/store/books/details?id=nWJ_xQEACAAJ.

41. Smith MA. 1926 *Monograph of the sea-snakes (Hydrophiidae)*. Printed by order of the Trustees of the British museum.

42. Hoffstetter R. 1939 Contribution à l'étude des Elapidae actuels et fossiles et de l'ostéologie des ophidiens. *Publications du musée des Confluences*

43. Storr GM. 1964 Some aspects of the geography of Australian reptiles. *Senckenb. Biol.* **45**, 577–589.

44. Keogh JS. 1998 Molecular phylogeny of elapid snakes and a consideration of their biogeographic history. *Biol. J. Linn. Soc. Lond.*

45. Sanders KL, Lee MSY. 2008 Molecular evidence for a rapid late-Miocene radiation of Australasian venomous snakes (Elapidae, Colubroidea). *Mol. Phylogenet. Evol.* **46**, 1165–1173. (doi:10.1016/j.ympev.2007.11.013)

46. Hall R. 2013 The palaeogeography of Sundaland and Wallacea since the Late Jurassic. *J. Limnol.* **72**, e1.

47. Yang T, Gurnis M, Zahirovic S. 2016 Mantle-induced subsidence and compression in SE Asia since the early Miocene. *Geophys. Res. Lett.* **43**, 1901–1909. (doi:10.1002/2016gl068050)

48. Esquerré D, Donnellan S, Brennan IG, Lemmon AR, Moriarty Lemmon E, Zaher H, Grazziotin FG,

Keogh JS. 2020 Phylogenomics, Biogeography, and Morphometrics Reveal Rapid Phenotypic Evolution in Pythons After Crossing Wallace's Line. *Syst. Biol.* **69**, 1039–1051. (doi:10.1093/sysbio/syaa024)

49. Heinicke MP, Greenbaum E, Jackman TR, Bauer AM. 2011 Phylogeny of a trans-Wallacean radiation (Squamata, Gekkonidae, Gehyra) supports a single early colonization of Australia. *Zool. Scr.* **40**, 584–602. (doi:10.1111/j.1463-6409.2011.00495.x)

50. Scanlon JD, Lee MSY, Archer M. 2003 Mid-Tertiary elapid snakes (Squamata, Colubroidea) from Riversleigh, northern Australia: early steps in a continent-wide adaptive radiation. *Geobios Mem. Spec.* **36**, 573–601. (doi:10.1016/S0016-6995(03)00056-1)

51. Heatwole H, Grech A, Marsh H. 2017 Paleoclimatology, Paleogeography, and the Evolution and Distribution of Sea Kraits (Serpentes; Elapidae; Laticauda). *Herpetological Monographs.* **31**, 1–17. (doi:10.1655/HERPMONOGRAPHS-D-16-00003)

52. O'Leary NA *et al.* 2016 Reference sequence (RefSeq) database at NCBI: current status, taxonomic expansion, and functional annotation. *Nucleic Acids Res.* **44**, D733–45. (doi:10.1093/nar/gkv1189)

53. Benson DA, Clark K, Karsch-Mizrachi I, Lipman DJ, Ostell J, Sayers EW. 2015 GenBank. *Nucleic Acids Res.* **43**, D30–5. (doi:10.1093/nar/gku1216)