Developmental and Comparative Immunology 176 (2026) 105560

Contents lists available at ScienceDirect

Developmental and Comparative Immunology

journal homepage: www.elsevier.com/locate/devcompimm

Discovery and classification of new reptile cathelicidins by genome mining: study of their structure and genomic organization in Testudines and Squamata

Andrea Otazo-Pérez [a,b], Manuel R. López [a,*], Sergio González-Acosta [a], Antonio Morales-delaNuez [a,1], José Manuel Pérez de la Lastra [a,**]

[a] Biotechnology of Macromolecules Research Group, IPNA, CSIC, Avda. Astrofísico Francisco Sánchez 3, 38206, San Cristóbal de La Laguna, Spain
[b] Escuela de Doctorado y Estudios de Posgrado, Universidad de La Laguna, Avda, Astrofísico Francisco Sánchez, SN. Edificio Calabaza-Apdo, 456, 38200, San Cristóbal de La Laguna, Spain

ARTICLE INFO

Keywords:
Antimicrobial peptides (AMPs)
Host defense
Bioinformatic
Drug discovery
Host defense peptides (HDPs)
Testudines
Squamata

ABSTRACT

Cathelicidins, a diverse family of host defence peptides (HDPs) in vertebrates, are recognized for their potential in combating resistant microorganisms and their varied biological functions. While extensive studies have focused on mammalian cathelicidins, those in reptiles remain largely unexplored. In this study, we conducted a genome mining analysis that identified 287 cathelicidin genes across reptilian orders Testudines and Squamata. Of these genes, we identified 219 complete cathelicidin protein sequences and 68 cathelicidin-like gene copies that appear to be pseudogenes or gene fragments lacking intact open reading frames. We established for the first time a classification for reptile cathelicidins of the cited orders based on their sequence and structure, observing six types for Testudines and six for Squamata, suggesting a common ancestral lineage. Furthermore, we investigated the genomic arrangement of these cathelicidin genes, uncovering that they are grouped into clusters with variability in the gene number and in their organization. Cathelicidin derived mature peptides were classified into nine groups based on their sequences and physicochemical properties. This comprehensive study enhances the understanding of the cathelicidin family in reptiles, clarifying their genomic organization and describing the different types present. These classifications pave the way for future studies on the functions and specialization of each identified cathelicidin group. Moreover, they enable potential structure-activity studies of the peptides, establishing a foundation for differentiating their key features.

1. Introduction

Antimicrobial peptides (AMPs) are a group of peptides derived from different protein families that perform defensive functions across all living organisms (Haney et al., 2017). In vertebrates these peptides are generally synthesized by immune cells like neutrophils and by some epithelial cells, especially in skin and intestines (Gallo and Hooper, 2012; Lei et al., 2019). However, new functions for these are constantly being discovered, for example, some peptides show antiviral and antifungal activity (Ballard et al., 2020; Hsieh and Hartshorn, 2016). Others also exhibit anticancer activity against the proliferation of tumour cells (Hoskin and Ramamoorthy, 2008; Kordi et al., 2023), and some have been shown to play a role in the regulation of the immune system (Guryanova and Ovchinnikova, 2022; Hancock et al., 2016; Steinstraesser et al., 2011). For this reason, AMPs are also commonly known as Host Defense Peptides (HDP) (Haney et al., 2017; Steinstraesser et al., 2011). In addition, they typically do not promote resistance in their targeted microorganisms, primarily attributed to their distinct mechanism of action (Veldhuizen et al., 2013). AMP families are distributed across a wide variety of organisms, such as thionins in plants (Tam et al., 2015), cecropins and attacins in Lepidopteran insects (e.g., *Hyalophora cecropia*) (Wu et al., 2018), piscidins in teleost fish such as *Morone saxatilis* (Asensio-Calavia et al., 2023; Silphaduang et al., 2006), defensins in various animal groups and cathelicidins in vertebrates

* Corresponding author.
** Corresponding author.
E-mail addresses: manuel.lopez.15@csic.es (M.R. López), jm.perezdelalastra@csic.es (J.M. Pérez de la Lastra).
[1] Current address: IUSA-ONEHEALTH 4, Animal Production and Biotechnology, Universidad de Las Palmas de Gran Canaria, 35413, Arucas, Spain.

https://doi.org/10.1016/j.dci.2026.105560
Received 20 October 2025; Received in revised form 20 January 2026; Accepted 29 January 2026
Available online 31 January 2026

(Izadpanah and Gallo, 2005).

Cathelicidins constitute one of the most significant families of HDPs in vertebrates and have been extensively studied due to their potent biological activity, which may have potential applications in biomedicine (Kościuczuk et al., 2012). Cathelicidins are first synthesized as a prepropeptide, with a signal peptide, a highly conserved region named cathelin domain, and the active peptide, located in the C-terminal region (Shinnar et al., 2003; Zanetti, 2004). The active peptides are cleaved from the rest of the protein by a serine protease. This enzyme is typically elastase, though other serine proteases, such as protease 3, also catalyse the release of human LL-37 (Andrault et al., 2015). At the genomic level, cathelicidin genes consist of four exons and three introns (Kościuczuk et al., 2012). The first three exons encode the signal peptide and the conserved cathelin domain, while the active peptide is encoded by exon 4 (Tossi et al., 2024). Usually, cathelicidin genes are located together in the genome forming a cluster (Kościuczuk et al., 2012).

The cathelin domain is generally well-conserved across distantly related animal groups. However, the resulting active peptides exhibit considerable variability, with even closely related species displaying different peptide sequences (Agier et al., 2015; Tomasinsig and Zanetti, 2005). Such sequence variations often correlate with differences in antimicrobial activity (Tossi et al., 2024). Additionally, the functional diversity of cathelicidins is likely influenced by the pathogens encountered by different species, which varies based on their habitats and diets (Hanson et al., 2019; van Dijk et al., 2023). Therefore, investigating the array of cathelicidins across various animal groups could provide a valuable repository of active compounds against a wide range of microorganisms. Particularly, the study of cathelicidins in reptiles could be highly beneficial as they are predicted to encode peptides with specialized biological functions, reflecting the interactions of these animals with their diverse ecological niches. Moreover, numerous studies have demonstrated their potent antimicrobial and antiviral activities (Kim et al., 2017; Zhao et al., 2008).

Although cathelicidins are present in all vertebrates, they have been mainly studied in mammals and birds (Kościuczuk et al., 2012). Within reptiles, snakes have garnered the most research interest, constituting the primary focus of most investigations in this domain (Wang et al., 2008; Zhang et al., 2010). Nevertheless, the underlying genetic features of these cathelicidins remain inadequately elucidated, and it is uncertain whether their genes are organised within a genomic cluster. Moreover, while variations in the sequence and functional activity of reptilian cathelicidins have been documented, a formal classification system analogous to that established for avian species has yet to be developed. In other groups of reptiles, such as lizards or turtles, there is a limited number of studies investigating their cathelicidins. Furthermore, reptile cathelicidins are often classified as 'avian-like' or variants, such as 'cathelicidin 2-like' or 'cathelicidin 3-like'. They may also be described as 'snake-like' or 'OH-like' referring to the cathelicidin 'OH' from the snake *Ophiophagus hannah*, one of the most extensively studied examples (Nash and Ryan, 2023; Zhao et al., 2008).

We used different bioinformatic tools to discover new reptile cathelicidins through genome mining and studied their main characteristics. Based on this, we propose a classification of cathelicidins and their derived peptides from the reptilian orders Testudines and Squamata into different types. We also present a model of the genomic organization of cathelicidin genes within these groups.

2. Materials and methods

2.1. Genome mining

To carry out the genome mining, a probe was initially designed to screen for cathelicidin genes across various reptile species (Table 1). This probe was comprised of three consensus sequences, which were derived from the multiple sequence alignment of the cathelin conserved domain from several annotated reptile cathelicidins housed in the NCBI database. These consensus sequences were specifically generated using Jalview software (version 10.0.5) from the alignment of cathelicidins annotated as 'cathelicidin 3-like', 'cathelicidin 2-like' and 'cathelicidin Oh-like' respectively.

This probe was employed as queries to identify putative cathelicidins genes by genome mining of available chromosomal-level genome assemblies of reptile taxa in the NCBI database (up to March 2024) by BLASTp and tBLASTn (Altschul et al., 1990). Searches were performed using the following parameters: E-value threshold $= 1 \times 10^{-5}$, word size $= 3$ (BLASTp) or 6 (tBLASTn), BLOSUM62 matrix, and default gap penalties. Only hits covering $\geq$60% of the query sequence and containing the four conserved cysteine residues typical of the cathelin domain were retained. Genomic hits were extended $\pm$5 kb to allow manual inspection of intron–exon boundaries and potential gene structure. Only species from Testudines (turtles) and Squamata (lizards, snakes, geckos, iguanas, and chameleons) taxonomic orders were analysed. Crocodylia (crocodiles, gharials, and alligators) and Sphenodontia (tuataras) was excluded from the analysis due to the limited availability of only two chromosomal-level genomes within both entire orders. We excluded the genomes of some species whose assemblies were poorly annotated, containing large portions of undetermined nucleotides (N's).

NCBI was selected as the primary data source because it currently provides the most taxonomically diverse and consistently annotated genome assemblies for reptiles, although we acknowledge that future long-read or chromosome-level assemblies will further refine the patterns described here.

After the BLAST analysis both the chromosome and the specific region of probe alignment were identified. We subsequently downloaded the complete chromosome sequence in FASTA format and selected a region of approximately 5×10^5 base pairs encompassing the probe alignment, using Unipro UGENE 49.0 (Okonechnikov et al., 2012). This sequence was then uploaded into gene prediction tools GeneScan (Burge and Karlin, 1998) and FGENESH (Solovyev et al., 2006) to search the exact location of all the cathelicidin genes within chromosome and the predicted protein sequences and coding sequences (CDS).

A gene predicted by these softwaretools was classified as complete if it fulfilled all of the following: (1) a four-exon/three-intron organization; (2) presence of a predicted signal peptide; (3) four conserved cysteine residues with spacing consistent with cathelin domains; (4) an exon 4 capable of encoding a plausible mature peptide; (5) presence of recognizable regulatory motifs, including all exon-intron splicing sequences, a TATA-like promoter upstream and a polyadenylation signal downstream when identifiable.

Conversely, a gene was classified as putatively defective if it displayed at least one of the following: (1) premature stop codons in any exon; (2) missing exons; (3) absence or sequence alteration of the ATG start codon; (4) disruption of canonical splice motifs (5′ GT or 3′ YAG); (5) absence of a stop codon in exon 4, resulting in an abnormally

Table 1
Consensus sequences that constitute the probe used in genome mining to search for new cathelicidin genes.

Type of annotated cathelicidin	Consensus sequence
Cathelicidin 3-like	FQEALAQAIYFYNQGPQVQNAFRLLDAAQIPRLNSSSGNLQPLNFTIQETVCPTSGGAAAELCDFKPDGLVKDCSGYFSNQQANSVIVITCD
Cathelicidin 2-like	VARMAVEEYNQESGGQAVFRLWKIRSTRKTRFDWGVHFSMDFTIKETLCQKTGKYRIGDCKHKPNGTIRDCSAEVSVHNFMQDSPLTSVNCHPLQANNPNSRNAR
Cathelicidin Oh-like	ALSYEEAVSLTAELYSQKAEVAFAFRLLEAKPQPDWDPRSEEPQDLEFTVKETTCSAAELLTAADCDFKEDGVVKECSGTFSIQQQAPVQLSCD

extended peptide; (6) incompletely sequenced genome region; (7) absence of a predicted signal peptide.

Because only one genome assembly per species was examined, these genes are referred to as putatively defective, and not definitively as pseudogenes. In addition, the Conserved Domain Search Tool from NCBI (Wang et al., 2023) was employed to confirm that the identified genes corresponded to cathelicidins.

Manual corrections of mispredicted gene models were performed by aligning the complete gene sequence with a cathelicidin CDS sequence from closely related species. We used the ClustalW multiple sequence alignment tool version 2.1, Thompson et al. (1994) with BLOSUM series of matrices (default = BLOSUM62), a gap opening penalty of 10 and gap extension penalty of 0.2.

These manual corrections addressed issues commonly found in *de novo* prediction, such as merged or skipped exons due to incorrect intron-exon boundaries detection, and misannotation of the predicted mature peptide in Exon 4. Finally, the identified cathelicidin sequences, along with their genomic locations, were compiled.

2.2. Cathelicidin classification

The classification system was developed based on the evolutionary dynamics observed in the cathelicidin family. All the putative cathelicidin genes were classified into types using a streamlined workflow summarised in Fig. 1.

First, we identified genes showing the characteristic features of cathelicidins, including a four-exon structure, a signal peptide, and the conserved cysteine framework of the cathelin-like domain. We then confirmed the integrity of this domain by examining its exon-derived structure and cysteine spacing. The C-terminal region encoded by exon 4 was subsequently inspected to delineate the putative mature peptide, focusing on the presence of a plausible cleavage site (V/I/A) (Anderson and Yu, 2003), expected peptide length, and general physicochemical properties.

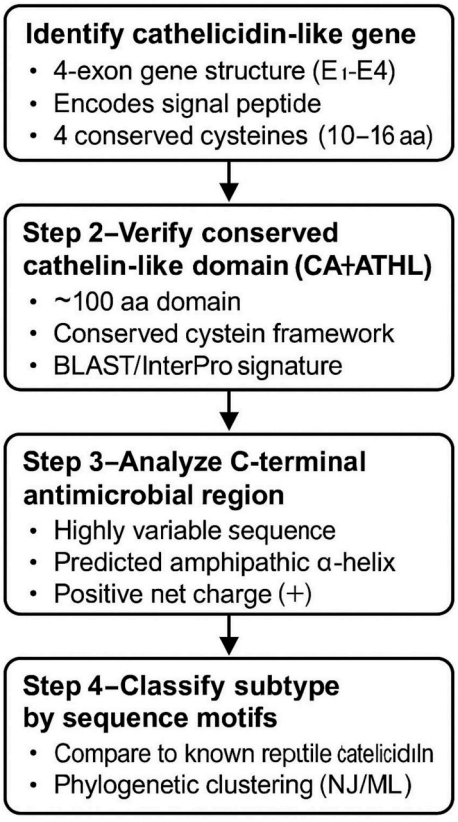


Fig. 1. Flowchart illustrating the streamlined classification pipeline for reptile cathelicidins. The process starts with gene identification based on hallmark structural signatures (four-exon gene structure, presence of a signal peptide, and four conserved cysteine residues). Identified sequences are then screened for conserved cathelin-like domains, analysed for precursor architecture (signal peptide, cathelin domain, mature peptide), and classified into cathelicidin subtypes according to sequence motifs and phylogenetic clustering. This workflow enables the systematic designation of cathelicidin-like genes in reptilian genomes.

The conserved domain protein sequences were initially aligned using the four conserved cysteines typical of the cathelin domain as anchoring points to ensure homology. Subtype assignment was based primarily on the presence and position of short type-specific motifs or signatures within the conserved domain. To aid in the better visualisation of the signatures for classification, the online tool WebLogo (Crooks et al., 2004) was used. When necessary, exon-length patterns and synteny with characteristic flanking genes (e.g. KLHL18, TBRG4, CDC25) were used to resolve ambiguous cases. Genes that could not be confidently assigned after applying these criteria were labelled as ambiguous.

2.3. Nomenclature of cathelicidin genes

To distinguish the identified genes, we established a nomenclature comprising the species name followed by a number that indicates the gene's position in the organism's genome relative to the other cathelicidins. Genes predicted as defective were labelled with the species name followed by the letter 'D' and a number indicating their position relative to the other defective cathelicidin genes.

2.4. Phylogenetic tree of cathelicidin sequences

To generate the phylogenetic trees first all the protein conserved domain sequences of the mining cathelicidin were identified and compiled using the Conserved Domain Search Tool from NCBI. These sequences were aligned using the MUSCLE algorithm implemented in the MEGA software version 11.0.13 (Tamura et al., 2021) which was also employed to construct a phylogenetic tree by the Neighbor-joining method using 1000 bootstraps and the Jones-Taylor-Thornton (JTT) substitution model with a Gamma Distributed (G) rate and a Gamma Parameter of 2. This model for constructing the phylogenetic tree was selected since it provided the best fit to the data, according to the analysis conducted using the 'Find Best DNA/Protein Models' function in MEGA.

2.5. Analysis of sequence and three-dimensional structure of cathelicidins

The percentage identity diagram of the aligned conserved domains was obtained using Unipro UGENE 49.0 (Okonechnikov et al., 2012).

The three-dimensional models of the complete cathelicidin and peptides sequences were generated using AlphaFold 3 (Abramson et al., 2024) and visualized with UCSF ChimeraX software version 1.7.1 (Pettersen et al., 2020). This tool was also used to identify regions of the peptide sequences forming α-helices. AlphaFold 3 predictions were computed using default settings for single-chain modelling without template enforcement. For all models, only the highest-ranked prediction was used. In ChimeraX (v1.7.1), structures were visualized with default energy-minimisation parameters, and secondary-structure elements (α-helix, β-strand, 3–10 helix) were assigned using the built-in 'Structure Analysis → Secondary Structure' tool.

Signal Peptide software version 6.09 (Teufel et al., 2022) was used to verifiy that all mined cathelicidins contained signal peptides.

2.6. Representation of the clusters of cathelicidin genes

The figures representing the cathelicidin gene clusters were generated using the 'ggplot2' and 'gggenes' packages in R software version 4.4.1 (Wilkins and Kurtz, 2023).

2.7. Identification of mature peptides

Following previously described cleavage patterns of cathelicidins by serine proteases (Anderson and Yu, 2003; Biniossek et al., 2011; Shinnar et al., 2003), we defined the putative mature peptide as the sequence starting from the first valine, alanine, or isoleucine residue in exon 4. This rule follows the canonical activation mechanism of vertebrate cathelicidins, in which neutrophil elastase and related serine proteases cleave immediately N-terminal to a small hydrophobic residue (V, I, or A). This criterion has been used in multiple functional studies of mammalian and reptilian cathelicidins and provides a consistent and biologically grounded definition of the mature peptide. In putatively defective genes, derived peptides were included only when a credible cleavage position and reasonable peptide length existed; these peptides were flagged in Table S5.

2.8. Analysis of mature peptides

The charge, hydrophobicity, and molecular weight were predicted using the Peptide Calculator server (Lear and Cobb, 2016). The alignment of the peptides was performed manually, focusing on the positions of apolar residues (leucine, isoleucine, valine, alanine, and phenylalanine), which are commonly conserved in α-helical antimicrobial peptides and contribute to their amphipathic structure. The peptides were encoded using the first letter of their genus and species, followed by an underscore, the first two letters corresponding to the first two amino acids in their sequence, and a number indicating the total amino acid count. This naming scheme follows the convention commonly used in other studies of antimicrobial peptides, such as the designation of the human cathelicidin as LL-37.

3. Results and discussion

3.1. Genome mining of reptiles

A total of 287 cathelicidin genes were identified across 38 reptile species, including 219 with intact open reading frames (ORFs) and 68 predicted to be nonfunctional due to disrupted coding regions or missing essential features. Among the 219 complete sequences, 129 were novel, while 90 corresponded to entries previously annotated in the NCBI database (Table 2). Several of these annotations were inaccurate, most commonly involving merged genes (e.g., XP_061441679.1) or the absence of exon 4 (e.g., XP_075775955.1). Two sequences from *Chelonia mydas* also differed from those deposited by Yu, H, et al (unpublished; QED55075.1, QED55073.1), likely reflecting allelic or assembly-related variation. From these gene models, we inferred 246 putative peptides based on exon 4 (described below).

All identified cathelicidins displayed the characteristic four-exon, three-intron organisation reported for vertebrate cathelicidins (Kościuczuk et al., 2012; Zanetti, 2004). Exons 1–3 encoded the signal peptide and the conserved cathelin-like domain, with exon sizes falling within narrow ranges: exon 1 (171–189 bp), exon 2 (108–111 bp), and exon 3 (72–87 bp; typically 84 bp) (Table 3, S1). Exon 4, which encoded the highly variable C-terminal peptide, showed the greatest size variation (69–222 bp). Promoter and termination elements were also detected, including a TATA-box–like motif located ~60 bp upstream of the start codon (Wiley et al., 1992) and a polyadenylation signal positioned ~200 bp downstream of the stop codon (Table S1).

Introns were comparatively large (Table S2). Intron 1 was the longest (mean ~2.5 kb), followed by intron 2 (~1.8 kb) and intron 3 (~1 kb). Splice boundaries were highly conserved: the canonical donor ‘GT' and acceptor ‘AG' were consistently present, in agreement with known splice-site architecture in vertebrates (Matera and Wang, 2014; Smith et al., 1993). The predominant 5′ donor signals were ‘GTGAG' (common in introns 1 and 3) and ‘GTAAG' (typical of intron 2), although several additional variants ('GTATG', ‘GTGGG', ‘GTAGG') also occurred, especially in Squamata. These last splicing signal variants reduced effectiveness of prediction softwares in distinguishing intron-exon boundaries. All of these sequences are consensus splicing sites previously described in the literature (Matera and Wang, 2014; Sibley et al., 2015). These signals were essential for manual correction of gene models when automated predictions failed to detect exon–intron boundaries accurately. At the 3′ end, the main acceptor motifs were ‘TAG' and ‘CAG', both consistent with vertebrate consensus splice sites and always preceded by a pyrimidine-rich tract (Smith et al., 1993; Matera and Wang, 2014).

Table 2
Summary of the cathelicidins identified during the genomic mining of reptiles.

Cathelicidin genes	Total		Testudines		Squamata	
	287		131		156	
Complete cathelicidin genes	219	76%	91	69%	128	82%
Defective cathelicidin genes	68	24%	40	31%	28	18%
Novel cathelicidin sequences	129	59%	48	53%	81	63%
Annotated cathelicidin sequences	90	41%	43	47%	47	37%
Predicted peptides	246		104		142	

3.2. Classification of reptile cathelicidins

To date, and except for the seven clusters proposed for Crocodylia by Santana et al. (2022), no formal classification has been available for cathelicidins from reptiles. To address this gap in Testudine and Squamata orders, we applied the workflow outlined in Fig. 1, aligning the identified sequences to search for short, exclusive sequence motifs. These motifs were required to be conserved across homologous cathelicidins from different species, yet mutually exclusive among paralogues within a given species, indicating divergence following gene duplication. This strategy successfully revealed characteristic signatures located at the beginning of the conserved domain (Table 3), which further corresponded with consistent exon-size patterns and conserved features in exon 4. Thus, these signatures were used as the primary criterion to define and name the cathelicidin types identified across reptiles. A summary of the signatures defining each type is provided in Table 4, and consensus sequences obtained for each type retained their group-specific motif (Fig. 2). In almost all cases, these signatures occupied consistent positions within the conserved domain (Table 3, Fig. 2).

In Testudines, the analysis of 91 complete cathelicidins revealed six distinct types. In the THEDA, YEKAI and VARSA types, signatures occurred at the start of the conserved domain, whereas in DMTS, DENSQ and AATNET types, they were located at the beginning of exon 2. The DMTS type also exhibited a second signature ('KAD') at the end of exon 1 (Tables 3 and 4).

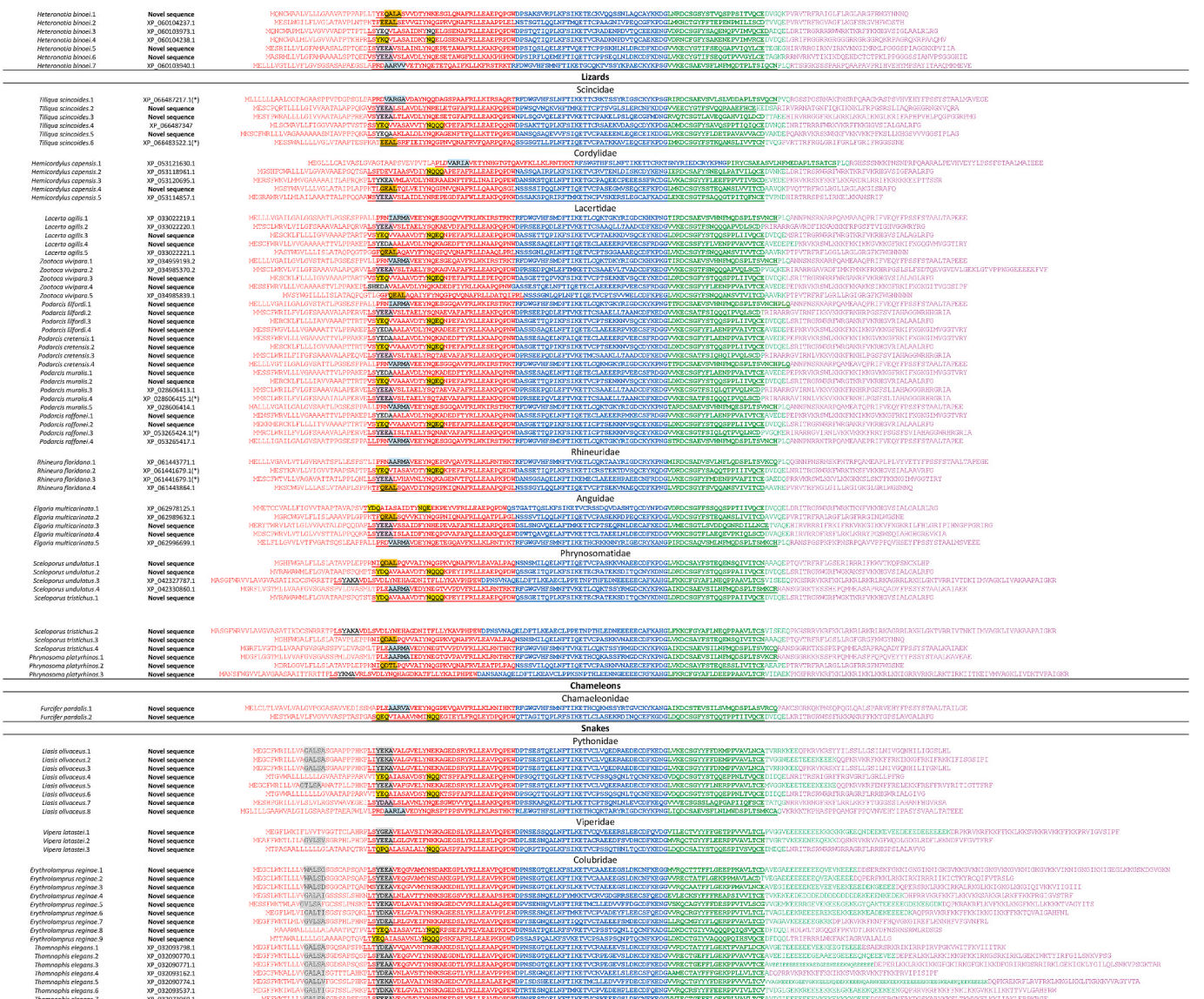
In Squamata, 128 complete cathelicidins were classified into six types as well, although with a greater diversity of signature variants and exon-length ranges (Tables 3 and 4). The characteristic motifs identified in this order were YEQ, GALSA, YEEA, QEAL, YEDA and VARMA. Notably, only YEQ- and GALSA-type cathelicidins were consistently present in snakes (suborder Serpentes), with the exception of *Liasis olivaceus*, which also possessed YEDA- and VARMA-type cathelicidins. GALSA-type cathelicidins were exclusive to snakes and correspond to previously characterised 'vipericidins', potent antimicrobial peptides found in pit-viper venom glands (Falcao et al., 2014). Within YEQ-type cathelicidins, the presence of the motif ‘YEQ' (or close variants) followed by ~8 residues and then ‘NQQ' (or variants) was required for reliable identification, since the short ‘YEQ' motif alone may occur in other types.

This classification provides a consistent nomenclature for reptilian cathelicidins, addressing the current inconsistency in numbering schemes across studies. For example, in *Pelodiscus sinensis*, Shi et al. (2019b) designated six cathelicidins as Cath 1–6, which correspond in our scheme to DMTS-type (Cath1), THEDA-type (Cath3–5) and

Table 3

Sequence of the 219 complete cathelicidins obtained by mining genomes of reptile species. The sequence encoded by exon 1 is marked in red, exon 2 in blue, exon 3 in green, and exon 4 in violet. The sequences are aligned based on the cysteines within the conserved domain, which are highlighted in yellow. The conserved domain of each sequence is shown in bold and underlined. In the 'Sequence ref. number' column, the access number of the sequence in the NCBI is indicated. This access number is accompanied by the symbol (*) in cases where the sequence annotated in the NCBI is different from the one found through genome mining. If the sequence identified through data mining has not been previously described or annotated in the NCBI, it is noted as a novel sequence. The characteristic signatures of each type of cathelicidin established in this article are presented in bold and shaded with the corresponding colours for each type: green for THEDA, dark blue for YEKAI, light blue for VARSA, light pink for DMTS, red for DENSQ, brown for AATNET, violet for YEEA, yellow for YEQ, bluish-grey for VARMA, orange for QEAL, light green for YEDA, and grey for GALSA.

Specie	Sequence ref. number	Sequence
Turtles		
Testudinidae		
Aldabrachelys gigantea.1	Novel sequence	METCWRCLLLLGIVAAAATAVPSQHKTLSYEEAIALAVDFYNQGSGIDHIFRLLKADPQPEWDMTSNPQQKLKFTVKETVCLVSENRHGTECDFIDNGVVRDCSGSFSTQQESPIVIISCNTVTQEDTHIRRPRSPRRKWRRPRWHLPGSYTLIAHGGGGGKGRGSRLQVA
Aldabrachelys gigantea.2	Novel sequence	METCLKVLLLVGVVTAAPAPSSSPLPSHEAAVLAAVQIYNQEPGTTLAYRLLEAEPQPDWDVSSKTIQSLTFSVKETVCLVSGNRDPNQCQFKEDGLVKDCSGFFSTEQDPPSIILKCEEASEEPNVVTRGRWKRFWRRTGNFVRQHGWGMAAHILKNVG
Aldabrachelys gigantea.3	Novel sequence	METCLEVLLLVGVVTAAPTPPSSPLPTYEDAVLAAVQIYNQEPGTTLAYRLLEAEPQPDWDVSSKTIQSLTFSVKETVCLVSQKRDPNQCEFKEDGLVKDCSGFFSTEQDPPSIIIKCEEASEEPNVVTRSRWRRWVRRISGIIHRIPLTNPLIRG
Aldabrachelys gigantea.4	Novel sequence	METCLKVLLLVGVVTAAPTPPSSPLPSHEAAVLAAVQIYNQEPGTTLAYRLLEAEPQPDWDVSSKTIQSLTFSVKETVCLVSEKRDPNQCEFKEDGLVKDCSGFFSTEQDPPSIMIKCEEASEEPKVITRSRWRKWGRIVSKALIFVG
Aldabrachelys gigantea.5	Novel sequence	METCLKVLLLVGVVTAAPAPSSSPLPSHEAAVLAAVQIYNQEPGTTLAYRLLEAEPQPDWDVSSKTIQSLTFSVKETVCLVSEKRDPNQCEFKEDGLVKDCSGFFSTEQDPPSIIIKCEEASEEPNVVTRSRWRKWGRKAGRFIHRNRWKIIRAGVQIILG
Aldabrachelys gigantea.6	Novel sequence	METCLKVLLLVGVVTAAPTPPSSPLPSHEAAVLAAVQIYNQEPGTTLAYRLLEAQPQPDWDVSSKTIQSLTFSVKETVCLVSQKRDPNQCEFKEDGLVKDCSGFFSTEQDPPSIIIKCEEASEETNVVTRSRWRRWGRRIGNFINRHPWILLGPG
Aldabrachelys gigantea.7	Novel sequence	METCLKVLLLVGVVTAAPTPPSSPLPSHEAAVLAAVQIYNQEPGTTLAYRLLEAQPQPDWDVSSKTIQSLTFSVKETVCLVSEKRDPNQCEFKEDGLVKDCSGFFSTEQDPPSVIIKCEEASEEPKVVTRWGGLGGIIRRILPYLG
Aldabrachelys gigantea.8	Novel sequence	METCLKVLLLVGIVTAAPAPSSSPLPSHEAAVLAAVQIYNQEPGTTLAYRLLEAEPQPDWDVSSKTIQSLTFSVKETVCLVSQKRDPNQCEFKEDGLVKDCSGFFSTEQDPPSVILKCEEASEKPKVVTRSRWGHWWSIIRPFILSLPG
Aldabrachelys gigantea.9	Novel sequence	METCLKVLLLIGVVTAAPTPPSSPLPSHEAAVLAAVQIYNQEPGTTLAYRLLEAQPQPDWDVSSKTIQSLTFSVKETVCLVSEKRDPNQCEFKEDGLVKDCSGFFSTEQDPPSIIIKCEEASEEPNVVTRGRWKRFWRKTGRFVRRHGWDILATGLKIAG
Aldabrachelys gigantea.10	Novel sequence	METCLEVLLLVGVVTAAPTPPSSPLPTYEDAVLAAVQIYNQEPGTTLAYRLLEAEPQPDWDVSSKTIQSLTFSVKETVCLVSQKRDPNQCEFKEDGLVKDCSGFFSTEQDPPSIIIKCEEASEEPNVVTRSRWRRWVRRISGIIRRIPLTYPLIRG
Aldabrachelys gigantea.11	Novel sequence	METCLKVLLLTRVVTAAPTPPSSPLPSHEDAVLAAVQIYNQDPGTTLAYRLLEAEPQPDWDVSSKTIQSLTFSVKETVCLVSEKRDPNQCEFKEDGLVKDCSGFFSTEQDPPSVIIKCEEASEKPKVVTRSRWRRWGRIVSKALIVLG
Aldabrachelys gigantea.12	Novel sequence	METCLKVLLLVGVVTAAPAPSSSPLPSHEAAVLAAVQIYNQEPGTTLAYRLLEAEPQPDWDVSSKTIQSLTFSVKETVCLVSEKRDPNQCEFKEDGLVKDCSGFFSTEQDPPSIIIKCEEASEEPNVVTRSRWRKWGRKAGRFIHRNRWKIIRAGVRIILG
Aldabrachelys gigantea.13	Novel sequence	METCLKVLLLVGVVTAAPTPPSSPLPTYEDAVLAAVQIYNQEPGTTLAYRLLEAEPQPDWDVSSKTIQSLTFSVKETVCLVSQKRDPNQCEFKEDGLVKDCSGFFSTEQDPPSIILKCEEASEKSKVVTRSRWRRWGKIVSKALIIFG
Aldabrachelys gigantea.14	Novel sequence	MASCWAVLLLVTVAAAVPVKTPQPYEKAIARAIDFYNQGPTVKYAFRLLTATAPPQVAQGTISNILQLNFTIMETTCPASDKAPVEQCAFKENGLVRDCSGHFSTQETCSVVAITCDVAPPQPVRVTRTWLPALSIGAKIVGHGLDYILQRG
Aldabrachelys gigantea.15	Novel sequence	MASCWAVLLLVTVAAAVPVKTPQPYEKAIARAIDFYNQGPTVKYAFRLLTATAPPQVAQGTISNTLQLNFTIMETTCPASDKAPVEQCAFKENGLVRDCSGHFSTQETCPVVAITCDVAPPQPIRVTRWWPLLRIGAKIVGHGIDYIRRRG
Aldabrachelys gigantea.16	Novel sequence	MAVVLFLVALLASLGVSVTASPLLEAQLLSPRDVARSAVEAYNQETGTQAVFKLLKLKSFHKTQFDWGTHFRINFTIRETNCRKSVASYRLEDCRYQPNGITRDCSADVSVLDFVQEAPLTSVKCTQLQRAKKPSKKPKPKAQTVPKVTVEYFPSSFSTAALTGPED
Gopherus evgoodei.1	XP_030410035.1(*)	METCLKVLLLVGVVTAAPTPPSSPQLTHEDAVLAAVQIYNQEPGTTLAYRLLEAEPQPDWDVSSKTIQSLKFTVKETVCPVSGNRNPNQCEFKEDGLVKDCSGFFSTEQDPPSIIIKCEEASEEPNVVTRGRWKRFWKGRGGGIVAAALKIIG
Gopherus evgoodei.2	Novel sequence	METSLKVLLLVGVVTAASTPPSSPFLPSHEDAVLAAVQIYNQEPGTTLAYRLLEAEPQPDWDVSSKTIQSLTFTVKETVCLVSEKRDPNQCEFKEDGLVKDCSGFFSTEQDPPSIIIKCEEASEKPKVVTRWGSWGRSIVRGLFGIIG
Gopherus evgoodei.3	XP_030405144.1	MASCWAVLLLVTVAAAVPVKTPQPYEKAIARAIAFYNQGPTVKYAFRLLTATPPPEVAQGTISNTLQLNFTIMETTCPASDKTPVEQCAFKENGLVRDCSGHFSTQETCSVFAITCDVAPPQPVRVTRWWPLVRIGAKILGHGIDYYRKRG
Gopherus evgoodei.4	XP_030405145.1	MASCWAVLLLVTVAAAVPVKTPQPYEKAIARAIAFYNQGPTVKYAFRLLTATPPPEVAQGTISNTLQLNFTIMETTCPASDKTPVEQCAFKENGLVRDCSGHFSTQETCSVFAITCDVAPPQPVRVTRWWPLVRIGAKILGHGIDYYRRRG
Gopherus evgoodei.5	XP_030405147.1	METCLKVLLLVGVVTAAPTPPSSPQLTHEDAVLAAVQIYNQEPGTTLAYRLLEAEPQPDWDVSSKTIQSLKFTVKETVCPVSGNRDPNQCEFKEDGLVKDCSGFFSTEQDPPSIIIKCEEASEEPNVVTRGRWKRFWKGRGGGIVAAALKIIG
Gopherus evgoodei.6	XP_030410037.1	METCLKVLLLVGVVTAALTPPSSPLPSHEDAVLAAVQIYNQEPGTTLAYRLLEAEPQPDWDVSSKTIQSLTFTVKETVCLVSEKRDPNQCEFKEDGLVKDCSGFFSTEQDPPSIIIKCEEASEKPKVVTRWGSWGRSIVRGLFGIIG
Gopherus evgoodei.7	XP_030410038.1	MASCWAVLLLVTVAAAVPVKTPQPYEKAIARAIAFYNQGPTVKYAFRLLTATPPPEVAQGTISNTLQLNFTIMETTCPASDKTPVEQCAFKENGLVRDCSGHFSTQETCPVFAITCDVAPPQPVRVTRWWPLVRIGAKILGHGIDYYRKRG
Gopherus evgoodei.8	XP_030410039.1(*)	MASCWAVLLLVTVAAAVPVKTPQPYEKAIARAIAFYNQGPTVKYAFRLLTATPPPEVAQGTISNTLQLNFTIMETTCPASDKTPVEQCAFKENGLVRDCSGHFSTQETCSVFAITCDVAPPQPVRVTRWWPLVRIGAKIVGHGIDYYRRRG
Gopherus evgoodei.9	XP_030405662.1	MAVVLFLVALLASLGVSVTASPLLEAQLLSPRDVARSAVEAYNQETGTQAVFRLFKLKSFHKTQFDWGTHFSINFTIKETNCWKSVASHRLEDCRYQPNGITRDCSADVSVLDFVQEAPLTSVKCTQQQRAKKPSKKPKPKAQTVPKVTVEYFPSFFSTAALTIPED
Gopherus flavomarginatus.1	XP_050798789.1	METCCRCLLLLGIVAAAATAVPSQHKTLSYEEAVALAVDFYNQESGIDHAFRLLKADPQPEWDMTSNPRQKLKFTVKETVCLVSENRHGTECDFIDNGVVRDCSGSFSTQQESPIVIISCNTVTQEDTHIRRPRSPRRKWRPPRRHLPGSYTLISHGGGKGKGKGSRLQVA
Gopherus flavomarginatus.2	XP_050798791.1	METCLKVLLLVGVVTAASTPPSSPQLTHEDAVLAAVQIYNQEPGTTLAYRLLEAEPQPDWDVSSKTIQSLKFTVKETVCPVSGNRDPNQCEFKEDGLVKDCSGFFSTQQDPPSIIIKCEEASEEPNVVTRGRWKRFWKGRGGGIVAAALKIIG
Gopherus flavomarginatus.3	XP_050798866.1	METSLKVLLLVGVVTAASTPPSSPLPSHEVAVLAAVQIYNQEPGTTLAYRLLEAEPQPDWDVSSKTIQSLTFTVKETVCLVSEKRDPSQCEFKEDGLVKDCSGFFSTEQDPPSIIIKCEEASEEPKVVTRWGTWGRSIVSGLFGFLG
Gopherus flavomarginatus.4	XP_050798794.1	MASCWAVLLLVTVAAAVPVKTPQPYEKAIARAIAFYNQGPTVKYAFRLLTATPLPEVAQGTISNAFQLNFTIMETTCPASDKTPVEQCAFKENGLVRDCSGHFSTQETCSVVAITCDVAPPQPVRVTRWWPLVRIGAKIVGHGIDYYRRRG
Gopherus flavomarginatus.5	XP_050798793.1	MASCWAVLLLVTVAAAVPVKTPQPYEKAIARAIAFYNQGPTVKYAFRLLTATPPPEVAQGTISNAFQLNFTIMETTCPASDKTPVEQCAFKENGLVRDCSGHFSTQETCPVFAITCDVAPPQPVRVTRWWPLVRIGAKIVGHGIDYYRRRG
Gopherus flavomarginatus.6	XP_050798790.1	MAVVLFLVALLASLGVSVTASPLLEAQLLSPRDVARSAVEAYNQETGTQAVFRLFKLKSFHKTQFDWGTHFSINFTIKETNCRKSAASYRLEDCRYQPNGITRDCSADVSVLDFVQEAPLTSVKCTQQQRAKKPSKKPKPKAQTVPKVTVEYFPSSFSTAALTIPED
Geoemydidae		
Mauremys mutica.1	XP_044860750.1	MAVLLFLVSLLASLGVSVTASPLLEAQLLSPRDVARSAVEAYNQETGTQAVFRLFKLKSFHKTQFDWGTHFSINFTIKETNCRKNAASYRLEDCRYRPNGITRDCSADVSVLDFVQEAPLTSVKCTQQQAKKPSKKPKPKAQTVPKVTVEYFPSSFSTAALTKPED
Mauremys mutica.2	XP_044862543.1	MASCWAVLLLVTVAAALPVKTPQPYEKAIAQAIDFYNQGPTVKYAFRLLTSTPPPEVAQGTISNILQLNFTIMETTCLASDKAAVEQCAFRENGLVRDCSGHVSTQETCPVVAITCDVAPAQPVRVTRWWPLVRLGAKILGHGISFFRKRG
Mauremys mutica.3	XP_044859061.1	METCLKVLLLVGVVTAAPAPSSSPLPTYEAAVLAAVEIYNQEPGTTLAYRLLEAEPQPDWDVSSKTIQSLTFTVKETVCPASEKRNPNQCEFKEDGLVKDCSGFFSTEQDPPSFIIKCEEASEEPNVVTRSRWKRWGRKAGRFIYRHRWKIVRVGLGFLG
Mauremys mutica.4	XP_044859218.1	METCLKVLLLVGVVTAAPAPSSSPLPTYEAAVLAAVEIYNQEPGTTLAYRLLEAEPQPDWDVSSKTIQSLTFTVQETVCPVSEKRNPNQCEFKEDGLVKDCSGFFSTEQDPPSFIIKCEEASEEPNVVTRSRWRRWGRKAGRFIYRHRWKIVRVGLGFLG
Mauremys mutica.5	Novel sequence	METCLKVLLLIGVVTAAPTPSSSPLPTHEDAVLAAVQIYNLEPGTTLGYRLLEAEPQPDWDASSKTIQSLTFTVKETVCLVSGNRDPSQCEFKEDGLVKDCSGFFSTEQDPPSIIIKCEEASEEPNVVTRSRWKRFWRKTGNFVRKRGWDILATGLKIFG
Mauremys mutica.6	Novel sequence	METCCRCLLLLGIVAAAATAVPSQHKTLSYEEAVALAVDFYNQGSGIDHAFRLLKADPQPEWDMTSNSRQQLKFTVKETVCPVSENLHGTKCDFTDNGVVRDCSGSFSTQQESPIVIISCNTVTQEDTHIRRPRSPRRSWPPRWHLPGSYTLIAHGRGKGKGKGSRLQVA
Mauremys mutica.7	XP_044860113.1(*)	MASCWAVLLLVTVAAALPVKTPQPYEKAIARAIDFYNQGPTVKYAFRLLTSTPPPEVAQGTISNILQLNFTIMETTCLASDKAAVEQCAFRENGLVRDCSGHVSTQETCPVVAITCDVAPAQPVRVTRWWPLVRLGAKILGHGISFFRKRG
Mauremys reevesii.1	Novel sequence	METCCRRLLLLGIVAAAATAVASQHKTLSYEEAIALAVDFYNQGSGIDHAFRLLKADPQPEWDMTSNSRQQLKFTVKETVCPVSENLHETKCDFTDNGVVRDCSGSFSTQQESPIVIISCNTVTQEDTHIRRPRSPRRSWRRRWHLPGSYTLIGHGGGKGKGKGSRLQVA
Mauremys reevesii.2	Novel sequence	METCLKVLLLVGVVTAAPTPSSSPLPTHEDAVLAAVQIYNLEPGTTLGYRLLEAEPQPDWDASSKTIQSLTFTVKETVCLVSGNRDPSQCEFKEDGLVKDCSGFFSTEQDPPSIIIKCEQASEEPNVVTRSRWKRFWRRTGNFVRRRGLDILATGLKIFG
Mauremys reevesii.3	XP_039381017.1	METCLKVLLLVGVVTAAPTPSSSPLPTHEAAVLAAVEIYNQEPGTTLAYRLLEAEPQPDWDVSSKTIQSLTFTVKETVCLVSEKRNPNQCEFKEDGLVKDCSGFFSTEQDPPSIIIKCEEASEEPNVVTRSRWRRWLRKAGSFLIHRNRSIFFVG
Mauremys reevesii.4	XP_039383706.1(*)	METCLKVLLLVGVVTAAPAPSSSPLPTHEAAVLAAVEIYNQEPGTTLAYRLLEAEPQPDWDVSSKTIQSLTFTAKETVCLVSEKRDPNQCEFKEDGLVKDCSGFFSTEQDPPSIIIKCEEASEEPNVVTRAWAAGYIHVRCPGSRNIQGRVPCSSNSWSWVSPFALPGAGPSPRRVFP
Mauremys reevesii.5	XP_039382276.1	MASCWAVLLLVTVAAALPVKTPQPYEKAIAQAIDFYNQGPTVKYAFRLLTSTPPPEVAQGTVLNILQLNFTIMETTCLASDKAAVEQCAFRENGLVRDCSGHVSTQETCPVVAITCDVAPAQPVRVTRWWPLLRLGAKIIGHGISFFRKRD
Mauremys reevesii.6	XP_039383632.1	MAVLLFLVSLLAPLGVSVTASPLLEAQLLSPRDVARSAVEAYNQETGTQAVFRLFKLKSFHKTQFDWGTHFSINFTIKETNCRKNAAGYRLEDCRYRPNGITRDCSADVSVLDFVQEAPLTSVKCTQQQAKKPSKKPKPKAQTGPKVTVEYFPSSFSTAALTKPED
Emydidae		
Emys orbicularis.1	XP_065253727.1	MAVVLFLVALLASLGVPVTASSLLEAQLLSPRDVARSAVEAYNQETGTQAVFRLFKLKSFHKTQFDWGTHFSINFTIKETNCRKSVASYRIEDCRYRPNGITRDCSANVSVLDFVQEAPLTSVKCTQQQRPKKPSKKPKPKAQTVPKVTVKYFTSSFSTAALTAPEED
Emys orbicularis.2	XP_065255936.1(*)	MVSCWAVLLLLVTAATAVPVKTPEPYEKAIARAIDFYNQGPTVKYAFRLLTATPPPEVAQGTQGTISNTLQLNFTIMETTCQVTDQVSNQVRVEQCAFKENGLVRDCSGHFSTQEACPVFAITCDVAPPQPVRVTRWWPLVRIGAKILGHGISFFHKRG
Emys orbicularis.3	XP_065255822.1	METCLKVLLLVGVVTAAPTPSSAPLPTHEDAILAAVLVYNQEPALTLAYRLLEAEPQPDWDVSSKTIQPLKFTVQETVCPVSENRDFNQCEFKEDGLVKDCSGFFSTVQDPPSVIIKCEEASEEPNVVTRGRWGRWRRKAGRFIKRNRWNIIATGLKLIG
Emys orbicularis.4	XP_065255823.1	METCLKVLLLVGVVTAAPTPSSAPLPTHEDAILAAVQVYNQEPGTTLAYRLLEAEPQPDWESDLGNRFDVSSKTIQPVTFTVKETVCLVSEKRDPNQCEFKEDGLVKDCSGFFSTEQDPPSIIIKCEEVSEEPNVVTRGRWKRFWRGTGNFFRKHGLNILSTGLKIIG
Emys orbicularis.5	Novel sequence	METCLKVLLLVGVVTAAPTPPSSSPLPTHEDAVLAAVQVYNQEPGTTLAYRLLEAEPQPDWDVSSKTIQSLPFVTFTVKETVCLVSEKRDPNQCEFKEDGLVKDCSGFFSTEQDPPSVIIKCEEASEEPNVVTRGRWRNVGRFIRRNAWDLIHVGLALTS
Emys orbicularis.6	Novel sequence	METCCRRVLLLVIVAAAATAVPSQHKTLSYEEAIALAVDFYNQGSGIDHAFRLLKADPQPEWDMTSNLRQELKFMVKETVCPVSENLNGTECDFRDNGVVRDCSGFFSTQQESPIVIINCNTVTKEDTHIRRSRSPRRNRRPRRWHLPGSYTLIAHGGGKGKGKGKGSRLQVA
Malaclemys terrapin.1	XP_053872754.1	MAVVLFLVALLASLGVSVTASSLLEAQLLSPRDVARSAVEAYNQETGTQAVFRLFKLKSFHKTQFDWGMHFSINFTIKETNCRKSMASYRLEDCKYRPNGITRDCSADVSVLDFVQEAPLTSVKCTQQQQRAKKPSKKPKPKAQTVPKVTVKYFTSSFSTAALTAPEEDKGQPA
Malaclemys terrapin.2	XP_053875054.1	MASCWAVLLLVTAATAVPLKTPEPYEKAIARAIDFYNQGPTVKYAFRLLTSSPPPEVAQVAISNTLQLNFTIKETTCQVSDQVRVEQCAFKENGLVRDCSGHFSIQEACPVVAITCDVAPPQPVRVTRWWPLVRIGAKILGHGISFFHKRG
Malaclemys terrapin.3	Novel sequence	METCLKVLLLVAVVTAAPTPSSAPLPTHEDAVLAAVLVYNQEPGLTLAYRLLEAEPQPDWDMSSKTIQPLKFTVQETVCPVSEQRDFNQCEFKEDGLVKDCSGFFSTEQDPPSVIIKCEEASEEPNIITRGRWGRWRRKAGRFIKRNRWNIIATGLKLIG
Malaclemys terrapin.4	Novel sequence	METCLKVLLLVGVVTAAPTQSSAPLPTHEDAVLAAVQVYNQEPGTTLAYRLLEAEPQPDWDVSSKTIQPVTFTVKETVCLVSEKRDPNQCEFKEDGLVKDCSGFFSTEQDPASVIIKCEEASEEPKVITRGRWGSVGRFIRRNAWDLIHVGIALTN
Malaclemys terrapin.5	Novel sequence	METCLKVLLLVGVVTAAPTPSSAPLPTHEDAVLAAVQAYNQEPGTTLAYRLLEAEPQPDWDVSSKTIQPVTFTVKETVCLVSEKRDPNQCEFKEDGLVKDCSGFFSTEQDPASVIIKCEEASEEPKVITRGRWGSVGRFIRRNAWDLIHVGIALTN
Malaclemys terrapin.6	Novel sequence	METCLKVLLLVGVVTAAPTPSSAPLPTHEDAVLAAVQAYNQEPGTTLAYRLLEAEPQPDWDVSSKTIQPVTFTVKETVCLVSEKRNPNQCEFKEDGLVKDCSGFFSTEQDPPSVIIKCEEASEEPKVITRGRWGSVGRFIRRNAWDLIHVGIALTN
Malaclemys terrapin.7	XP_053872755.1	METCCRRVLLLVTVAAAATAVPSQHKTLSYEEAIALAVDFYNQGSGLDRAFRLLKADPQPEWDMTSNPRQELKFMVKETVCPVSENLNGTECDFRDNGVVRDCSGFFSTQQESPIVIINCNTVTKEDPHIRRSRSPRRSRWPRRWHLPGSYTLIAHGGGKGKGKGSRLQMA
Trachemys scripta.1	Novel sequence	METCCRRVLLLVSVAAVATAVPSQHKTLSYEEAIALAVDFYNQGSGLDRAFRLLKADPQPEWDMTSNPRQELKFMVKETVCPVSENLNGTECDFRDNGVVRDCSGFFSTQQESPIVIINCNTVTKEDPHIRRSRSPRRSRWPRRWYLPGSYTLIAHGGGKGKGKGSRLQMA
Trachemys scripta.2	Novel sequence	MANCWAVLLLVVTAATAVPVKTPEPYEKAIARAIDFYNQGPTVKYAFRLLTSSPPPEVAQVAISNTLQLNFTIKETTCQVSDQVRVEQCAFKENGLVRDCSGHFSTQEVCPVVAITCDVAPPQPVRVTRWWPLLRIGAKILGHGISFFHNKG
Trachemys scripta.3	Novel sequence	MAVVLFLVALLASLGVSVTASSLLEAQLLSPRDVARSAVEAYNQETGTQAVFRLFKLKSFHKTQFDWGTHFSINFTIKETNCRKSMASYRLEDCRYRPNGITRDCSADVSVLDFVQEAPLTSVKCTQQQQRAKKPSKKPKPKAQTVPKVTVKYFTSSFSTAALTAPEEDKGQPA
Dermochelyidae		
Dermochelys coriacea.1	XP_038247013.1	MVVVLFLVALLASLAVSVTASPLLEAQLLSPRDVARSAVEAYNQETGTQAVFRLFKLKSFHKTQFDWGTHFSINFTIKETNCWKSVAGYRLEDCRYRPNGITRDCSANISVLDFVQEAPLTSVKCTQEQRAKKTSKKPKPKAQSVPKVTVEYFPSSFSTAALTVPEEG
Dermochelys coriacea.2	XP_038246568.2	MESCWGVIMMLGVAVATTSPPPPQQMLSYEEAVSLAIDLYNQEPHVDFAFRLLEAEAQPDWAATNETLQLLEFTVKETECPVSENLLLDECDFRSNGVVRVCSGTISTEPEAPLVLLTCDTVAQERIRVRRSIFRKLRRKIKKGLKKGIRILLAGGRQGPPRGGRPGMI
Dermochelys coriacea.3	XP_038246569.1(*)	MERGWGVIMMLGVAVATTSPLPQTRSYEVAVSHAVDLYNQKAPGDSLFRLLEADLQPGWDENSQSAQELNFLIKETVCPVSEKAVTEQCDFKEDGLVRDCSGYIFPEQQLATILITCDTVAEEPTRVRRSRFGRFFKKVRKQLGRVLRHSRITLGGRMRF
Dermochelys coriacea.4	XP_038246570.1	METCLKVLLLIGVVTAAPAPSSSPLPTHEDAVLAAVQTYNQEPDITLAYRLLEAEPQPDWDVSSKTIQPLTFTVQETVCPVKEKRDISQCEFKEDGLVKDCSGFFSTEQDPASVIIKCEEASEEPNLVTRGRWSRFRRRVGRFFRRHKEKIIRAGIGIALG
Dermochelys coriacea.5	Novel sequence	METCLKVLLLIGVVTAAPAPSSSPLPTHEDAVLAAVQLYNQEPDITLAYRLLEAEPQPDWDVSSKTIQPLTFTVQETVCPVKEKRDISQCEFKEDGLVKDCSGFFSTEQDPASVIIKCEEASEEPNIITRGRWSRFRKRAGRFLRRHKGKIIRAGIGIALG
Dermochelys coriacea.6	Novel sequence	METCLKVLLLIGVVTAAPVPSLSPLPTHKDAVLAAVQIYNQEPGTTLAYRLLEAEPQPDWIMSSKTIQPLRFTVKETVCLVSEKRDINQCEFKEDGLVKDCSGFFSTEQEPASVIIKCEEASEEPNLVTRGRWKRFWRGAGNFFRQHGLNILNASLKLIG
Dermochelys coriacea.7	XP_038248398.1(*)	METCCRPVLLLVIVAAAATAVPSQHKMLSYEEAIALAVDFYNKGSGIDHAFRLLKADPQPEWDMTSNPQQELKFMVKETVCSVSENLPRTECDFRDNGVVRDCSGFFSTLQKSPIVLISCNTVTQERTHIRRSRSPRRNWRRRWHLPGSYTLISHFGSKGKGKGKGSGLQVA
Cheloniidae		
Caretta caretta.1	Novel sequence	METCCRPVLLLVIVAAAATAVPSQHKTLSYEEAIALAVDFYNKGSGIDHAFRLLKADPQPEWDMTSNPRQALKFTVKETVCPVSENLPGTECDFRDNGVVRDCSGFFSTLQKSPIVLISCNTVTQEGAHIRRSRSPKGRWWRRRWHLPGSYTLIAQGGRKGKGKGKGSRLQMV
Caretta caretta.2	Novel sequence	METCLKVLLLIVGVVTAAPAPSSSPLPTHKDAVLAAVQLYNQEPDITLAYRLLEAEPQPDWDMSSKTIQTLRFTVKETVCLVSEKRDINQCEFKEDGLVKDCSGFFSTEKDPPSVIIKCEEASEEPNIVTRGRWKRFWRGAGKFFRQHGLNILTATLKMVG
Caretta caretta.3	Novel sequence	METCLKVLLLIGVVTAAPAPSSSPLPTHEDAVLAAVQLYNQEPDITLAYRLLEAEPQPDWDVSSKTIQPLTFTVQETVCPVKEKRDISQCEFKEDGLVKDCSGFFSTEQDPASVIIKCEEASEEPNIVTRGRWSRFRKRAGRFLRRHKGKIIRAGIGIALG
Caretta caretta.4	Novel sequence	METCLKVLLLIGVVTAAPTPSSSPLPTHEDAVLAAVQLYNQEPDITLAYRLLEAEPQPDWDVSSKTIQPLTFTVQETVCPVKEKRDISQCEFKEDGLVKDCSGFFSTEQDPASVIIKCEEASEEPNIVTRGRWSRFRRKVGRFFRRHKEKIIRAAVDIVFP
Caretta caretta.5	XP_048693294.1(*)	MERGWGVIMMLGVAVATTSPLPQTRSYEVAVSHAVDLYNQKAPGDSLFRLLEADPQPGWDENSQSTQELNFLIKETVCPVSEKAVTEQCDFKEDGLVRDCSGYIFPEQQLATILITCDTVAEEPTRVRRSRFGRFFKKVRKQLGRVLRHSRITLGGRMRF
Caretta caretta.6	XP_048693293.1	MESCWGVIMMLGVAVATTASPPPPQQMLSYEEAVSLAIDLYNQEPHIDFAFRLLEAEAQPDWATTNETLQLLEFTVKETVCPVSENLLLDECDFRSNGVVRACSGTVSTEPEAPLVLLTCDTVAQERTRVRRSIFRKLRRKIKKGLKKGIQHLLAGGRQGLLQGGRPGMI
Caretta caretta.7	XP_048693295.1	MAVVLFLVALLASLGVSVTASPLLEAQLLSPRDVARSAVEAYNQETGTQAVFRLFKLKSFHKTQFDWGMHFSINFTIKETSCWKSVASYRLEDCRYRPNGMKRDCSANVSVLDFVQEAPLTSVKCTQEQRAKKNSKKPKPKAQSVPKVTVQYFPSSFSTAALTAPEEG
Chelonia mydas.1	XP_037747220.1	MAVVLFLVALLASLGVSVTASPLLEAQLLSPRDVARSAVEAYNQETGTQAVFRLFKLKSFHKTQFDWGTHFSINFTIKETNCWKSVASYRLEDCRYRPNGMKRDCSANVSVLDFVQEAPLTSVKCTQEQRAKKTSKKPKPKAQSVPKVTVQYFPSSFSTAALTAPEEG
Chelonia mydas.2	QED55073.1(*)	MESCWGVIMMLGVAVATTTSPPPPQQQMLSYEEAVSLAIDLYNQEPHIDFAFRLLEAEAQPDWAATNETLQLLEFTVKETECPVSENLLLDECDFRSNGVVRACSGTISTEPEAPLVLLTCDTVAQERTRVRRSIFRKLRRKIKKGLKKGIQHLLAGGRQGLPQGGRPGMI
Chelonia mydas.3	QED55074.1	MERGWGVIMMLGVAVATTTSPLPQTRSYEVAVSHAVDLYNQKAPGDSLFRLLEADPQPGWDENSQSTQELNFLIKETVCPVSEKAVTEQCDFKEDGLVRDCSGYIFPEQQPATILITCDTVAEEPTRVRRSRFGRFFKKVRKQLGRVLRHSRITVGGRMRF
Chelonia mydas.4	QED55075.1(*)	METCLKVLLLIVGAVTAAPAPSSSPLPSHEDAVLAAVQIYNQDPDITLAYRLLEAEPQPDWDVSSKTIQPLTFTVQETVCPVKEKRDISQCEFKEDGLVKDCSGFFSTEQDPASVIIKCEEASEEPNIVTRGRWKRFWRGAGRFFRRHKEKIIRAAVDIVLS
Chelonia mydas.5	Novel sequence	METCLKVLLLIVGAVTAAPAPSSSPLPSHEDAVLAAVQIYNQDPDITLAYRLLEAEPQPDWDVSSKTIQPLTFTVQETVCPVKEKRDISQCEFKEDGLVKDCSGFFSTEQDPASVIIKCEEASEEPNIVTRGRWSRFRKRAGRFLRRHKGKIIRAGIGIALG
Chelonia mydas.6	Novel sequence	METCLKVLLLIVGAVTAAPAPSSSPLPSHEDAVLAAVQIYNQDPDITLAYRLLEAEPQPDWDVSSKTIQPLTFTVQETVCPVKEKRDISQCEFKEDGLVKDCSGFFSTEQDPASVIIKCEEASEEPNIVTRGRWSRFRKRAGRFLRRHKGKIIRAGIGIALG
Chelonia mydas.7	Novel sequence	METCLKVLLLIVGAVTAAPAPSSSPLPSHEDAVLAAVQIYNQDPDITLAYRLLEAEPQPDWDVSSKTIQPLTFTVQETVCPVKEKRDISQCEFKEDGLVKDCSGFFSTEQDPASVIIKCEEASEEPNIVTRGRWSRFRKRAGRFLRRHKGKIIRAGIGIALG
Chelonia mydas.8	Novel sequence	METCLKVLLLIVGAVTAAPAPSSSPLPSHEDAVLAAVQIYNQDPDITLAYRLLEAEPQPDWDVSSKTIQPLTFTVQETVCPVKEKRDISQCEFKEDGLVKDCSGFFSTEQDPASVIIKCEEASEEPNIVTRGRWSRFRKRAGRFLRRHKGKIIRAGIGIALG
Chelonia mydas.9	Novel sequence	METCLKVLLLIVGAVTAAPAPSSSPLPSHEDAVLAAVQIYNQDPDITLAYRLLEAEPQPDWDVSSKTIQPLTFTVQETVCPVKEKRDISQCEFKEDGLVKDCSGFFSTEQDPASVIIKCEEASEEPNIVTRGRWSRFRKRAGRFLRRHKGKIIRAGIGIALG
Chelonia mydas.10	Novel sequence	METCLKVLLLIVGAVTAAPAPSSSPLPSHEDAVLAAVQIYNQDPDITLAYRLLEAEPQPDWDVSSKTIQPLTFTVQETVCPVKEKRDISQCEFKEDGLVKDCSGFFSTEQDPASVIIKCDEASEEPNIVTRGRWSRFRKRAGRFLRRHKGKIIRAGIGIALG
Chelonia mydas.11	Novel sequence	METCLKVLLLIVGAVTAAPAPSSSPLPSHEDAVLAGVQLYNQEPGTTLAYRLLEAEPQPDWDMSSKTIQTLRFTVKETVCLVSEKRDINQCEFKEDGLIKDCSGFFSTEKDPPSVIIKCEEASEEPNLVTRSRWKRFWRGAGKFFRQHGLNILTATLKIIS
Chelonia mydas.12	XP_043397659.1(*)	METCCRPVLLLVIVAAAATAVPSQHKTLSYEEAIALAVDFYNKGSGIDHAFQLLKADPQPEWDMTSNSRQALKFTVKETVCPVSEKLPGTECDFRDDGVVRDCSGFFSTLQKSPIVLISCNTVTQERTHIRRSRSPRRSWRRRWHRPGSYTLISQGGRKGKGKGSRLQMA
Trionychidae		
Rafetus swinhoei.1	Novel sequence	MELLLGVLLVSLGVSVTASPLLEAQLLSPRDVARAAVEAHNQETGTQAVFRLLKLKSFHKTQFDWGRHFSMNFTIKETNCRKSMASYRLEDCRYRPNGVTKDCSADVSVLYFLQEAPLTSVKCTQQQGVKKTSKKPKPQAMTFPKVTVEYFPSSFSTAALTVPED
Rafetus swinhoei.2	Novel sequence	MASYWACLLLLVAVASAVPLQTPLQTPQQTSLQTPQPYEQATTQAIDFYNQGPTVTNAFRLLFAAPPPQVAQDTAANLLQQLNFTIMETTCLASGNPVVEQCAFKENGLVRDCSGHFSTQEACPVVAITCDVAPPRPVRVTRWWIPALRLGAKIAGHAISLFRKRN
Rafetus swinhoei.3	Novel sequence	MKSCLKILLLIGVVTAAPTPPASSLPSHEDAILSAVQVYNQEPGLTLAYRLLEAEPQPDWDVTSKTVQPLKFTIKETVCLISEKRDINQCDFKEDGLIKDCSGFFSTEQDPPSAIIKCEEASEEPDIVTRGRWGRFKRRAGRFLRRNRWQIISTGLKLIG
Rafetus swinhoei.4	Novel sequence	METCLKILLLVGVVTAAPTPPASSLPSHEDAILAAVKMYNQEPGLTLAYRLLEAEPQPDWDMTSKTIQPLKFTIKETVCLISEKRDINQCDFKEDGLIKDCSGFFSTEQDPPSAMIECEEASEEPNLVTRSRWRRFTRKVGGFVRKNRWNIISTALKWIG
Rafetus swinhoei.5	Novel sequence	MESGRRHLLLLVVVATVAAAEPSPHKTLSYEEAVALAIDFYNQGSGIDHAFRLLKATPLPKWDMTANPRQELKFTVKETVCLVAENLPGTECDFRDDGVVRDCSGFFSTQQVSPIVMVSCNSVAQEPTHIRRSRSPRRRWTWMPRRRPHIGSYTLITHGGRKGKHNGLQMA
Geckos		
Eublepharidae		
Eublepharis macularius.1	Novel sequence	MGLFLVGAVLVLLGASGSVSSPPDGSMSPRDAARIVVETYNQETETQAVFKLLKFRSTHKTMFDWGVHFSMNFTIKETSCQKTMTSYKLGECRYKPNGEVKECSSEVSFLNFMQDTPLSSVQCSPLQQGNGSSSTKHKSSSNPKVMPAMFRIYVEHYMPSAYSTAAQMAMEGE
Eublepharis macularius.2	Novel sequence	MESWILLLLGLAVAAVALPPPQEELSYEEAVSSVVDLYNQEPGVELAFRLLEAKPQPEWDPSGQSLQELEFTVQETTCPPAKQLNLDECDFKDDGVVKECHGTIFTEQGAPVIQYLCETADQGHIRVRRGLGKIVKKIKKNKIKKGIKKVLPGGGSSIAQGPGGSKTHYA
Eublepharis macularius.3	Novel sequence	MENWILLLILSLAVAATALSPPQEGLSYEEAISLAVDRYNQEPGLELAFRLQEAKPQPEWDPSGQSLQELEFTVQETTCPPAKQLNLDECDFKDDGVVKECFGTIFSEQGAPVLQLNCETAGQGHIRVRRGGFRKFVKKVKKAFKKATRPLRRG
Eublepharis macularius.4	XP_054847725.1	MQSCWALVMLLVGVASAAPTRSTPLSYEQVLASAVSTYNQELGQENAFRLLEAEPLPDWDPSAKTIQSLKFSIKETVCSTAENLDVSQCDYKDDGVDRDCSGFYSAEQDPPLIIVQCEDMEQELNRITRSRWRKFWRKTKKFVKKHGVSIGLAALRFPG
Eublepharis macularius.5	XP_054847726.1	MQSCWALVMLLVGVASAAPTRSTPLSYEQVLASAVSTYNQELGQENAFRLLEAEPLPDWDPSAKTIQSLKFSIKETVCSTAENLDLSQCDYKDDGVDRDCSGFYSADQDPPLIIVQCEDMDQELNRITRSRWRRFWRKTKNFVKQHGVSIGLAALRLRG
Eublepharis macularius.6	Novel sequence	MQTSWAMLLLFAGAAVAHPAPPPPLSYEQVLASAIETYNKESGLENAFRLLEPEPQPDWDPSALTIQPLKFSVRETVCKAAESSNIDQCDYKDDGLDRDCSGFFSYAAQSPPLMVVQCEGVAQELERIKRGRFKKWVKKIGAFIKKYGPRIASVLCNIAG
Eublepharis macularius.7	Novel sequence	MRSAWGVLFLLGFGTAVPLEAPKTFEEFLPQVVKIYNQGPGVQNAFRLLEATPPLGMDSTSGTLQQLNFTMQETTCPASEDVVAEQCEFKPDGLMKDCIGYFSTKPENPMILINCDTAIQEPVHVTRFRGLGGFLKGFSRGVQLWDAHH
Sphaerodactylidae		
Euleptes europaea.1	XP_056713328.1	MDLLLVGAVLVFLGVSGSASPPAEGSPAPRDAARIVVDMYNQESETQAVFKLLKFRNTHKTRFDWGVHFSLNFTIKETNCRKTMTNYKPGNCKYKPNGAVNECSAEVSFLNFMQDTPLTSVQCTPLQRAGGKKSNSPSRLQAVPSVPRIHVEHYMPSAYITAAQMALDVE
Euleptes europaea.2	XP_056713132.1	MGSWVLLMLCLAVAARASPATQREFSYEEAISLAIDLYNQEPEVELAFRLLEAKPQPDWDPSMQSLQELEFTVQETTCPPAEQLNLDKCDFKDKGVVKECYGTIFSEQGAPVIRYLCETAGQGHIRVKRCIWRGIRKGIKKNKKEIKEIKNLPPCEKSRKEIKEPFPGSGSPIGKVPVGNEHIA
Euleptes europaea.3	XP_056713133.1	MEGWFLLMLSLAVVATASPPAQEELSYEEAVSVAVELYNQEPGLKLAFRLLEAKPQPEWDPAIHSLQELEFTVQETTCPPSEQLNLDKCDFKNKGVVKECSGIIFSEQGAPVIELDCESAGQVHVRVRRNFFHTAYNKVKETLQRLSCHCEWRN
Euleptes europaea.4	XP_056713784.1	MQSCWALMLLVGVVAGAPTQPTLLSYEQVLASAVATYNQELGPENAFQLLEAEPQPDWDPSAKTIQSLKFSIKETVCRAEENPDVTQCEPKEDGVDRDCSGFFSAEQNPPLIIVQCEDVDQELSRITRGRWRRFWRKTKRFVKKHGVSIALGAARFG
Euleptes europaea.5	XP_056713134.1	MANCWAMLLLMVGAAIAIPTVPTRLSYEQVLASAVKTYNQESGLDSAFRLLEAEPQQDWDSSAQAVQPLKFSVKETVCQATESSDIDQCDYKDDGLDQDCSGFFSTAQSPPLIVVQCEDVDQELERITRGRFKKWKKWKKKLRKVLKKQGPSIAVGAFTG
Euleptes europaea.6	XP_056713135.1	MERAWGILFLAGLAAAVPLDTPKTLEEALSQVIEIYNRGPRVQNAFRLLEATAPLELDPTTGTLQQLTFTMQETTCPNSENVVTEQCEFKPDGLIKGCTGYFSTEPGNPAILVSCDNVVQEPVRVTRFRGLGGFVRGFSRGIRFWDSNQ
Euleptes europaea.7	XP_056713136.1	MANCWATLLLMVGVAIAIPTVPPQLSYDQALASAVKTYNQESGLESIFRLLEAEPKQNWDPTTGNLQQLTFTMQETTCPNSENVVTEQCEFKPDGLIKGCTGYFSTEPGNPKILVSCENAVQKPVRVTRFRGLGLLLRGLGRGFGRLWGSNQ
Gekkonidae		



VARSA-type (Cath6). Likewise, in *Chelonia mydas*, Qiao et al. (2019) described four cathelicidins corresponding to AATNET- (Cath1), DENSQ- (Cath2), THEDA- (Cath3) and YEKAI-type (Cath4) cathelicidins.

The contrasting distribution of cathelicidin types across Testudines and Squamata suggests lineage-specific evolutionary trajectories. The apparent restriction of DENSQ and AATNET types to marine turtles, and the absence of VARMA type in snakes, may reflect differences in ecological pressures, pathogen exposure or broader immune-system diversification. The exclusive presence of GALSA-type cathelicidins in snakes may relate to physiological specialisations, including venom-associated immunity or adaptation to distinct microbial environments. In contrast, turtles retained and duplicated several types, consistent with long-term gene-family stability. Although the classification framework is robust within Testudines and Squamata, its applicability to other reptilian orders will require additional genomes and may reveal new signature types or structural variants. Together, these patterns highlight the influence of ecological context and genomic history on the diversification of cathelicidins in reptiles.

3.3. Phylogenetic tree analysis

When constructing the phylogenetic tree using the sequences of the conserved domain from 91 cathelicidins within the order Testudines, they grouped according to our previously established classification rather than by their phylogenetic relationships (Fig. 3). Cathelicidins of

Table 4
Distinctive features used to classify reptilian cathelicidins. Each type was primarily identified by a type-specific motif and its position within the cathelin domain, supported by typical exon length patterns (especially exons 1 and 4), and, when needed, by genomic synteny (e.g., position relative to KLHL18 flanking gene).

Cathelicidin type	Taxonomic Order	Position and type-specific motif	Typical Exon size (bp)				Additional observations
			Exon 1	Exon 2	Exon 3	Exon 4	
THEDA	Testudines	"THEDA" motif or close variants at start of conserved domain	180	108	84	93–111	Often multiple tandem copies
YEKAI	Testudines	"YEKAI" motif at start of conserved domain	174	111	84	90	Produces small peptides (<30 aa)
VARSA	Testudines	"VARSA" motif at start of conserved domain	189	111	84/87	117/120	Larger introns, cysteine spacing 11–10–19 aa, near KLHL18
DMTS	Testudines	"DMTS" motif at start of exon 2 + "KAD" at end of exon 1	186	108	84	138/144	Produces extended peptides, often Pro/Arg-rich
DENSQ	Testudines	"DENSQ" motif at start of exon 2	180	108	84	114	Only found without potential expression defects in sea turtles
AATNET	Testudines	"AATNET" motif at start of exon 2	186/189	108	84	135/159	Lys/Arg-rich mature peptides
YEQ	Squamata	"YEQ … NQQ" motif or close variants at start of conserved domain	171–183	108/111	84	99–111	Present in all Squamata suborders
GALSA	Squamata	"GALSA" motif at signal peptide + "YEEA" at start of conserved domain	186/189	108	117–267	111–135	Exclusive in Snakes; exon 3 extension (117–267 bp)
YEEA	Squamata	"YEEA" motif at start of conserved domain	177–189	108	69–83	96–144	Absent in most snake species
QEAL	Squamata	"QEAL" motif or close variants at start of conserved domain	171/174	108	84	81–105	Produces small peptides (<30 aa)
YEDA	Squamata	"YEDA" motif or close variants at start of conserved domain	186–195	108	84	114/126	Related to GALSA (phylogenetically close)
VARMA	Squamata	"VARMA" motif or close variants at start of conserved domain	183–195	108/111	84/87	132–144	Large introns, cysteine spacing 11–10–19 aa, near KLHL18

the same type clustered together, regardless of the species from which they originated. Furthermore, all the clusters in which the different types of cathelicidins were grouped were established with a robust probability, greater than a value of 90% (values circled in red in Fig. 3). This supports the presence of homology among turtle (Testudines) cathelicidin genes and supports the validity of our previously established classification. It is noteworthy that VARSA-type cathelicidins were more distant to the other groups (with a probability of 100%), which indicated they derived from an older common ancestor (Fig. 3).

In the order Squamata, the results were similar, the conserved domains clustered by cathelicidin type rather than by species, following our previously established classifications (Fig. 4). These findings may indicate the existence of six cathelicidin-type genes. It is noteworthy that the variability found in the cathelicidin sequences within this order was significantly greater than that observed in Testudines.

Significant differentiation was observed between snakes (suborder Serpentes) and the other suborders of Squamata. Consequently, within the branch corresponding to YEQ-type cathelicidins (present across the order), a clear division emerged, separating snakes from the other species. In contrast, compared to the order Testudines, the observed high number of cathelicidin variants led to less robust clustering, with a lower likelihood (typically around 70%) of sharing a common ancestor (values circled in red in Fig. 4).

By reconstructing a phylogenetic tree using the consensus sequences of each type of cathelicidin identified in reptiles (including both Testudines and Squamata orders), we assessed which of these may have derived from a common ancestor (Fig. 5). This analysis suggests a high probability that VARSA-type and VARMA-type cathelicidins originated from a different ancestral gene than other reptilian cathelicidins (Fig. 5). These types have undergone significantly less diversification, whereas the other 10 cathelicidin types have emerged from a single ancestral gene. Due to their high sequence conservation, VARSA and VARMA cathelicidins may be considered a single type conserved across different reptilian orders. Additionally, YEDA and GALSA-type cathelicidins likely originated from a common gene, as their branches exhibited probabilities above 90%. Similarly, QEAL and YEKAI-type cathelicidins in Squamata and Testudines, respectively, showed a probability above 75% (Fig. 5). In contrast, the high sequence diversification observed in the other cathelicidin types led to low probability values for tree branches, making the evolutionary history of different cathelicidin types unclear.

3.4. Structural analysis

Based on the phylogenetic tree results, we conducted a homology analysis of the cathelicidin types found in the two reptilian orders studied. After independently aligning the conserved domains of each cathelicidin type, six identity diagrams were generated for Testudines and six for Squamata (Fig. S1). Each diagram displayed a consensus sequence at the top, followed by a bar chart showing the percentage of identity i.e., the proportion of sequences matching the consensus at each position. A colour map highlighted the degree of homology, with intense blue indicating higher sequence similarity. The analysis revealed a high level of homology for each cathelicidin type, such as THEDA-type cathelicidins in Testudines, where identical amino acid residues consistently appeared at specific positions, presenting an identity greater than 80% across 84% of their sequence positions. This trend was observed across all types in both orders (Fig. S1).

Next, we compared the homology between the six cathelicidin types from Testudines and Squamata. Consensus sequences from the initial analysis were aligned, and new homology diagrams were created (Fig. 6). These diagrams showed that less than 40% of positions exhibited an identity exceeding 60%, and only a few positions reached 100%, indicating significant differences between the types in each order. These results constitute additional evidence supporting the validity of the classification of cathelicidins within reptiles. However, certain segments with nearly 100% homology were present not only within their respective orders but also between both, suggesting that these conserved segments may play important structural or functional roles, possibly contributing to the overall fold of the cathelin domain.

In general, the conserved domain follows the typical structure found in cathelicidins, consisting of a β-sheet formed by four anti-parallel β-strands connected by loops, along with an α-helix (Figs. 7 and 8). In the predicted model, this structure is stabilized by two sulphide bridges formed between the four cysteine residues characteristic of the conserved cathelin-like domain. These cysteines are typically found separated by distances of 10, 10, and 16 amino acid residues, respectively (Sanchez et al., 2002b). In reptiles, this pattern also seems to hold, except for the VARSA-type and VARMA-type cathelicidins, where the distances between cysteines change to 11, 10, and 19 residues, respectively (Table 3). However, this alteration in the separation of the cysteines does not imply a loss of structure (Figs. 7C and 8C). The β-sheet is also stabilized by van der Waals interactions between certain nonpolar

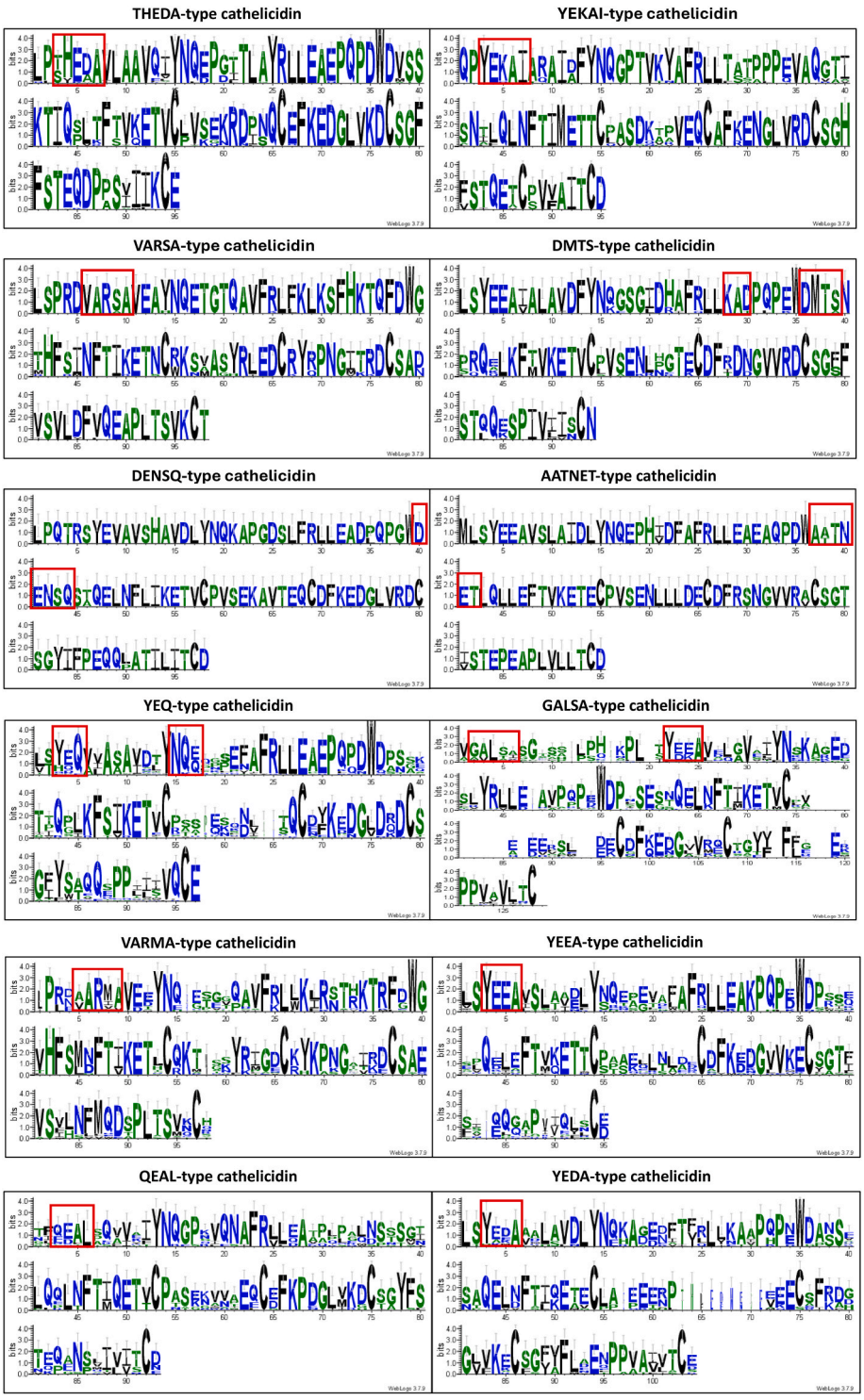


Fig. 2. Sequence conservation logos for the novel reptilian cathelicidin classification. The diagrams display the most conserved amino acids at each position following the alignment of the conserved domains within each cathelicidin type. The height of the letters indicates the frequency of amino acid appearance at that specific position. The characteristic motifs (signatures) used to identify and name the different cathelicidin types are highlighted with a red box. For all cathelicidin types, the conserved domains were aligned, except for the GALSA type, for which part of the signal peptide was included, as the "GALSA" motif is located within this region. Figures were generated using the server online WebLogo (Crooks et al., 2004).

amino acids (Sanchez et al., 2002a), which are generally conserved in these new reptilian cathelicidins (squared residues in Fig. 6).

Overall, the conserved domain structure of the six distinct types of cathelicidins found in each order is very similar; therefore, the main differences that allow for the differentiation of these groups lie in their sequences. However, there are some differences, for example it is most common for cathelicidins to exhibit a 3–10 helix in the appending domain (located between β-strands 2 and 3), but certain types of cathelicidins, such as THEDA, DMTS, and YEDA, form within this region two helices instead of one (Fig. 7A, D, 8F). It is noteworthy that QEAL-type cathelicidins contain an α-helix within the appendage domain and a second 3–10 helix located between β-strands 4 and 5 (Fig. 8E).

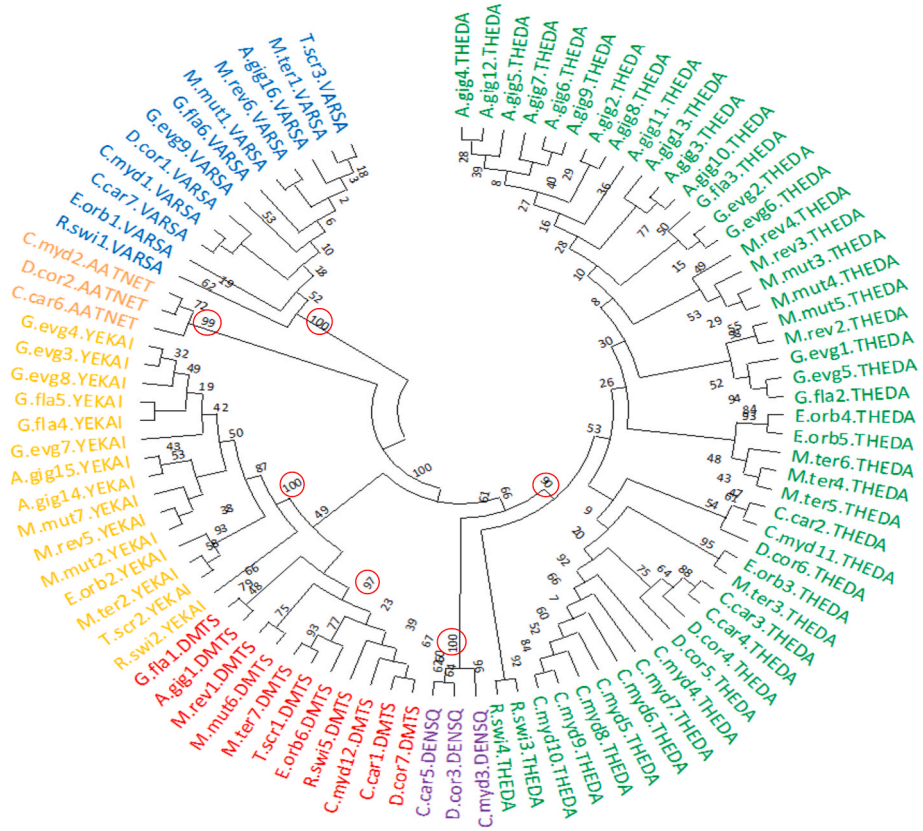


Fig. 3. Phylogenetic tree of the conserved domain in 91 complete cathelicidins obtained from genome mining of Testudines. The various cathelicidins were labelled with the first letter of the genus, the first three letters of the species, the occurrence number in the genome, and the name of the cathelicidin group to which each protein was classified. The colours indicated different types of cathelicidins: green for THEDA, violet for DENSQ, red for DMTS, yellow for YEKAI, orange for AATNET, and blue for VARSA. The numbers surrounded by red indicate the probability that the cathelicidins from each of the six distinct groups derived from a common ancestor. Tree obtained using MEGA software version 11.0.13 (Tamura et al., 2021) by Neighbor-joining method with 1000 bootstraps and JTT substitution model.

Additionally, VARSA, VARMA and QEAL-type cathelicidins feature a β-sheet composed of five β-strands rather than four. Typically, the sequence encoded by exon 1 gives rise to one β-strand, the sequence encoded by exon 2 gives rise to another β-strand, and the last two strands are encoded by exon 3. In the three types of cathelicidins with five β-strands, an additional strand is formed from the sequence encoded at the end of exon 1 and the beginning of exon 2 (Figs. 7C, 8C and 8E).

Furthermore, VARSA and VARMA-type cathelicidins exhibit several characteristics that distinguish them from other types. Firstly, while VARSA-type retain the typical structure of cathelicidins, their sequences yield a classification of ‘cystatin superfamily' in the conserved domain search server, unlike those of other cathelicidins, which are classified as part of the ‘cathelicidin superfamily'. In contrast, VARMA-type cathelicidins, which share a common ancestor with VARSA types (Fig. 5), are identified by this server as belonging to the ‘cathelicidin superfamily'. In addition to those differences in sequence and structure, their exon 4 presents alanine as a possible cleavage site instead of valine or isoleucine, as we observed in the other reptile cathelicidins. Phylogenetic analysis indicates that they are positioned more anciently and distantly from other cathelicidins (Figs. 3–5). These differences raise the question of whether they represent proteins with similar structures but different from cathelicidins. However, Shi et al. (2019b) studied a VARSA-type cathelicidin (designated by them as Cath6) and demonstrated that its RNA is expressed in response to bacterial infection, with the spleen being its primary site of expression. Furthermore, they tested the activity of the derived active peptide, which exhibited antimicrobial and immunomodulatory activity. All these findings appear to confirm that, despite their specific characteristics, VARSA-type cathelicidins function in a manner analogous to other cathelicidins. Moreover, their presence in most studied species (excluding snakes) suggests they fulfil an important function, as we did not observe gene-disrupting features in most VARSA copies, with the exception of a second, apparently defective, VARMA-type copy in *Podarcis muralis* (Fig. S2). Further research is warranted, as reptiles, being an evolutionarily ancient group of animals, could enhance our understanding of the origins of cathelicidins and their potential emergence from cystatin evolution (Zhu, 2008).

GALSA-type cathelicidins exhibit a distinctive extension in exon 3 caused by a downstream acceptor splice site. Whereas most reptile cathelicidins show the expected intron–exon boundary approximately 84 bp from the start of exon 3, GALSA-type genes splice further downstream, producing a substantially larger exon (117–267 bp). This region is enriched in glutamic acid residues, with occasional valine or lysine, and begins immediately after the fourth conserved cysteine of the cathelin domain; therefore, it does not interfere with the canonical β-sheet arrangement or cysteine spacing of the conserved domain (Table 3).

Structural predictions indicate that this elongated segment can adopt different conformations. In some sequences it appears as an extended, flexible region (Fig. 9A), whereas in others it forms a defined α-helix (Fig. 9B). In several cases, the extension bends toward the conserved domain, positioning the C-terminal peptide closer to the β-sheet core. Frequently, the peptide's C-terminal region is predicted to form an additional β-strand that aligns with the β-sheet of the cathelin domain (Fig. 9C), a type of interaction previously described for porcine protegrin (Sanchez et al., 2002b). Similar observations were also noted by Wang et al. (2008) in snake cathelicidins, where extended regions

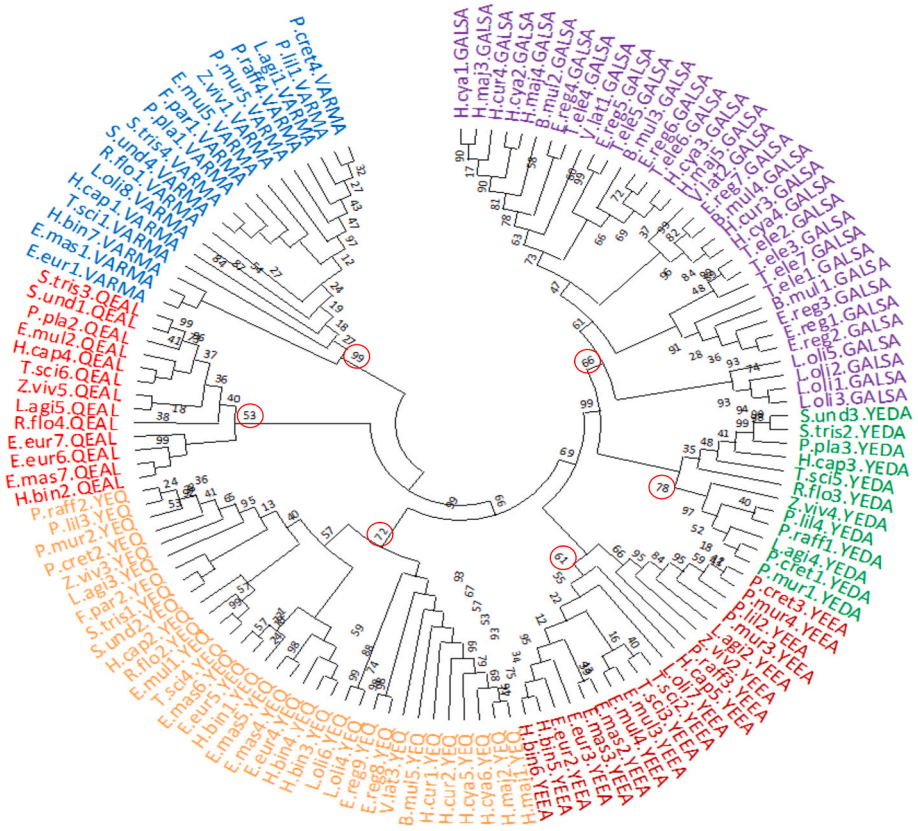


Fig. 4. Phylogenetic tree of the conserved domain in 128 complete cathelicidins obtained from genome mining of Squamata. The various cathelicidins were labelled with the first letter of the genus, the first three letters of the species, the occurrence number in the genome, and the name of the cathelicidin group to which each protein was classified. The colours indicated different types of cathelicidins: blue for VARMA, yellow for YEQ, red for QEAL, brown for YEEA, green for YEDA and purple for GALSA. The numbers surrounded by red indicate the probability that the cathelicidins from each of the 6 distinct groups derive from a common ancestor. Tree obtained using MEGA software version 11.0.13 (Tamura et al., 2021) by Neighbor-joining method with 1000 bootstraps and JTT substitution model.

modulated the structural arrangement of the precursor.

Not all GALSA-type cathelicidins possess this extension, as some maintain the shorter exon 3 typical of other reptile cathelicidins (Fig. 9D). Based on (i) the presence or absence of the exon-3 extension, (ii) its predicted conformation, and (iii) whether it interacts with the conserved domain, we tentatively classified GALSA sequences into four subgroups: (A) extended but non-structured regions, (B) α-helical extensions, (C) extensions forming a β-strand that pairs with the cathelin domain, and (D) sequences lacking the exon-3 extension (Table S3).

3.5. Defective cathelicidin genes

In many reptile cathelicidin genes, we detected anomalies that may compromise gene expression or protein function. Of the 287 genes identified, 68 were classified as putative defective (Table S4). The most common alterations were the loss of the initial ATG codon—frequently replaced by alternative start codons such as CTG—and the presence of premature stop codons within internal exons. Other defects included missing exons or sequence alterations in canonical splice-site motifs (5′ GT/3′ AG), which are essential for accurate intron removal in vertebrates (Matera and Wang, 2014; Smith et al., 1993). Disruption of these motifs would likely cause intron retention, frame shifts and termination at downstream stop codons.

Several exon-4 regions lacked a proper termination codon, resulting in peptides substantially longer than functional cathelicidins typically described in vertebrates (Kościuczuk et al., 2012). In one case (*Hydrophis curtus*.D1), the exon-1 sequence was anomalous and lacked a detectable signal peptide, an essential motif for secretion (Teufel et al., 2022).

Only copies that appear unlikely to be transcribed or processed correctly should be considered potential pseudogenes (Miyata and Hayashida, 1981). However, given the known variability and assembly limitations of reptile genomes—particularly in Squamata (Mongiardino Koch and Gauthier, 2018)—some defects may reflect technical artefacts rather than genuine loss of function. The higher number of defective copies observed in Testudines compared to Squamata may reflect lineage-specific differences in gene-family turnover, selective constraints or tolerance for the accumulation of nonfunctional duplicates.

3.6. Genomic organization of cathelicidin genes in reptiles

In Testudines, cathelicidin gene copy numbers were consistently high but varied markedly across species, ranging from five to nineteen genes (Table 5; Fig. S2). As in mammals and birds, these genes formed a cluster on the same chromosome (Cheng et al., 2015; Whelehan et al., 2014) —typically chromosome 2, except in *Rafetus swinhoei*, where the cluster was located on chromosome 1— and appeared in either the forward or reverse orientation depending on the species. The clusters frequently contained transposons and retroviral insertions, which were sometimes found between genes or even within cathelicidin exons (Fig. S2).

Flanking genes were identified as the nearest non-cathelicidin loci, excluding mobile elements. KLHL18, a regulator of mitotic entry (Shi et al., 2019a), consistently served as the downstream flanking gene (upstream in reverse-oriented clusters) and was located immediately adjacent to the VARSA-type cathelicidin. The upstream flanking gene (downstream in reverse-oriented clusters) showed greater variability. The most frequent upstream gene was Transforming Growth Factor Beta Regulator 4 (TBRG4), a member of the FAS serine/threonine kinase

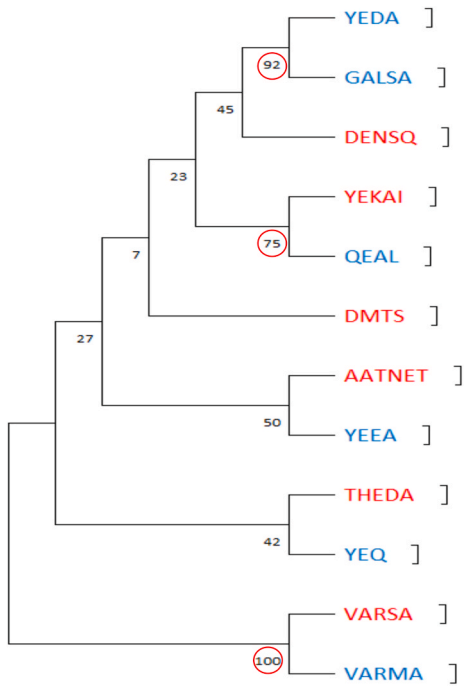


Fig. 5. Phylogenetic tree of the consensus sequences from the conserved domains of the twelve cathelicidin types in reptiles. The consensus sequences are labelled with the name of their type of cathelicidin. The numbers indicate the probability that the sequences grouped in the same branch of the tree are derived from a common ancestor. The cathelicidin-types belonging to the order Testudines are coloured red, while those belonging to Squamata are coloured blue. Numbers highlighted in red indicate a high probability. Tree obtained using MEGA software version 11.0.13 (Tamura et al., 2021) by Neighbor-joining method with 1000 bootstraps and JTT substitution model.

family (Zhang et al., 2022), although alternative flanking genes were observed in *Emys orbicularis* and *Malaclemys terrapin* (Table 5; Fig. S2). The upstream flanking region was generally located farther from the cluster than KLHL18 and often contained multiple transposon insertions. A comparable flanking-gene configuration has been described in most avian species (Cheng et al., 2015). In contrast, placental mammals show a different arrangement, with CDC25 and NME6 flanking the cluster and KLHL18 positioned further away, along with a 15-kDa leucocyte-protein-like gene outside the cluster but sharing a conserved domain with cathelicidins (Tossi et al., 2024).

Despite differences among species, Testudines exhibited a consistent organisational pattern that was broadly correlated with phylogeny, allowing us to summarise a generalised cluster structure based on gene types and relative positions (Fig. 10). In most species, the upstream region (near TBRG4) began with a DMTS-type cathelicidin, followed by a predicted defective THEDA-type gene. This arrangement was conserved across all species except *Rafetus swinhoei*, which lacked the defective THEDA-type copy (Fig. S2). Several THEDA-type copies followed, most commonly three, although species such as *Aldabrachelys gigantea* carried up to twelve copies, while *Rafetus swinhoei* possessed only two.

Downstream, an additional THEDA-type cathelicidin was typically present and predicted to be defective, except in marine turtles (Dermochelyidae and Cheloniidae) and in *Rafetus swinhoei*. A DENSQ-type cathelicidin followed, defective in most species but functional in sea turtles, and absent in Geoemydidae (Fig. S2). A larger genomic interval separated this gene from the YEKAI-type cathelicidin, which occurred in multiple copies in Testudinidae but was predicted defective in sea turtles. An AATNET-type gene appeared next in sea turtles and in Emydidae, where it was likely defective.

At the distal end of the cluster, the VARSA-type cathelicidin consistently appeared adjacent to KLHL18. This gene was approximately twice the size of other cathelicidins due to its large introns and retained both its integrity and genomic position across all species analysed. With the exception of VARSA and the flanking genes, which were always oriented opposite to the rest of the cluster, most cathelicidins shared the same transcriptional orientation within each species (Fig. 10).

Two notable exceptions deviated from the general Testudines pattern. First, *Gopherus evgoodei* exhibited a complete duplication of the segment spanning from the first THEDA-type gene to the second YEKAI-type gene (Fig. S2). Second, *Mauremys mutica* possessed an additional YEKAI-type cathelicidin outside the main cluster, approximately 5×10^6 bp downstream on the same chromosome. This gene was flanked by Thioredoxin 11, the same flanking gene observed for the cathelicidin cluster in the Iguania suborder of Squamata (Fig. S2).

In contrast to the more uniform organisation observed in Testudines, the arrangement of cathelicidin genes in Squamata showed extensive variation across suborders (Fig. S2). To reflect this diversity, clusters are described below by suborder.

Gekkota. Species of Gekkota generally carried seven cathelicidin genes. KLHL18 was the most consistent upstream flanking gene (Fig. 11A). Immediately downstream, a VARMA-type cathelicidin was positioned in the same orientation and exhibited a level of positional conservation comparable to the VARSA-type in Testudines. Two YEEA-type cathelicidins followed, and a cytochrome P450 gene was found within the cluster, sometimes present in multiple copies (Fig. 11A; Fig. S2). Further downstream, two YEQ-type cathelicidins were identified, followed by two QEAL-type genes. The downstream flanking gene was CDC25, involved in cell-cycle regulation (Lee et al., 2013), located ~300 kb from the last cathelicidin gene, with retroviral and

Fig. 6. Identity percentage diagram of the conserved domains among the six groups of cathelicidins in Testudines and Squamata. The alignment of the six consensus sequences for each of the proposed cathelicidin groups in (A) Testudines and (B) Squamata is presented. The colour blue represents the percentage of sequence identity, with darker shades of blue indicating 100% identity. The histogram at the top also represents the percentage of identity of the protein at each position in the sequence. Amino acids highlighted in red indicate the residues that stabilize the van der Waals interactions of the conserved domain described by Sanchez et al. (2002b). Figure obtained using Unipro UGENE 49.0 (Okonechnikov et al., 2012).

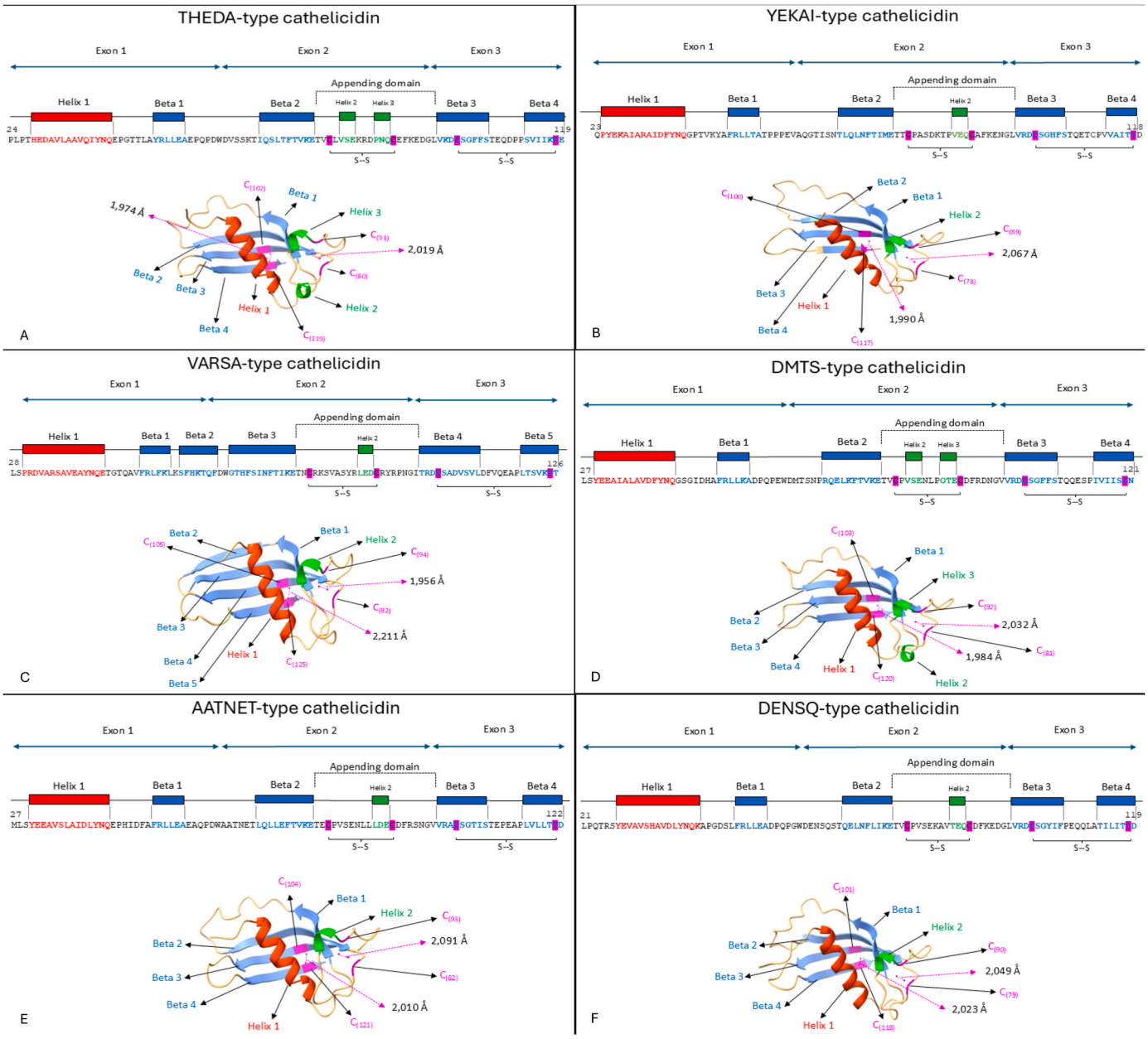


Fig. 7. Sequence and three-dimensional structure of the conserved domain of the six **groups of cathelicidins in Testudines.** For each cathelicidin group, the consensus sequence of the conserved domain is presented, with the sequence color-coded according to its arrangement in the three-dimensional structure. The α-helix is represented in red, β-strands in blue, short α-helices of the appending domain in green, and cysteines that stabilize the structure are highlighted in pink. The brackets beneath the cysteines in the sequence indicate the presence of disulfide bridges in the structure, while the numbers on either side of the sequence represent the positions of the first and last amino acids of the conserved domain in the cathelicidin. In the three-dimensional representation of each conserved domain, various conserved motifs are indicated. Additionally, the distance in angstroms between the sulphur atoms of the cysteines forming the disulfide bridges is noted. Figure based on Sanchez et al. (2002b). Three-dimensional models were generated using AlphaFold 3 (Abramson et al., 2024) and visualized with UCSF ChimeraX software version 1.7.1 (Pettersen et al., 2020).

mobile-element insertions between them (data not shown). In *Heterononia binoei*, however, the downstream flanking gene was TBRG4, as observed in Testudines (Fig. S2).

Laterata. Laterata shared a broadly similar organisation (Fig. 11B), with notable differences. Most species contained a single YEEA-type gene, except *Rhineura floridana*, which lacked it. One of the two YEQ-type cathelicidins (the forward-oriented copy) was replaced by a YEDA-type gene. A single QEAL-type gene was present, absent in Podarcis. *Podarcis muralis* also showed duplication of the cluster region spanning from cytochrome P450 to the VARMA-type gene (Fig. S2). Only occasional defective copies were detected in Gekkota and Laterata, contrasting with the higher number observed in Testudines; this likely reflects the smaller number of paralogues and stronger purifying selection in Squamata.

Serpentes. Serpentes displayed the greatest structural variability. Multiple GALSA-type cathelicidins (3–7 copies) occurred immediately downstream of KLHL18 (Fig. 11C), and many species also carried predicted defective copies (Fig. S2). Typically, one to three YEQ-type genes were present, although in some species, such as *Thamnophis elegans*, they were absent (Fig. S2). Importantly, this suborder lacked VARSA/VARMA-type cathelicidins entirely. *Liasis olivaceus* was exceptional, containing YEEA, VARMA and defective QEAL-type genes, consistent

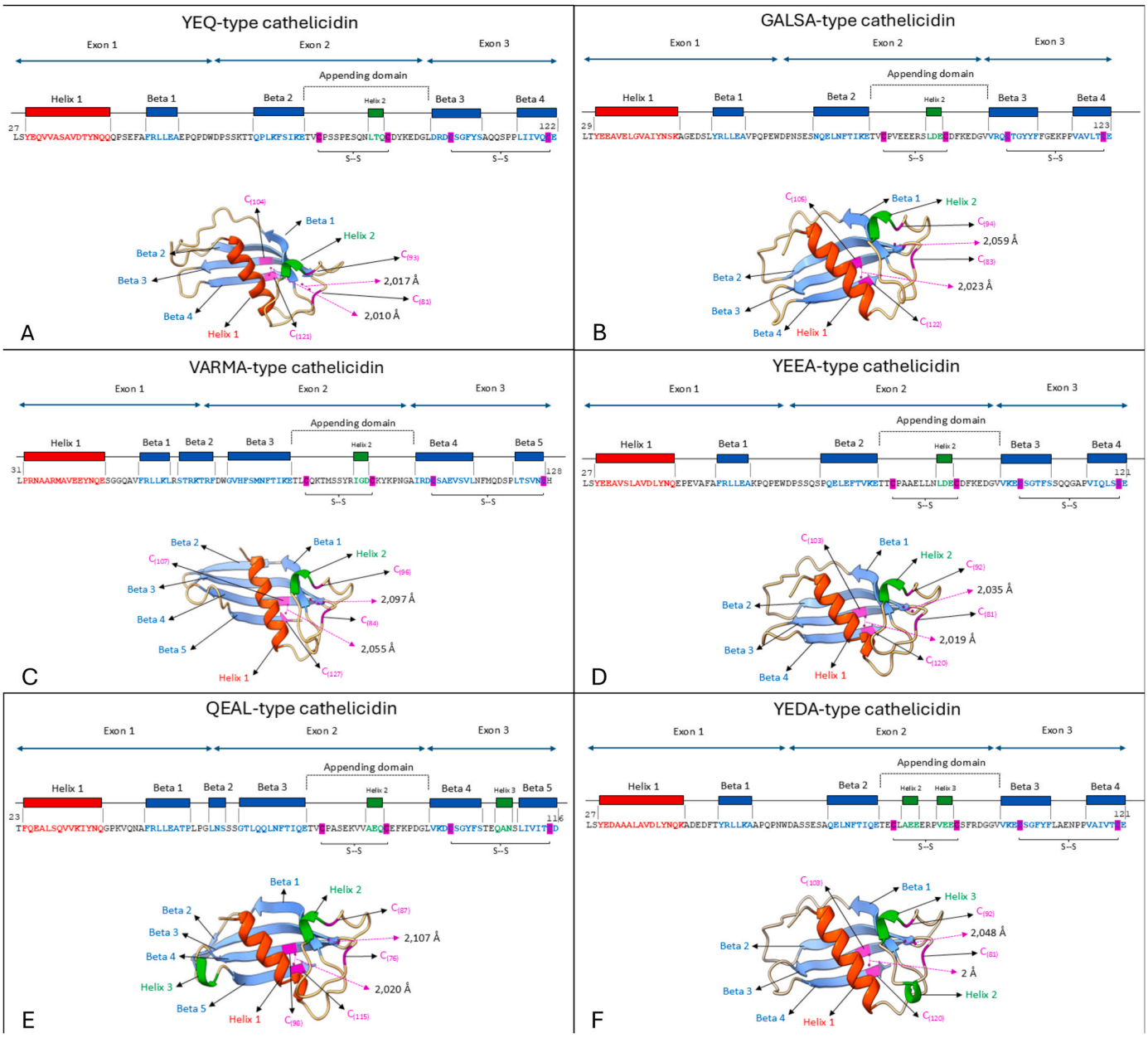


Fig. 8. Sequence and three-dimensional structure of the conserved domain of the six **groups of cathelicidins in Squamata.** For each cathelicidin group, the consensus sequence of the conserved domain is presented, with the sequence color-coded according to its arrangement in the three-dimensional structure. The α-helix is represented in red, β-strands in blue, short α-helices of the appending domain in green, and cysteines that stabilize the structure are highlighted in pink. The brackets beneath the cysteines in the sequence indicate the presence of disulfide bridges in the structure, while the numbers on either side of the sequence represent the positions of the first and last amino acids of the conserved domain in the cathelicidin. In the three-dimensional representation of each conserved domain, various conserved motifs are indicated. Additionally, the distance in angstroms between the sulphur atoms of the cysteines forming the disulfide bridges is noted. Figure based on Sanchez et al. (2002b). Three-dimensional models were generated using AlphaFold 3 (Abramson et al., 2024) and visualized with UCSF ChimeraX software version 1.7.1 (Pettersen et al., 2020).

with its phylogenetic position within Pythonidae, which diverged earlier than most snake families (Yin et al., 2016).

Scincomorpha, Anguimorpha and Iguania. Consensus clusters were not generated for these suborders due to the limited number of available genomes (one or two per group) and high structural variability. Nonetheless, their individual clusters are shown in Fig. S2. Scincomorpha resembled Laterata in general arrangement. In Anguimorpha and Iguania, KLHL18 and a VARMA-type gene were located separately by ~1 × 10^6 bp upstream or downstream of the cluster, depending on the species (Fig. S2). In *Elgaria multicarinata* (Anguimorpha), these genes were accompanied by two YEEA-type copies. The downstream flanking gene CDC25 was retained only in *Furcifer pardalis* (Iguania), while in other species it was replaced by Thioredoxin 11 (Table 5; Fig. S2). Most species contained combinations of QEAL, YEQ and YEDA types, although gene order varied even among closely related species (e.g., *Sceloporus undulatus* vs. *S. tristichus*; Mongiardino Koch and Gauthier, 2018), reflecting the high genomic plasticity of Iguania.

Chromosomal distribution. The chromosomal location of the cathelicidin cluster was generally conserved within each suborder (Fig. 11), with some exceptions (Table 5). In most snakes, the cluster was located on the Z chromosome, whereas in *Liasis olivaceus* it was on chromosome 4. In *Thamnophis elegans*, six GALSA-type genes occurred between CDC25 and KLHL18 on chromosome Z, with an additional copy on chromosome 4, flanked by microneme/rhoptry antigen genes. As in

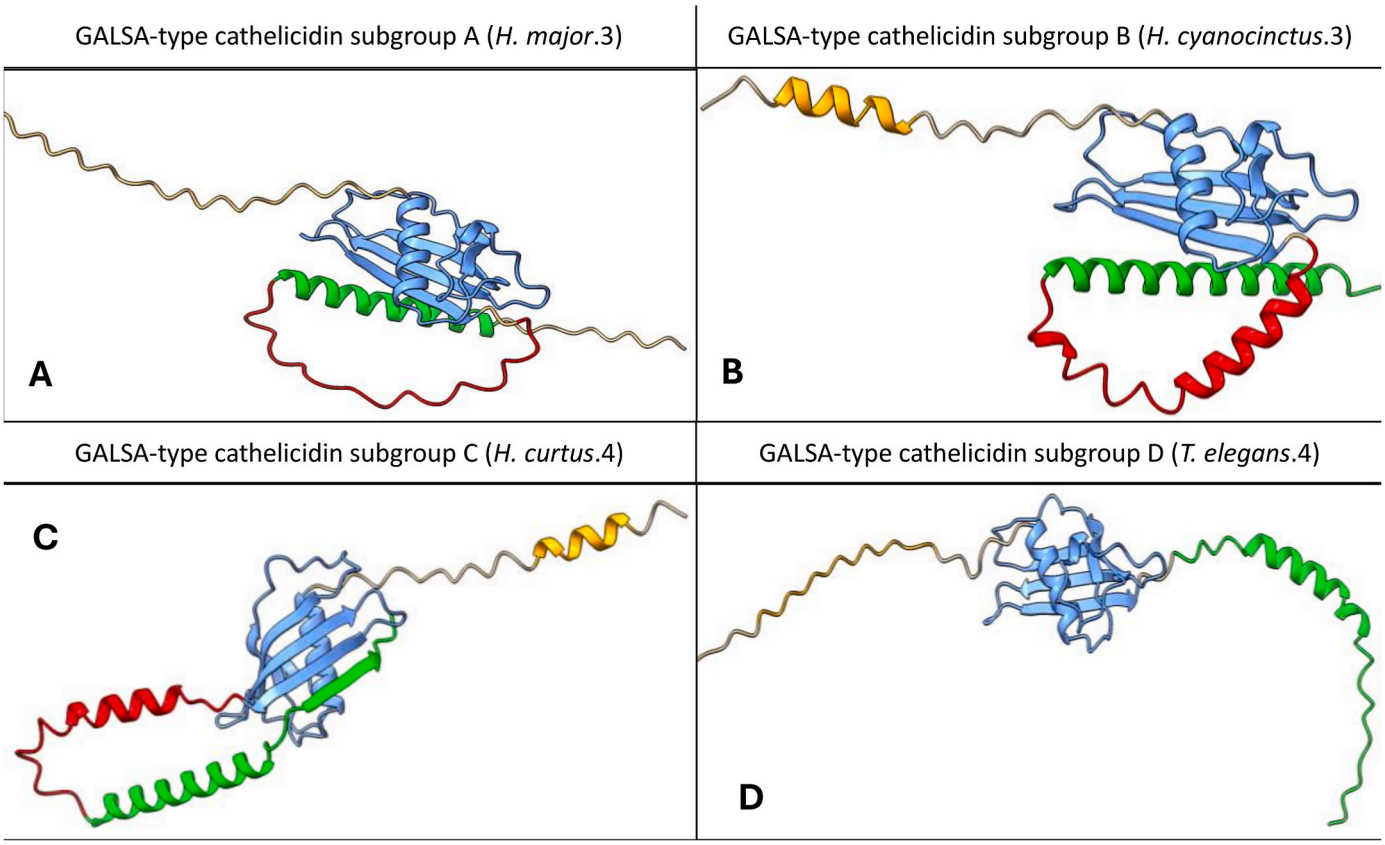


Fig. 9. Sequence and three-dimensional structure of four complete GALSA-type cathelicidins. The represented GALSA are part of each of the four subgroups within this type of cathelicidin. The various regions of the protein are colour-coded: orange for the signal peptide, blue for the β-sheet, and green for the active peptide. Three of the four cathelicidins exhibit an extension at the end of the exon 3 (represented in red), which forms an extended structure (A), an α-helix (B) or (C) facilitates interaction between the peptide and the β-sheet. The remaining GALSA-subgroup (D) does not possess this sequence elongation. Three-dimensional models were generated using AlphaFold 3 (Abramson et al., 2024) and visualized with UCSF ChimeraX software version 1.7.1 (Pettersen et al., 2020).

Testudines, several species contained retroviral insertions within or near the cluster (Fig. S2).

VARSA/VARMA locus and evolutionary implications. The consistent presence of the VARMA-type cathelicidin locus adjacent to KLHL18 across Squamata and its mirroring location (VARSA) in Testudines (Figs. 10 and 11, S2) prompted investigation into its wider evolutionary conservation. BLAST analysis using the conserved domain sequence of the *Caretta caretta* VARSA-type cathelicidin further revealed that the results not only showed high homology with VARSA cathelicidins from other turtles and VARMA types from some Squamata but also included hits with a high query cover (approx. 98%) yet a low identity (approx. 45%) to uncharacterised proteins from several bird species, such as *Gallus gallus* (XP_046789495.1) and *Athene noctua* (XP_074754302.1). Intriguingly, these sequences are sometimes predicted by Gnomon as 'cathelicidin-related antimicrobial peptides' (like *Onychostruthus taczanowskii* (XP_041264461.1) and are consistently located at the same genomic locus as the reptilian VARMA/VARSA genes (adjacent to the flanking KLHL18 gene). This uncharacterised avian protein aligns with a sequence previously reported as a 'sequence related to the cathelicidin cluster (srcc)' (González-Acosta et al., 2025). This is consistent with the possibility that it represents a cathelicidin-related gene, potentially sharing an ancestral origin with the reptilian VARSA/VARMA types. This is especially relevant considering that the VARSA type cathelicidin expression, as well as its antimicrobial and immunomodulatory activities, have already been confirmed by PCR and other *in vitro* assays (Shi et al., 2019b). Whether this avian cathelicidin related sequence is functional remains uncertain, necessitating further investigation to clarify their evolutionary origin.

3.7. Classification of cathelicidin derived peptides

Of the 287 cathelicidin genes identified through genome mining, 246 peptides were generated by cleaving at the first valine, isoleucine, or alanine within exon 4, as is accepted for cathelicidins (Anderson and Yu, 2003; Shinnar et al., 2003; Wilkins and Kurtz, 2023). Of these, 212 are unique sequences, as some species possess multiple gene copies with identical exon 4, and different species can share the same gene, resulting in identical peptides. The peptides ranged in size from 15 to 62 amino acids (Table S5), predominantly exhibiting cationic character and an α-helical secondary structure. This article proposes a classification of the peptides based on sequence, structure, and charge, extending the framework established by Koehbach and Craik, 2019 (Tables 6 and 7, S5).

Peptides were classified based on their structure, as α-helices, helical hairpins (pairs of α-helices linked by a β-turn), or extended forms. Within these categories, cationic peptides (charge 1–9) and super-cationic peptides (charge >9) were distinguished. Some peptide sequences were noted for specific amino acid enrichments, leading to classifications such as lysine-arginine rich, proline-arginine rich, arginine rich, lysine-proline rich, or lysine rich (Tables 6 and 7, S5).

The classification of cathelicidins is closely associated with their respective peptide types. In Testudines, THEDA-type cathelicidins were predicted to form cationic α-helical peptides of three distinct categories: type I, characterised by cationic α-helices; type II, consisting of cationic hairpin α-helices; and type III, which encompasses super-cationic hairpin α-helices. In contrast, YEKAI-type cathelicidins consistently yielded cationic α-helices, representative of type I peptides. AATNET-type cathelicidins were notable for producing α-helices that are

Table 5

Summary of the organisation of cathelicidin clusters from the order Testudines. The analysed turtle species are presented, along with the NCBI assembly code of the genome employed for the mining, the chromosome in which the cluster was found, and the total number of cathelicidin genes that were identified. Genes that are apparently capable of giving rise to complete and functional cathelicidins are listed in the 'complete cathelicidins' column. In contrast, genes that exhibit some form of error that may lead to a loss of expression, structure, or function of the cathelicidin are listed in the 'defective cathelicidins' column. Finally, for the species *Trachemys scripta*, it was not possible to accurately identify a gene due to sequencing errors in that genome (presence of gaps), and it was thus annotated in the 'partial cathelicidins' column. Flanking genes on both sides of the cluster are also indicated. The abbreviation TBRG4 corresponds to Transforming growth factor β-regulator 4, KLHL18 to Kelch-like Protein 18 and CDC25 to Cell Division Cycle 25.

Species	Genome assembly	Chromosome	Complete cathelicidins	Partial cathelicidins	Defective cathelicidins	Total genes	Upstream flanking gene	Downstream flanking gene
Order Testudines								
Testudinidae								
Aldabrachelys gigantea	AldGig_1.0	2	16	–	3	19	TBRG4	KLHL18
Gopherus evgoodei	rGopEvg1_v1.p	2	9	–	6	15	TBRG4	KLHL18
Gopherus flavomarginatus	rGopFla2.mat.asm	2	6	–	3	9	TBRG4	KLHL18
Geoemydidae								
Mauremys mutica	ASM2049712v1	2	7	–	4	11	KLHL18	TBRG4
Mauremys reevesii	ASM1616193v1	2	6	–	2	8	TBRG4	KLHL18
Emydidae								
Emys orbicularis	rEmyOrb1.hap1	2	6	–	5	11	KLHL18	Seven-transmembrane G protein-coupled receptor (7TM-GPCR)
Malaclemys terrapin	rMalTer1.hap1	2	7	–	6	13	KLHL18	Class III nucleotidyl cyclase (Nucleotidyl_cyc_III)
Trachemys scripta	CAS_Tse_1.0	2	3	1	4	8	TBRG4	KLHL18
Dermochelyidae								
Dermochelys coriacea	rDerCor1.pri.v4	2	7	–	2	9	KLHL18	Transforming growth factor beta regulator 4 (TBRG4)
Cheloniidae								
Caretta caretta	GSC_CCare_1.0	2	7	–	2	9	TBRG4	KLHL18
Chelonia mydas	rCheMyd1.pri.v2	2	12	–	2	14	KLHL18	TBRG4
Trionychidae								
Order Squamata								
Rafetus swinhoei	ASM1942577v1	1	5	–	0	5	KLHL18	TBRG4
Eublepharidae								
Eublepharis macularius	*MPM_Emac_v1.0*	11	7	–	1	8	KLHL18	CDC25
Sphaerodactylidae								
Euleptes europaea	rEulEur1.hap1	11	7	–	–	7	KLHL18	CDC25
Gekkonidae								
Heteronotia binoei	APGP_CSIRO_Hbin_v1	10	6	–	1	7	TBRG4	KLHL18
Scincidae								
Tliqua scincoides	rTilSci1.hap2	5	6	–	–	6	KLHL18	CDC25
Cordylidae								
Hemicordylus capensis	rHemCap1.1.pri	6	5	–	–	5	KLHL18	CDC25
Lacertidae								
Lacerta agilis	rLacAgi1.pri	12	5	–	–	5	KLHL18	CDC25
Zootoca vivipara	rZooViv1.1	12	5	–	–	5	KLHL18	CDC25

(*continued on next page*)

Table 5 (*continued*)

Species	Genome assembly	Chromosome	Complete cathelicidins	Partial cathelicidins	Defective cathelicidins	Total genes	Upstream flanking gene	Downstream flanking gene
Podarcis lilfordi	rPodLil1.2	12	4	–	–	4	KLHL18	CDC25
Podarcis cretensis	rPodCre2.1	12	4	–	–	4	CDC25	KLHL18
Podarcis muralis	PodMur_1.0	12	5	–	1	6	CDC25	KLHL18
Podarcis raffonei	rPodRaf1.pri	12	4	–	–	4	KLHL18	CDC25
Rhineuridae								
Rhineura floridana	rRhiFlo1.hap2	10	4	–	–	4	KLHL18	CDC25
Anguidae								
Elgaria multicarinata	rElgMul1.1.pri	1	5	–	2	7	Blood Coagulation Factor FA58C	Nucleoside Triphosphate Hydrolase
Phrynosomatidae								
Sceloporus undulatus	SceUnd_v1.1	6	4	1	–	5	Thioredoxin 11	104 kDa Microneme/rhoptry antigen
Sceloporus tristichus	ASM1680106v1	6	4	–	1	5	Thioredoxin 11	104 kDa Microneme/rhoptry antigen
Phrynosoma platyrhinos	MUOH_PhPlat_1.1	6	3	–	2	5	104 kDa Microneme/rhoptry antigen	Thioredoxin 11
Chamaeleonidae								
Furcifer pardalis	ASM3044067v1	8	2	–	2	4	KLHL18	CDC25
Pythonidae								
Liasis olivaceus	rLiaOli1.hap1	4	8	–	1	9	CDC25	KLHL18
Viperidae								
Vipera latastei	rVipLat1.pri	Z	3	–	3	6	CDC25	KLHL18
Colubridae								
Erythrolamprus reginae	rEryReg1.hap1	Z	9	–	3	12	CDC25	KLHL18
Thamnophis elegans	rThaEle1.pri	Z/4	7	–	–	7	CDC25	KLHL18
Elapidae								
Bungarus multicinctus	ASM2365372v1	4	5	–	2	7	CDC25	KLHL18
Hydrophis curtus	HCur_v2	Z	4	–	3	7	KLHL18	CDC25
Hydrophis cyanocinctus	HCya_v2	Z	6	–	2	8	CDC25	KLHL18
Hydrophis major	HMAJ_1.0	Z	5	–	4	9	KLHL18	CDC25

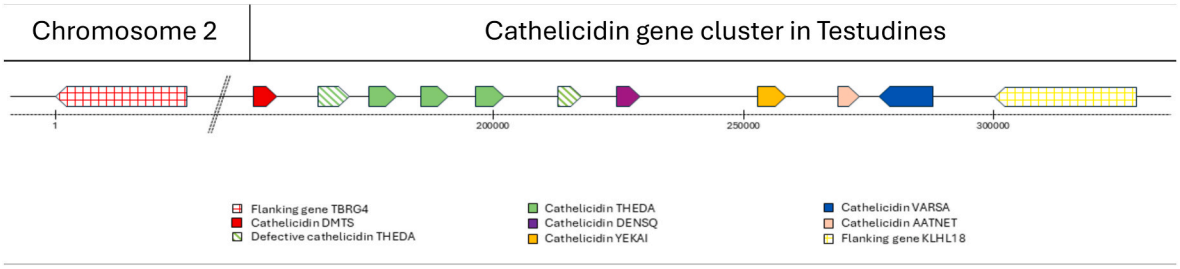


Fig. 10. Cathelicidin cluster organization in the order Testudines. This figure presents a comprehensive representation of the organization of the cathelicidin cluster in the order Testudines. It includes all cathelicidin types, as well as the modal number of copies of each type, the average size of each cathelicidin gene and its flanking genes, and the average distance between genes. Genes are represented by arrowheads, indicating their orientation within the DNA. Flanking genes are shaded with squares, and each type of cathelicidin is represented by a solid colour. Putatively defective genes are depicted with stripes, with the colour corresponding to their specific type. Figure was generated using the 'ggplot2' and 'gggenes' packages in R software version 4.4.1 (Wilkins and Kurtz, 2023).

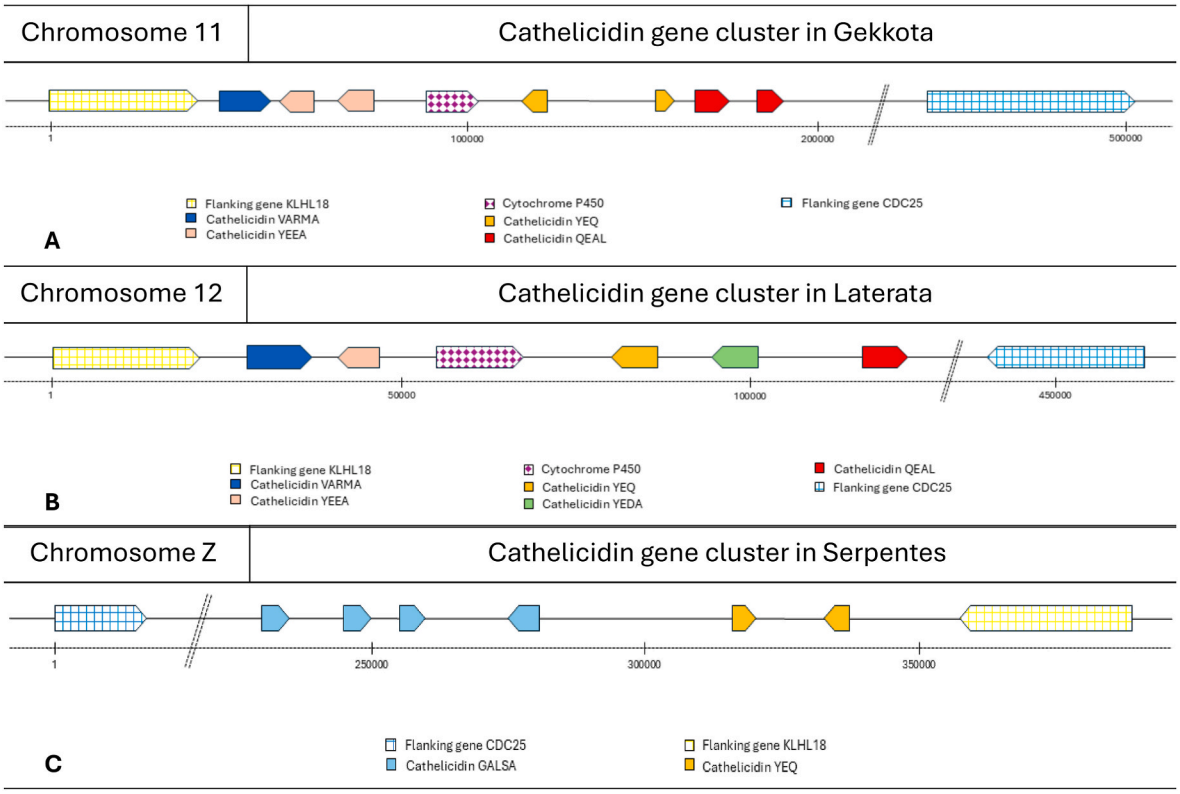


Fig. 11. Cathelicidin cluster organization in the order Squamata. This figure presents a comprehensive representation of the organization of the cathelicidin cluster in three suborders of Squamata: (A) Gekkota, (B) Laterata, and (C) Serpentes. It includes all cathelicidin types, as well as the modal number of copies of each type, the average size of each cathelicidin gene and its flanking genes, and the average distance between genes. Genes are represented by arrowheads, indicating their orientation within the DNA. Flanking genes are shaded with squares, and each type of cathelicidin is represented by a solid colour. Putatively defective genes are depicted with stripes, with the colour corresponding to their specific type. Figure generated using the 'ggplot2' and 'gggenes' packages in R software version 4.4.1 (Wilkins and Kurtz, 2023).

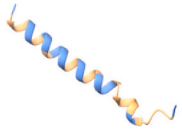
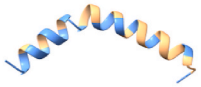
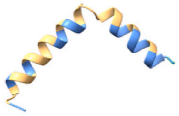
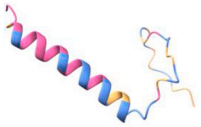
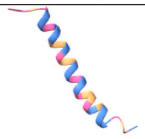
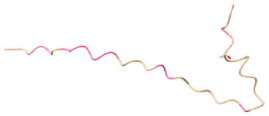
enriched in lysine and arginine, classified as type IV peptides. Moreover, DENSQ-type cathelicidins were characterised by their arginine-rich α-helical structures, classified as type V peptides. Conversely, both DMTS-type and VARSA-type cathelicidins produced extended structure peptides, with DMTS-type peptides being rich in proline and arginine (type VI), and VARSA-type peptides are rich in proline and lysine (type VII) (Table 6, S5).

In the order Squamata, YEQ-type cathelicidins were predicted to generate peptides type I, II, and III. In contrast, QEAL-type cathelicidins exclusively yield peptides type I, while YEEA and GALSA-type cathelicidins are associated with the production of type IV peptides and can also lead to the formation of arginine-lysine-rich super cationic helices (type VIII) peptides (Table 7, S5).

Finally, VARMA-type cathelicidins display a peptide structure (type IX) that is markedly distinct from those identified in other reptiles, characterised by an extended structure with a slight anionic charge (Table 7). Moreover, these peptides are particularly enriched in polar uncharged amino acid residues, especially serine and asparagine.

In summary, reptiles may exhibit nine distinct types of peptides based on their sequence, charge, and structure. Types I, II, III, and IV peptides are common to both orders of reptiles, while types V, VI, and VII are exclusive to Testudines, and types VIII and IX are exclusive to Squamata (Tables 6 and 7). It is noteworthy that cationic peptides Type V (Testudines) and Type VIII (Squamata) are structurally similar, differing primarily in their predominant cationic amino acid residues—arginine and lysine, respectively. This similarity may indicate that

Table 6
Description of the main types of cathelicidin-derived peptides present in Testudines. Each proposed type of peptide is assigned a number as follows: (I) Cationic helix, (II) Cationic hairpin, (III) Super cationic hairpin, (IV) Arginine-lysine rich super cationic helix, (V) Arginine rich super cationic helix, (VI) Proline-arginine rich extended structure, and (VII) Proline-lysine rich extended structure. The consensus sequence for each type of peptide is also presented, with the region forming an α-helix in bold and underlined. Finally, the charge of each of these sequences and their three-dimensional structures is displayed. In the latter, the most abundant apolar amino acid residues in this type of peptide (leucine, isoleucine, phenylalanine, alanine, and valine) are marked in blue, while proline, lysine, and arginine residues in the peptides rich in these amino acids are indicated in pink. Three-dimensional models of peptides were generated using AlphaFold 3 (Abramson et al., 2024) and visualized with UCSF ChimeraX software version 1.7.1 (Pettersen et al., 2020).

Peptide type	Sequence	Charge	Structure
I	VTRS**RVTRWWPGLVRRGAWKILGHGISFFRKR**G	9	
II	VT**RGRWKRFWRGTGNFFRR**H**GLNILATGLKII**G	8	
III	VT**RGRWGRWRRKAGRFIRR**N**RWKIIRAGLKILL**G	13	
IV	RRKS**KFKKRRKKIKKGLKKVIQHLL**AGGRQGLPQGGRPGMISPGLPPH	16	
V	RRS**RFGRFFKKVRKQLGRVLRHSRITLGG**RMRF	13	
VI	RRSRSPRRSWWPRRWHLPGSYTLIAHGGRKGKGKGKGSRLQVA	13	
VII	RAKKPSKKPKPKAQTVPKVTVEYFPSSFSTAALTAPEDG	5	

both types share a common target and were evolutionarily selected in these suborders. Hb_PR34 from *Heteronotia binoei* exhibits a unique structure that does not correspond to any of the previously established groups, characterized by an extended structure rich in glycine and arginine (Table S5). Due to putatively defective genes escaping evolutionary pressure when they cease to be expressed (Miyata and Hayashida, 1981), the peptides that we predicted from these genes have often exhibit certain variations, such as reduced or excessive size, or the addition of anomalous sequences in comparison to other peptides of the same type (for example, Fp_FA28 from *Furcifer pardalis*).

This peptide diversity likely reflects multiple immune functions; for instance, proline and arginine-rich peptides may inhibit protein translation by binding to and inhibiting bacterial ribosomes (Gagnon et al., 2016), while α-helical peptides mostly interact with biological membranes (Sun et al., 2016).

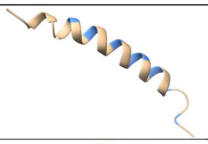
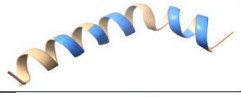
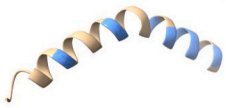
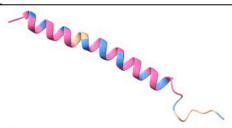
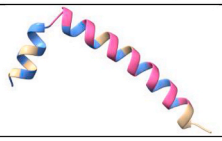
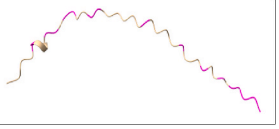
3.8. Evolutionary and functional insights from reptilian cathelicidin diversity

This study provides the most comprehensive genomic overview of reptilian cathelicidins to date, revealing an unexpectedly high diversity of gene types, copy numbers and genomic organisations in Testudines and Squamata. The identification of twelve conserved types, each defined by short diagnostic motifs within the cathelin domain and supported by exon structure and synteny, highlights that reptile cathelicidins have undergone substantial lineage-specific diversification. This contrasts with the more reduced repertoires reported in birds (González-Acosta et al., 2025) and mammals (Kościuczuk et al., 2012), suggesting that the common sauropsid ancestor may have possessed a broader cathelicidin set that was subsequently streamlined in the avian lineage.

Differences across reptilian lineages provide insight into the evolutionary pressures acting on this gene family. Turtles retain multiple

Table 7

Description of the main types of cathelicidin-derived peptides present in Squamata. Each proposed type of peptide is assigned a number as follows: (I) Cationic helix, (II) Cationic hairpin, (III) Super cationic hairpin, (IV) Arginine-lysine rich super cationic helix, (VIII) Lysine rich super cationic helix and (IX) Polar uncharged amino acids rich anionic extended structure. The consensus sequence for each type of peptide is also presented, with the region forming an α-helix in bold and underlined. Finally, the charge of each of these sequences and their three-dimensional structures is displayed. In the latter, the most abundant apolar amino acid residues in this type of peptide (leucine, isoleucine, phenylalanine, alanine, and valine) are marked in blue, while proline, lysine, and arginine residues in the peptides rich in these amino acids are indicated in pink. Polar uncharged amino acids are also marked in pink in peptide type IX. Three-dimensional models of peptides were generated using AlphaFold 3 (Abramson et al., 2024) and visualized with UCSF ChimeraX software version 1.7.1 (Pettersen et al., 2020).

Peptide type	Sequence	Charge	Structure
I	RVTRFRGLGGLLRGFGRGFGRLWGSNNNN	6	
II	TRHRWKRAVRKVGRFVRRYGPLIATVVGG	9	
III	TRGRWRRFWRKTKRFVKKHGVSIALAALRLFG	11	
IV	KRFKKLRKRVRRGGFRKIVKKLKKKKLKKKFRKPLPGSGSA	22	
VIII	KRSWKFKKFFKKIKKGVKKRFKKFFKKKPIVIGVSIRF	18	
IX	NSGNNKNKRNARPQAMEAAQPRIFVEQYFPSSYSTAALTAPEEE	0	

types—such as DENSQ and AATNET—that are absent in squamates and multiple copies of THEDA type, whereas snakes show exclusive presence of GALSA-type cathelicidins and a loss of VARSA/VARMA types. These contrasting patterns may reflect ecological specialization, differences in pathogen exposure, or distinct immunophysiological demands. For example, GALSA-type peptides, previously associated with venom glands and strong antimicrobial activity (Falcao et al., 2014), are present exclusively in snakes, consistent with the unique microbial and tissue environments linked to venom systems. Conversely, turtles exhibit high copy numbers and frequent gene duplications, which may reflect long-term conservation of multiple paralogues in stable ecological niches (Nash and Ryan, 2023).

The genomic organisation also reveals shared structural themes across reptiles. In both Testudines and Squamata, cathelicidin genes form clusters flanked by conserved regulatory genes such as KLHL18, TBRG4 and CDC25, a pattern also observed in birds (Cheng et al., 2015). This conserved synteny suggests deep evolutionary stability in the genomic positioning of the cluster, despite considerable turnover in gene number and composition. Evidence for frequent insertions of retroelements and transposons within and around clusters—especially in turtles—may further contribute to the birth-and-death dynamics characteristic of antimicrobial peptide families (Hanson et al., 2019).

The presence of numerous defective or truncated copies, particularly in Testudines, is consistent with this model. However, only copies that are unlikely to be expressed or properly processed should be considered potential pseudogenes (Miyata and Hayashida, 1981). Some defects may instead reflect assembly limitations typical of reptile genomes (Mongiardino Koch and Gauthier, 2018), highlighting the need for improved sequencing quality and transcriptomic validation.

The classification of mature peptides into nine groups reveals additional functional diversity. Variation in net charge, hydrophobicity and structural predictions—especially the distinctive exon-3 extension found in GALSA-type genes—points to a wide spectrum of antimicrobial, immunomodulatory or context-specific activities (Gallo and Hooper, 2012; Steinstraesser et al., 2011). This framework provides a basis for future structure–function studies, comparative immunology and biochemical characterisation.

4. Conclusions

Our genome mining approach, combined with manual refinement of exon predictions, enabled the identification of 287 putative cathelicidin genes, most of which were previously unreported. This manual curation was essential in cases where splicing signals diverged from the common GTGAG or GTAAG motifs, a pattern particularly frequent in Squamata.

This study proposes the first classification of reptilian cathelicidins into distinct types, each defined by a characteristic motif within the cathelin domain and supported by exon size and genomic context. This framework provides a practical basis for future investigations of tissue-specific expression, functional diversification and lineage-specific

patterns, such as the apparent restriction of AATNET and DENSQ types to sea turtles and the exclusive occurrence of GALSA in snakes.

We also examined their genomic organisation and found that cathelicidin genes occur in clusters with substantial variation in copy number and arrangement across species. In addition, the diversity of reptilian cathelicidins enabled the classification of their predicted peptides into nine groups, which may support future structure–activity analyses and help determine whether these types play distinct roles in innate immune defence.

Since all results are based on bioinformatic predictions from single genome assemblies, experimental validation of gene expression, proteolytic processing and antimicrobial activity will be essential in future studies. Although the classification system defined here is robust across Testudines and Squamata, extending it to other reptilian orders (e.g., Crocodylia or Sphenodontia) will require additional high-quality genomes. Such comparative analyses will clarify whether the twelve types identified here reflect the ancestral reptile repertoire or whether additional lineage-specific variants remain to be discovered.

CRediT authorship contribution statement

Andrea Otazo-Pérez: Writing – review & editing, Writing – original draft, Visualization, Methodology, Investigation, Data curation, Conceptualization. **Manuel R. López:** Writing – review & editing, Writing – original draft, Visualization, Supervision, Methodology, Conceptualization. **Sergio González-Acosta:** Writing – original draft. **Antonio Morales-delaNuez:** Writing – original draft, Methodology, Conceptualization. **José Manuel Pérez de la Lastra:** Writing – review & editing, Writing – original draft, Visualization, Supervision, Funding acquisition, Conceptualization.

Funding

This research was conducted in the framework of the project ProID2020010134," funded by the Canarian Agency for Research, Innovation and Information Society (Agencia Canaria de Investigación, Innovación y Sociedad de la Información, ACIISI) of the Government of the Canary Islands, Spain.

Conflict of interest

The authors declare no conflicts of interest.

Acknowledgments

Andrea Otazo-Pérez is recipient of pre-doctoral fellowships from the Canary Islands Agency for Research, Innovation, and the Information Society of the Department of Universities, Science, Innovation, and Culture, and by the European Social Fund Plus (ESF+) Integrated Operational Programme of the Canary Islands 2021–2027, Priority Axis 3 Topic 74 (85%).

Manuel R. López is a beneficiary of the contract funded by the Honourable Cabildo of Tenerife for the development of the 'TALENTUM INNOVATION CSIC' programme.

We acknowledge support of the publication fee by the CSIC Open Access Publication Support Initiative through its Unit of Information Resources for Research (URICI).

We are grateful to CEAMED S.A for providing support in carrying out this research under the agreement with IPNA-CSIC.

We would like to express our gratitude to Dr. Mariano Nicolás Hernández Ferrer for his assistance in the development and interpretation of the phylogenetic trees, which facilitated the establishment of the various types of cathelicidins present in the groups of reptiles studied in this article.

Appendix A. Supplementary data

Supplementary data to this article can be found online at https://doi.org/10.1016/j.dci.2026.105560.

Data availability

Data will be made available on request.

References

Abramson, J., Adler, J., Dunger, J., Evans, R., Green, T., Pritzel, A., Jumper, J.M., 2024. Accurate structure prediction of biomolecular interactions with AlphaFold 3. Nature 630 (8016), 493–500. https://doi.org/10.1038/s41586-024-07487-w.

Agier, J., Efenberger, M., Brzezińska-Blaszczyk, E., 2015. Cathelicidin impact on inflammatory cells. Cent. Eur. J. Immunol. 40 (2), 225–235. https://doi.org/10.5114/ceji.2015.51359.

Altschul, S.F., Gish, W., Miller, W., Myers, E.W., Lipman, D.J., 1990. Basic local alignment search tool. J. Mol. Biol. 215 (3), 403–410. https://doi.org/10.1016/S0022-2836(05)80360-2.

Anderson, R.C., Yu, P.L., 2003. Isolation and characterisation of proline/arginine-rich cathelicidin peptides from ovine neutrophils. Biochem. Biophys. Res. Commun. 312 (4), 1139–1146. https://doi.org/10.1016/j.bbrc.2003.11.045.

Andrault, P.M., Samsonov, S.A., Weber, G., Coquet, L., Nazmi, K., Bolscher, J.G., Lalmanach, A.C., Jouenne, T., Brömme, D., Pisabarro, M.T., Lalmanach, G., Lecaille, F., 2015. Antimicrobial peptide LL-37 is both a substrate of cathepsins S and K and a selective inhibitor of Cathepsin L. Biochemistry 54 (17), 2785–2798. https://doi.org/10.1021/acs.biochem.5b00231.

Asensio-Calavia, P., González-Acosta, S., Otazo-Pérez, A., López, M.R.R., Morales-delaNuez, A., Pérez de la Lastra, J.M., 2023. Teleost Piscidins—In silico perspective of natural peptide antibiotics from marine sources. Antibiotics 12 (5), 855. https://doi.org/10.3390/antibiotics12050855.

Ballard, E., Yucel, R., Melchers, W.J.G., Brown, A.J.P., Verweij, P.E., Warris, A., 2020. Antifungal activity of antimicrobial peptides and proteins against Aspergillus fumigatus. J. Fungi 6 (2). https://doi.org/10.3390/jof6020065.

Biniossek, M.L., Nägler, D.K., Becker-Pauly, C., Schilling, O., 2011. Proteomic identification of protease cleavage sites characterizes prime and non-prime specificity of cysteine cathepsins B, L, and S. J. Proteome Res. 10 (12), 5363–5373. https://doi.org/10.1021/pr200621z.

Burge, C.B., Karlin, S., 1998. Finding the genes in genomic DNA. Curr. Opin. Struct. Biol. 8 (3), 346–354. https://doi.org/10.1016/s0959-440x(98)80069-9.

Cheng, Y., Prickett, M.D., Gutowska, W., Kuo, R., Belov, K., Burt, D.W., 2015. Evolution of the avian -defensin and cathelicidin genes. BMC Evol. Biol. 15. https://doi.org/10.1186/s12862-015-0465-3.

Crooks, G.E., Hon, G., Chandonia, J.M., Brenner, S.E., 2004. WebLogo: a sequence logo generator. Genome Res. 14 (6), 1188–1190. https://doi.org/10.1101/gr.849004.

Falcao, C.B., de La Torre, B.G., Pérez-Peinado, C., Barron, A.E., Andreu, D., Rádis-Baptista, G., 2014. Vipericidins: a novel family of cathelicidin-related peptides from the venom gland of South American pit vipers. Amino Acids 46 (11), 2561–2571. https://doi.org/10.1007/s00726-014-1801-4.

Gallo, R.L., Hooper, L.V., 2012. Epithelial antimicrobial defence of the skin and intestine. Nat. Rev. Immunol. 12 (7), 503–516. https://doi.org/10.1038/nri3228.

Gagnon, M.G., Roy, R.N., Lomakin, I.B., Florin, T., Mankin, A.S., Steitz, T.A., 2016. Structures of proline-rich peptides bound to the ribosome reveal a common mechanism of protein synthesis inhibition. Nucleic Acids Res. 44 (5), 2439–2450. https://doi.org/10.1093/nar/gkw018.

González-Acosta, S., Baca-González, V., Asensio-Calavia, P., Otazo-Pérez, A., López, M. R., Morales-delaNuez, A., de la Lastra, J.M.P., 2025. Exploring diversity in avian immune defence: insights from cathelicidin clusters. Dev. Comp. Immunol. 166. https://doi.org/10.1016/j.dci.2025.105363.

Guryanova, S.V., Ovchinnikova, T.V., 2022. Immunomodulatory and allergenic properties of antimicrobial peptides. Int. J. Mol. Sci. 23 (5). https://doi.org/10.3390/ijms23052499.

Hancock, R.E.W., Haney, E.F., Gill, E.E., 2016. The immunology of host defence peptides: beyond antimicrobial activity. Nat. Rev. Immunol. 16 (5), 321–334. https://doi.org/10.1038/nri.2016.29.

Haney, E.F., Mansour, S.C., Hancock, R.E.W., 2017. Antimicrobial peptides: an introduction. Methods Mol. Biol. 1548, 3–22. https://doi.org/10.1007/978-1-4939-6737-7_1 (Nota: Referencia a Capítulo de Libro).

Hanson, M.A., Lemaitre, B., Unckless, R.L., 2019. Dynamic evolution of antimicrobial peptides underscores trade-offs between immunity and ecological fitness. Front. Immunol. 10. https://doi.org/10.3389/fimmu.2019.02620.

Hoskin, D.W., Ramamoorthy, A., 2008. Studies on anticancer activities of antimicrobial peptides. Biochim. Biophys. Acta Biomembr. 1778 (2), 357–375. https://doi.org/10.1016/j.bbamem.2007.11.008.

Hsieh, I.N., Hartshorn, K.L., 2016. The role of antimicrobial peptides in influenza virus infection and their potential as antiviral and immunomodulatory therapy. Pharmaceuticals 9 (3). https://doi.org/10.3390/ph9030053.

Izadpanah, A., Gallo, R.L., 2005. Antimicrobial peptides. J. Am. Acad. Dermatol. 52 (3), 381–390. https://doi.org/10.1016/j.jaad.2004.08.026.

Kim, D., Soundrarajan, N., Lee, J., Cho, H.S., Choi, M., Cha, S.Y., Ahn, B., Jeon, H., Le, M. T., Song, H., Kim, J.H., Park, C., 2017. Genomewide analysis of the antimicrobial peptides in Python bivittatus and characterization of cathelicidins with potent

antimicrobial activity and low cytotoxicity. Antimicrob. Agents Chemother. 61 (9), e00530. https://doi.org/10.1128/AAC.00530-17.
Koehbach, J., Craik, D.J., 2019. The vast structural diversity of antimicrobial peptides. Trends Pharmacol. Sci. 40 (7), 517–528. https://doi.org/10.1016/j.tips.2019.04.012.
Kościuczuk, E.M., Lisowski, P., Jarczak, J., Strzałkowska, N., Jóźwik, A., Horbańczuk, J., Krzyżewski, J., Zwierzchowski, L., Bagnicka, E., 2012. Cathelicidins: family of antimicrobial peptides. A review. Mol. Biol. Rep. 39 (12), 10957–10970. https://doi.org/10.1007/s11033-012-1997-x.
Kordi, M., Borzouyi, Z., Chitsaz, S., Asmaei, M.H., Salami, R., Tabarzad, M., 2023. Antimicrobial peptides with anticancer activity: today status, trends and their computational design. Arch. Biochem. Biophys. 733. https://doi.org/10.1016/j.abb.2022.109484.
Lear, S., Cobb, S.L., 2016. Pep-Calc.com: a set of web utilities for the calculation of peptide and peptoid properties and automatic mass spectral peak assignment. J. Comput. Aided Mol. Des. 30 (3), 271–277. https://doi.org/10.1007/s10822-016-9902-7.
Lee, K.H., Tsutsui, T., Honda, K., Asano, R., Kumagai, I., Ohtake, H., Omasa, T., 2013. Generation of high-producing cell lines by overexpression of cell division cycle 25 homolog A in Chinese hamster ovary cells. J. Biosci. Bioeng. 116 (6), 754–760. https://doi.org/10.1016/j.jbiosc.2013.05.032.
Lei, J., Sun, L., Huang, S., Zhu, C., Li, P., He, J., Mackey, V., Coy, D.H., He, Q., 2019. The antimicrobial peptides and their potential clinical applications. Am. J. Transl. Res. 11 (7), 3919–3931.
Matera, A.G., Wang, Z., 2014. A day in the life of the spliceosome. Nat. Rev. Mol. Cell Biol. 15 (2), 108–121. https://doi.org/10.1038/nrm3742.
Miyata, T., Hayashida, H., 1981. Extraordinarily high evolutionary rate of pseudogenes: evidence for the presence of selective pressure against changes between synonymous codons. Proc. Natl. Acad. Sci. U. S. A 78 (9), 5739–5743. https://doi.org/10.1073/pnas.78.9.5739.
Mongiardino Koch, N., Gauthier, J.A., 2018. Noise and biases in genomic data may underlie radically different hypotheses for the position of Iguania within Squamata. PLoS One 13 (8), e0202729. https://doi.org/10.1371/journal.pone.0202729.
Nash, A., Ryan, E.J., 2023. Immunity in sea turtles: review of a host-pathogen arms race millions of years in the running. Animals 13 (4), 556. https://doi.org/10.3390/ani13040556.
Okonechnikov, K., Golosova, O., Fursov, M., 2012. Unipro UGENE: a unified bioinformatics toolkit. Bioinformatics 28 (8), 1166–1167. https://doi.org/10.1093/bioinformatics/bts091.
Pettersen, E.F., Goddard, T.D., Huang, C.C., Meng, E.C., Couch, G.S., Croll, T.I., Morris, J. H., Ferrin, T.E., 2020. UCSF ChimeraX: structure visualization for researchers, educators, and developers. Protein Sci. 30 (1), 70–82. https://doi.org/10.1002/pro.3943.
Qiao, X., Yang, H., Gao, J., Zhang, F., Chu, P., Yang, Y., Zhang, M., Wang, Y., Yu, H., 2019. Diversity, immunoregulatory action and structure-activity relationship of green sea turtle cathelicidins. Dev. Comp. Immunol. 98, 189–204. https://doi.org/10.1016/j.dci.2019.05.005.
Sanchez, J.F., Hoh, F., Strub, M.P., Aumelas, A., Dumas, C., 2002a. Structure of the cathelicidin motif of protegrin-3 precursor: structural insights into the activation mechanism of an antimicrobial protein. Structure 10 (10), 1363–1370. https://doi.org/10.1016/s0969-2126(02)00859-6.
Sanchez, J.F., Wojcik, F., Yang, Y., Strub, M.P., Strub, J.M., Dorsselaer, Av, Martin, M., Lehrer, R., Ganz, T., Chavanieu, A., Calas, B., Aumelas, A., 2002b. Overexpression and structural study of the cathelicidin motif of the Protegrin-3 precursor. Biochemistry 41 (1), 21–30. https://doi.org/10.1021/bi010930a.
Santana, F.L., Estrada, K., Alford, M.A., Wu, B.C., Dostert, M., Pedraz, L., Akhoundsadegh, N., Kalsi, P., Haney, E.F., Straus, S.K., Corzo, G., Hancock, R.E.W., 2022. Novel alligator cathelicidin As-CATH8 demonstrates anti-infective activity against clinically relevant and crocodylian bacterial pathogens. Antibiotics 11 (11), 1603. https://doi.org/10.3390/antibiotics11111603.
Shi, N., Cai, S., Gao, J., Qiao, X., Yang, H., Wang, Y., Yu, H., 2019b. Roles of polymorphic cathelicidins in innate immunity of soft-shell turtle, Pelodiscus sinensis. Dev. Comp. Immunol. 92, 179–192. https://doi.org/10.1016/j.dci.2018.11.010.
Shi, X., Xiang, S., Cao, J., Zhu, H., Yang, B., He, Q., Ying, M., 2019a. Kelch-like proteins: physiological functions and relationships with diseases. Pharmacol. Res. 148. https://doi.org/10.1016/j.phrs.2019.104404.
Shinnar, A.E., Butler, K.L., Park, H.J., 2003. Cathelicidin family of antimicrobial peptides: proteolytic processing and protease resistance. Bioorg. Chem. 31 (6), 425–436. https://doi.org/10.1016/s0045-2068(03)00080-4.
Sibley, C.R., Emmett, W., Blazquez, L., Faro, A., Haberman, N., Briese, M., Trabzuni, D., Ryten, M., Weale, M.E., Hardy, J., Modic, M., Curk, T., Wilson, S.W., Plagnol, V., Ule, J., 2015. Recursive splicing in long vertebrate genes. Nature 521 (7552), 371–375. https://doi.org/10.1038/nature14466.
Silphaduang, U., Colorni, A., Noga, E.J., 2006. Evidence for widespread distribution of piscidin antimicrobial peptides in teleost fish. Dis. Aquat. Org. 72 (3), 241–252. https://doi.org/10.3354/dao072241.
Smith, C.W.J., Chu, T.T., Nadal-Ginard, B., 1993. Scanning and competition between AGs are involved in 3' splice site selection in mammalian introns. Mol. Cell Biol. 13 (8), 4939–4952. https://doi.org/10.1128/mcb.13.8.4939-4952.1993.
Solovyev, V., Kosarev, P., Seledsov, I., Vorobyev, D., 2006. Automatic annotation of eukaryotic genes, pseudogenes and promoters. Genome Biol. 7 (Suppl. 1), 10.1–10.12. https://doi.org/10.1186/gb-2006-7-s1-s10.
Steinstraesser, L., Kraneburg, U., Jacobsen, F., Al-Benna, S., 2011. Host defense peptides and their antimicrobial-immunomodulatory duality. Immunobiology 216 (3), 322–333. https://doi.org/10.1016/j.imbio.2010.07.003.
Sun, S., Zhao, G., Huang, Y., Cai, M., Shan, Y., Wang, H., Chen, Y., 2016. Specificity and mechanism of action of -helical membrane-active peptides interacting with model and biological membranes by single-molecule force spectroscopy. Sci. Rep. 6, 29145. https://doi.org/10.1038/srep29145.
Tam, J.P., Wang, S., Wong, K.H., Tan, W.L., 2015. Antimicrobial peptides from plants. Pharmaceuticals 8 (4), 711–757. https://doi.org/10.3390/ph8040711.
Tamura, K., Stecher, G., Kumar, S., 2021. MEGA11: molecular evolutionary genetics analysis version 11. Mol. Biol. Evol. 38 (7), 3022–3027. https://doi.org/10.1093/molbev/msab120.
Teufel, F., Almagro Armenteros, J.J., Johansen, A.R., Gíslason, M.H., Pihl, S.I., Tsirigos, K.D., Winther, O., Brunak, S., von Heijne, G., Nielsen, H., 2022. SignalP 6.0 predicts all five types of signal peptides using protein language models. Nat. Biotechnol. 40 (7), 1023–1025. https://doi.org/10.1038/s41587-021-01156-3.
Thompson, J.D., Higgins, D.G., Gibson, T.J., 1994. Clustal W: improving the sensitivity of progressive multiple sequence alignment through sequence weighting, position-specific gap penalties and weight matrix choice. Nucleic Acids Res. 22 (22), 4673–4680. https://doi.org/10.1093/nar/22.22.4673.
Tomasinsig, L., Zanetti, M., 2005. The Cathelicidins-structure, function and evolution. Curr. Protein Pept. Sci. 6 (1), 23–34. https://doi.org/10.2174/1389203053027520.
Tossi, A., Gerdol, M., Caporale, A., Pacor, S., Mardirossian, M., Scocchi, M., Prickett, M. D., Manzini, G., Gennaro, R., 2024. Cathelicidins—A rich seam of antimicrobial peptides waiting for exploitation. Front. Drug Discov. 4. https://doi.org/10.3389/fddsv.2024.1458057.
van Dijk, A., Guabiraba, R., Bailleul, G., Schouler, C., Haagsman, H.P., Lalmanach, A.C., 2023. Evolutionary diversification of defensins and cathelicidins in birds and primates. Mol. Immunol. 157, 53–69. https://doi.org/10.1016/j.molimm.2023.03.011.
Veldhuizen, E.J.A., Brouwer, E.C., Schneider, V.A.F., Fluit, A.C., 2013. Chicken cathelicidins display antimicrobial activity against multiresistant bacteria without inducing strong resistance. PLoS One 8 (4). https://doi.org/10.1371/journal.pone.0061964.
Wang, J., Chitsaz, F., Derbyshire, M.K., Gonzales, N.R., Gwadz, M., Lu, S., Marchler, G. H., Song, J.S., Thanki, N., Yamashita, R.A., Yang, M., Zhang, D., Zheng, C., Lanczycki, C.J., Marchler-Bauer, A., 2023. The conserved domain database in 2023. Nucleic Acids Res. 51 (D1), D384–D388. https://doi.org/10.1093/nar/gkac1096.
Wang, Y., Hong, J., Liu, X., Yang, H., Liu, R., Wu, J., Wang, A., Lin, D., Lai, R., 2008. Snake cathelicidin from Bungarus fasciatus is a potent peptide antibiotic. PLoS One 3 (9), e3217. https://doi.org/10.1371/journal.pone.0003217.
Whelehan, C.J., Barry-Reidy, A., Meade, K.G., Eckersall, P.D., Chapwanya, A., Narciandi, F., Lloyd, A.T., O'Farrelly, C., 2014. Characterisation and expression profile of the bovine cathelicidin gene repertoire in mammary tissue. BMC Genom. 15. https://doi.org/10.1186/1471-2164-15-128.
Wilkins, D., Kurtz, Z., 2023. Gggenes: draw gene arrow maps in 'ggplot2' [software]. R package version 0.5.0. https://wilkox.org/gggenes/.2023.(Nota:ReferenciaaSoftware/PaqueteR.
Wiley, S.R., Kraus, R.J., Mertz, J.E., 1992. Functional binding of the "TATA" box binding component of transcription factor TFIID to the -30 region of TATA-less promoters. Proc. Natl. Acad. Sci. U. S. A 89 (13), 5814–5818. https://doi.org/10.1073/pnas.89.13.5814.
Wu, Q., Patočka, J., Kuča, K., 2018. Insect antimicrobial peptides, a mini review. Toxins 10 (11). https://doi.org/10.3390/toxins10110461.
Yin, W., Wang, Zj, Li, Qy, Lian, Jm, Zhou, Y., Lu, Bz, Jin, Lj, Qiu, Px, Zhang, P., Zhu, Wb, Bo, W., Huang, Yj, Lin, Zl, Qiu, Bt, Yang, Hm, Zhang, Gj, Yan, Gm, Zhou, Q., 2016. Evolutionary trajectories of snake genes and genomes revealed by comparative analyses of five-pacer viper. Nat. Commun. 7, 13107. https://doi.org/10.1038/ncomms13107.
Zanetti, M., 2004. Cathelicidins, multifunctional peptides of the innate immunity. J. Leukoc. Biol. 75 (1), 39–48. https://doi.org/10.1189/jlb.0403147.
Zhang, H., Wong, J.P., Ni, G., Cano, P., Dittmer, D.P., Damania, B., 2022. Mitochondrial protein, TBRG4, modulates KSHV and EBV reactivation from latency. PLoS Pathog. 18 (11), e1010990. https://doi.org/10.1371/journal.ppat.1010990.
Zhang, Y., Zhao, H., Yu, G.Y., Liu, X.D., Shen, J.H., Lee, W.H., Zhang, Y., 2010. Structure-function relationship of king cobra cathelicidin. Peptides 31 (8), 1488–1493. https://doi.org/10.1016/j.peptides.2010.05.005.
Zhao, H., Gan, T.X., Liu, X.D., Jin, Y., Lee, W.H., Shen, J.H., Zhang, Y., 2008. Identification and characterization of novel reptile cathelicidins from elapid snakes. Peptides 29 (10), 1685–1691. https://doi.org/10.1016/j.peptides.2008.06.008.
Zhu, S., 2008. Did cathelicidins, a family of multifunctional host-defense peptides, arise from a cysteine protease inhibitor? Trends Microbiol. 16 (8), 353–360. https://doi.org/10.1016/j.tim.2008.05.007.