

RESEARCH ARTICLE

Open Access



Co-option of the same ancestral gene family gave rise to mammalian and reptilian toxins

Agneesh Barua^{1*}, Ivan Koludarov² and Alexander S. Mikheyev^{3*}

Abstract

Background: Evolution can occur with surprising predictability when organisms face similar ecological challenges. For most traits, it is difficult to ascertain whether this occurs due to constraints imposed by the number of possible phenotypic solutions or because of parallel responses by shared genetic and regulatory architecture. Exceptionally, oral venoms are a tractable model of trait evolution, being largely composed of proteinaceous toxins that have evolved in many tetrapods, ranging from reptiles to mammals. Given the diversity of venomous lineages, they are believed to have evolved convergently, even though biochemically similar toxins occur in all taxa.

Results: Here, we investigate whether ancestral genes harbouring similar biochemical activity may have primed venom evolution, focusing on the origins of kallikrein-like serine proteases that form the core of most vertebrate oral venoms. Using syntenic relationships between genes flanking known toxins, we traced the origin of kallikreins to a single locus containing one or more nearby paralogous kallikrein-like clusters. Additionally, phylogenetic analysis of vertebrate serine proteases revealed that kallikrein-like toxins in mammals and reptiles are genetically distinct from non-toxin ones.

Conclusions: Given the shared regulatory and genetic machinery, these findings suggest that tetrapod venoms evolved by co-option of proteins that were likely already present in saliva. We term such genes ‘toxipotent’—in the case of salivary kallikreins they already had potent vasodilatory activity that was weaponized by venomous lineages. Furthermore, the ubiquitous distribution of kallikreins across vertebrates suggests that the evolution of envenomation may be more common than previously recognized, blurring the line between venomous and non-venomous animals.

Keywords: Evolution, Venom, Phylogenetics, Kallikreins, Comparative genomics

Background

The extent to which shared history determines repeated evolution of traits remains an important and open question in evolutionary biology. Experiments replaying the

tape of life showed that phenotypes can arise through a combination of deterministic forces like natural selection and stochastic, non-deterministic forces like mutation and genetic drift [1]. The historical nature of evolution gives it a certain degree of ‘contingency’, such that past events can drastically alter evolutionary trajectories [1]. The role of contingency and chance in shaping evolution is substantial, so much so that a single positive mutation might allow a genetic system to thrive and tolerate less favourable mutations or even create scenarios where similar selection pressures might not lead to the same

* Correspondence: agneesh.barua@oist.jp; alexander.mikheyev@anu.edu.au
Agneesh Barua and Ivan Koludarov are co-first authors.

¹Ecology and Evolution Unit, Okinawa Institute of Science and Technology Graduate University, Okinawa, Japan

³Research School of Biology, Australian National University, Canberra, ACT, Australia

Full list of author information is available at the end of the article



© The Author(s). 2021 **Open Access** This article is licensed under a Creative Commons Attribution 4.0 International License, which permits use, sharing, adaptation, distribution and reproduction in any medium or format, as long as you give appropriate credit to the original author(s) and the source, provide a link to the Creative Commons licence, and indicate if changes were made. The images or other third party material in this article are included in the article's Creative Commons licence, unless indicated otherwise in a credit line to the material. If material is not included in the article's Creative Commons licence and your intended use is not permitted by statutory regulation or exceeds the permitted use, you will need to obtain permission directly from the copyright holder. To view a copy of this licence, visit <http://creativecommons.org/licenses/by/4.0/>. The Creative Commons Public Domain Dedication waiver (<http://creativecommons.org/publicdomain/zero/1.0/>) applies to the data made available in this article, unless otherwise stated in a credit line to the data.

evolutionary outcome [2, 3]. Therefore, tracing the evolutionary trajectory of genes can offer valuable information regarding the role of contingency and chance in shaping phenotypes. Selection on homologous and deeply conserved genetic mechanisms can repeatedly produce diverse phenotypes. For example, developmental toolkit genes regulate animal development and are involved in controlling differentiation among body axes, generating the extensive diversity in animal forms [4]. In plants, modifications of a shared developmental network have repeatedly led to the evolution of bilateral floral symmetry from a radially symmetric ancestor [5]. However, most traits are not controlled by such master regulators but emerge from complex interactions within polygenic networks. Yet, how regulatory complexity yields phenotypic novelty remains poorly understood.

To fully reveal the course of evolutionary changes, it is essential to have a good understanding of the link between genotype and the phenotype they produce [6–8]. But due to the complex nature of most biological traits, this link is rarely clear. Thus, while short-term evolution via quantitative genetic models is relatively easy to predict, how qualitatively novel traits arise repeatedly is less clear. Exceptionally, reptilian and mammalian oral venoms are proteinaceous cocktails where each constituent toxin can be traced to a specific locus, providing an unprecedented level of genetic tractability [9–12]. Venoms primarily evolve through sequence and gene expression changes of their constituent toxins, the phenotypic effects of which are clearly understood [10, 13–16]. Venoms are also excellent examples of convergent traits where individual toxins are believed to have been convergently recruited [11, 17, 18]. This high degree of convergence coupled with the genetic tractability of venom has allowed researchers to uncover genetic changes that contributed to the convergence of venom components, particularly in reptiles. For example, snake venom metalloproteinases (SVMP), which make up the primary component of viperid venoms, evolved through a series of deletions and tandem duplication from a single deeply conserved adam28 disintegrin [19]. Similarly, deletion and lineage specific expansion of phospholipase A2 (PLA2) lead to the evolution of novel venom phenotypes in some viperids [20, 21]. However, a similar tracing of genetic origins is still incomplete for the most ubiquitous toxin family in venom—the serine proteases.

Found in all kingdoms of cellular life as well as in viruses, serine proteases are perhaps the most widely distributed group of proteolytic enzymes [22]. Although best characterized in snakes, kallikrein-like (KLK-like) serine proteases are the main components in mammalian venom like that in *Blarina* shrews and *Solenodon*, as well as reptilian venoms in *Heloderma* lizards [11, 23, 24]. Yet, given the diversity of kallikrein types within

and between organisms, researchers recognized early on that “the kallikreins from different sources are not identical molecules, as originally assumed” [17]. This view has persisted to the present day, and even within mammals, co-option of KLK-like serine proteases into venom is believed to represent convergence [11]. By contrast, Fry and colleagues hypothesized the recruitment of kallikreins into reptile and mammal venoms could have occurred from a phylogenetically common source [25, 26]. Yet, distinguishing these hypotheses has been difficult until now given the vast number of serine proteases found in vertebrate genomes. Phylogenetic studies have not yet adequately sampled genes from reptilian and mammalian taxa and their phylogenetic relationships remain unresolved [27, 28]. Specifically, Hargreaves et al. [29] noted that “the orthology of previously published Toxiciferan Kallikrein genes is currently unclear”.

Here, we benefit from recent advances in genomics, which allowed us to reconstruct syntenic relationships between KLK-like toxins and their flanking genes in order to correctly identify paralogs dating back to a common tetrapod ancestor. We were then able to use phylogenetics to resolve the evolutionary origins of venom KLK-like genes. Our results show that mammalian and reptilian venom serine proteases have an origin distinct from other non-venomous KLKs and have been recruited into venom in parallel. This is in line with previous results that the repeated evolution of venom in vertebrates has occurred due to exaptation of already existing components rather than independent evolution of the similar components in different lineages.

Results

Genomic organization of the snake-venom like (SVL) and KLK loci

To determine the genetic history of the venom KLK-like toxins, we identified homologues of the kallikreins in the genomes of mammals, reptiles, amphibians. We specifically focused on tissue kallikreins (TKLs) which are abundant in tissues like pancreas, kidney, as well as in saliva. They have functions ranging from mediating blood pressure and muscle contraction to inflammatory cascades and pain induction [28]. Since they are also the gene family associated with toxicity of various animal venoms we restricted ourselves to only TKLs [30]. Mammalian kallikrein toxins are closely related to the KLK1 gene [11], and we will refer to them as KLK-like toxins. The reptilian counterparts are highly syntenic to snake venom serine protease (SVSP) in vipers (Additional file 1: Fig. S1). Therefore, we refer to their reptilian counterparts as snake venom-like (SVL) toxins.

In humans, TKLs are located in a cluster comprising 15 copies (genes KLK1 through KLK15) on the 19th chromosome (19q13.4). TKL clusters are also found in

other mammalian genomes, though the degree of synteny differs considerably. The KLK1 and KLK15 genes underwent tandem duplications in venomous mammals like solenodon and blarina [11, 31]. The expanded KLK1 genes contribute to the major toxin component of solenodon salivary and venomous secretions [11] (Fig. 1A). Unlike mammalian genomes, where KLK-like genes are contiguous, reptilian genomes have 2–3 gene clusters separated by several hundred kilobases and interrupted by other types of genes. One of these clusters contains genes that gave rise to viperid SVSPs (Fig. 1A). In highly venomous snakes like vipers, the expansion of snake venom serine protease (SVSP) genes is linked to the diversification of the venom phenotype [16, 32], paralleling expansions associated with the evolution of mammalian venoms. Thus, in both reptiles and mammals a single gene cluster gave rise to kallikrein-like serine protease toxins. However, the relationship between these genes is difficult to ascertain based on synteny alone and detailed phylogenetic analysis was needed.

Phylogeny of SVL and mammalian KLK genes

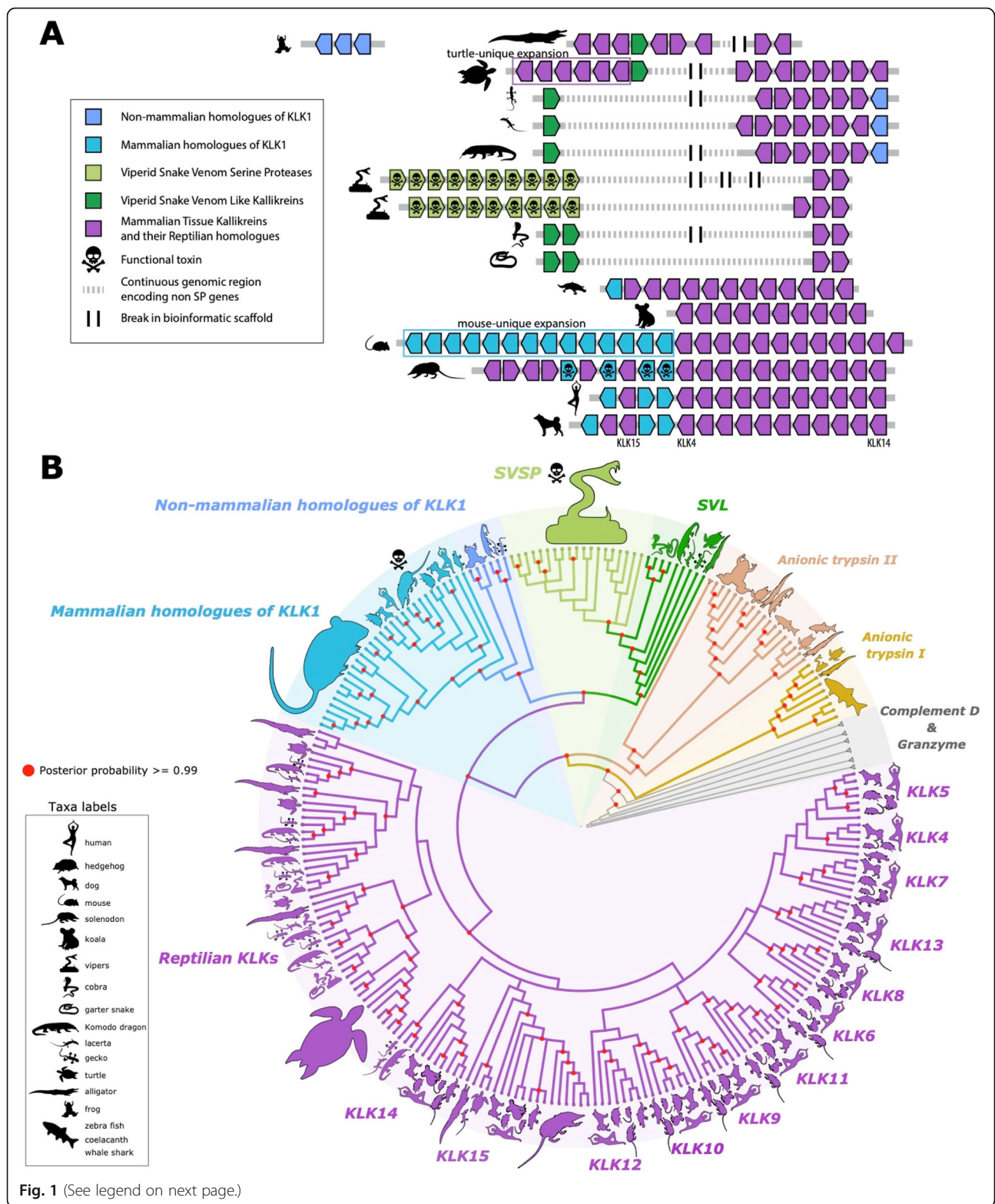
We conducted phylogenetic analyses to better understand relationships between and with TKL genes and to identify the likely origin of these genes. Since the TKL-like genes represent a large and diverse gene family, they were essential that we sample a wide repertoire of genes across a wide taxonomic distribution. To do this, we searched for sequences closely related to KLKs in mammals, reptiles, amphibians, and fish, as classified by NCBI. NCBI's classifications rely on a combination of calculated orthology and similarity in protein architectures based on sequences in the RefSeq database. This gene set included many non-KLK serine proteases like anionic trypsins, plasminogen, granzyme, and complement D, along with a list of all possible KLK-related sequences that are available in NCBI (with a combined total of a few thousand sequences). In order to isolate phylogenetically comparable genes, we used this large gene set (see the “Methods” section) as input for OrthoFinder. OrthoFinder classified genes into several large orthogroups. We isolated the orthogroup that contained TKL, SVL, and SVSP genes (Additional file 2) and resolved the phylogenetic relationship between genes within this group. This approach also allowed us to appropriately root our tree and reconstruct the early evolutionary history of TKLs.

We used a maximum-likelihood as well as a Bayesian approach to construct the phylogeny (see the “Methods” section). Both approaches yielded the same structure at each key nodes (discussed below) as well as comparable levels of support (Additional file 3 and Additional file 4). For the sake of brevity, we only display the Bayesian phylogeny (Fig. 1B) with Bayesian node supports at key

nodes. Using complement D and granzyme (Fig. 1B; grey branches) as outgroups, we observed a clear origin of TKLs from two groups of anionic trypsins that are shared between reptiles, amphibians, and fish. After the divergence from anionic trypsins, the TKLs split into two separate lineages. While most of the mammalian KLK branching is consistent with previously published mammalian TLK phylogenies [27, 28], our tree has better overall support; for instance, in Koumandou et al. [28], the divergence of mammalian KLK1-KLK2-KLK3 (mKLK1,2,3, includes KLK toxins) has a Bayesian node support of ~ 0.80 whereas our trees have a support > 0.99 . Additionally, we observe several new relationships between genes that were previously not described. First, the SVSP-SVL and mKLK1,2,3 genes formed a monophyletic clade sister to the other KLKs (Fig. 1B). This topology has high posterior probability (> 0.99) and was further supported by stepping-stone sampling (Bayes Factor of 111.0 in favour of monophyly between KLK1/2/3 and SVL-like genes vs. the monophyly of all KLK-like genes excluding SVSP-like genes). Within the SVL-mKLK1,2,3 clade, the reptilian and mammalian genes form their own sub-clades. The SVL genes appear to group according to the toxicofera classification, with SVL in cobra (*Naja naja*) and garter snake (*Thamnophis sirtalis*) forming a sister clade to the SVSP in elapids and vipers, while non-toxicofera like the leopard gecko (*Eublepharis macularius*) and the sand lizard (*Lacerta agilis*) forming individual lineages (Fig. 1B). Second, KLK15 and KLK14 in reptiles formed a clade with their mammalian homologs; however, several reptile KLKs formed separate reptile specific clades.

Selection analysis of SVL and mammalian KLK genes

The SVL genes in reptiles are homologous to SVSPs and could have a potential role in imparting toxicity to salivary secretions, as suggested for example in Anguimorph lizards [24]. Under this assumption, we would expect selection to vary in species believed to have toxic oral secretions, i.e. species belonging to the clade Toxicofera, as compared to non-toxicofera. To test the toxicofera hypothesis, we performed branch selection analysis using Phylogenetic Analysis by Maximum Likelihood (PAML) [33]. We applied a ‘free ratio’ model for branches leading up to toxicofera and compared its fit to a uniform ‘one ratio’ model for all branches. For a better representation of the toxicofera clade, we obtained additional anguimorpha kallikrein sequences from NCBI. We only included coding sequences that encoded for a mature protein and formed a monophyletic clade with our already identified SVL genes (Additional file 1: Fig.S4). We did not include venomous snakes in our test because higher selection for toxin genes in venomous snakes is already an established fact and could bias analyses [10,



34, 35]. The two-rate model fits significantly better (likelihood ratio test (LRT), $p < 0.001$) than the uniform one rate model suggesting that toxicoferan SVL genes experienced different selective pressures as compared to non-toxicoferans.

We performed the same analysis to test whether venomous mammals experienced different selection as compared to non-venomous ones. We use the KLK toxins in *Solenodon* genes and their homologs

(See figure on previous page.)

Fig. 1 Origins and diversification of tissue kallikreins (TKL). **A** TKL genes are located at a single genomic locus. In mammals, TKL genes are found in a single cluster, but in reptiles, they are scattered across two to three nearby clusters located several hundred kilobases apart. Venom evolution is associated with expansions of toxin-containing gene clusters, but there are also lineage-specific expansions that are not linked to venom evolution (e.g. turtles and mice). In existing genomic assemblies, the TKL clusters are often fragmented (represented by dashed line) across different scaffolds, but they share many common genes and are clearly contiguous (Supplementary Figure 1 and 2). **B** Phylogenetic analysis revealed that tetrapod TKLs originated from a common ancestor with vertebrate anionic trypsin, which are commonly expressed in the pancreas and are found elsewhere in the genome. TKLs diverged into two distinct clades, one comprising the KLK4-KLK15 lineages and the other the KLK1/2/3-SVSP/SVL lineage that contains toxipotent genes. Species silhouettes represent members of entire clades rather than a strict node to species demarcation. For a more conventional format, please refer to phylogeny (Supplementary Figure 2 and supplementary dataset 1) in supplementary. Serine protease-based toxins are homologs deriving from the same ancestral gene, implying that these toxins originated in parallel venoms in reptiles and mammals

in humans, dogs, and hedgehogs. The branches leading up to venomous mammals *Solenodon* and *Blarina* experienced selective forces significantly different from the rest of the tree (LRT, $p < 0.001$). While it is difficult to attribute positive selection as the reason for differences in selective pressures from this simple test, some branches (both in toxicofera and venomous mammals) did show high ω values (> 1) that are indicative of positive diversifying selection (Additional file 5 and 6). To get a better picture of the selective forces driving the evolution of the toxicofera and venomous mammals' clade, we performed several branch-specific tests using the Datamonkey server [36].

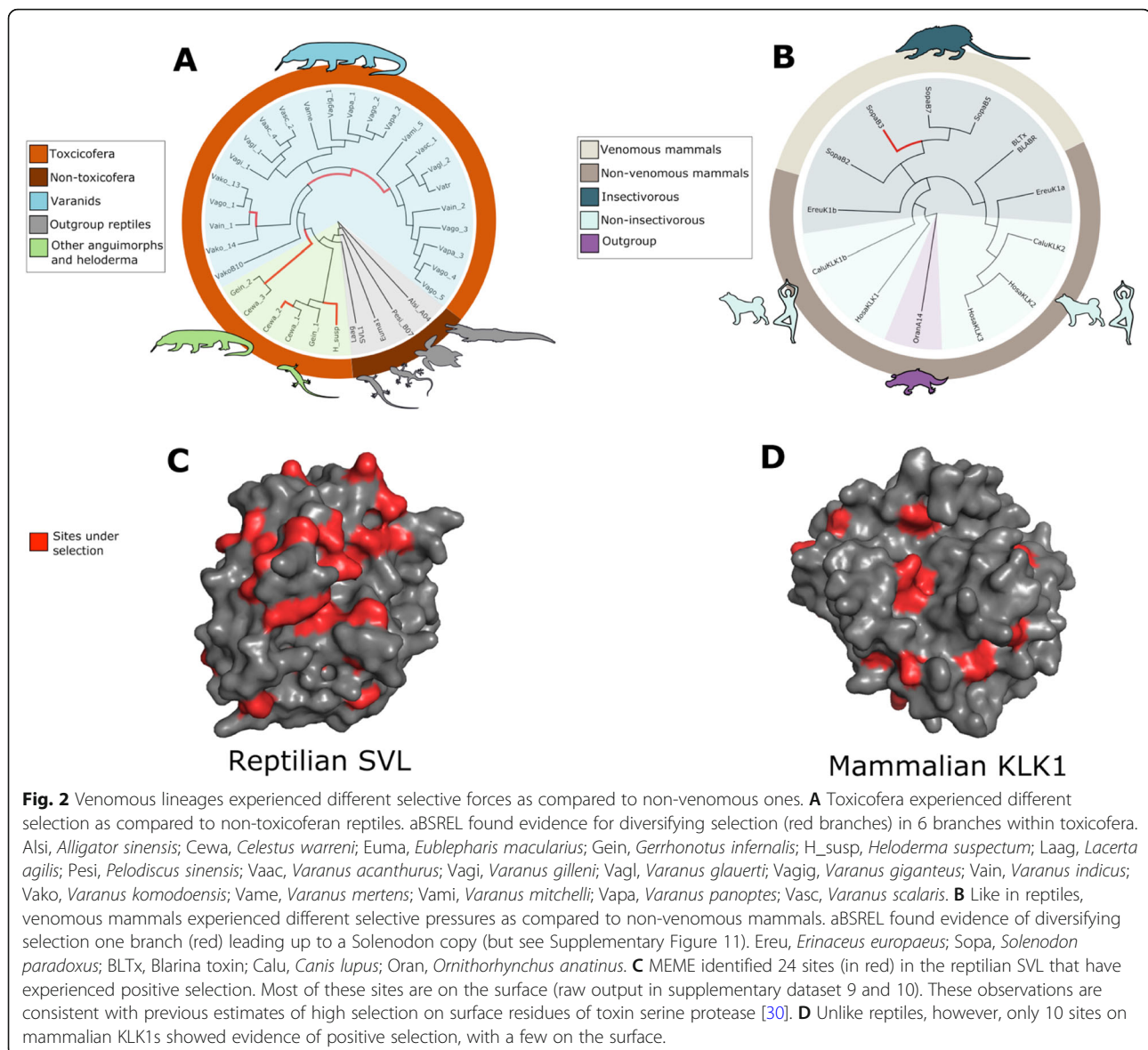
We first used the branch-site unrestricted statistical test for episodic selection (BUSTED) to check for evidence of episodic diversifying selection on any site in the gene along any of the branches of toxicofera and venomous mammals [37]. For both mammals and reptiles, BUSTED found evidence for diversifying selection in at least one site on at least one test branch (Additional file 1: Fig.S7, Fig.S8). Since BUSTED revealed joint evidence of branch and site-specific selection, we used the adaptive branch site random effects model (aBSREL) and mixed effects model of evolution (MEME) to get a better resolution of positive selection in branches of the phylogeny and sites along the gene respectively [38, 39]. Testing the same toxicofera and venomous mammal lineages, aBSREL found evidence for episodic diversifying selection in 1 branch leading to one of the *Solenodon* KLK1 copies, while in toxicofera, it found evidence in 6 branches, one of them leading to the heloderma gilatoxin, another leading to a SVL copy in Haitian giant galliwasp (the lizard *Celestus warreni*), and the rest in branches leading up to the radiation of varanids (Fig. 2A, B).

The MEME model identified several sites in reptilian SVL genes and mammalian KLK genes that showed significant evidence of positive selection ($p < 0.05$). In reptile SVLs, MEME identified 24 sites experiencing positive selection, while in the mammalian KLKs, 10 sites were identified (Fig. 2C, D). While some of these sites were in the internal structure of the proteins, the majority of them were on surface residues.

We did not include the mouse-specific KLK1 in our main analyses as they are an expansion exclusive to mice and form a clade separate from the other mammalian KLKs, including those believed venomous in *Solenodon* and *Blarina* (Fig. 1B). However, for the sake of consistency, we performed selection tests using PAML, BUSTED, aBSREL, and MEME using the mouse-specific KLK1s. Overall, PAML, BUSTED, and MEME produced the same results as the previous analysis; venomous mammals experienced different rates of selection. In addition to evidence of selection along the same *Solenodon* branch, aBSREL found evidence along the *Blarina* branch as well. The new results of selection analysis using the mouse sequences are found in the supplementary material (Additional file 1: Fig. S11, Fig. S12, Additional file 7-12). The large expansion of KLK1 in a lineage of mammals that are not venomous was fascinating. Using BUSTED and aBSREL, we tested for selection on venomous mammal lineages and the mouse expansion. Interestingly, both models found evidence of selection; BUSTED found evidence at the gene level and aBSREL showed evidence of selection in several specific mouse branches (Additional file 1: Fig. S13, Fig. S14). The functional relevance of this heightened selection is not clear, although there is evidence of sex-limited expression in mouse, suggesting a potential adaptive role in sex interactions [40].

Discussion

Non-deterministic forces can give rise to evolutionary novelties de novo. Several well characterized mechanisms like gene duplication, gene fusion, and horizontal gene transfer are responsible for the birth of new genes [41]. These new genes in turn contribute to species specific processes and generate morphological and physiological diversity [42]. Although non-deterministic processes produce genetic variation (on which natural selection acts), many adaptive traits can be exapted through modifications of already pre-existing characters [43]. Such exaptation has led to the origin of vertebrate oral venoms on at least two levels. Recent work has shown that the ancestral salivary gland gene regulatory



mechanisms were exapted in snake venom glands [44]. We now show that individual serine protease-based toxins used by diverse lineages share a common ancestor distinct from the ancestor of other non-toxin serine proteases. Thus, vertebrate venoms have evolved in parallel, at both the regulatory and also the genetic levels. This suggests that ancient shared history, namely salivary gland regulatory architecture and the presence of homologous genes biochemically suitable for toxicity, have facilitated venom evolution in distantly related taxa.

To determine the role of exaptation in venom evolution, it is important to understand the genetic makeup of adaptive traits, and how they lead to biochemical activity suitable for the envenomation. KLK1 genes in mammals and their reptilian homologs share kininogenase activity, which results in the release of bradykinin, a

potent hypotensive agent, when injected into the bloodstream [23, 45]. This is true even of salivary kallikreins of non-venomous mammals, such as mice, which can induce hypotension and even death [46–48]. Hypotension is also one of two major strategies which venomous snakes use to immobilize their prey [49]. The biochemical link between bradykinin-producing enzymes in mammals and snakes was evident to researchers who first characterized kallikrein-like properties of a snake venom enzymes, calling them “the salivary kallikrein of the snake” [50]. That being said, biochemical similarity does not imply homology. Schachter [17] wrote in an early review that “kallikreins from different sources are not identical molecules, as originally assumed, nor is it likely that they are derived from a parent molecule”. While the biochemical homology of kallikrein venoms is

now an accepted fact, the genetic homology and its role in the evolution of venoms was never extensively elaborated. Our analysis shows that genes underlying KLK-venom evolution in mammals and reptiles are homologous. Indeed, all KLK1 and SVL-like, and non-toxin KLK genes shared a common origin at the dawn of the tetrapods when they perhaps formed nearby gene clusters (Fig. 1A). However, even from within this family of paralogous proteases, venoms evolved from more closely related homologous genes as compared to the non-toxin KLKs (Fig. 1B).

Evolution of tetrapod venoms by kallikrein exaptation

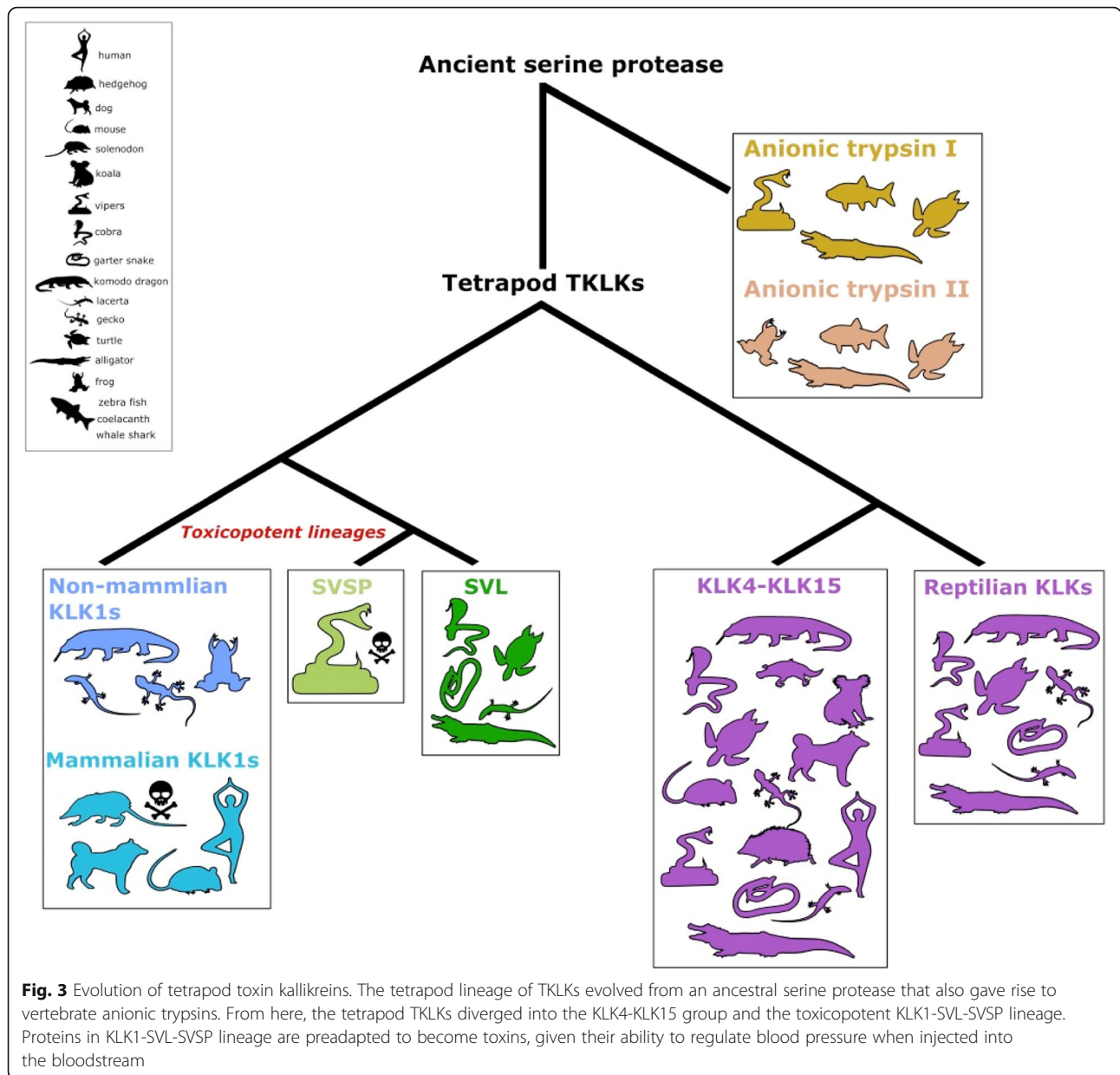
Most exaptations have bifunctional intermediates where both the old and new functions are preserved [51, 52]. This bifunctional nature likely allows for a gradual transition from one phenotypic state to another. For example, after gene duplication one or both the gene copies can perform its original function; or one copy can randomly acquire a new function in the course of accumulating neutral mutations [53]. This is the standard model of snake toxin evolution, which presuppose gene duplication prior to the acquisition of novel function (toxicity) [54, 55]. This is indeed observed in a recent study reconstructing the evolution of metalloproteinase toxins, which evolved from adam28 disintegrin by duplication and modification, such as the loss of a transmembrane domain improving solubility [19]. However, it appears that kallikreins already possess biochemical activity suitable for envenomation (vasodilation via bradykinin production); we have called such genes 'toxipotent'. Interestingly, serine protease genes in viperid snake venoms have undergone extensive duplication, with no clear distinction (like the loss of disintegrin domain for SVMP, or deletion of PLA2 genes in viperids [19, 21]) between an ancestral gene and its derived toxic counterparts, which is at odds with the classical venom evolutionary model. However, while there does not seem to be substantial differences in nucleotide or amino acid sequences between the gene copies, variations in gene expression, protein expression, or biochemical activity might exist. So far, genomes of venomous mammals and *Heloderma* lizards are insufficiently well characterized to test whether a specific genetic modification(s) gave rise to the toxin serine proteases.

In a previous publication, we proposed a unified model of early venom evolution in mammals and reptiles, suggesting that venoms evolved when kallikreins already present in saliva increased (via higher copy number) and became more effective (via sequence level changes) [44]. In this study, we were able to reconstruct the evolution of ubiquitous kallikrein-based toxins via phylogenetics based on extensive taxonomic sampling and gene orthologs accurately selected from the wide range of serine

proteases found in the genome based on phylogenetic and syntenic proximity. First, we found that copy number changes accompany the evolution of venom (e.g. snakes and *Solenodon*), but some lineages experience copy number expansions without evolving venom (mice and turtles, Fig. 1A). Second, we found that venomous taxa (*Gila monster* and *Solenodon*) indeed have a higher rate of nonsynonymous changes in the rates of venom evolution, consistent with selection for novel function (Fig. 2). Intriguingly, we also find evidence of selection in reptilian members of the Toxicofera clade, such as varanid lizards, where the existence of venom is debated (Fig. 2A) [56, 57]. From our results, the functional relevance of selection in the varanid lineage is not clear, though some studies have suggested a role of varanid oral secretions in prey procurement [29, 58, 59]. However, the presence of toxipotent genes in the saliva of many animals makes the line between venomous and non-venomous animals less clear. As most tetrapods already possess the requisite machinery for venom evolution, there could indeed be many taxa that lie on the continuum between what we currently perceive as venomous and non-venomous. Thus, the presence of serine proteases in saliva, and even sequence-level data suggesting past selection, may be insufficient to identify which animals are venomous. In order to do that, we need ecological evidence that animals, in fact, use their saliva for envenomation.

Conclusion

In this study, we expanded our knowledge on the phylogeny of kallikreins (KLKs) and, for the first time, with high certainty, resolved the relationship between tissue kallikreins (TKLKs) and their venomous counterparts in tetrapods. The tetrapod lineage of TKLKs evolved from an ancient serine protease that also gave rise to vertebrate anionic trypsins. From here, the tetrapod TKLKs diverged into the KLK4-KLK15 group and the toxipotent KLK1-SVL-SVSP lineage (Fig. 3). These toxipotent homologs eventually diversified and became a part of venom in snakes, some lizards, as well as some shrews and solenodon. We add to a long held belief that venoms primarily originate through a combination of constraint and convergence and show that shared history and parallel evolution (parallelism) can explain the repeated evolution of toxins in venoms. Parallelism is sometimes considered a process that led to the rise of phenotypic similarity in closely related species [8]. While this perspective can account for a shared molecular basis and history, the numerous exceptions to this prevents it from being definitive [60, 61]. It is more appropriate to consider parallelism as the use of shared molecular mechanisms to produce convergent phenotypes, irrespective of their taxonomic proximity [62]. We illustrate



this by showing that venom in mammals and reptiles originated multiple times in parallel by modifying the same gene family despite 300 million years separating these lineages. Thus, ancient conserved molecular mechanisms and building blocks can continue to be a source of adaptive novelty, allowing nature to replay the tape of life, albeit with a new perspective.

Methods

Genomic analysis

We used publicly available vertebrate genomes of good quality (Additional file 13) to establish location and synteny of the Kallikrein clusters. We used genomes for which RNA-seq verified genomic annotations were

available as a reference point and created an extensive map of the genes that flank SVL and TKL in those genomes. These include HPN, SCN1B, GRAMD1A, PSMC4, RBM42, HAUS5, and MAG (Additional file 1: Fig.S1, Fig.S2). That allowed us to establish syntenic relationships of those regions in different genomes. We then proceeded to use those flanking genes as a database to BLAST (NCBI-BLAST v.2.7.1+ suite, blastn, e-value cutoff of 0.05, default restrictions on word count and gaps) the genomes if they were less well annotated. That gave us a number of genomic scaffolds that potentially contained KLK genes. We used those for the second round of BLAST (tblastx, e-value cutoff of 0.01) against a database of exons extracted from well-annotated

mammalian TKL and viper SVL genes. Positive hits were checked by eye in Geneious v11 (<https://www.geneious.com>), and any complete exons were manually annotated and later merged into CDS of newly annotated genes if the exon order and count was in accordance with existing reliable KLK annotations. All resulting genes that produced viable mature peptides were then used for the phylogenetic analysis.

Phylogenetic analysis

All viable genes located in the previous step were translated into proteins and aligned with selected publicly available sequences of interest using L-INS-i method of MAFFT software v7.305 (Katoh and Standley 2013) with 1000 iterations (--localpair --maxiterate 1000). These parameters were used for all subsequent alignments. The publicly available serine protease sequences were obtained from NCBI. Using human KLK1 (gene ID: 3816) as a search query we obtained a list of all similar genes that were estimated based on synteny information and conserved protein domains. We selected sequences from Human (*Homo sapiens*), mouse (*Mus musculus*), dog (*Canis lupus familiaris*), hedgehog (*Erinaceus europaeus*), Lacerta (*Lacerta agilis*), garter snake (*Thamnophis elegans*), habu (*Protobothrops mucrosquamatus*), Chinese soft-shell turtle (*Pelodiscus sinensis*), alligator (*Alligator sinensis*), frog (*Xenopus tropicalis*), zebra fish (*Danio rerio*), coelacanth (*Latimeria chalumnae*), and whale shark (*Rhincodon typus*). These gene sets were used as input for OrthoFinder (OF). Using an mcl threshold of 1.2 OF grouped closely related genes into several orthogroups. We selected the orthogroup that contained SVSP-SVL-KLK1 sequences for a more rigorous phylogenetic analysis (Additional file 2). We selected complement D and granzyme (which were not present in the orthogroup mentioned above) as outgroups. Alignments were observed in Geneious v11 (<https://www.geneious.com>). As a sanity check, we made sure that known homologous parts of the molecule (like the cysteine backbone which is a prominent, highly conserved feature of serine proteases [30]) were aligned properly. A final alignment with 50% masked gaps was used to make the tree (Additional file 14). We constructed the Bayesian Phylogeny using MrBayes (v3.2.3) [63]. The analysis used a mixed amino acid model and was carried out across two parallel runs for 200 million generations [64], by which point the standard deviation of split frequencies reached 0.0065. Half of the trees were removed as burn-in and the rest summarized to compute posterior probabilities. We also computed Bayes factor support for monophyly of SVSPs and KLK1/2/3 vs. the monophyly of all KLK genes by stepping-stone sampling of tree space with corresponding backbone constraints for 50 million generations [65].

The maximum-likelihood phylogeny was constructed using PhyML (v3.3.2) [66]. PhyML selected the WAG +G+I model based on Akaike Information Criteria [67]. Branch supports were calculated using aBayes [68]. aBayes is a Bayesian-like transformation of approximate likelihood-ratio test (aLRT) that offers the highest power compared to other methods to estimate node support and values that have similar interpretation to Bayesian posterior probabilities [68].

Selection analysis

Alignments for sequence analysis were carried out using the MAFFT alignment tool, implementing the E-INS-i algorithm with BLOSUM62 as the scoring matrix [69]. All alignments were trimmed to remove signal peptide. The phylogeny was constructed based on a neighbour-joining tree using the Jukes-Cantor model. Additional anguimorpha kallikrein can be found in (Additional file 15). To test for selection on branches leading to venomous animals we used maximum likelihood models implemented in CodeML of the PAML package [33]. The log likelihood was compared between test branches (venomous animals) vs. background branches (non-venomous animals), and significant difference in models was determined using a log likelihood ratio test. Tests for adaptive evolution using BUSTED, aBSREL, and MEME analysis were carried out on the Datamonkey server [36]. The three-dimensional protein models for SVL and KLK1 were generated using a homology search implemented on the Phyre2 server [70] using consensus sequences obtained from the alignment of reptile SVLs and mammalian KLKs used in the selection analysis. PyMOL was used for visualization (PyMOL Molecular Graphics System, Schrödinger, LLC).

Abbreviations

aBSREL: Adaptive branch site random effects likelihood; BUSTED: Branch-site unrestricted statistical test for episodic selection; KLK: Kallikreins; LRT: Likelihood-ratio test; MEME: Mixed effects model of evolution; mKLK1,2,3: Mammalian kallikreins 1, 2, and 3; PAML: Phylogenetic analysis using maximum likelihood; SVL: Snake-venom like; SVSP: Snake venom serine protease; TKLs: Tissue kallikreins

Supplementary Information

The online version contains supplementary material available at <https://doi.org/10.1186/s12915-021-01191-1>.

Additional file 1: Figures S1-S14. Fig. S1- Synteny of SVSP and SVL genes. **Fig. S2-** Synteny of Kallikreins across mammals and reptiles. **Fig. S3-** Bayesian phylogeny of SVL-SVSP-KLKs. **Fig. S4-** Additional anguimorpha kallikrein sequences. **Fig. S5-** Test and background branches for reptiles. **Fig. S6-** Test and background branches for mammals. **Fig. S7-** Evidence ratio for BUSTED model in reptiles. **Fig. S8-** Evidence ratio for BUSTED model in mammals. **Fig. S9-** Phylogenetic tree showing aBSREL result for branch specific selection in reptiles. **Fig. S10-** Phylogenetic tree showing aBSREL result for branch specific selection in mammals. **Fig. S11-** Evidence ratio for BUSTED model in mouse KLK copies analysis. **Fig. S12-** aBSREL result in mouse KLK copies analysis. **Fig. S13-** Evidence ratio

for BUSTED model in venomous mammals with mouse KLK copies. **Fig. S14** - aBSREL result in venomous mammals with mouse KLK copies.

Additional file 2. – List of orthogroups.

Additional file 3. – PhyML tree file.

Additional file 4. – MrBayes tree file.

Additional file 5. – PAML results reptiles.

Additional file 6. – PAML results mammals.

Additional file 7. – BUSTED results mouse KLK copies.

Additional file 8. – aBSREL results mouse KLK copies.

Additional file 9. – MEME results mouse KLK copies.

Additional file 10. – PAML mouse KLK copies H1 test.

Additional file 11. – PAML mouse KLK copies H0 null.

Additional file 12. – Alignment of mouse KLK copies and other sequences.

Additional file 13. – Genome accessions.

Additional file 14. – Alignment of all sequences for main phylogeny.

Additional file 15. – Anguimorpha sequences.

Additional file 16. – BUSTED results reptiles.

Additional file 17. – BUSTED results mammals.

Additional file 18. – aBSREL results reptiles.

Additional file 19. – aBSREL results mammals.

Additional file 20. – MEME results reptiles.

Additional file 21. – MEME results mammals.

Additional file 22. – BUSTED results venomous mammals and mouse KLK copies.

Additional file 23. – aBSREL results venomous mammals and mouse KLK copies.

Additional file 24. – PhyML stats and standard out file.

Acknowledgements

A.B. would like to acknowledge the Scientific Computing and Data Analysis Section of the Okinawa Institute of Science and Technology Graduate University and Prof Vincent Laudet for his overall support.

Authors' contributions

A.B., I.K., and A.S.M conceptualized the study. I.K. annotated the sequences and performed the synteny analysis. All authors analysed the data and wrote the manuscript. All authors read and approved the final manuscript.

Authors' information

Twitter

Agneesh Barua: @AgneeshBarua

Alexander S Mikheyev: @SashaMikheyev

Funding

This study was funded by The Okinawa Institute of Science and Technology Graduate University.

Availability of data and materials

All data including sequences and output of phylogenetic programs is included within the article and its additional files.

Declarations

Ethics approval and consent to participate

Not applicable.

Consent for publication

Not applicable.

Competing interests

The authors declare no competing interests.

Author details

¹Ecology and Evolution Unit, Okinawa Institute of Science and Technology Graduate University, Okinawa, Japan. ²Animal Venomics Group, Justus Leibniz University, Giessen, Germany. ³Research School of Biology, Australian National University, Canberra, ACT, Australia.

Received: 26 August 2021 Accepted: 11 November 2021

Published online: 23 December 2021

References

- Blount ZD, Lenski RE, Losos JB. Contingency and determinism in evolution: replaying life's tape. *Science*. 2018;362(6415). <https://doi.org/10.1126/science.aam5979>.
- Vignogna RC, Buskirk SW, Lang GI. Exploring a local genetic interaction network using evolutionary replay experiments. *Mol Biol Evol*. 2021;38(8):3144–52. <https://doi.org/10.1093/molbev/msab087>.
- Xie VC, Pu J, Metzger BP, Thornton JW, Dickinson BC. Contingency and chance erase necessity in the experimental evolution of ancestral proteins. *Elife*. 2021;10. <https://doi.org/10.7554/eLife.67336>.
- Holland PWH. Evolution of homeobox genes. *Wiley Interdiscip Rev Dev Biol*. 2013;2(1):31–45. <https://doi.org/10.1002/wdev.78>.
- Citerne H, Jabbour F, Nadot S, Damerval C. The evolution of floral symmetry. In: Kader J-C, Delseny M, editors. *Advances in Botanical Research*. Academic Press; 2010. p. 85–137.
- Hoekstra HE, Coyne JA. The locus of evolution: evo devo and the genetics of adaptation. *Evolution*. 2007;61(5):995–1016. <https://doi.org/10.1111/j.1558-5646.2007.00105.x>.
- Stern DL, Orgogozo V. The loci of evolution: how predictable is genetic evolution? *Evolution*. 2008;62(9):2155–77. <https://doi.org/10.1111/j.1558-5646.2008.00450.x>.
- Conte GL, Arnegard ME, Peichel CL, Schluter D. The probability of genetic parallelism and convergence in natural populations. *Proc Biol Sci*. 2012;279(1749):5039–47. <https://doi.org/10.1098/rspb.2012.2146>.
- Rokyta DR, Lemmon AR, Margres MJ, Aronow K. The venom-gland transcriptome of the eastern diamondback rattlesnake (*Crotalus adamanteus*). *BMC Genomics*. 2012;13(1):312. <https://doi.org/10.1186/1471-2164-13-312>.
- Sunagar K, Moran Y. The rise and fall of an evolutionary innovation: contrasting strategies of venom evolution in ancient and young animals. *PLoS Genet*. 2015;11(10):e1005596. <https://doi.org/10.1371/journal.pgen.1005596>.
- Casewell NR, Petras D, Card DC, Suranse V, Mychajliw AM, Richards D, et al. Solenodon genome reveals convergent evolution of venom in eulipotyphlan mammals. *Proc Natl Acad Sci U S A*. 2019;116(51):25745–55. <https://doi.org/10.1073/pnas.1906117116>.
- Walker AA. The evolutionary dynamics of venom toxins made by insects and other animals. *Biochem Soc Trans*. 2020;48(4):1353–65. <https://doi.org/10.1042/BST20190820>.
- Rodríguez de la Vega RC, Giraud T. Intragenome diversity of gene families encoding toxin-like proteins in venomous animals. *Integr Comp Biol*. 2016;56:938–49.
- Safavi-Hemami H, Li Q, Jackson RL, Song AS, Boomsma W, Bandyopadhyay PK, et al. Rapid expansion of the protein disulfide isomerase gene family facilitates the folding of venom peptides. *Proc Natl Acad Sci U S A*. 2016;113(12):3227–32. <https://doi.org/10.1073/pnas.1525790113>.
- Margres MJ, Wray KP, Hassinger ATB, Ward MJ, McGivern JJ, Lemmon EM, et al. Quantity, not quality: rapid adaptation in a polygenic trait proceeded exclusively through expression differentiation. *Mol Biol Evol*. 2017;34(12):3099–110. <https://doi.org/10.1093/molbev/msx231>.
- Barua A, Mikheyev AS. Many options, few solutions: over 60 million snakes converged on a few optimal venom formulations. *Mol Biol Evol*. 2019;36(9):1964–74. <https://doi.org/10.1093/molbev/msz125>.
- Schachter M. Kallikreins and kinins. *Physiol Rev*. 1969;49(3):509–47. <https://doi.org/10.1152/physrev.1969.49.3.509>.
- Aminetzach YT, Srouji JR, Kong CY, Hoekstra HE. Convergent evolution of novel protein function in shrew and lizard venom. *Curr Biol*. 2009;19(22):1925–31. <https://doi.org/10.1016/j.cub.2009.09.022>.
- Giorgianni MW, Dowell NL, Griffin S, Kassner VA, Selegue JE, Carroll SB. The origin and diversification of a novel protein family in venomous snakes. *Proc Natl Acad Sci U S A*. 2020;117(20):10911–20. <https://doi.org/10.1073/pnas.1920011117>.

20. Chijiwa T, Deshimaru M, Nobuhisa I, Nakai M, Ogawa T, Oda N, et al. Regional evolution of venom-gland phospholipase A2 isoenzymes of *Trimeresurus flavoviridis* snakes in the southwestern islands of Japan. *Biochem J*. 2000;347(2):491–9. <https://doi.org/10.1042/bj3470491>.
21. Dowell NL, Giorgianni MW, Kassner VA, Selegue JE, Sanchez EE, Carroll SB. The deep origin and recent loss of venom toxin genes in rattlesnakes. *Curr Biol*. 2016;26(18):2434–45. <https://doi.org/10.1016/j.cub.2016.07.038>.
22. Di Cera E. Serine proteases. *IUBMB Life*. 2009;61(5):510–5. <https://doi.org/10.1002/iub.186>.
23. Kita M, Nakamura Y, Okumura Y, Ohdachi SD, Oba Y, Yoshikuni M, et al. Blarina toxin, a mammalian lethal venom from the short-tailed shrew *Blarina brevicauda*: isolation and characterization. *Proc Natl Acad Sci U S A*. 2004; 101(20):7542–7. <https://doi.org/10.1073/pnas.0402517101>.
24. Fry BG, Winter K, Norman JA, Roelants K, Nabuurs RJA, van Osch MJ, et al. Functional and structural diversification of the *Anguimorpha* lizard venom system. *Mol Cell Proteomics*. 2010;9(11):2369–90. <https://doi.org/10.1074/mcp.M110.001370>.
25. Fry BG. From genome to “venome”: molecular origin and evolution of the snake venom proteome inferred from phylogenetic analysis of toxin sequences and related body proteins. *Genome Res*. 2005;15(3):403–20. <https://doi.org/10.1101/gr.3228405>.
26. Fry BG, Roelants K, Champagne DE, Scheib H, Tyndall JDA, King GF, et al. The toxicogenomic multiverse: convergent recruitment of proteins into animal venoms. *Annu Rev Genomics Hum Genet*. 2009;10(1):483–511. <https://doi.org/10.1146/annurev.genom.9.081307.164356>.
27. Pavlopoulou A, Pampalakis G, Michalopoulos I, Sotiropoulou G. Evolutionary history of tissue kallikreins. *PLoS One*. 2010;5(11):e13781. <https://doi.org/10.1371/journal.pone.0013781>.
28. Koumandou VL, Scorilas A. Evolution of the plasma and tissue kallikreins, and their alternative splicing isoforms. *PLoS One*. 2013;8(7):e68074. <https://doi.org/10.1371/journal.pone.0068074>.
29. Hargreaves AD, Swain MT, Logan DW, Mulley JF. Testing the Toxicofera: comparative transcriptomics casts doubt on the single, early evolution of the reptile venom system. *Toxicon*. 2014;92:140–56. <https://doi.org/10.1016/j.toxicon.2014.10.004>.
30. Fry BG. *Venomous reptiles and their toxins: evolution, pathophysiology, and biodecovery*: Oxford University Press; 2015.
31. Hanf ZR, Chavez AS. A comprehensive multi-omic approach reveals a relatively simple venom in a diet generalist, the northern short-tailed shrew *Blarina brevicauda*. *Genome Biol Evol*. 2020;12(7):1148–66. <https://doi.org/10.1093/gbe/evaa115>.
32. Barua A, Mikheyev AS. Toxin expression in snake venom evolves rapidly with constant shifts in evolutionary rates. *Proceedings of the Royal Society B: Biological Sciences*. 2020;287(1926):20200613. <https://doi.org/10.1098/rspb.2020.0613>.
33. Yang Z. PAML 4: phylogenetic analysis by maximum likelihood. *Mol Biol Evol*. 2007;24(8):1586–91. <https://doi.org/10.1093/molbev/msm088>.
34. Juárez P, Comas I, González-Candelas F, Calvete JJ. Evolution of snake venom disintegrins by positive Darwinian selection. *Mol Biol Evol*. 2008; 25(11):2391–407. <https://doi.org/10.1093/molbev/msn179>.
35. Aird SD, Arora J, Barua A, Qiu L, Terada K, Mikheyev AS. Population genomic analysis of a pitviper reveals microevolutionary forces underlying venom chemistry. *Genome Biol Evol*. 2017;9(10):2640–9. <https://doi.org/10.1093/gbe/evx199>.
36. Weaver S, Shank SD, Spielman SJ, Li M, Muse SV, Kosakovsky Pond SL. Datamonkey 2.0: a modern web application for characterizing selective and other evolutionary processes. *Mol Biol Evol*. 2018;35(3):773–7. <https://doi.org/10.1093/molbev/msx335>.
37. Murrell B, Weaver S, Smith MD, Wertheim JO, Murrell S, Aylward A, et al. Gene-wide identification of episodic selection. *Mol Biol Evol*. 2015;32(5): 1365–71. <https://doi.org/10.1093/molbev/msv035>.
38. Smith MD, Wertheim JO, Weaver S, Murrell B, Scheffler K, Kosakovsky Pond SL. Less is more: an adaptive branch-site random effects model for efficient detection of episodic diversifying selection. *Mol Biol Evol*. 2015;32(5):1342–53. <https://doi.org/10.1093/molbev/msv022>.
39. Murrell B, Wertheim JO, Moola S, Weighill T, Scheffler K, Kosakovsky Pond SL. Detecting individual sites subject to episodic diversifying selection. *PLoS Genet*. 2012;8(7):e1002764. <https://doi.org/10.1371/journal.pgen.1002764>.
40. Karn RC, Laukaitis CM. Positive selection shaped the convergent evolution of independently expanded kallikrein subfamilies expressed in mouse and rat saliva proteomes. *PLoS One*. 2011;6(6):e20979. <https://doi.org/10.1371/journal.pone.0020979>.
41. Van Oss SB, Carvunis A-R. De novo gene birth. *PLoS Genet*. 2019;15(5): e1008160. <https://doi.org/10.1371/journal.pgen.1008160>.
42. Khalturin K, Hemmrich G, Fraune S, Augustin R, Bosch TCG. More than just orphans: are taxonomically-restricted genes important in evolution? *Trends Genet*. 2009;25(9):404–13. <https://doi.org/10.1016/j.tig.2009.07.006>.
43. Gould SJ, Vrba ES. Exaptation—a missing term in the science of form. *Paleobiology*. 1982;8(1):4–15. <https://doi.org/10.1017/S0094837300004310>.
44. Barua A, Mikheyev AS. An ancient, conserved gene regulatory network led to the rise of oral venom systems. *Proc Natl Acad Sci U S A*. 2021;118(14): e2021311118. <https://doi.org/10.1073/pnas.2021311118>.
45. Komori Y, Nikai T. Chemistry and biochemistry of kallikrein-like enzymes from snake venoms. *J Toxicol Toxin Rev*. 1998;17(3):261–77. <https://doi.org/10.3109/15569549809040394>.
46. Huang JC, Hoshino K, Kim YT, Chebib FS. Species and strain differences in the lethal factor of the mouse submandibular gland. *Can J Physiol Pharmacol*. 1977;55(5):1107–11. <https://doi.org/10.1139/y77-152>.
47. Hiramatsu M, Hatakeyama K, Minami N. Male mouse submaxillary gland secretes highly toxic proteins. *Experientia*. 1980;36(8):940–2. <https://doi.org/10.1007/BF01953804>.
48. Dean DH, Hiramamoto RN. Lethal effect of male rat submandibular gland homogenate for rat neonates. *J Oral Pathol*. 1985;14(8):666–9. <https://doi.org/10.1111/j.1600-0714.1985.tb00544.x>.
49. Aird SD. Ophidian envenomation strategies and the role of purines. *Toxicon*. 2002;40(4):335–93. [https://doi.org/10.1016/S0041-0101\(01\)00232-X](https://doi.org/10.1016/S0041-0101(01)00232-X).
50. Iwanaga S, Sato T, Mizushima Y, Suzuki T. Studies on snake venoms. XVII. Properties of the bradykinin releasing enzyme in the venom of *Agkistrodon halys blomhoffii*. *J Biochem*. 1965;58(2):123–9. <https://doi.org/10.1093/oxfordjournals.jbchem.a128173>.
51. Manley GA, Fay RR, Popper AN, editors. Springer, New York, NY: Evolution of the vertebrate auditory system; 2004. <https://doi.org/10.1007/978-1-4419-8957-4>.
52. Luo Z-X. Transformation and diversification in early mammal evolution. *Nature*. 2007;450:1011–1019, 7172, DOI: <https://doi.org/10.1038/nature06277>.
53. Innan H, Kondrashov F. The evolution of gene duplications: classifying and distinguishing between models. *Nat Rev Genet*. 2010;11(2):97–108. <https://doi.org/10.1038/nrg2689>.
54. Fry BG, Vidal N, van der Weerd L, Kochva E, Renjifo C. Evolution and diversification of the Toxicofera reptile venom system. *J Proteomics*. 2009; 72(2):127–36. <https://doi.org/10.1016/j.jprot.2009.01.009>.
55. Hargreaves AD, Swain MT, Hegarty MJ, Logan DW, Mulley JF. Restriction and recruitment-gene duplication and the origin and evolution of snake venom toxins. *Genome Biol Evol*. 2014;6(8):2088–95. <https://doi.org/10.1093/gbe/evu166>.
56. Fry BG, Wroe S, Teeuwisse W, van Osch MJ, Moreno K, Ingle J, et al. A central role for venom in predation by *Varanus komodoensis* (Komodo Dragon) and the extinct giant *Varanus* (Megalania) priscus. *Proc Natl Acad Sci U S A*. 2009;106(22):8969–74. <https://doi.org/10.1073/pnas.0810883106>.
57. Koludarov I, Jackson TN, den Brouw BO, Dobson J, Dashevsky D, Arbuckle K, et al. Enter the dragon: the dynamic and multifunctional evolution of *Anguimorpha* lizard venoms. *Toxins*. 2017;9. <https://doi.org/10.3390/toxins9080242>.
58. Sweet SS. Chasing flamingos: Toxicofera and the misinterpretation of venom in varanid lizards. Institute for Research and Development. Bangkok: Suan Sunandha Rajabhat University; 2016. https://www.researchgate.net/profile/Samuel_Sweet2/publication/307612937_Chasing_Flamingos_Toxicofera_and_the_Misinterpretation_of_Venom_in_Varanid_Lizards/links/57cee09e08ae582e0693853f.pdf
59. Borek HA, Charlton NP. How not to train your dragon: a case of a Komodo dragon bite. *Wilderness Environ Med*. 2015;26(2):196–9. <https://doi.org/10.1016/j.wem.2014.12.014>.
60. Steiner CC, Rompler H, Boettger LM, Schoneberg T, Hoekstra HE. The genetic basis of phenotypic convergence in beach mice: similar pigment patterns but different genes. *Molecular Biology and Evolution*. 2008;26(1): 35–45. <https://doi.org/10.1093/molbev/msn218>.
61. Manceau M, Domingues VS, Linnen CR, Rosenblum EB, Hoekstra HE. Convergence in pigmentation at multiple levels: mutations, genes and function. *Philosophical Transactions of the Royal Society B: Biological Sciences*. 2010;365(1552):2439–50. <https://doi.org/10.1098/rstb.2010.0104>.
62. Rosenblum EB, Parent CE, Brandt EE. The molecular basis of phenotypic convergence. *Annu Rev Ecol Evol Syst*. 2014;45(1):203–26. <https://doi.org/10.1146/annurev-ecolsys-120213-091851>.

63. Ronquist F, Teslenko M, Van Der Mark P, Ayres DL, Darling A, Höhna S, et al. MrBayes 3.2: efficient Bayesian phylogenetic inference and model choice across a large model space. *Syst Biol.* 2012;61(3):539–42. <https://doi.org/10.1093/sysbio/sys029>.
64. Altekar G, Dwarkadas S, Huelsenbeck JP, Ronquist F. Parallel Metropolis coupled Markov chain Monte Carlo for Bayesian phylogenetic inference. *Bioinformatics.* 2004;20(3):407–15. <https://doi.org/10.1093/bioinformatics/btg427>.
65. Xie W, Lewis PO, Fan Y, Kuo L, Chen M-H. Improving marginal likelihood estimation for Bayesian phylogenetic model selection. *Syst Biol.* 2011;60(2):150–60. <https://doi.org/10.1093/sysbio/syq085>.
66. Guindon S, Dufayard J-F, Lefort V, Anisimova M, Hordijk W, Gascuel O. New algorithms and methods to estimate maximum-likelihood phylogenies: assessing the performance of PhyML 3.0. *Syst Biol.* 2010;59(3):307–21. <https://doi.org/10.1093/sysbio/syq010>.
67. Lefort V, Longueville J-E, Gascuel O. SMS: smart model selection in PhyML. *Mol Biol Evol.* 2017;34(9):2422–4. <https://doi.org/10.1093/molbev/msx149>.
68. Anisimova M, Gil M, Dufayard J-F, Dessimoz C, Gascuel O. Survey of branch support methods demonstrates accuracy, power, and robustness of fast likelihood-based approximation schemes. *Syst Biol.* 2011;60(5):685–99. <https://doi.org/10.1093/sysbio/syr041>.
69. Katoh K, Standley DM. MAFFT multiple sequence alignment software version 7: improvements in performance and usability. *Mol Biol Evol.* 2013;30(4):772–80. <https://doi.org/10.1093/molbev/mst010>.
70. Kelley LA, Mezulis S, Yates CM, Wass MN, Sternberg MJE. The Phyre2 web portal for protein modeling, prediction and analysis. *Nat Protoc.* 2015;10(6):845–58. <https://doi.org/10.1038/nprot.2015.053>.

Publisher's Note

Springer Nature remains neutral with regard to jurisdictional claims in published maps and institutional affiliations.

Ready to submit your research? Choose BMC and benefit from:

- fast, convenient online submission
- thorough peer review by experienced researchers in your field
- rapid publication on acceptance
- support for research data, including large and complex data types
- gold Open Access which fosters wider collaboration and increased citations
- maximum visibility for your research: over 100M website views per year

At BMC, research is always in progress.

Learn more biomedcentral.com/submissions

