

The Natterin Proteins Diversity

Subjects: Biochemistry & Molecular Biology

Contributor: Carla Lima

The Natterin proteins were first revealed in the venom of the medically significant Brazilian toadfish *Thalassophryne nattereri* (VTn) in five orthologs named Natterin (1–4, and -P). They were identified as being responsible for the main toxic non-lethal effects of the VTn, such as local edema and excruciating pain, that evolved to necrosis. Following the first identification, the group of Natterin-like proteins has been expanded over time. Several sequences homologous to Natterin have been identified in different teleostean genomes, such as the venomous catfish *Plotosus lineatus* and non-venomous fish, including lampreys *Lampetra japonica* or *Lethenteron camtschaticum* and *Lampetra morii*, arctic charr *Salvelinus alpinus*, zebrafish *Danio rerio*, atlantic cod *Gadus morhua*, and ovate pompano *Trachinotus ovatus*.

Keywords: *Thalassophryne nattereri* ; Natterin ; aerolysin ; protein evolution ; immune function ; bioinformatics

1. Phylogenetic Analysis of Natterin-like Proteins

A large number of proteins containing the Natterin domain are distributed throughout most of the kingdoms of life, except Prokaryotes, Protists, Amphibians, and Mammals, which corroborates what has been described for other sequences. Many types of species also contain a substantial number of proteins with few or no known metazoan homologs [1][2][3][4][5][6].

The phylogenetic analysis with all sequences yielded 16 major clades with considerable diversity of Eukaryotic organisms, including Plants, Fungi, and sessile marine animals with a primitive structure and anatomical organization, in addition to a large variety of fish that might be the animals presenting higher complexity within this group of proteins. The groups with more species included in our analysis were Insects, Birds, and Fish (**Figure 1**), in contrast to the majority (90%) of the aerolysin proteins found in Proteobacteria, Firmicutes, and Fungi [7].

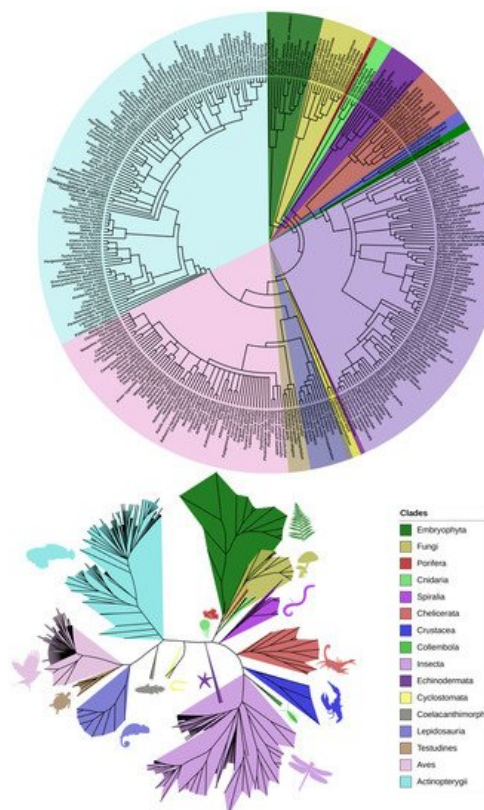


Figure 1. Phylogenetic trees generated using the software PhyloT that represent all the species found sharing any Natterin or Natterin-like protein in the tree of life. The circular tree with the corresponding species (**top**) and the unrooted tree with the main clades (**bottom**) demonstrate the species' evolutionary relationship.

The group of Natterin-like proteins might have somehow appeared in a distant ancestor of the Plants and Animals at least 400 million years ago. This is assumed once it is the most reasonable estimation of the first lineages' divergence that presented any putative Natterin protein in our research. In addition, horizontal transfer might have helped to spread the proteins and increase their variability. Given its appearance in such a variety of taxons, as plants or venomous animals, we can attribute to the members of the Natterin group a decisive evolutionary role.

Based on the species from different taxons that present Natterin proteins, phylogenetic trees, also known as cladograms, were generated to represent evolutionary relationships among organisms based on clade grouping and their paths throughout the evolutionary process [8]. The most accurate phylogenetic tree will have the fewest nodes. This is called parsimony, which means that the best tree is the simplest [9].

It is essential to understand that phylogenetic trees are nested hierarchies, i.e., any individual set of branches is also part of a larger set of branches. This is easily seen in **Figure 1** and **Figure 2**, where, for example, *Thalassophryne amazonica* and *T. nattereri* are grouped in the clade of the Batrachoididae family. However, they are part of the Teleostei and Chordata clades as well. The Darwinian hypothesis of descent with modification predicts that a set of nested hierarchies would represent the organisms' evolutionary history.

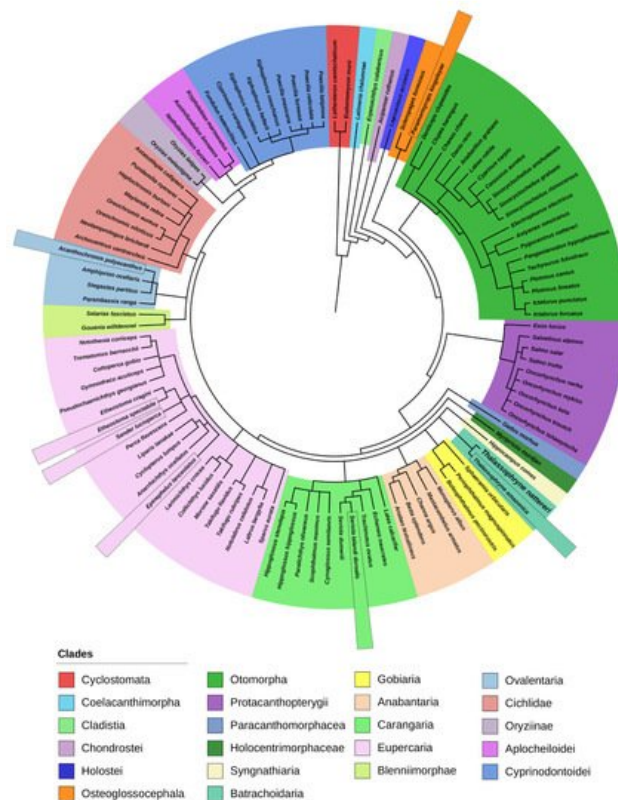


Figure 2. Phylogenetic tree generated using the software PhyloT to represent exclusively the fish species included in the group of Natterin-like proteins, 109 representatives. The species that presented the protein sequences with the higher identity percentage with the founder members Natterin-1–4 are highlighted.

In the general phylogenetic tree (**Figure 1**), it is notable that the Embryophyta and Fungi are in the tree base. Together, with all the other metazoan species, they make the representatives of the group of Natterin-like proteins. The relatedness comparison might be evidence that Natterin proteins arose long in the past, possibly in a pluricellular eukaryote. That inference came out because the most primitive species in the tree is *Selaginella moellendorffii*, which belongs to the Lycopodiophyta clade, considered the oldest division of existing vascular plants, including some of the most basal living species known that first appeared in the fossil record around 400 million years ago. The lycophyte *S. moellendorffii* is an important model organism in comparative genomics [10].

The Natterin domain was reported in ancestral species that belong to the older diverged lineages in Metazoans, e.g., the sponge *Amphimedon queenslandica*; an obligate biotrophic arbuscular mycorrhizal fungus *Rhizophagus irregularis*; the common Indo-Pacific scleractinian coral *Acropora digitifera*; the invasive bivalve species zebra mussel *Dreissena*

polymorpha; common liver fluke *Fasciola hepatica*, blood flukes *Schistosoma japonicum* and *S. haematobium*; the blacklegged tick *Ixodes scapularis*; the salmon louse *Lepeophtheirus salmonis*; and the springtail *Orchesella cincta*.

Moreover, the tremendous contribution of the insects in the group of Natterin-like proteins was noticeable. Even though they are not closely related to fish, insects are the most diverse animals on the planet ^[11]. This would be enough to justify their overrepresentation in the search. As the largest and most widely distributed group of arthropod animals, invertebrates represent more than 70% of all species of living beings described. Insects were among the first animals to colonize and exploit terrestrial and freshwater ecosystems. These characteristics are undoubtedly related to their diversification ^[12].

Insects can be found in almost every ecosystem on the planet, and the most diverse orders are Odonata, Orthoptera, Lepidoptera, Diptera, Hemiptera, Coleoptera, and Hymenoptera. Our analysis showed the presence of Natterin-like proteins in four species of the *Apis* genus and in 16 species of *Drosophila*, except for *D. melanogaster*.

All the ray-finned fish share common ancestors with other groups like Lepidosauria, Testudines, and Aves. Although there were not many reptile species (14) in the whole group, the birds were very well represented, which might be explained by this evolutionary relationship. In addition, the birds form a diverse vertebrate group found all over the globe, from equatorial to polar regions. According to a study, the birds' biodiversity is severely underestimated, and the authors determined that there are around 18,000 species worldwide, nearly twice as many as previously thought ^[13].

In the present study, we explored the fish clade since they are the main focus of our studies and experimental models. Indeed, in the aquatic species phylogeny, 86.5% of the organisms represented are fish (**Figure 2**). They represent more than half of the world's known vertebrate species. Fish heterogeneity is based on many aspects of their biology and habitats. These differences evolved in parallel with the fact that fish have undergone a second WGD event (2R), following the ancient genome duplication that occurred in early vertebrates (1R) and a further one in the teleostean lineage (3R), all of these leading to the subsequent duplication or deletion of various genome parts ^{[14][15][16][17]}.

The general classification of fishes is considered a paraphyletic assemblage, including the classes Myxini (hagfishes) and Petromyzontida (lampreys), from the superclass Agnatha; Chondrichthyes (sharks, rays, and chimeras); Sarcopterygii (coelacanths and lungfishes), and Actinopterygii (ray-finned fishes), from the superclass Osteichthyes (bony fish) ^{[15][18]}.

Other species from aquatic environments but fish are members of the phylum Porifera, Cnidaria, Protostomia, and Echinodermata. The Porifera representative is *A. queenslandica*, a sponge native to the Great Barrier Reef, the world's largest coral reef system. Its genome was the first from a sponge to be sequenced, and it provides insights into the evolution of animal complexity and evolution of metazoan development ^[19].

Cnidaria is an Animalia phylum containing over 11,000 species; they are more complex than sponges and are found predominantly in marine environments. They mostly have two basic body forms: swimming medusae and sessile polyps, both radially symmetrical with mouths surrounded by tentacles that bear cnidocytes ^[20]. In our survey, four species of polyps were registered to contain Natterin-like. Although animal venoms have evolved at least a hundred times independently ^[21], evidence for the implication of horizontal gene transfer ^[22], including from parasitic fungi in the evolutionary origin of Natterin-like in the coral *A. digitifera*, has been issued by Gacesa et al. ^[23]. This type of mechanism that provides a quick channel for the evolution of novelty through the exploitation of bacterial or fungal weapons in animal venoms has also been shown to be crucial to centipede (Chilopoda) venoms ^[24], one of the oldest terrestrial venomous lineages, with a fossil record going back 418 million years.

The other aquatic non-fish clade composed of seven species in the Natterin group is the Protostomia, comprising animals with bilateral symmetry and three germ layers ^[25]. This group includes animals such as arthropods, annelids, and mollusks. Among the seven species shown here, there are rotifer, mussel, oyster, chelicerate arthropod, copepod, shrimp, and crab, indicating a great variety in the type of organisms. The Natterin domain was also found in one species of the Echinodermata phylum, the sea cucumber *Apostichopus japonicus*.

Interestingly, in the aquatic cladogram, the Cyclostomata and Coelacanthimorpha groups, considered the most primitive kind of living fish on Earth, share a common ancestor, not just with the modern ray-finned fish (Actinopterygii), but also with the Lepidosauria (one of the most prominent Reptilia groups, represented mostly by lizards and snakes), Testudines (an order of some of the earliest reptile alive, commonly known as turtles, tortoises, and terrapins), and the diverse class of Aves. These groups did not descend from each other but share ancestors and diverged through life's evolution instead.

The Cyclostomata is a group of agnathans that comprise the living jawless fishes, with horny epidermal structures that function as teeth and branchial arches that are internally positioned instead of externally, as in jawed fish ^[17]. In the fish tree (**Figure 2**), Natterin-like sequences were found in Cyclostomates represented by the Arctic lamprey, also known as

the Japanese river lamprey (*Lethenteron camtschaticum*, synonym *Lampetra japonica*) and the Korean lamprey (*Eudontomyzon morii*, synonym *Lampetra morii*), from the order Petromyzontiformes. These species represent the oldest fish to present Natterin-like proteins in this review. Most lampreys are ectoparasites on fish, using a circular, sucker-like mouth to clamp onto their hosts [26]. Unlike bony fish, lampreys lack scales, fins, and gill covers, however, like sharks, their skeletons are made of cartilage.

The lamprey clade likely diverged from a common ancestor in the Silurian Period, from 443 million to 416 million years ago [26]. It also corroborates with a time estimation of the raising of the Natterins since the Lycophytes and the Coelacanth divergence, according to genetic analysis of current species, are thought to have occurred about 390–420 million years ago [27][28][29].

Our search resulted in only one Sarcopterygii lineage species presenting a Natterin-like protein, the West Indian Ocean coelacanth *Latimeria chalumnae* considered phylogenetically closer to lungfish and tetrapods than ray-finned fish (Actinopterygii) [30]. The group's most important characteristic is paired fins (pectorals and pelvic), whose bases are muscular peduncles that resemble the members of terrestrial vertebrates and move in the same way.

Together, Sarcopterygii and Actinopterygii form the group of bone-fish (Osteichthyes), which are more related to each other than to the lamprey group, and share a more distant ancestor and present very distinctive physical attributes. Natterin domain sequences were identified in only one species of the groups Cladistia, Chondrostei, Holostei, Paracanthomorphacea, Holocentromorphacea, and Syngnatharia, which are ramifications of the Actinopterygii (**Figure 2**).

However, it is crucial to notice that the fish in the bottom of the ray-finned clade are less related to most current fish since they diverged from the common ancestor long ago. The reedfish *Erpetoichthys calabaricus*, which lacks pelvic fins, is a member of the clade Cladistia; it consists of a few anguilliform (i.e., eel-shaped) remnants of an ancient diversity. The sterlet *Acipenser ruthenus* is the only member of the Chondrostei group to present Natterin-like genes. This is a group of essentially cartilaginous fish presenting some degree of ossification. Its members share with the Elasmobranchii (sharks and rays) certain features, such as the possession of spiracles, a heterocercal tail, and the absence of scales.

Holostei, an infraclass of the Neopterygii subclass, is represented in our cladogram by the presence of a Natterin-like sequence in the spotted gar *Lepisosteus oculatus*, restricted to the freshwaters of eastern North America [31]. Holosteans are closer to Teleosts than are the Chondrosteans. The spiracles are reduced to vestigial remnants (in gars, the spiracles do not even open to the outside), and the bones are lightly ossified. The thick ganoid scales of the gars are more primitive than those of the bowfin. A thin layer of bone covers a mostly cartilaginous skeleton in the bowfins, and they have many-rayed dorsal fins. In gars, the tail is still heterocercal but less so than in the Chondrosteans.

The Teleostei is the most diverse lineage of the Neopterygii and by far the largest infraclass in the class Actinopterygii, from the 109 fish in the group of Natterin-like proteins, 106 are part of this group, as seen from the clade Osteoglossocephala (**Figure 2**). Teleosts are the most abundant aquatic vertebrates living today, containing over 30,000 named species [18], which is more than all living mammals, birds, reptiles, and amphibians combined. They comprise around 96% of all extant fishes and nearly half of all vertebrate species, which perhaps represents the most extensive adaptive radiation in vertebrate evolution [31][32]. The difference between Teleosts and other bony fish lies notably in their jawbones; they have a movable premaxilla and corresponding modifications in the jaw musculature, making it possible for them to protrude their jaws outwards from the mouth. Another difference is that the caudal tail fin's upper and lower lobes are about equal in size. The spine ends at the caudal peduncle, distinguishing this group from other fish in which the spine extends into the upper lobe of the tail fin [33].

Resolution of the phylogenetic relationships of Teleosts is critical to understanding the timing of their diversification. There is currently a discordance between the estimated age of divergence for Teleosts, as inferred from the fossil record and molecular studies. It is estimated that crown teleosts' lineage first diverged during the Carboniferous to early Permian (333.0–285.8 Ma), following the Devonian Age of Fishes [34].

More recent works connect the duplicate genomes of Teleosts as the driver of their prolific phenotypic diversification, concordant with the more general hypothesis that increased morphological complexity and innovation is an expected consequence of WGDs. This process provided entire sets of genes with increased biological complexity and the origin of evolutionary novelties [14][31][35]. The Teleost-specific (TS) WGD event, the third round in fish's evolution, took place in the common ancestor of all extant Teleosts shaping this group's history. The Teleost lineage split from basal ray-finned fishes and started to diverge after a WGD event that happened around 320–350 mya [16].

After WGD, duplicate genes (ohnologs) may follow different fates. The most likely outcome is that one member becomes a pseudogene and disappears (non-functionalization) due to the lack of selective constraint on preserving both, or the copies persist as a result of complementation [16][36]. Mechanisms that act on the preservation of duplicates are (1) subfunctionalization, which is the partitioning of different functions subsets of an ancestral among ohnologs, providing an attractive explanation for why so many duplicated genes exist in eukaryotes, without requiring each duplication event to have conferred a selective advantage [16][37]; and (2) neofunctionalization, a process where one ohnolog mutates into a function that was not observed before duplication, leading to the retention of both copies [16][38][39]. The new function must be positively selected; if only one ohnolog evolves a new beneficial function, however, it must also lose an essential ancestral function that the complementing ohnolog maintains; otherwise, the second copy will disappear because it is no longer positively selected [36]. Regarding the Natterin-like proteins, the last phenomenon might explain the broader number and diversity of functionalities potentially designated, primarily exhibited by fish.

Teleost fishes are adapted to widely varied habitats from cold Arctic and Antarctic oceans to desert hot springs; from fast, rock-laden torrential mountain streams to the lightless depths of ocean trenches [40]. Regarding their wide morphological variation, including not only torpedo-shaped fish built for speed, Teleosts can also be flattened vertically or horizontally, be elongated cylinders or take specialized shapes as in anglerfish and seahorses. The last example present in the Natterin group is the tiger tail seahorse *Hippocampus comes* (Syngnatharia).

Classic explanations for Teleost success include key innovations in feeding (e.g., protrusible jaws and pharyngeal jaws), reproduction, and the modes that they use to take up, transport, and deliver oxygen to the tissues, and features that enhance the capacitance of blood for O₂ (βb): the Bohr and Root effects, RBC β-adrenergic sodium proton exchangers (RBC β-NHE), and the retia mirabilia (teleost vascular countercurrent exchangers) [31][32]. As a consequence of the genomic rearrangements during the TS-WGD event, some immune molecular families have expanded tremendously in some species, providing important functional effects against pathogens to which different species have been exposed. Proteins of the Natterin-like group may be evidence of this diversification of function. The conservation of the Natterin domain in different species may have been crucial for the evolution of species.

Teleosts have adopted a range of reproductive strategies. Most use external fertilization without any further parental involvement. A fair proportion of Teleosts are sequential hermaphrodites, starting life as females and transitioning to males at some stage, with a few species reversing this process. The green swordtail shown in this research (*Xiphophorus helleri*) tends to undergo sex reversal under certain environmental conditions. Another example of species containing Natterin-like proteins, the mangrove killifish (*Kryptolebias marmoratus*), is the only naturally occurring vertebrate known to be capable of self-fertilization; most populations consist primarily or exclusively of hermaphroditic individuals or males, and females do not seem to exist [41].

Another curious example of a Natterin group member is the Amazon molly (*Poecilia formosa*), an all-female species thought to have originated due to hybridization between two other species in the genus. It reproduces gynogenetically, meaning once the sperm has penetrated the egg membrane, it takes no further part in the embryo's development. The reproduction is triggered by copulation and stimulation by sperm from the males of other species in the genus [42].

According to the biological aspects of fish and environmental distribution, we observed that most species containing Natterin-like proteins are present in freshwater (~29%); followed by freshwater and brackish (~23%); marine (~21%); marine, brackish, and freshwater (~15%); or marine and brackish environments (~12%). It can be seen that half of the species occupy more than one environment throughout the life cycle. This is due to the tolerance to physical–chemical variations that some fish exhibit, as well as the behavior of moving into brackish or freshwater to spawn, such as *Morone saxatilis*. Still, some occupy different habitats at different life stages, such as *Hippoglossus stenolepis*, which when young are found near the shore, moving out to deeper waters as they grow older; besides this, no other freshwater fish are found as far north as the arctic charr *S. alpinus*, and among marine fish, it is notably a recurrent reef-associated behavior.

Within these aquatic environments, the species members of the Natterin-like group predominantly occupy the benthopelagic and demersal zones (~70%). Both are ecological regions associated with the lowest water body level, where the species live and feed near the bottom. The fish distribution through the climate zones was, in descending order, Tropical (~40%), Subtropical (32%), Temperate (~21%), and Polar (~6%), demonstrating the circumglobal distribution [43].

Regarding the potential to provoke envenomation in human victims, only four species are venomous and present venom apparatus: *Plotosus canius*, *Plotosus lineatus*, *Thalassophryne nattereri*, and *Thalassophryne amazonica*. Venoms, by definition, require a method by which their bearer can introduce them into the body of a target organism; this is accomplished via spiny elements associated with the fins or opercular and cleithral bones that contain grooves that facilitate the flow of venom along with the spin; in most cases, the glandular tissue rests within the groove itself [44].

The venom glands of catfishes (Plotosidae) are composed of aggregations of glandular cells associated with bony spines in the dorsal and pectoral fins. The spines of many species are additionally armed with retrorse serrations along one or both of the spine margins surrounded by a tegumentary sheath with specialized glands. When the spine enters a potential predator, the glands are torn, releasing the largely proteinaceous venom into the wound [44]. However, members of the genus *Thalassophryne* have a complete venom inoculation apparatus composed of two dorsal canaliculated spines and one on each side covered by a membrane and all connected to the venom glands at the base of their fins [45][46][47]. Two other species are poisonous when eaten; *Takifugu rubripes* and *Takifugu flavidus* contain lethal amounts of the poison tetrodotoxin in the internal organs, especially the liver and ovaries. Moreover, *Myripristis murdjan* and *Seriola dumerili* have been reported to provoke ciguatera poisoning, a foodborne illness.

Natterin-like sequences are observed in the three different species that can cause traumatogenic injury through bites *Pygocentrus nattereri*, *Epinephelus lanceolatus*, and *Anarrhichthys ocellatus* (**Figure 2**). Lastly, the electric eel, *Electrophorus electricus*, a South American electric fish and the only species in its genus, presents voltage electric organs that can discharge electric shocks [43].

Even though Natterin-like genes are widely distributed among different organisms, they do not appear homogeneously throughout evolution; mammals, for example, were not included in our results. Furthermore, bacteria did not show up in our search, despite the fact that the Aerolysin is the founding member of a major class of pore-forming toxins. Even the Chondrichthyes did not present Natterin-like sequences up to date, leading to the understanding that the Natterin-like proteins follow a distribution pattern where some groups and many species that descend from the same common ancestor do not have this type of gene.

Whether or not fish have evolved independently, the question of central importance is whether they preferentially retained Natterin-like as one more common solution to challenges of infections despite their exploitation of widely divergent trophic ecologies, consistent with continuity of function and adaptive value.

2. Multiple Alignments of the 15 Most Similar Members of Natterin-like Proteins

Next, we generated a multiple sequence alignment of Natterin-like sequences from the predicted aerolysin conserved domain limited to the inner β -barrel and the outer β -barrel region of the pore structure (residues 190–315 amino acids) of Natterin founder members to evaluate the conservation of protein domains, as well as individual amino acids or nucleotides [48][49].

Initially, the alignment of all the Natterin-like protein sequences from the group of fish clustered together generated a wide range of sequence identity from 12% to 87% between domains. When we narrow the search for the top 15 protein sequences with the high level of identity with the Natterins from *T. nattereri*, we verified that except the *Sander lucioperca* (Natterin-1, Natterin-2, and Natterin-3) *Acanthochromis polyacanthus* (Natterin-1, Natterin-2, and Natterin-4), *Epinephelus lanceolatus* (Natterin-3), and *Paramormyrops kingsleyae* (Natterin-4), the species *Thalassophryne amazonica*, *Seriola lalandi dorsalis*, and *Etheostoma spectabile* showed sequences with a high identity with the all Natterin founder members (**Figure 3**). As expected, *T. amazonica* presented sequences with the highest percentages of identity, and all 15 sequences analyzed showed a greater identity with Natterin-3.

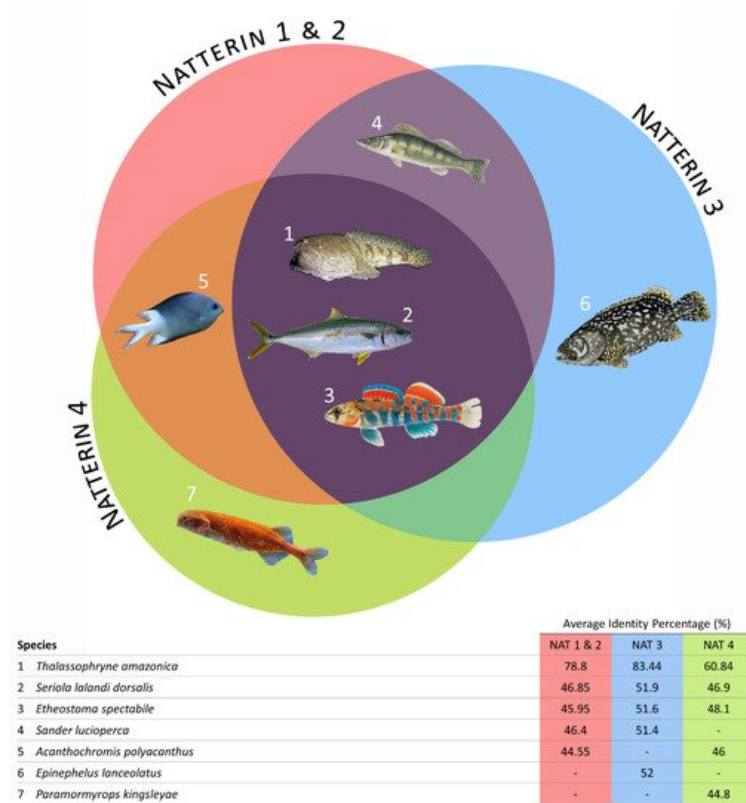


Figure 3. The seven fish species presenting the Natterin proteins (the topmost 15 sequences) with a higher percentage of identity (pid) compared with founder members Natterins-1–4 (**top**); The queried sequences' average identity percentage comparing to the founder members Natterins-1–4 (**bottom**).

Thalassophryne is a genus of venomous toadfish found in the western Atlantic Ocean, and *T. amazonica* is found in the Amazon River and some of its tributaries, while *T. nattereri* has been found in Northeastern Brazil. *T. amazonica* presents two putative Natterin genes and 24 Natterin-like genes, four of them with two isoforms, totalizing 30 proteins. This finding is surprising because both descend from a common ancestor (**Figure 2**).

We observed 64 species with two to five genes, 25 species with six to nine genes, nine species with more than ten genes, and two species had more than 20 genes, suggesting evolutionary-driven gene duplication. Specifically, *T. amazonica* (26); *Anabas testudineus* (25); *Sinocyclocheilus grahami* (14); *Salmo salar* (13); *Oreochromis niloticus* (12); *Sinocyclocheilus rhinoceros* (12); *Notolabrus celidotus* (11); *Oncorhynchus mykiss* (11); *Perca flavescens* (11); *Acanthochromis polyacanthus* (10); and *Danio rerio* (10) are the species with the highest numbers of genes. In contrast, eight species presented only one Natterin-like gene, as follows: *Boleophthalmus pectinirostris*, *Latimeria chalumnae*, *Paramormyrops kingsleyae*, *Ictalurus furcatus* (*Pimelodus furcatus*), *Lethenteron camtschaticum* (*Lampetra japonica*), *Eudontomyzon morii* (*Lampetra morii*), *Plotosus canius*, and *Trachinotus ovatus*.

This inconsistency in the number of copies is probably because it is estimated that 75% of the genes from the TS-WGD event may revert to singletons [50]. Contrarily, the duplication-degeneration-complementation (DDC) hypothesis [51][52] might help to explain the unexpectedly high retention of duplicate genes, which suggests that genes with simple tissue- and time-specific regulatory elements would be more likely to revert to singletons than those with complex regulation.

To further characterize the Natterin domain, we explored separately Natterin-like sequences of *Thalassophryne amazonica* (XP_034025386.1; XP_034025387.1; XP_034025391.1; XP_034025388.1; XP_034025389.1; XP_034025400.1; XP_034025394.1; XP_034025495.1; XP_034025390.1; XP_034025393.1; XP_034025496.1; XP_034025402.1; XP_034027055.1; XP_034027056.1; XP_034027058.1; XP_034023581.1; XP_034017473.1; XP_034022621.1; XP_034023585.1; XP_034023580.1; XP_034023577.1; XP_034021588.1; XP_034023579.1; XP_034023576.1; XP_034021587.1; XP_034017472.1), *Seriola lalandi dorsalis* (XP_023250786.1), *Sander lucioperca* (XP_031161243.1), *Etheostoma spectabile* (XP_032371798.1), *Paramormyrops kingsleyae* (XP_023657493.1), *Epinephelus lanceolatus* (XP_033476485.1), and *Acanthochromis polyacanthus* (XP_022070844.1) to compare with Natterin-1 (Q66S25), Natterin-2 (Q66S21), Natterin-3 (Q66S17), and Natterin-4 (Q66S13) of *T. nattereri* (**Figure 4**).

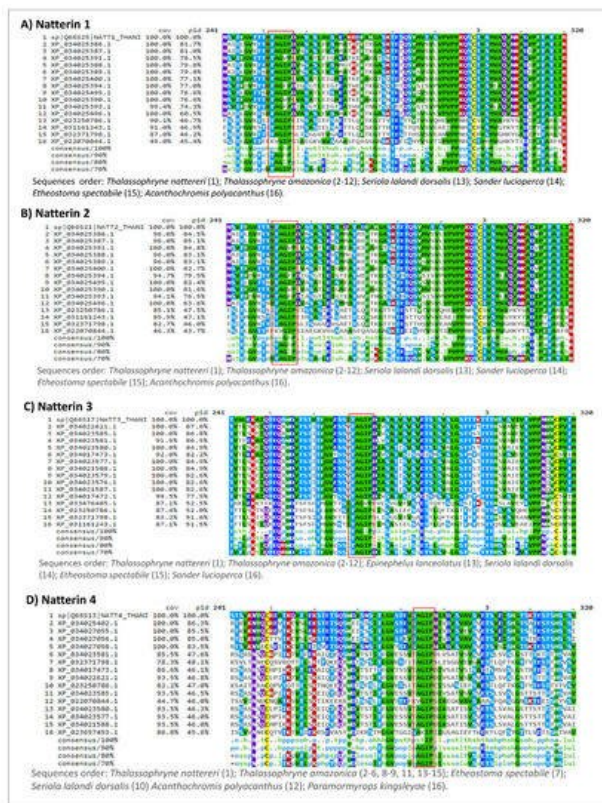


Figure 4. Multiple alignment and conserved residues found on the 15 topmost similar fish sequences compared with the founder members Natterins-1–4, based on the percentage of identity (pid). The conserved residues “TAGIP” and “AGIP” are highlighted in the pictures, picture (A–D), respectively.

Natterin-1 and Natterin-2 contain six evolutionarily conserved motifs in the interval (241–320 aa) GV, AGIP, QSY, VPVPP, MVA, and PFTATLIR. Natterin-3 shows eight evolutionarily conserved motifs QTEQRWDV, TST, GV, SS, AGIP, ETSLSVLGST, TTTHSV, and VTVPNN. In contrast, few motifs were found in Natterin-4, TK, VTL, WD, GV, AGIP, and ETS. Among all motifs, two motifs that remained present in all four Natterin sequences were GV and AGIP (**Figure 4**).

The AGIP motif is composed of four non-polar hydrophobic residues, which show metabolically inexpensive features. Alanine (A) and isoleucine (I) residues show a preference to be in regions inside the regular secondary structure. As expected, both glycine (G) and proline (P) also show a preference for regions outside the regular secondary structure and play important roles in many turn types. Proline (P) disrupts the secondary structure and is often found as a capping residue. Glycine (G) is often found in loop regions, probably because of a conformation with positive ϕ (ϕ , torsion angle around the N–C $_{\alpha}$ bond) that is often required to complete a turn, as reviewed by Shapovalov, Vucetic, and Dunbrack [53].

Natterins-1–4 show four glycine (G) residues along the length of the pore-forming region, including one of them within the AGIP motif and a second along with the non-polar hydrophobic valine (V) residue forming the short GV motif (**Figure 4**), described as the main residues that act as hinge located on the membrane-binding domain involved in the pre-pore to pore conformation [54]. GV and AGIP motifs remained preserved in all top 15 Natterin-like proteins as in most of them. Then, we reasoned that both motifs represent to members of the group of Natterin-like proteins the pore-forming loop for membrane anchoring ability and transmembrane barrel insertion.

A special feature of all β -PFPs is alternating serine (S) and threonine (T) polar residues found in the insertion loop, as well as throughout the rest of the pore-forming modules, such as the lumen of the β -barrel. These residues are thought to participate in membrane binding [55] and oligomerization [56], and help the amphipathic loops in transmembrane pore formation [57].

Interestingly, the sequence consensus AGIP present in Natterins-1–3 founder members is immediately flanked by threonine (T) and aspartate (D) residues on each side and beyond to the end of the C-terminal by one residue of serine (S), and farther away from this site surrounding by other flexibility-inducing amino acids (serine or threonine), characterizing a region rich in carboxyl (aspartate) and hydroxyl (serine or threonine) groups allowing interactions that guarantee the flexibility and stabilization of the loop conformation. However, in Natterin-4, the threonine (T) and aspartate (D) flexible residues surrounding the AGIP core have been replaced by serine (S) and polar asparagine (N), and the

presence of threonine (T), aspartate (D), and serine (S) residues were noticed to be flanking this structure on each side. Asparagine (N) and aspartate (D) are known to adopt conformations in the left-handed α -helical region and other partially allowed regions of the Ramachandran plot more readily than any other non-glycyl amino acids [58].

Then, we found that the two hydroxylated amino acid residues of threonine (T) and serine (S) that flanked the AGIP motif located in the insertion membrane of Natterin-1 and Natterin-2 was conserved in all sequences, except in Natterin-like sequences of *Acanthochromis polyacanthus* that was replaced by lysine (K) and glutamic acid (E) residues, which demonstrated low and intermediate flexibility features, respectively [59]. In comparison to Natterin-4, the threonine (T) residues were replaced by serine (S) in seven Natterin-like sequences of *T. amazonica* and one sequence of *Paramormyrops kingsleyae*. The serine (S) residues were replaced by lysine (K), glycine (G), and mainly threonine (T) in sequences of *T. amazonica* (7 sequences), *Etheostoma spectabile* (1 sequence), *Acanthochromis polyacanthus* (1 sequence), and *Paramormyrops kingsleyae* (1 sequence). However, all Natterin-like sequences conserved both polar residues when compared to Natterin-3.

References

1. Marchler-Bauer, A.; Bryant, S.H. CD-Search: Protein domain annotations on the fly. *Nucleic Acids Res.* 2004, 32, W327–W331.
2. Von Reumont, B.M.; Undheim, E.A.B.; Jauss, R.T.; Jenner, R.A. Venomics of remipede crustaceans reveals novel peptide diversity and illuminates the venom's biological role. *Toxins* 2017, 9, 234.
3. Drukewitz, S.H.; Bokelmann, L.; Undheim, E.A.B.; von Reumont, B.M. Toxins from scratch? Diverse, multimodal gene origins in the predatory robber fly *Dasypogon diadema* indicate a dynamic venom evolution in dipteran insects. *GigaScience* 2019, 8, giz081.
4. Walker, A.A.; Madio, B.; Jin, J.; Undheim, E.A.B.; Fry, B.G.; King, G.F. Melt with this kiss: Paralyzing and liquefying venom of the assassin bug *Pristhesancus plagipennis* (Hemiptera: Reduviidae). *Mol. Cell. Proteom.* 2017, 16, 552–566.
5. Özbek, R.; Wielsch, N.; Vogel, H.; Lochnit, G.; Foerster, F.; Vilcinskis, A.; von Reumont, B.M. Proteo-transcriptomic characterization of the venom from the endoparasitoid wasp *Pimpla turionellae* with aspects on its biology and evolution. *Toxins* 2019, 11, 721.
6. Fingerhut, L.C.H.W.; Strugnelli, J.M.; Faou, P.; Labiaga, A.R.; Zhang, J.; Cooke, I.R. Shotgun proteomics analysis of saliva and salivary gland tissue from the common octopus *Octopus vulgaris*. *J. Proteome Res.* 2018, 17, 3866–3876.
7. Szczesny, P.; Iacovache, I.; Muszewska, A.; Ginalska, K.; van der Goot, F.G.; Grynberg, M. Extending the aerolysin family: From bacteria to vertebrates. *PLoS ONE* 2011, 6, e20349.
8. Baum, D. Reading a Phylogenetic Tree: The Meaning of Monophyletic Groups. *Nat. Educ.* 2008, 1, 190.
9. Baum, D. Trait evolution on a phylogenetic tree: Relatedness, similarity, and the myth of evolutionary advancement. *Nat. Educ.* 2008, 1, 191.
10. Banks, J.A.; Nishiyama, T.; Hasebe, M.; Bowman, J.L.; Gribskov, M.; dePamphilis, C.; Albert, V.A.; Aono, N.; Aoyama, T. The Selaginella Genome Identifies Genetic Changes Associated with the Evolution of Vascular Plants. *Science* 2011, 332, 960–963.
11. Nigel, E.S. How many species of insects and other terrestrial arthropods are there on earth? *Annu. Rev. Entomol.* 2018, 63, 31–45.
12. Misof, B.; Liu, S.; Meusemann, K.; Peters, R.S.; Donath, A.; Mayer, C.; Kjer, K.M.; Zhou, X. Phylogenomics resolves the timing and pattern of insect evolution. *Science* 2014, 346, 763–767.
13. Barrowclough, G.F.; Cracraft, J.; Klicka, J.; Zink, R.M. How many kinds of birds are there and why does it matter? *PLoS ONE* 2016, 11, e0166307.
14. van de Peer, Y.; Maere, S.; Meyer, A. The evolutionary significance of ancient genome duplications. *Nat. Rev. Genet.* 2009, 10, 725–732.
15. Castro, R.; Tafalla, C. Overview of fish immunity. In *Mucosal Health in Aquaculture*; Beck, B.H., Peatman, E., Eds.; Academic Press: Cambridge, MA, USA, 2015; pp. 3–54.
16. Glasauer, S.M.K.; Neuhauss, S.C.F. Whole-genome duplication in teleost fishes and its evolutionary consequences. *Mol. Genet. Genom.* 2014, 289, 1045–1060.
17. Kuraku, S.; Meyer, A.; Kuratani, S. Timing of genome duplications relative to the origin of the vertebrates: Did cyclostomes diverge before or after? *Mol. Biol. Evol.* 2009, 26, 47–59.

18. Nelson, J.S. *Fishes of the World*, 4th ed.; John Wiley & Sons: New York, NY, USA, 2006.
19. Srivastava, M.; Simakov, O.; Chapman, J.; Fahey, B.; Gauthier, M.E.A.; Mitros, T.; Richards, G.S.; Conaco, C.; Dacre, M.; Hellsten, U.; et al. The *Amphimedon queenslandica* genome and the evolution of animal complexity. *Nature* 2010, 466, 720–726.
20. Park, E.; Hwang, D.; Lee, J.; Song, J.; Seo, T.; Won, Y. Estimation of divergence times in cnidarian evolution based on mitochondrial protein-coding genes and the fossil record. *Mol. Phylogenet. Evol.* 2012, 62, 329–345.
21. Schendel, V.; Rash, L.D.; Jenner, R.A.; Undheim, E.A.B. The diversity of venom: The importance of behavior and venom system morphology in understanding its ecology and evolution. *Toxins* 2019, 11, 666.
22. Moran, Y.; Fredman, D.; Szczesny, P.; Grynberg, M.; Technau, U. Recurrent Horizontal Transfer of Bacterial Toxin Genes to Eukaryotes. *Mol. Biol. Evol.* 2012, 29, 2223–2230.
23. Gacesa, R.; Hung, J.Y.; Bourne, D.G.; Long, P.F. Horizontal transfer of a natterin-like toxin encoding gene within the holobiont of the reef building coral *Acropora digitifera* (Cnidaria: Anthozoa: Scleractinia) and across multiple animal lineages. *J. Venom Res.* 2020, 10, 7–12.
24. Undheim, E.; Jenner, R.A. Phylogenetic analyses suggest centipede venom arsenals were repeatedly stocked by horizontal gene transfer. *Nat. Commun.* 2021, 12, 818.
25. Hejnal, A.; Obst, M.; Stamatakis, A.; Ott, M.; Rouse, G.W.; Edgecombe, G.D. Assessing the root of bilaterian animals with scalable phylogenomic methods. *Proc. R. Soc. Biol. Sci.* 2009, 276, 4261–4270.
26. Shimeld, S.M.; Donoghue, P.C.J. Evolutionary crossroads in developmental biology: Cyclostomes (lamprey and hagfish). *Development* 2012, 139, 2091–2099.
27. Spencer, V.; Venza, Z.N.; Harrison, C.J. What can lycophytes teach us about plant evolution and development? Modern perspectives on an ancient lineage. *Evol. Dev.* 2020, 9, e12350.
28. Friedman, M.; Coates, M.I.; Anderson, P. First discovery of a primitive coelacanth fin fills a major gap in the evolution of lobed fins and limbs. *Evol. Dev.* 2007, 9, 329–337.
29. Johanson, Z.; Long, J.A.; Talent, J.A.; Janvier, P.; Warren, J.W. Oldest coelacanth, from the Early Devonian of Australia. *Biol. Lett.* 2006, 2, 443–446.
30. Holder, M.T.; Erdmann, M.V.; Wilcox, T.P.; Caldwell, R.L.; Hillis, D.M. Two living species of coelacanths? *Proc. Natl. Acad. Sci. USA* 1999, 96, 12616–12620.
31. Clarke, J.T.; Lloyd, G.T.; Friedman, M. Little evidence for enhanced phenotypic evolution in early teleosts relative to their living fossil sister group. *Proc. Natl. Acad. Sci. USA* 2016, 113, 11531–11536.
32. Harter, T.S.; Brauner, C.J. The O₂ and CO₂ transport system in teleosts and the specialized mechanisms that enhance Hb–O₂ unloading to tissues. *Fish Physiol.* 2017, 36, 1–106.
33. Witten, P.W.; Hall, B.K. Teleost skeletal plasticity: Modulation, adaptation, and remodelling. *Copeia* 2015, 103, 727–739.
34. Near, T.J.; Eytan, R.I.; Dornburg, A.; Kuhn, K.L.; Moore, J.A.; Davis, M.P.; Wainwright, P.C.; Friedman, M.; Smith, W.L. Resolution of ray-finned fish phylogeny and timing of diversification. *Proc. Natl. Acad. Sci. USA* 2012, 109, 13698–13703.
35. Sato, Y.; Nishida, M. Teleost fish with specific genome duplication as unique models of vertebrate evolution. *Environ. Biol. Fishes* 2010, 88, 169–188.
36. Braasch, I.; Bobe, J.; Guiguen, Y.; Postlethwait, J.H. Reply to: ‘Subfunctionalization versus neofunctionalization after whole-genome duplication’. *Nat. Genet.* 2018, 50, 910–911.
37. Cusack, B.P.; Wolfe, K.H. When gene marriages don’t work out: Divorce by subfunctionalization. *Trends Genet.* 2007, 23, 270–272.
38. Rastogi, S.; Liberles, D.A. Subfunctionalization of duplicated genes as a transition state to neofunctionalization. *BMC Evol. Biol.* 2005, 14, 5–28.
39. Sandve, S.R.; Rohlfs, R.V.; Hvidsten, T.R. Subfunctionalization versus neofunctionalization after whole-genome duplication. *Nat. Genet.* 2018, 50, 908–909.
40. Weitzman, S.H. *Teleost Fish Encyclopedia Britannica*. 2018. Available online: <https://www.britannica.com/animal/teleost> (accessed on 12 April 2020).
41. Avise, J.C.; Tatarenkov, A. Population genetics and evolution of the mangrove rivulus *Kryptolebias marmoratus*, the world’s only self-fertilizing hermaphrodite vertebrate. *J. Fish Biol.* 2015, 87, 519–538.

42. Lampert, K.P.; Scharl, M. The origin and evolution of a unisexual hybrid: *Poecilia formosa*. *Philos. Trans. R. Soc. B* 2008, 363, 2901–2909.
43. Froese, R.; Pauly, D. FishBase 2000: Concepts, Design and Data Sources; ICLARM: Los Baños, Philippines, 2000; p. 344.
44. Wright, J.J. Evolutionary History of Venom Glands in the Siluriformes. *Evol. Venom. Anim. Their Toxins* 2015, 1–19.
45. Fróes, H.P. Sur un poisson toxiphore brésilien: Le “niquim” *Thalassophryne maculosa*. *Rev. Sudam. Med. Cirurgia* 1932, 3, 871–878.
46. Fróes, H.P. Studies on venomous fishes of tropical countries. *J. Trop. Med. Hyg.* 1933, 36, 134–135.
47. Lopes-Ferreira, M.; Grund, L.Z.; Lima, C. *Thalassophryne nattereri* fish venom: From the envenoming to the understanding of the immune system. *J. Venom. Anim. Toxins Incl. Trop. Dis.* 2014, 20, 1–12.
48. Chowdhury, B.; Garai, G. A review on multiple sequence alignment from the perspective of genetic algorithm. *Genomics* 2017, 109, 419–431.
49. Chatzou, M.; Magis, C.; Chang, J.M.; Kemena, C.; Bussotti, G.; Erb, I.; Notredame, C. Multiple sequence alignment modeling: Methods and applications. *Brief. Bioinform.* 2016, 17, 1009–1023.
50. Postlethwait, J.H.; Woods, I.G.; Ngo-Hazelett, P.; Yan, Y.L.; Kelly, P.D.; Chu, F.; Huang, H.; Hill-Force, A.; Talbot, W.S. Zebrafish comparative genomics and the origins of vertebrate chromosomes. *Genome Res.* 2000, 10, 1890–1902.
51. Force, A.; Lynch, M.; Pickett, F.B.; Amores, A.; Yan, Y.L.; Postlethwait, J. Preservation of duplicate genes by complementary, degenerative mutations. *Genetics* 1999, 151, 1531–1545.
52. MacCarthy, T.; Bergman, A. The limits of subfunctionalization. *BMC Evol. Biol.* 2007, 7, 213.
53. Shapovalov, M.; Vucetic, S.; Dunbrack, R.L., Jr. A new clustering and nomenclature for beta turns derived from high-resolution protein structures. *PLoS Comput. Biol.* 2019, 15, e1006844.
54. Cirauqui, N.; Abriata, L.A.; van der Goot, F.G.; Dal Peraro, M. Structural, physicochemical and dynamic features conserved within the aerolysin pore-forming toxin family. *Sci. Rep.* 2017, 7, 13932.
55. Akiba, T.; Abe, Y.; Kitada, S.; Kusaka, Y.; Ito, A.; Ichimatsu, T.; Katayama, H.; Akao, T.; Higuchi, K.; Mizuki, E.; et al. Crystal structure of the parasporin-2 *Bacillus thuringiensis* toxin that recognizes cancer cells. *J. Mol. Biol.* 2009, 386, 121–133.
56. Cole, A.R.; Gibert, M.; Popoff, M.; Moss, D.S.; Titball, R.W.; Basak, A.K. *Clostridium perfringens* epsilon-toxin shows structural similarity to the pore-forming toxin aerolysin. *Nat. Struct. Mol. Biol.* 2004, 11, 797–798.
57. Degiacomi, M.T.; Iacovache, I.; Pernot, L.; Chami, M.; Kudryashev, M.; Stahlberg, H.; van der Goot, F.G.; Dal Peraro, M. Molecular assembly of the aerolysin pore reveals a swirling membrane-insertion mechanism. *Nat. Chem. Biol.* 2013, 9, 623–629.
58. Deane, C.M.; Allen, F.H.; Taylor, R.; Blundell, T.L. Carbonyl-carbonyl interactions stabilize the partially allowed Ramachandran conformations of asparagine and aspartic acid. *Protein Eng.* 1999, 12, 1025–1028.
59. Mayorov, A.; Dal Peraro, M.; Abriata, L.A. Active site-induced evolutionary constraints follow fold polarity principles in soluble globular enzymes. *Mol. Biol. Evol.* 2019, 36, 1728–1733.