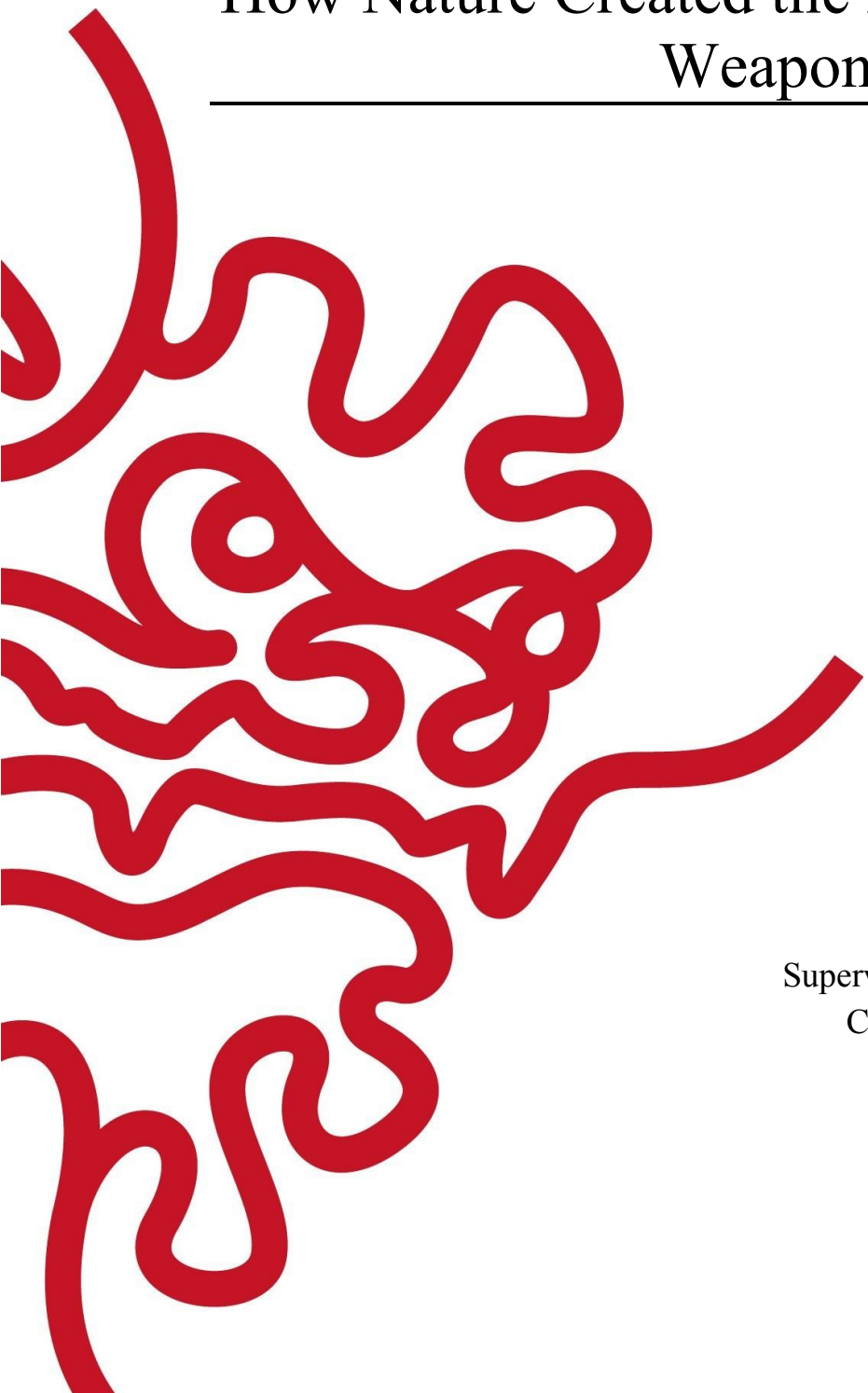
Okinawa Institute of Science and Technology
Graduate University

Thesis submitted for the degree

Doctor of Philosophy

The Evolutionary Genetics of Venoms: How Nature Created the Perfect Chemical Weapon



by

Agneesh Barua

Supervisor: **Alexander S. Mikheyev**
Co-Supervisor: **Vincent Laudet**

June 16th 2022

Declaration of Original and Sole Authorship

I, Agneesh Barua, declare that this thesis entitled "The evolutionary genetics of venoms: How nature created the perfect chemical weapon" and the data presented in it are original and my own work.

I confirm that:

- No part of this work has previously been submitted for a degree at this or any other university.

- References to the work of others have been clearly acknowledged. Quotations from the work of others have been clearly indicated, and attributed to them.

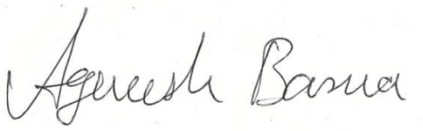
- In cases where others have contributed to part of this work, such contribution has been clearly acknowledged and distinguished from my own work.

None of this work has been previously published elsewhere, with the exception of the following:

  - **Barua A**, Koludarov I, Mikheyev AS. Co-option of the same ancestral gene family gave rise to mammalian and reptilian toxins**.** *BMC Biology*. 2021;19: 268.
  - **Barua A**, Mikheyev AS. An ancient, conserved gene regulatory network led to the rise of oral venom systems. *Proceedings of the National Academy of Sciences USA*. 2021;118:2021311118
  - **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;652;287:20200613.
  - **Barua A**, Mikheyev AS. Many Options, Few Solutions: Over 60 Million Snakes Converged on a Few Optimal Venom Formulations. *Molecular Biology and Evolution*. 2019;36:1964–74.

- Authorization of release has been obtained from all co-authors.

Date: 16th June 2022

Signature: 

Abstract

Venomous animals have fascinated humans for millennia. How nature shaped a simple biological secretion into a potent chemical weapon is a testament to evolution's power and versatility. However, the early origins and genetic mechanisms of venom evolution are not clearly understood. Venoms consist of proteinaceous cocktails where each protein can be mapped to a specific gene; I utilized this genetic tractability to uncover the molecular and genetic mechanisms behind its evolution. Using a combination of quantitative genetics, transcriptomics, and phylogenetics, I have identified specific mechanisms that led to the origin of oral venoms in mammals and reptiles. Oral venoms originated from an ancient conserved gene regulatory network whose primary role was maintaining cellular homeostasis during increased protein production. This ancient system could tolerate high protein loads, facilitating the parallel recruitment of various diverse protein families into the ancient venom. Venom complexity then increased by sequence and copy number variation of toxins. High copy numbers contributed to this system's phenotypic flexibility, allowing it to further diversify through changes in evolutionary rates and by altering the combinations of toxins used. These features enabled evolution to refine venom cocktails to form optimal formulations. I provide the first unified and deep evolutionary model describing the early steps in forming a venom system and show how millions of years of evolution produced venom phenotypes in extant lineages. All chapters of this thesis have been peer-reviewed and published.

Acknowledgements

I would like to acknowledge my supervisor Alexander Mikheyev for inspiring me to think beyond the norms and strive for openness, rigor, and thoroughness in science. I would also like to thank my co-supervisor and mentor for the last few months of my PhD, Vincent Laudet, whose mentorship and academic outlook have shaped my ideals and motivations for science. I would like to thank the Scientific Computing and Data Analysis (SCDA, HPC) specifically Jan Moren for his patience, guidance, and encouraging enthusiasm. I would also like to extend my gratitude to the Sequencing Section (SQC) for their help with sequencing. Lastly, I would like to thank my friends, colleagues, and family, without whom this entire journey would've been impossible.

List of Abbreviations

Bayesian analysis of macroevolutionary mixture (BAMM)

Cysteine rich secretory protein (CRISP)

Endoplasmic reticulum associated degradation (ERAD)

Gene regulatory networks (GRNs)

Kallikreins (KLK)

Phospholipase A2 (PLA2)

Snake venom metalloproteinase (SVMP)

Snake venom like (SVL)

Snake venom serine protease (SVSP)

Three finger toxin (TFTx)

Unfolding protein response (UPR)

List of publications

I. **Barua A**, Koludarov I, Mikheyev AS. Co-option of the same ancestral gene family gave rise to mammalian and reptilian toxins**.** *BMC Biology*. 2021;19: 268.

Contributions
AB and ASM conceptualized and designed the study. IK annotated sequences and constructed synteny map. AB and ASM constructed the phylogeny. AB performed selection analysis. All three authors wrote the manuscript.

II. **Barua A**, Mikheyev AS. An ancient, conserved gene regulatory network led to the rise of oral venom systems. *Proceedings of the National Academy of Sciences USA*. 2021;118:2021311118.

Contributions
AB and ASM conceptualized and designed the study. AB carried out the analysis. All authors wrote the manuscript.

III. **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;652;287:20200613.

Contributions
AB and ASM conceptualized and designed the study. AB carried out the analysis.. All authors wrote the manuscript.

IV. **Barua A**, Mikheyev AS. Many Options, Few Solutions: Over 60 Million Snakes Converged on a Few Optimal Venom Formulations. *Molecular Biology and Evolution*. 2019;36:1964–74.

Contributions
AB and ASM conceptualized and designed the study. AB carried out the analysis.. All authors wrote the manuscript.

Nomenclature

Metavenom network: The metavenom network represents groups of housekeeping genes whose expression is tightly associated with venom genes. The network primarily consists of genes involved in the UPR and ERAD pathways, functioning to maintain cellular homeostasis during periods of high protein load. This thesis, and its respective publication, provides the first description and characterization of this network which was later confirmed to be a feature of diverse venom systems across the animal kingdom [1,2].

Toxipotent: We introduce this term to describe lineages of ancient kallikreins that possessed a biochemical activity capable of producing a toxic effect when injected into another animal. This ancient lineage represents the ancestral states of KLK1, SVL, and SVSP toxins found in extant venomous mammals and reptiles, respectively.

Table of Contents

List of figures:

Introduction

Genetics and the evolution of complex traits

John Maynard Smith, one of the pioneers of evolutionary theory, was fascinated by the bee orchid. This seemingly inconspicuous flower had evolved petals that resembled and even smelled like a female bee. Males bees would be attracted to these flowers and try to mate. Rather than successfully copulating with the flower, the males were instead covered with pollen. When the male bees moved to another flower, they would transfer the carried pollen and aid in cross-pollination. John Maynard Smith wondered how nature could produce something so complex and utterly unexpected. Indeed, this puzzle is at the heart of genetics and evolutionary biology. The intricate shape of the bee orchid is an example of a complex trait.

The origins of specific adaptations with a simple genetic basis have been described in wild populations of animals like mice, lizards, and birds [3–5]. In some, the effects of particular mutations have also been functionally verified [6,7]. However, a single gene coding for a specific trait is the exception rather than the rule; most traits in nature are complex traits.
Complex or quantitative traits are controlled by many genes and are influenced by the environment. The study of complex traits is vital to our understanding of biology and has several practical applications. For example, body shape and body size are complex traits, and studying their evolution will help us understand how novelties arise in form and function. Like diabetes and cancer, most diseases are also complex traits whose genetic characterization is paramount in developing new therapies. Due to their importance, complex traits have been extensively studied, and through the advancements in next-generation sequencing, our understanding of the genetic underpinnings of complex traits has improved manifold. Despite this, the way complex traits first arise and evolve is still largely a mystery.

The early 2000s saw an extensive debate regarding the mechanistic origins of complex traits and adaptations. Proponents of evo-devo regarded *cis*-regulatory changes as the most important evolutionary force behind the origin of new traits [8]. Whereas evolutionary geneticists implored that more direct genetic changes like gene duplication, alternate splicing, or recruitment of coding domains were the primary forces driving the formation of new traits [9]. The scales have shifted in favor of both theories throughout the years, resulting in a stalemate.
A feature uniting this dichotomy of verisimilitude is the importance of gene expression, specifically gene expression variation. The evo-devo framework postulates that mutations in *cis*-regulatory regions can produce the differences in timing or location of gene expression needed to form evolutionary innovations [8]. However, as pointed out by proponents of evolutionary genetics, genetic variation in cis-regulatory regions is not the only way to alter gene expression. Gene duplication increases gene dosage [9,10]. Indeed, gene expression variation is responsible for forming many traits and has been well studied.

Gene expression variation and adaptation

Gene expression is closely tied to various aspects of animal biology, such as physiological and ecological states. In honey bees, the differences in gene expression influence their role in the colony. Non-reproductive worker bees have a higher expression of genes involved in metabolism, foraging behavior, visible light detection, and synaptic transmission. In contrast,

fertile workers have a higher expression of oogenesis-related genes [11]. The alternate expression profiles suggest that non-reproductive worked bees play a greater role in energy-intensive and high-risk behaviors like foraging; contrastingly, the reproductive females restrict themselves to brooding activities [11]. Gene expression variation not only influences complex traits like behavior but can produce phenotypic diversity that increases adaptive potential. In plants, gene expression variation alters signaling pathways that affect multiple processes to produce diverse phenotypes [12]. By changing signaling pathways without affecting normal circadian rhythms, gene expression variation can overcome the constraints imposed by pleiotropy and produce novel phenotypes [9,12]. The above examples align with the idea that changes in gene expression are brought about by naturally occurring allelic variation; however, changes in non-coding regions can also influence gene expression and produce advantageous phenotypes.

In northern European populations of *Drosophila melanogaster*, a deletion in the non-coding region of the *MtnA* gene leads to increased gene expression imparting a higher resistance to oxidative stress [13]. This is an instance where a modification in the *cis*-regulatory region impacts gene expression leading to an adaptive advantage. Adaptive advantage due to *cis*-regulatory variation is not restricted to only *Drosophila*. *Cis*-regulatory expression quantitative trait loci for *Adam17* and *Bcat2* genes were associated with variations in body mass in mice [14]. A deletion in the cis-regulatory region for *SWS1* opsin gene and multiple deletions in *trans*-regulatory regions contribute to visual diversity in African cichlids [15]. Interestingly, *cis*-regulatory variations have also been linked to changes in protein abundance, thereby acting as a source of phenotypic variation in humans [16].

The functional relevance of gene expression variation is easy to determine for large-effect alleles; however, establishing functional relevance gets more complicated when traits comprise many genes of minor effects [17]. Furthermore, unlike gene sequences whose evolution can be compared and tested against a well-developed null model, several features like transcriptomic noise, a weak mutation-gene expression relationship, and the absence of ancestral gene expression information have prevented the formulation of efficient models for gene expression [18]. The evolution of complex phenotypes is no doubt complicated, and the discordance between genotype and phenotypes poses an added challenge. Given this situation, traits with high genetic tractability, i.e. traits whose genetic origins can be clearly traced, can offer valuable insight into how phenotypes originate and the genetic mechanisms driving their evolution.

Venom and its unique features

Venoms and venomous animals have fascinated humans for millennia, and for good reason. “Is it venomous?” is an age-old question humans and other animals had to answer quickly to survive. Venoms are biological substances produced by an organism that comprise toxins that cause an adverse physiological effect when injected into other organisms; they are produced by or stored in specialized organs and are delivered through external injury. Venoms have independently evolved at least a hundred times throughout the animal kingdom to accomplish several functions such as predation, defense, and feeding [19]. Venoms are also highly convergent traits. Several protein (toxins) families were convergently recruited into the venoms of animals as diverse as snakes, scorpions, spiders, shrews, cone snails, and sea anemones [20,21]. The various toxins in venom have unique biochemical functions and often work in an agonistic manner, causing a cascade of physiological reactions that subdue the envenomed animal [20]. Although proteins form the main bioactive component in venoms, they are just

one class of molecules present in venom. Apart from proteins, venoms comprise small peptides, salts, organic molecules like polyamines, and neurotransmitters [20,22–24]. In addition to being biochemical cocktails, venoms also have a strong ecological component. As venom primarily functions when injected into another, an interaction between organisms is vital, without which a venomous context ceases to exist [25].

As a trait, venom has several useful features that make it an ideal system to study complex trait evolution. First, venom activity is primarily determined by the composition of venom. Therefore, venom is a complex trait whose phenotype depends on several genes coding for different proteins. Defensive venoms like those in fish and insects have low complexity and often exhibit a single activity causing localized pain [21]. By contrast, venoms used for predation are highly complex, with a myriad of biochemical activities [21] Individual components in the venom and their relative abundance influence the biochemical activity of the venom. For example, venoms of elapids like cobras and kraits are predominantly neurotoxic owing to the presence of three-finger toxins (TFTx) which bind the nicotinic acetylcholine receptor. In contrast, the major component of viper venoms like Phospholipase A2 (PLA2) and snake venom metalloprotease (SVMP) makes it hemotoxic and often leads to necrosis in their prey [26]. (One of the chapters of this thesis delves more deeply into the taxonomic distribution of toxin families and its implication for venom evolution).
Second, the individual components that make up the venom can be traced to distinct genetic loci [27–30]. This degree of genetic tractability helps overcome the discordance between genotype and phenotype, enabling researchers to classify venom as a polygenic phenotype and linking changes in venom activity to genetic changes at specific loci. The high genetic tractability also makes it possible to trace specific genetic modifications that increase the adaptive potential of venoms. For instance, intragenic deletions and domain loss (membrane-tethering domain, cysteine-rich domain, and disintegrin domain) in viperid SVMPs accelerated the evolution of novel paralogs and gave rise to the three major classes of secreted SVMP toxins [31,32]. The ancestor of rattlesnakes possessed a specialized heterodimeric neurotoxin that got independently deleted in lineages like the Eastern and Western Diamondback rattlesnakes; in contrast, the Mojave rattlesnake retained the neurotoxic gene while losing the PLA2 myotoxin gene [33]. These studies show how specific genetic changes in toxin gene loci can explain venom activity in different species.
Third, venoms primarily evolve through changes in gene expression, especially in snakes. Studies revealed a substantial concordance between mRNA and toxin protein levels in the venom, implying that post-transcriptional mechanisms have a minor role in the phenotypic variation of snake venoms [34,35]. Furthermore, the expression of highly abundant toxins is typically conserved between closely related species, although there is evidence suggesting that more abundant proteins tend to evolve rapidly in a microevolutionary time scale [29,36,37]. While the link between gene expression and venom variation is clear in snakes, it is not as straightforward in other animals. For example, spiders show highly divergent patterns of toxin gene expression where the expression of venom gland biased genes is not more conserved than the expression of more broadly expressed gene families [38]. Additionally, protein sequence variations were high in venom gland specific transcripts, and positive selections did not appear to be preferentially higher in venom gland biased genes [38]. Lacking a centralized venom system, cnidarians offer unique insights into how venom systems can function and evolve. Cnidarians show an impressive degree of spatial and ontogenetic heterogeneity in the expression of venom toxins [39–41]. For instance, different tissues in Nematostella modulate the expression of various toxins in response to other biotic interactions [40]. It has also been hypothesized that differences in regulatory variation alter toxin expression levels to generate multiple unique venom profiles [41]. The above examples show that (while there are

exceptions) gene expression variation is a major mechanism that produces variability in venom systems, thereby acting as an important target for natural selection.

In summary, the compositional nature of the venom phenotype, high genetic tractability, and importance of gene expression variation makes venom an ideal trait to study the molecular mechanisms that cause complex trait evolution. With the advent of the high-throughput sequencing era, the genomes of many venomous animals have been sequenced, providing unprecedented insight into their biology. The surge in large scale genomic data has opened up entirely new research directions and have enabled researchers to answer questions that previously seemed impossible.

Evolution of venom

Historically, most research on venom evolution focused on individual toxin families. Research on individual toxins have provided a treasure trove of knowledge into the ways evolution shapes venoms and have laid the conceptual foundations for future studies to decipher the molecular origins of venom systems as a whole.
One of the most essential aspects of venom evolution is forming the tissue system that produces venom. A central paradigm in the evolution of venom systems is that toxin proteins are recruited into specialized venom-secreting cells [20]. Therefore, it is crucial to understand the distinct origins of various venom systems. The developmental origins of venom systems are highly diverse. In hymenopteran (wasps, ants, and bees), venom glands are believed to have originated from a sexual accessory gland in ancestral hymenopterans, owing to its striking homology to the ovipositor [42]. Unlike in hymenopterans, venom systems of spiders, scorpions, and centipedes evolved from specific groups of cells that formed internalized glands in the chelicerae (fangs), telson (tail-segment), and forcipules (modified legs), respectively [43]. In snakes, the venom system typically consists of a gland that produces venom and a delivery system comprising musculature and hollow fangs [44,45]. Fangs in snakes, especially in colubrids, have undergone multiple periods of loss due to dietary shifts and prey capture mode [46]. Several studies supported that front and rear fangs are homologous and likely evolved from a rear-fanged ancestor [45–47]. However, the development and evolution of the venom gland in snakes is not very clear. Early evidence suggested that venom glands evolved from salivary glands owing to their shared origin from oral epithelial tissue [48]. An alternative hypothesis proposed that venom glands in snakes originated from the pancreas, primarily supported by the expression of a microRNA (miR-375) [49]. More evidence is needed to verify this hypothesis, especially considering miR-375 has activity in diverse tissues, not only the pancreas [50]. A recent study successfully cultured snake venom gland organoids that propagated in response to factors that promote mammalian salivary tissue proliferation, implying a developmental origin similar to salivary glands [51]. One of the chapters of this thesis offers insight into the gene regulatory networks of snake venom glands and helps resolve their evolutionary origins. Developmental origins of venom apparatus can also vary between life stages, where the formations of different types of venom systems coincide with transitions in life-cycle. In *Hydra,* cnidocytes originate from interstitial cells found throughout the mid-gastric regions of the endoderm; these interstitial cells differentiate into the cnidocytes from within a post-Golgi vesicle [52,53]. The starlet sea anemone *N. vectensis* possesses three kinds of cnidocytes that develop at varying densities between the planula and early polyp stage [47]. Venom glands may develop differently in cone snails depending on whether the cone snails feed at the early larval stage. *Conus anemone* develops venom glands from the ventral glandular region of the foregut if the typically non-feeding larval stage consumes prey [54].

Venom systems have unique developmental dynamics, which are only just being realized. The many unanswered questions, the advent of genomic resources, and the development of new model organisms set the stage for new and exciting discoveries in evo-devo of venom systems.

Venom's exceptional genetic tractability provides an excellent opportunity to investigate the relationship between natural selection, genetics, and molecular processes that generate diversity in complex traits. Many toxins in venom evolve through the 'birth and death' process of gene evolution, mediated by gene duplication [21]. Gene duplication is one of the most widespread mechanisms leading to genetic novelty [55]. Although several models describe evolution after gene duplication, it is believed venom evolution follows Onho's model of neofunctionalization [55,56]. The main principle behind neofunctionalization is the loss of selective constraint followed by the acquisition of a new function [56]. After gene duplication, the original copy (or one of the copies) maintains its original function. At the same time, the other is free to evolve, accumulate variation, alter its expression, and eventually develop a new role. Non-toxin physiological genes undergo gene duplication, after which one of the copies is recruited into the venom system; once a part of the venom system, additional gene duplication can occur, resulting in large multi-locus gene families encoding a variety of toxins [21]. Recent genomic studies provided an in-depth look into the chromosomal arrangements and structural genetics of toxin genes. The importance of gene duplication was highlighted in studies that observed the prevalence of toxins in genes on microchromosomes of snakes, which are highly prone to recombination and producing gene copies [57,58].

Following gene duplication, toxins experience accelerated rates of evolution and positive selection [59]. Indeed, positive selection appears to be ubiquitous among venomous animals, including snakes, scorpions, spiders, and cone snails [21]. Positive selection also predominantly acts on surface-exposed amino acid residues [60,61]. Restricting selection (or variation) to surface-exposed residues has two critical features. First, it ensures that the structural core of the protein is preserved, ensuring stability. Second, changes in surface-exposed residues can produce new and increased affinities to receptors, thereby increasing toxin activity and producing novel effects. Although positive selection is the primary force behind venom evolution, genetic drift acts as a combinatorial force, especially in snake venom [29].

Another fascinating feature of venoms is that they are highly convergent. Several protein families have been convergently recruited for venom use in numerous lineages [20]. Interestingly, some venom proteins are also recruited by hematophagous insects that use them in their feed secretions, having much of their neurotoxicity and hematotoxicity in common with venoms [20]. Convergent evolution usually occurs at two levels. There is a high degree of convergence at the level of biochemical targeting, where venom targets the primary physiological process and tissues accessible via the bloodstream [21]. The other level of convergence occurs in the molecules selected for use as toxins. At least fourteen unique protein families have been convergently recruited into the venoms of taxa as diverse as cnidarians, cephalopods, fish, reptiles, and insects. For instance, PLA2s has been recruited into the venoms of cephalopods, cnidarians, insects, snakes, and scorpions, while Kunitz-type toxins were independently recruited into venoms of reptiles, cone snails, spiders, and cnidarians [20]. The diversity of venomous lineages and the high degree of convergence has prompted a recent interest in understanding the degree of convergence in biological components beyond just the toxins [62]. Questions regarding the convergence of both venomous and non-venomous components is one of the focus areas of this thesis.

Thesis outline

Despite the incredible diversity of venomous animals, venom systems in snakes are the most well studied. The preference for studying the venom system in snakes is primarily due to its medical importance. Indeed, envenomation through snake bites affects almost 1.5 million people a year, which prompted the World Health Organization to classify snakebite as a neglected tropical disease [63]. There is substantial next-generation sequencing data available from snakes, including high-quality genomes and RNA-seq of venom glands, allowing large-scale comparative analysis of venom evolution. Although this thesis primarily focuses on the evolutionary genetics of venom in snakes, the results can be extrapolated to any oral venom system in tetrapods. Additionally, the mechanistic insights can be used to understand the ways complex traits can evolve, improving our understanding of the ways evolution created diversity in form and function.

The first chapter of this thesis looks at the origin of oral venom systems in snakes. Rather than focusing on toxins in the venom, we instead characterized the gene regulatory network of the entire venom gland. Studying the non-venom components of the venom system provided information on the biological processes associated with venom expression. We go on to compare the gene expression of this network in salivary tissues from other amniotes to identify any conserved patterns in expression. Lastly, we show how components of this network experienced higher rates of selection in venomous snakes, highlighting their role in venom evolution. Chapter two explores the origin of one of the most ubiquitous venom components in vertebrates, serine proteases. We traced the evolutionary history of kallikreins and serine proteases using genomic and phylogenetic approaches to reveal a common origin. Because of their wide distribution across vertebrates and the presence of an already potent vasodilatory activity, salivary kallikreins were likely one of the first salivary components to become co-opted into the vertebrate oral venom system. Chapter three provides a macroevolutionary perspective by estimating the tempo and mode of evolution of the snake venom phenotype. We show how evolution has constantly changed the venom phenotype through shifts and alteration to the combination of toxins used to formulate the venom. We provide a link between molecular evolution and macroevolutionary processes that worked together to shape extant venom phenotypes. Lastly, the conclusion of the thesis combines the results from all chapters to propose a unified deep-evolutionary model of venom evolution and provides new conceptual frameworks for future research directions.

Chapter 1

The beginning - An ancient, conserved gene regulatory network led to the rise of oral venom systems.

This chapter has been published as:

- Barua A, Mikheyev AS. **An ancient, conserved gene regulatory network led to the rise of oral venom systems**. *Proceedings of the National Academy of Sciences USA*. 2021;118:2021311118. (Appendix I)

Introduction

How oral venom systems originated is still poorly understood. Most studies typically focus on the evolution of toxins comprising venom. However, using toxins present in modern-day animals to trace the evolutionary origins of venom systems is challenging; toxins tend to evolve rapidly, have complex expression patterns, and were typically incorporated into the venom system after it was formed. Instead, we focused on the gene regulatory network associated with venom production in snakes.

Gene regulatory networks (GRNs) aim to identify interacting genes based on a common expression profile [64]. Gene co-expression networks are widely used to construct GRNs because of their reliability in capturing biologically relevant interaction between genes, as well as their high power in reproducing known protein-protein interactions [65,66]. Genes are identified using clustering methods (like hierarchical clustering) and placed into "modules" [67]. Comparing the expression profiles of modules across taxa can identify pivotal drivers of phenotypic change and identify the earliest targets of natural selection [68,69].

Summary of results

We used a gene co-expression network characterized from the venom gland of *Protobothrops mucrosquamatus* to identify the genes associated with venom production. We term this network the "metavenom network". The metavenom network comprises genes involved in the unfolding protein response (UPR) and endoplasmic reticulum-associated degradation (ERAD) pathways. This network was unique to the venom gland as module preservation (a metric to assess the conservation of network structure) was high for venom glands, like that from cobra, and low for other tissues in snakes.

UPR and ERAD pathways were highly conserved across taxa. To check whether the metavenom network was also conserved, we compared gene expression of metavenom orthologs from different tissues in various taxa. The expression patterns of orthologs between the venom gland in snakes and the salivary glands in mammals were surprisingly well conserved. The metavenom network modules were significantly preserved in mammalian salivary glands but not in tissues like the kidney. This suggests that while mammalian salivary and snake venom glands have diverged considerably, they still share a common regulatory core.

Several genes in the metavenom network belonged to gene families with several gene copies. Increasing gene copy number is crucial in bringing about evolutionary novelty, especially in venom systems. We determined whether gene families evolved more rapidly in venomous lineages than in other taxa in our dataset. Our results showed that gene families not only

evolved more rapidly in lineages leading to venomous snakes, but they had also undergone more significant expansions as compared to other taxa. Gene copies that experienced significant expression were involved in protein ubiquitination and protein modification.

Discussion

Genes with evolutionarily conserved expression patterns represent functionally critical genes where coregulation is advantageous [70]. As such, the metavenom network had a vital role in the evolution of venoms. The UPR and ERAD pathways act as "quality control" machinery, ensuring that proteins undergo proper folding and maturation [71]. Several components of the metavenom network work synergistically to ensure proper protein folding. The UPR anticipates, detects, and correctly folds misfolded proteins; the ERAD ensures that misfolded proteins are degraded, preventing cellular toxicity, and the ubiquitin ligase system provides an overall level of regulation to fine-tune these processes [72]. The metavenom network improved the tenacity of the ancient secretory system, enabling it to tolerate an increase in complexity through composition and changes in concentrations of secreted proteins. Therefore, the metavenom network likely primed the transition from a simple ancestral secretory system to complex venom systems found in extant species. Recent studies found that the UPR and ERAD pathway are conserved across venom systems of diverse taxa, validating our hypothesis that the pathways are essential regulatory components of venoms [1,2].

We also described two stages of oral venom formation from an ancestral salivary GRN. The first stage involved the exaptation of components already present in the saliva. Considering their large scale taxonomic distribution and well characterized hypotensive activity, we hypothesized that kallikreins (KLK) were likely one of the first secretory proteins to be co-opted into the primitive venom system. After the initial co-option, different protein families were recruited into the venom systems, increasing their complexity and toxicological effect. In the following chapter, we trace the evolution of the kallikrein protein family to better resolve the relationship between venomous kallikrein and their non-toxin counterparts.

Chapter 2

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

This chapter has been published as:

- Barua A, Koludarov I, Mikheyev AS. **Co-option of the same ancestral gene family gave rise to mammalian and reptilian toxins.** *BMC Biology*. 2021;19: 268. (Appendix II)

Introduction

We previously hypothesized that kallikreins were likely one of the first proteins recruited into the venom. Kallikrein-like serine proteases are major components of venoms in mammals like *Blarina* shrew and *Solenodon,* and reptilian venoms like those in snakes and *Heloderma* lizards [73–75]. However, the genetic origins of kallikreins and their relationship with venomous serine proteases like snake venom serine protease (SVSP) are unclear. Several hypotheses exist. Some suggested that "the kallikreins from different sources are not identical molecules, as originally assumed" [76]; this idea has persisted to the present day with mammalian KLK-like serine proteases believed to be independently recruited into mammalian venom [75]. On the other hand, Fry and colleagues suggested that the recruitment of KLKs into mammalian and reptilian venoms could have occurred from a phylogenetically common source [20,77]. Here we use genomic and phylogenetic data to distinguish between these hypotheses. By comparing genomes across vertebrates and phylogenetically sampling all known serine proteases, we resolved the relationships between the multiple tissue KLKs, their venomous counterparts, and SVSPs.

Summary of results

Using a combination of conserved synteny and sequence alignment, we identified tissue KLK orthologs across mammalian, reptilian, and amphibian genomes. In reptiles, we identified groups of KLK genes that were highly syntenic and had high sequence similarity with SVSP found in vipers. We termed these genes as snake venom like (SVL) toxin genes. Venomous mammals and vipers expanded KLK1 and SVSP genes, respectively. Although their expansion is indicative of their use in venom, the relationships between these genes cannot be determined from the synteny alone.

We conducted a phylogenetic analysis to resolve the relationships between the genetic homologs of KLKs (the toxin and non-toxin ones). To improve the power of our analysis, we included a broad sampling of KLK, KLK-like, and non-KLK serine proteases from mammals, reptiles, amphibians, and fish. Both the maximum likelihood and Bayesian phylogenies had high concordance and could resolve relationships (with high support nodes) between groups of KLKs. Our phylogeny revealed that all tissue KLKs share a common ancestor with anionic trypsins of reptiles, amphibians, and fish. The tissue KLKs then split into two separate lineages, forming the non-toxin KLKs (KLK4 to KLK15, and reptilian specific KLKs) and the other forming a group comprising KLK1/2/3, SVL, and SVSP. Because of their high similarity and relatedness with venomous KLKs like KLK1 in shrews and *Solenodon* and SVSP in snakes,

we term the latter lineage as “toxipotent”. The topology of KLK1/2/3 forming a sister clade with SVL and SVSP had posterior support of >0.99. It was further supported by stepping-stone sampling, with a Bayes Factor of 111.0 favoring monophyly between KLK1/2/3, SVL, and SVSP, *vs* the monophyly of all KLK-like genes excluding the SVSP genes. In other words, we found evidence that KLK1/2/3 and SVL-SVSP are significantly more likely to form a monophyletic clade than any other combination of branches.

Lastly, we carried out tests for selection to determine whether the SVL genes and the toxicofera clade, in general, experienced different selection compared to non-toxicofera. We found that toxicoferan SVL genes did experience significantly different selective pressures than non-toxicofera. Using various branch and site-specific selection models, we identified branches and sites that experienced higher selection. We observed the same trends in genes from venomous mammals. Our selection analysis highlights the unique evolutionary patterns of genes with a toxin potential and shows how they already experience selective forces different from their non-venomous counterparts.

Discussion

Adaptive traits can evolve through modifications of pre-existing characters [78]. We show that individual serine protease-based toxins used by diverse lineages share a common ancestor distinct from other non-toxin serine proteases. Coupled with the existence of a pre-existing regulatory core, we concluded that vertebrate oral venoms evolved in parallel [72]. The shared history of a salivary regulatory architecture and the presence of homologous genes that were biochemically suitable for toxicity facilitated the evolution of oral venoms in distantly related taxa.

Our conclusions are, to an extent, in line with the “restriction” hypothesis proposed by Hargreaves [79]. The hypothesis suggests that rather than being recruited from different tissues into the venom system, venom-encoding genes may have originated from genes that were ancestrally expressed in multiple tissues, including salivary and ancestral venom glands [79,80]. However, further diversifications through either subfunctionalization or neofunctionalization are still expected once recruited into the venom gland. Although the restriction hypothesis does highlight a current uncertainty regarding the timing of duplication events (prior to or following recruitment), it doesn't explicitly contradict the theory of recruitment but instead provides an alternate and seemingly parsimonious explanation of the early origins of venom systems [58,80].

Chapter 3

The phase of increasing complexity - Changes in evolutionary rates and combinations of toxins produced optimal venom formulations.

This chapter has been published as:

- 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;652;287:20200613. (Appendix III)
- Barua A, Mikheyev AS. **Many Options, Few Solutions: Over 60 Million Snakes Converged on a Few Optimal Venom Formulations**. *Molecular Biology and Evolution*. 2019;36:1964–74. (Appendix IV)

Introduction

The first two chapters focus on the origins and early evolution of venom systems. The current chapter looks at the macroevolution of the venom phenotype as a whole. We classified the venom phenotype as a combination of the ten most abundant toxins characterized in the venom gland transcriptome across 52 different snake species, spanning the three families of venomous snakes. We sought to answer two important questions through a combination of quantitative genetics and phylogenetic comparative methods. First, are there any evolutionary constraints in the toxins used to form the venom phenotype? Second, what is the tempo and mode of phenotypic evolution of snake venom?

Summary of results

Interactions (or constraints) between genes have a substantial effect in channeling adaptive responses, and covariance between genes play a vital role in shaping complex traits by determining evolutionary trajectories through natural selection [81]. Using a modern rendition of a phylogenetic generalized linear mixed model based on the animal model in quantitative genetics, we estimated the degree of covariance (which is a measure of constraint) between gene expression levels of toxins comprising the venom phenotype [82,83]. We found that toxin combinations are not constrained, which suggests that throughout evolution, venoms have the potential to comprise any combination of any toxin. However, despite this lack of constraint, the phylomorphic space had low dimensionality, with two principal components explaining >74% of the data. On the phylomorphic space, snakes clustered around four toxins, TFTx, SVSP, SVMP, and PLA2, which were driving most of the variation in the dataset. We termed these four toxins ‘major’ venom components representing distinct envenomation strategies. Vipers followed multiple strategies, from a strategy primarily centered around SVMP to one employing a mixture of PLA2 and SVSP. Most elapids followed a strategy primarily around TFTx, while a few used a combination of TFTx and PLA2. Lastly, colubrids (non-front fanged snakes in the dataset) have at least one species using the above strategies. Clustering of distantly related snakes around similar strategies hint at the importance of parallelism. The

'many options/few solutions' pattern of snake venom evolution has been observed and verified in by other researchers [84,85].

Key innovations are adaptations that provide ecological opportunity by enabling the utilization of previously unexplored niches [86–88]. Typically, key innovations are believed to contribute to species diversifications or adaptive radiation; however, key innovations should not be considered the sole reason for differential rates of species diversification [89]. Instead, key innovations should be regarded as adaptations allowing entry into novel ecological niches or adaptive zones. Studies should aim to identify specific shifts in tempo and mode of phenotypic evolution of the assumed key innovation [89]. We used the Bayesian analysis of macroevolutionary mixtures (BAMM) to determine the tempo and mode of venom evolution in our comparative venom expression dataset [90]. We observed several rate shifts along the venomous snake phylogeny, indicating that evolutionary rates for toxins have changed throughout their evolution. All the toxin families analyzed have experienced at least one rate shift since becoming part of the venom arsenal. Toxins like SVMP, TFTx, and Cysteine-rich secretory protein (CRISP) showed a larger distribution of high evolutionary rates in the ancestral lineages, following which they experienced subsequent slowdowns in extant species. SVMP experienced higher rates in vipers, while TFTx experienced higher rates in elapids. A '*rate through time'* analysis which determined the rates of toxin expression evolution revealed that the most abundant toxins such as SVMP, SVSP, PLA2, and TFTx had higher rates of evolution than minor components. We fitted several trait evolution models to our data to better understand which evolutionary process best describes snake venom evolution. We tested the Brownian motion, Ornstein-Uhlenbeck, early-burst, and jump models (pulsed models) using the pulsR package [91]. Based on weighted Akaike scores, the jump model best fitted our data.

Discussion

Our results showed that the snake venom phenotype evolves rapidly, and the lack of constraint facilitated the exploration of a vast phenotypic space. These two mechanisms give venomous snakes an edge in the evolutionary arms race, enabling them to evolve novel venom formulations to counteract resistance developed in prey species. The results also highlight the importance of venom as a key innovation but question its role in adaptive radiation and speciation (a view shared by other experts like Luke Harmon [92]. More research on how venom influences processes like niche partitioning and reproductive isolation can provide clues as to the direct role (if any) of venom in the evolution of species.

Conclusion

By combining the results from the above chapters, we propose a model of venom evolution (Fig. 1). Although the model was informed from results using reptiles, and oral venom glands, the evolutionary mechanisms can apply to any kind of predatory venom that immobilizes or kills its prey even ones that have diverse developmental origins like venom glands in insects and cone snails [1,2].

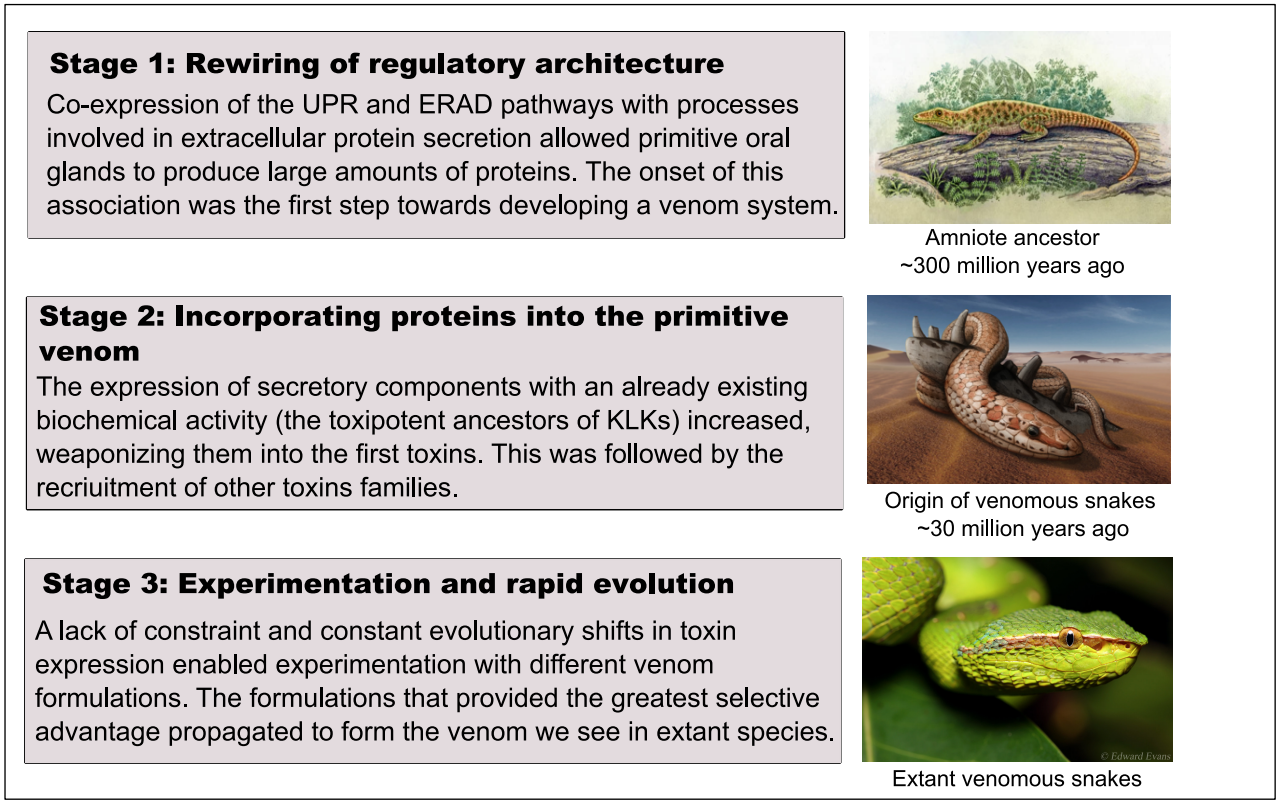


Fig. 1: Schematic describing the various stages of venom evolution in snakes and tetrapods. The right panel showing time and ancestral lineages are approximations based on current knowledge about the ancestors of amniotes and venomous snakes. All images in the above figure are in the public domain.

Step 1: Rewiring the regulatory architecture of ancient secretory glands.
Increased expression and association of UPR and ERAD pathways with other processes and pathways involved in extracellular protein secretion allowed primitive (oral) glands to produce large amounts of proteins. This association laid the molecular foundations to develop a complex venom system. The convergence in the expression of genes associated with the UPR and ERAD pathways across metazoan venom systems suggests that the onset of this regulatory association was likely one of the first steps towards developing a venom system [1,72].

Step 2: Incorporating proteins with toxin potential into primitive venom.
Following the rewiring of the metavenom network, the expression of secretory components with an already existing biochemical activity increased, weaponizing them into toxins. Toxipotent lineages that gave rise to mammalian KLKs, SVL, and SVSP were the first components to become a part of the primitive venom systems, at least in tetrapods [93]. The

increased expression of kallikrein-like toxins in the primitive venom increased adaptive potential, setting the stage for further diversification and recruitment of different protein families (novel components like PLA2) via gene duplications.

Step 3: Experimentation and rapid evolution to form extant venom compositions.
A lack of constraint and constant shifts in rates of toxin expression enabled experimentation with different venom formulations, allowing venoms to explore an expansive phenotypic space [94,95]. The formulations that provided the greatest selective advantage propagated to form the venom we see in extant species [95].

Although this thesis describes the early steps in venom evolution, additional genetic and molecular mechanisms are likely behind the transition from a non-venomous state to a venomous one. To identify these other mechanisms, we need to study non-venomous taxa and closely related venomous taxa with different forms of venom systems. These taxa provide alternate conditions for comparison. For instance, closely related non-venomous taxa represent a null condition to test hypotheses regarding features that are essential for venom formation, while comparing alternate venom systems can tell us how specific features evolve. Such comparisons can also be used to ask important questions about the evolution of complex traits.
All evolutionary lineages have experienced different degrees of trait gain and trait loss, such that extant lineages exhibit various configurations of ancestral and derived characters [96]. Forces shaping these configurations, the genetic modifications behind their origins, and the ecological forces driving their formation are vital to understanding complex trait evolution. Reptiles like *Heloderma* and snakes belong to the toxicofera clade but have evolved very different venom systems. Venom systems of *Heloderma* comprise slotted teeth and venom glands located on their lower jaw; at the same time, snakes have specialized fangs that deliver venom from a venom gland located on their upper jaw [97]. Studying these alternate venom systems can shed light on the development of homologous tissues, the origin of novel cell types, and the molecular mechanisms behind tissue organization. These comparisons can also be used to study how differences in ecologies might influence different envenomation strategies. Varanid oral secretions, *Heloderma* venom, and snake venom have several components in common, like Phospholipases, metalloproteases, serine proteases, and L-amino oxidases [97,98]; however, the potency of venom from these three lineages vary considerably likely in response to their different feeding ecologies [98].
Similarly, rear-fanged snakes from the superfamily Colubroidae have developed multiple strategies for prey procurement despite evolving from an ancestor believed to be venomous [99]. Rear fanged snakes use constriction, envenomation, or a combination of both to immobilize prey [25]. Venom systems in rear-fanged snakes are phenotypically and ecologically very different from that of front-fanged snakes such as Viperidae and Elapidae; studying them can shed light on the evolutionary mechanisms shaping a phenotype into alternate forms [25,100]. Furthermore, it is important to study the genetic backdrop of such alternate venom systems to reveal how the ancestors of venomous snakes evolved venom systems over evolutionary time. “Long before venom could rapidly subdue prey, it may have only slightly weakened it” [25]. This was likely the case considering that evolution occurs in gradual steps, even in lineages believed to evolve rapidly. Since some rear-fanged snakes combine envenomation and constriction to immobilize prey, they may represent a (an ancestral) phenotype at the interface of non-venomous, slightly venomous, and highly venomous.

In addition to the model describing the early origins of venom systems, another major contribution of this thesis is providing evidence of molecular conservation between oral secretory tissues of venomous and non-venomous tetrapods. Oral secretory systems like mammalian salivary gland and reptilian venom glands share essential housekeeping functions and secrete similar components like kallikreins. This surprising similarity blurs the line between venomous and non-venomous, suggesting that animals may lie at different positions on the venomous-non-venomous spectrum. However, here it is important to remember the historical context of evolution; the characters we see today result from millions of years of biological change in response to particular environmental pressures. Just because two organisms started from the same starting point doesn't mean they'll end up at the same destination. The molecular conservation between such disparate oral glands highlights the importance of vital and frequently overlooked aspect of venom systems, their close association with ecology [19,25].
Venoms are usually the primary means of prey procurement. Therefore the relevance of a venom system is only realized when there is a predator-prey relationship. Additionally, the importance of venom systems makes them an integral part of an animal's ecology. For example, shrews have exceedingly high metabolic rates, which means they have to feed constantly. Shrews use their venom to immobilize prey in a comatose state and hoard them for future consumption, thereby ensuring a constant source of nutrition [101]. On the other hand, rats in an urban environment can scavenge and survive on human leftovers. Therefore, despite sharing similar molecular components, the ecological pressure for a rat to develop a venom system is low, whereas it is crucial for the shrew. Venoms also play a broader role beyond just predation and defense, although this remains understudied [19]

The high genetic tractability, close association with ecology, and multiple sources of functional data make venoms an appealing system to study various facets of complex trait evolution. Interest is growing in expanding the direction of venom research by focusing on phylogenetic comparative methods and the developmental underpinning of venom systems [62,102]. These new directions have the advantage of decades of fundamental research in toxinology, ecology, wildlife biology, and natural history to inform their results and observations. The future of the venom field is bright, and perhaps in a few years, we can see venom systems joining the ranks of anoles and Darwin's finches as textbook examples of phenotypic evolution.

References

1. Zancolli G, Reijnders M, Waterhouse RM, Robinson-Rechavi M. Convergent evolution of venom gland transcriptomes across Metazoa. Proc Natl Acad Sci U S A. 2022;119. doi:10.1073/pnas.2111392119

2. Pardos-Blas JR, Tenorio MJ, Galindo JCG, Zardoya R. Comparative Venomics of the Cryptic Cone Snail Species Virroconus ebraeus and Virroconus judaeus. Mar Drugs. 2022;20: 149. doi:10.3390/md20020149

3. Hoekstra HE, Hirschmann RJ, Bundey RA, Insel PA, Crossland JP. A single amino acid mutation contributes to adaptive beach mouse color pattern. Science. 2006;313: 101–104. doi:10.1126/science.1126121

4. Rosenblum EB, Römpler H, Schöneberg T, Hoekstra HE. Molecular and functional basis of phenotypic convergence in white lizards at White Sands. Proc Natl Acad Sci U S A. 2010;107: 2113–2117. doi:10.1073/pnas.0911042107

5. Lamichhaney S, Han F, Berglund J, Wang C, Almén MS, Webster MT, et al. A beak size locus in Darwin's finches facilitated character displacement during a drought. Science. 2016;352: 470–474. doi:10.1126/science.aad8786

6. Yokoyama S, Tada T, Zhang H, Britt L. Elucidation of phenotypic adaptations: Molecular analyses of dim-light vision proteins in vertebrates. Proc Natl Acad Sci U S A. 2008;105: 13480–13485. doi:10.1073/pnas.0802426105

7. Geffeney SL, Fujimoto E, Brodie ED, Ruben PC. Evolutionary diversification of TTX-resistant sodium channels in a predator–prey interaction. Nature. 2005;434: 759–763. doi:10.1038/nature03444

8. Carroll SB. Endless forms: the evolution of gene regulation and morphological diversity. Cell. 2000;101: 577–580. doi:10.1016/S0092-8674(00)80868-5

9. Hoekstra HE, Coyne JA. The locus of evolution: evo devo and the genetics of adaptation. Evolution. 2007;61: 995–1016. doi:10.1111/j.1558-5646.2007.00105.x

10. Kondrashov FA. Gene Dosage and Duplication. Evolution after Gene Duplication. 2011. pp. 57–76. doi:10.1002/9780470619902.ch4

11. Cardoen D, Wenseleers T, Ernst UR, Danneels EL, Laget D, DE Graaf DC, et al. Genome-wide analysis of alternative reproductive phenotypes in honeybee workers. Mol Ecol. 2011;20: 4070–4084. doi:10.1111/j.1365-294X.2011.05254.x

12. Montague MJ, Li G, Gandolfi B, Khan R, Aken BL, Searle SMJ, et al. Comparative analysis of the domestic cat genome reveals genetic signatures underlying feline biology and domestication. Proc Natl Acad Sci U S A. 2014;111: 17230–17235. doi:10.1073/pnas.1410083111

13. Catalán A, Glaser-Schmitt A, Argyridou E, Duchen P, Parsch J. An indel polymorphism in the MtnA 3' untranslated region is associated with gene expression variation and local adaptation in Drosophila melanogaster. PLoS Genet. 2016;12. Available: https://www.ncbi.nlm.nih.gov/pmc/articles/PMC4847869

14. Mack KL, Ballinger MA, Phifer-Rixey M, Nachman MW. Gene regulation underlies environmental adaptation in house mice. Genome Res. 2018;28: 1636–1645. doi:10.1101/gr.238998.118

15. Nandamuri SP, Conte MA, Carleton KL. Multiple trans QTL and one cis-regulatory deletion are associated with the differential expression of cone opsins in African cichlids. BMC Genomics. 2018;19: 945. doi:10.1186/s12864-018-5328-z

16. Wu L, Candille SI, Choi Y, Xie D, Jiang L, Li-Pook-Than J, et al. Variation and genetic control of protein abundance in humans. Nature. 2013;499: 79–82. doi:10.1038/nature12223

17. Martin A, Orgogozo V. The Loci of repeated evolution: a catalog of genetic hotspots of phenotypic variation. Evolution. 2013;67: 1235–1250. doi:10.1111/evo.12081

18. Harrison PW, Wright AE, Mank JE. The evolution of gene expression and the transcriptome-phenotype relationship. Semin Cell Dev Biol. 2012;23: 222–229. doi:10.1016/j.semcdb.2011.12.004

19. Schendel V, Rash LD, Jenner RA, Undheim EAB. The Diversity of Venom: The Importance of Behavior and Venom System Morphology in Understanding Its Ecology and Evolution. Toxins . 2019;11: 666. doi:10.3390/toxins11110666

20. 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: 483–511. doi:10.1146/annurev.genom.9.081307.164356

21. Casewell NR, Wüster W, Vonk FJ, Harrison RA, Fry BG. Complex cocktails: the evolutionary novelty of venoms. Trends Ecol Evol. 2013;28: 219–229. doi:10.1016/j.tree.2012.10.020

22. Olivera BM. Conus Venom Peptides: Reflections from the Biology of Clades and Species. Annu Rev Ecol Syst. 2002;33: 25–47. doi:10.1146/annurev.ecolsys.33.010802.150424

23. Escoubas P, King GF. Venomics as a drug discovery platform. Expert Rev Proteomics. 2009;6: 221–224. doi:10.1586/epr.09.45

24. Fry BG, Casewell NR, Wüster W, Vidal N, Young B, Jackson TNW. The structural and functional diversification of the Toxicofera reptile venom system. Toxicon. 2012;60: 434–448. doi:10.1016/j.toxicon.2012.02.013

25. Jackson TNW, Jouanne H, Vidal N. Snake Venom in Context: Neglected Clades and Concepts. Frontiers in Ecology and Evolution. 2019;7: 332. doi:10.3389/fevo.2019.00332

26. Fry BG, Scheib H, van der Weerd L, Young B, McNaughtan J, Ramjan SFR, et al. Evolution of an arsenal: structural and functional diversification of the venom system in the advanced snakes (Caenophidia). Mol Cell Proteomics. 2008;7: 215–246. doi:10.1074/mcp.M700094-MCP200

27. Rokyta DR, Lemmon AR, Margres MJ, Aronow K. The venom-gland transcriptome of the eastern diamondback rattlesnake (Crotalus adamanteus). BMC Genomics. 2012;13: 312. doi:10.1186/1471-2164-13-312

28. Rokyta DR, Wray KP, Margres MJ. The genesis of an exceptionally lethal venom in the timber rattlesnake (Crotalus horridus) revealed through comparative venom-gland transcriptomics. BMC Genomics. 2013;14: 394. doi:10.1186/1471-2164-14-394

29. 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: 2640–2649. doi:10.1093/gbe/evx199

30. Shibata H, Chijiwa T, Oda-Ueda N, Nakamura H, Yamaguchi K, Hattori S, et al. The habu genome reveals accelerated evolution of venom protein genes. Sci Rep. 2018;8: 11300. doi:10.1038/s41598-018-28749-4

31. Casewell NR, Wagstaff SC, Harrison RA, Renjifo C, Wüster W. Domain loss facilitates accelerated evolution and neofunctionalization of duplicate snake venom metalloproteinase toxin genes. Mol Biol Evol. 2011;28: 2637–2649. doi:10.1093/molbev/msr091

32. 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: 10911–10920. doi:10.1073/pnas.1920011117

33. 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: 2434–2445. doi:10.1016/j.cub.2016.07.038

34. Aird SD, Watanabe Y, Villar-Briones A, Roy MC, Terada K, Mikheyev AS. Quantitative high-throughput profiling of snake venom gland transcriptomes and proteomes (Ovophis okinavensis and Protobothrops flavoviridis). BMC Genomics. 2013;14: 790. doi:10.1186/1471-2164-14-790

35. Rokyta DR, Margres MJ, Calvin K. Post-transcriptional Mechanisms Contribute Little to Phenotypic Variation in Snake Venoms. G3 . 2015;5: 2375–2382. doi:10.1534/g3.115.020578

36. Aird SD, Aggarwal S, Villar-Briones A, Tin MM-Y, Terada K, Mikheyev AS. Snake venoms are integrated systems, but abundant venom proteins evolve more rapidly. BMC Genomics. 2015;16: 647. doi:10.1186/s12864-015-1832-6

37. Margres MJ, Wray KP, Seavy M, McGivern JJ, Herrera ND, Rokyta DR. Expression Differentiation Is Constrained to Low-Expression Proteins over Ecological Timescales. Genetics. 2016;202: 273–283. doi:10.1534/genetics.115.180547

38. Haney RA, Clarke TH, Gadgil R, Fitzpatrick R, Hayashi CY, Ayoub NA, et al. Effects of Gene Duplication, Positive Selection, and Shifts in Gene Expression on the Evolution of the Venom Gland Transcriptome in Widow Spiders. Genome Biol Evol. 2016;8: 228–242. doi:10.1093/gbe/evv253

39. Macrander J, Broe M, Daly M. Tissue-Specific Venom Composition and Differential Gene Expression in Sea Anemones. Genome Biol Evol. 2016;8: 2358–2375. doi:10.1093/gbe/evw155

40. Columbus-Shenkar YY, Sachkova MY, Macrander J, Fridrich A, Modepalli V, Reitzel AM, et al. Dynamics of venom composition across a complex life cycle. Elife. 2018;7. doi:10.7554/eLife.35014

41. Surm JM, Smith HL, Madio B, Undheim EAB, King GF, Hamilton BR, et al. A process of convergent amplification and tissue-specific expression dominates the evolution of toxin and toxin-like genes in sea anemones. Mol Ecol. 2019;28: 2272–2289. doi:10.1111/mec.15084

42. Robertson PL. A morphological and functional study of the venom apparatus in representatives of some major groups of Hymenoptera. Aust J Zool. 1968;16: 133–166. doi:10.1071/zo9680133

43. Walker AA. The evolutionary dynamics of venom toxins made by insects and other animals. Biochem Soc Trans. 2020;48: 1353–1365. doi:10.1042/BST20190820

44. Kochva E. The origin of snakes and evolution of the venom apparatus. Toxicon. 1987;25: 65–106. doi:10.1016/0041-0101(87)90150-4

45. Vonk FJ, Admiraal JF, Jackson K, Reshef R, de Bakker MAG, Vanderschoot K, et al. Evolutionary origin and development of snake fangs. Nature. 2008;454: 630–633. doi:10.1038/nature07178

46. Westeen EP, Durso AM, Grundler MC, Rabosky DL, Davis Rabosky AR. What makes a fang? Phylogenetic and ecological controls on tooth evolution in rear-fanged snakes. BMC Evol Biol. 2020;20: 80. doi:10.1186/s12862-020-01645-0

47. Surm JM, Moran Y. Insights into how development and life-history dynamics shape the evolution of venom. Evodevo. 2021;12: 1. doi:10.1186/s13227-020-00171-w

48. Kochva E. DEVELOPMENT OF THE VENOM GLAND AND TRIGEMINAL MUSCLES IN VIPERA PALAESTINAE. Cells Tissues Organs. 1963;52: 49–89. doi:10.1159/000142338

49. Vonk FJ, Casewell NR, Henkel CV, Heimberg AM, Jansen HJ, McCleary RJR, et al. The king cobra genome reveals dynamic gene evolution and adaptation in the snake venom system. Proc Natl Acad Sci U S A. 2013;110: 20651–20656. doi:10.1073/pnas.1314702110

50. Bhinge A, Namboori SC, Bithell A, Soldati C, Buckley NJ, Stanton LW. MiR-375 is essential for human spinal motor neuron development and may be involved in motor neuron degeneration. Stem Cells. 2016;34: 124–134. doi:10.1002/stem.2233

51. Post Y, Puschhof J, Beumer J, Kerkkamp HM, de Bakker MAG, Slagboom J, et al. Snake Venom Gland Organoids. Cell. 2020;180: 233–247.e21. doi:10.1016/j.cell.2019.11.038

52. David CN. Interstitial stem cells in Hydra: multipotency and decision-making. Int J Dev Biol. 2012;56: 489–497. doi:10.1387/ijdb.113476cd

53. Beckmann A, Özbek S. The nematocyst: a molecular map of the cnidarian stinging organelle. Int J Dev Biol. 2012;56: 577–582. doi:10.1387/ijdb.113472ab

54. Ball DA. Foregut ontogeny of the Neogastropoda: comparison of development in Nucella lapillus and Conus anemone. Boll Malacol. 2002; 51–78. Available: https://dialnet.unirioja.es/servlet/articulo?codigo=6372475

55. Innan H, Kondrashov F. The evolution of gene duplications: classifying and distinguishing between models. Nat Rev Genet. 2010;11: 97–108. doi:10.1038/nrg2689

56. Ohno S. Evolution by gene duplication. Berlin, Germany: Springer Science; 1970. Available: https://market.android.com/details?id=book-5SjqCAAAQBAJ

57. Schield DR, Card DC, Hales NR, Perry BW, Pasquesi GM, Blackmon H, et al. The origins and evolution of chromosomes, dosage compensation, and mechanisms underlying venom regulation in snakes. Genome Res. 2019. doi:10.1101/gr.240952.118

58. Koludarov I, Jackson TNW, Suranse V, Pozzi A, Sunagar K, Mikheyev AS. Reconstructing the evolutionary history of a functionally diverse gene family reveals complexity at the genetic origins of novelty. Cold Spring Harbor Laboratory. 2020. p. 583344. doi:10.1101/583344

59. Kordis D, Gubensek F. Adaptive evolution of animal toxin multigene families. Gene. 2000;261: 43–52. doi:10.1016/s0378-1119(00)00490-x

60. Juárez P, Comas I, González-Candelas F, Calvete JJ. Evolution of snake venom disintegrins by positive Darwinian selection. Mol Biol Evol. 2008;25: 2391–2407. doi:10.1093/molbev/msn179

61. Gao B, Peigneur S, Dalziel J, Tytgat J, Zhu S. Molecular divergence of two orthologous scorpion toxins affecting potassium channels. Comp Biochem Physiol A Mol Integr Physiol. 2011;159: 313–321. doi:10.1016/j.cbpa.2011.03.027

62. Zancolli G, Casewell NR. Venom systems as models for studying the origin and regulation of evolutionary novelties. Mol Biol Evol. 2020. doi:10.1093/molbev/msaa133

63. World Health Organization; Neglected tropical disease: Snakebite envenoming. Available: https://www.who.int/health-topics/snakebite

64. Barabási A-L, Oltvai ZN. Network biology: understanding the cell's functional organization. Nat Rev Genet. 2004;5: 101–113. doi:10.1038/nrg1272

65. Allen JD, Xie Y, Chen M, Girard L, Xiao G. Comparing statistical methods for constructing large scale gene networks. PLoS One. 2012;7: e29348. doi:10.1371/journal.pone.0029348

66. Guido Sanguinetti VAH-T, editor. Gene Regulatory Networks.pdf. 233 Spring Street, New York, NY 10013, USA: Springer Science and Business Media; 2019.

67. Miller JA, Horvath S, Geschwind DH. Divergence of human and mouse brain transcriptome highlights Alzheimer disease pathways. Proc Natl Acad Sci U S A. 2010;107: 12698–12703. doi:10.1073/pnas.0914257107

68. Oldham MC, Horvath S, Geschwind DH. Conservation and evolution of gene coexpression networks in human and chimpanzee brains. Proc Natl Acad Sci U S A. 2006;103: 17973–17978. doi:10.1073/pnas.0605938103

69. Filteau M, Pavey SA, St-Cyr J, Bernatchez L. Gene coexpression networks reveal key drivers of phenotypic divergence in lake whitefish. Mol Biol Evol. 2013;30: 1384–1396. doi:10.1093/molbev/mst053

70. Stuart JM, Segal E, Koller D, Kim SK. A gene-coexpression network for global discovery of conserved genetic modules. Science. 2003;302: 249–255. doi:10.1126/science.1087447

71. Hwang J, Qi L. Quality Control in the Endoplasmic Reticulum: Crosstalk between ERAD and UPR pathways. Trends Biochem Sci. 2018;43: 593–605. doi:10.1016/j.tibs.2018.06.005

72. 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. doi:10.1073/pnas.2021311118

73. 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: 7542–7547. doi:10.1073/pnas.0402517101

74. Fry BG, Winter K, Norman JA, Roelants K, Nabuurs RJA, van Osch MJP, et al. Functional and structural diversification of the Anguimorpha lizard venom system. Mol Cell Proteomics. 2010;9: 2369–2390. doi:10.1074/mcp.M110.001370

75. 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: 25745–25755. doi:10.1073/pnas.1906117116

76. Schachter M. Kallikreins and kinins. Physiol Rev. 1969;49: 509–547. doi:10.1152/physrev.1969.49.3.509

77. 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: 403–420. Available: http://genome.cshlp.org/content/15/3/403.short

78. Gould SJ, Vrba ES. Exaptation—a Missing Term in the Science of Form. Paleobiology. 1982;8: 4–15. doi:10.1017/S0094837300004310

79. 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: 2088–2095. doi:10.1093/gbe/evu166

80. Sunagar K, Casewell NR, Varma S, Kolla R, Antunes A, Moran Y. Deadly innovations: Unraveling the molecular evolution of animal venoms. Venom genomics and proteomics. Springer, Dordrecht; 2016. pp. 1–27. doi:10.1007/978-94-007-6416-3_27

81. Arnold SJ, Pfrender ME, Jones AG. The adaptive landscape as a conceptual bridge between micro- and macroevolution. Genetica. 2001;112-113: 9–32. Available: https://www.ncbi.nlm.nih.gov/pubmed/11838790

82. Wilson AJ, Réale D, Clements MN, Morrissey MM, Postma E, Walling CA, et al. An ecologist's guide to the animal model. J Anim Ecol. 2010;79: 13–26. doi:10.1111/j.1365-2656.2009.01639.x

83. de Villemereuil and Shinichi Nakagawa P. General Quantitative Genetics Methods for Comparative Biology. In: Zsolt GL, editor. Modern phylogenetic comparative methods and their application in evolutionary biology: Concepts and practice. Berlin, Germany: Springer; 2014. pp. 287–304. doi:10.1007/978-3-662-43550-2

84. Manceau M, Marin J, Morlon H, Lambert A. Model-based inference of punctuated molecular evolution. Mol Biol Evol. 2020. doi:10.1093/molbev/msaa144

85. Mora-Obando D, Salazar-Valenzuela D, Pla D, Lomonte B, Guerrero-Vargas JA, Ayerbe S, et al. Venom variation in Bothrops asper lineages from North-Western South America. J Proteomics. 2020;229: 103945. doi:10.1016/j.jprot.2020.103945

86. Galis F. Key Innovations and Radiations. In: Wagner GP, editor. The Character Concept in Evolutionary Biology. San Diego: Academic Press; 2001. pp. 581–605. doi:10.1016/B978-012730055-9/50038-0

87. Hunter JP. Key innovations and the ecology of macroevolution. Trends Ecol Evol. 1998;13: 31–36. doi:10.1016/s0169-5347(97)01273-1

88. Yoder JB, Clancey E, Des Roches S, Eastman JM, Gentry L, Godsoe W, et al. Ecological opportunity and the origin of adaptive radiations. J Evol Biol. 2010;23: 1581–1596. doi:10.1111/j.1420-9101.2010.02029.x

89. Rabosky DL. Phylogenetic tests for evolutionary innovation: the problematic link between key innovations and exceptional diversification. Philos Trans R Soc Lond B Biol Sci. 2017;372. doi:10.1098/rstb.2016.0417

90. Rabosky DL. Automatic detection of key innovations, rate shifts, and diversity-dependence on phylogenetic trees. PLoS One. 2014;9: e89543. doi:10.1371/journal.pone.0089543

91. Landis MJ, Schraiber JG. Pulsed evolution shaped modern vertebrate body sizes. Proc Natl Acad Sci U S A. 2017;114: 13224–13229. doi:10.1073/pnas.1710920114

92. McDermott A. Snake venom evolved in fits and spurts. [cited 21 Feb 2022]. Available: https://blog.pnas.org/2020/05/snake-venom-evolved-in-fits-and-spurts/

93. Barua A, Koludarov I, Mikheyev AS. Co-option of the same ancestral gene family gave rise to mammalian and reptilian toxins. BMC Biol. 2021;19: 268. doi:10.1186/s12915-021-01191-1

94. 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: 20200613. doi:10.1098/rspb.2020.0613

95. Barua A, Mikheyev AS. Many Options, Few Solutions: Over 60 Million Snakes Converged on a Few Optimal Venom Formulations. Mol Biol Evol. 2019;36: 1964–1974. doi:10.1093/molbev/msz125

96. Strathmann RR, Eernisse DJ. What Molecular Phylogenies Tell Us about the Evolution of Larval Forms. Integr Comp Biol. 1994;34: 502–512. doi:10.1093/icb/34.4.502

97. Lino-López GJ, Valdez-Velázquez LL, Corzo G, Romero-Gutiérrez MT, Jiménez-Vargas JM, Rodríguez-Vázquez A, et al. Venom gland transcriptome from Heloderma horridum horridum by high-throughput sequencing. Toxicon. 2020;180: 62–78. doi:10.1016/j.toxicon.2020.04.003

98. Koludarov I, Jackson TN, Brouw BO den, Dobson J, Dashevsky D, Arbuckle K, et al. Enter the Dragon: The Dynamic and Multifunctional Evolution of Anguimorpha Lizard Venoms. Toxins . 2017;9. doi:10.3390/toxins9080242

99. Xie B, Dashevsky D, Rokyta D, Ghezellou P, Fathinia B, Shi Q, et al. Dynamic genetic differentiation drives the widespread structural and functional convergent evolution of snake venom proteinaceous toxins. BMC Biol. 2022;20: 4. doi:10.1186/s12915-021-01208-9

100. Junqueira-de-Azevedo ILM, Campos PF, Ching ATC, Mackessy SP. Colubrid Venom Composition: An -Omics Perspective. Toxins. 2016;8. doi:10.3390/toxins8080230

101. Kowalski K, Rychlik L. Venom Use in Eulipotyphlans: An Evolutionary and Ecological Approach. Toxins . 2021;13. doi:10.3390/toxins13030231

102. Arbuckle K. Phylogenetic Comparative Methods can Provide Important Insights into the Evolution of Toxic Weaponry. Toxins . 2018;10. doi:10.3390/toxins10120518

Appendices

Appendix I:
Barua A, Mikheyev AS. An ancient, conserved gene regulatory network led to the rise of oral venom systems. *Proceedings of the National Academy of Sciences USA*. 2021;118:2021311118.

An ancient, conserved gene regulatory network led to the rise of oral venom systems

Agneesh Barua[a,1] and Alexander S. Mikheyev[a,b]

[a]Ecology and Evolution Unit, Okinawa Institute of Science and Technology Graduate University, Okinawa, 904-0495, Japan; and [b]Evolutionary Genomics Group, Australian National University, Canberra, ACT 0200, Australia

Edited by Günter P. Wagner, Yale University, New Haven, CT, and approved February 11, 2021 (received for review October 27, 2020)

Oral venom systems evolved multiple times in numerous vertebrates enabling the exploitation of unique predatory niches. Yet how and when they evolved remains poorly understood. Up to now, most research on venom evolution has focused strictly on the toxins. However, using toxins present in modern day animals to trace the origin of the venom system is difficult, since they tend to evolve rapidly, show complex patterns of expression, and were incorporated into the venom arsenal relatively recently. Here we focus on gene regulatory networks associated with the production of toxins in snakes, rather than the toxins themselves. We found that overall venom gland gene expression was surprisingly well conserved when compared to salivary glands of other amniotes. We characterized the "metavenom network," a network of ∼3,000 nonsecreted housekeeping genes that are strongly coexpressed with the toxins, and are primarily involved in protein folding and modification. Conserved across amniotes, this network was coopted for venom evolution by exaptation of existing members and the recruitment of new toxin genes. For instance, starting from this common molecular foundation, *Heloderma* lizards, shrews, and solenodon, evolved venoms in parallel by overexpression of kallikreins, which were common in ancestral saliva and induce vasodilation when injected, causing circulatory shock. Derived venoms, such as those of snakes, incorporated novel toxins, though still rely on hypotension for prey immobilization. These similarities suggest repeated cooption of shared molecular machinery for the evolution of oral venom in mammals and reptiles, blurring the line between truly venomous animals and their ancestors.

venom | evolution | gene regulatory networks | transcriptomics | complex traits

Venoms are proteinaceous mixtures that can be traced and quantified to distinct genomic loci, providing a level of genetic tractability that is rare in other traits (1–4). This advantage of venom systems provides insights into processes of molecular evolution that are otherwise difficult to obtain. For example, studies in cnidarians showed that gene duplication is an effective way to increase protein dosage in tissues where different ecological roles can give rise to different patterns of gene expression (2, 5). Studies of venom in snakes have allowed comparisons of the relative importance of sequence evolution vs. gene expression evolution, as well as how a lack of genetic constraint enables diversity in complex traits (6, 7).

Despite the wealth of knowledge venoms have provided about general evolutionary processes, the common molecular basis for the evolution of venom systems themselves is unknown. Even in snakes, which have perhaps the best studied venom systems, very little is known about the molecular architecture of these systems at their origin (8, 9). Using toxin families present in modern snakes to understand evolution at its origin is difficult because toxins evolve rapidly, both in terms of sequence and gene expression (10, 11). Toxins experience varying degrees of selection and drift, complicating interpretations of evolutionary models (12), and estimation of gene family evolution is often inconsistent, varying with which part of the gene (exon or intron) is used to construct the phylogeny (13). Most importantly, present-day toxins became a part of the venom over time; this diminishes their utility in trying to understand events that lead to the rise of venom systems in the nonvenomous ancestors of snakes (14, 15).

A gene coexpression network aims to identify genes that interact with one another based on common expression profiles (16). Groups of coexpressed genes that have similar expression patterns across samples are identified using hierarchical clustering and are placed in gene "modules" (17). Constructing a network and comparing expression profiles of modules across taxa can identify key drivers of phenotypic change, as well as aid in identifying initial genetic targets of natural selection (18, 19). Comparative analysis using gene coexpression networks allows us to distinguish between ancient genetic modules representing core cellular processes, evolving modules that give rise to lineage-specific differences, and highly flexible modules that have evolved differently in different taxa (20). Gene coexpression networks are also widely used to construct gene regulatory networks (GRNs) owing to their reliability in capturing biologically relevant interactions between genes, as well as their high power in reproducing known protein–protein interactions (21, 22).

Here we focus on gene coexpression networks involved in the production of snake venom, rather than the venom toxins themselves. Using a coexpression network we characterized the genes associated with venom production, which we term the "metavenom network," and determine its biological role. We traced the origin of this network to the common ancestor of amniotes, which

Significance

Although oral venom systems are ecologically important characters, how they originated is still unclear. In this study, we show that oral venom systems likely originated from a gene regulatory network conserved across amniotes. This network, which we term the "metavenom network," comprises over 3,000 housekeeping genes coexpressed with venom and play a role in protein folding and modification. Comparative transcriptomics revealed that the network is conserved between venom glands of snakes and salivary glands of mammals. This suggests that while these tissues have evolved different functions, they share a common regulatory core, that persisted since their common ancestor. We propose several evolutionary mechanisms that can utilize this common regulatory core to give rise to venomous animals from their nonvenomous ancestors.

Author contributions: A.B. and A.S.M. designed research, performed research, analyzed data, and wrote the paper.

The authors declare no competing interest.

This article is a PNAS Direct Submission.

[1]To whom correspondence may be addressed. Email: agneesh.barua@oist.jp.

This article contains supporting information online at https://www.pnas.org/lookup/suppl/doi:10.1073/pnas.2021311118/-/DCSupplemental.

Published March 29, 2021.

suggests that the venom system originated from a conserved gene regulatory network. The conserved nature of the metavenom network across amniotes suggests that oral venom systems started with a common gene regulatory foundation, and underwent lineage-specific changes to give rise to diverse venom systems in snakes, lizards, and even mammals.

Results

The Metavenom Network Is Involved in Toxin Expression in Snakes. Previously published RNA libraries from Taiwan habu (*Protobothrops mucrosquamatus*) were used to construct the network (12). Weighted gene coexpression network analysis (WGCNA) was used to construct the coexpression network (23). WGCNA estimates correlations between genes across samples (libraries) and clusters genes with similar profiles into modules (23). This clustering is based solely on similarities in expression levels and does not imply any association based on ecological roles of genes (or toxins).

Using data from venom gland samples, WGCNA clustered 18,313 genes into 29 modules ranging in size from 38 to 3,380 genes. All secreted venom toxins were found in the largest module (module 1), which we term the metavenom network (Fig. 1*A*). Therefore, the metavenom network represents an assemblage of housekeeping genes that are strongly associated with toxin genes. This forms an ensemble that is the GRN involved in expression of toxin genes. The genes in the metavenom network have a higher functional relevance than genes that are simply up-regulated in the venom gland. For example, some genes involved in formation of musculature of the venom gland might be highly expressed in the venom gland as compared to say kidney, but it might not necessarily be involved in the expression of toxin genes themselves. WGCNA makes this distinction, and has been consistently shown to provide robust functional relationships between genes (20–22). We performed module preservation analysis to determine whether within-module characteristics like gene density and connectivity between genes are conserved between venom gland and other tissues like heart, kidney, liver. In other words, module preservation statistics were used to determine whether the characteristics of genes and their modules identified in one (reference) tissue were present in another (test) tissue. A module preservation $Z_{summary} > 2$ implies that module characteristics within a module are preserved in other tissues, while a score <2 denotes no preservation (24). $Z_{summary}$ statistic (Dataset S1 *A–C*) revealed that the metavenom network module is not preserved in the heart or liver, but has borderline preservation in the kidney ($Z_{summary} = 2.000522$). This implies that much of the expression pattern of the metavenom network is unique to the venom gland and bears only a slight similarity in kidneys.

After defining the metavenom network, which comprises genes that are tightly associated with toxin expression, we identified the biological processes involved using Gene Ontology (GO) enrichment. The metavenom network is primarily involved in protein modification, and protein transport (Dataset S2*C*). GO terms associated with the unfolding protein response (UPR): GO:0006986, GO:0034620, and GO:0035966, and endoplasmic reticulum associated protein degradation (ERAD): GO:0034976, GO:0030968, and GO:0036503 were the most significantly enriched biological processes in the metavenom (Fig. 1*B*).

Since the metavenom network has over 3,000 genes, visualizing the entire network topology would be impractical. Therefore, we selected the top 20 highly expressed nonvenom genes, and the top 10 highly expressed toxin genes for visualization and to identify the levels of connection between them (Fig. 1*A*). An interactive visualization can be found in *SI Appendix*, Fig. S1. The network diagram revealed that almost all of the highly expressed venom toxins have strong links with each other, as well as directly with the nonvenom genes. Zinc metalloproteinase (SVMP: 107298299) and snake venom serine protease serpentokallikrein-2 (SVSP: 107287553) were the exceptions, which have links with only a few toxin genes and nonvenom genes (namely DLG1, CANX, HSP90, RPLP0, PDIA4, and LOC8828).

Several network characteristics can be used to identify genes integral to a network. One of these characteristics is module membership, which represents connectivity of a gene with other genes within a module and is used to define centralized hub genes (23). Module membership (MM) has values between 0 and 1, where values closer to 1 represent high connectivity within a module, and values closer to 0 represent low connectivity. We estimated module membership of genes in the metavenom network and identified sets of differentially expressed genes (DEGs) (Dataset S3*B*). An ANOVA-like test for gene expression in venom gland, heart, liver, and kidney of habu revealed that out of 3,380 genes that make up the metavenom network, 1,295 were significantly differentially expressed ($P < 0.05$) (Dataset S3*B*). To identify genes most specific to the venom gland, we filtered the DEGs associated with the UPR and ERAD that had high module membership (MM > 0.9) and high average expression across all venom gland libraries. We obtained a list of 149 genes (Dataset S3*E*). On an average, most of these genes were up-regulated in the venom gland, with a few up-regulated in the nonvenom tissues (Fig. 1*C*, only 8 shown, full dataset in Dataset S3*C*), implying that these genes are of greater functional relevance in the venom gland.

External validation of module preservation. To confirm that modules identified in this study, particularly the metavenom network module, represent technically reproducible and evolutionarily meaningful features, we assessed the extent of module preservation between our work and a WGCNA investigation of the human salivary gland (25). Other than the WGCNA algorithm, this study employed different methodologies, such as microarray gene expression measurements, and the inclusion of samples from patients with salivary gland pathogenesis. Nonetheless, there were significant overlaps in modules detected in both studies, supporting the method's robustness (*SI Appendix*, Fig. S2).

The Metavenom Network Is Conserved across Amniotes. Conserved gene expression profiles between taxa are indicative of a shared ancestry that can be used to provide insights into key drivers of phenotypic change as well as revealing molecular organization of a trait at its origin (17, 20). The metavenom network is significantly enriched for genes belonging to the UPR and ERAD pathways. These families of housekeeping genes are widely conserved across the animal kingdom (26). This high level of conservation encouraged the search for orthologs in other taxa. Once the list of orthologs was obtained we carried out comparative transcriptomic analysis to determine if the expression of metavenom network was conserved across taxa. We identified 546 one-to-one orthologs of the metavenom network, that were expressed in four tissue groups of nine species: human, chimpanzee, mouse, dog, anole, habu, cobra, chicken, and frog. To do this we first obtained one-to-one orthologs from the National Center for Biotechnology Information (NCBI)'s eukaryotic genome annotation pipeline and combined them with phylogenetically inferred orthologs from OrthoFinder (27, 28). In addition to the substantial overlap between estimated orthologs, both approaches estimated orthologs with conserved synteny (*SI Appendix*, Fig. S3). Public RNA datasets from four tissues (heart, kidney, liver, and salivary glands) were used for comparative transcriptomic analysis (*Materials and Methods*). We obtained expression data for cobra tissues, including that of venom gland from Suryamohan et al. (29).

To get an overview of metavenom network gene expression patterns between species, we performed a principal component analysis (PCA) using a comparative dataset of the one-to-one metavenom network orthologs. PCA clustered gene expression by tissue and despite the over 300 million years' divergence between

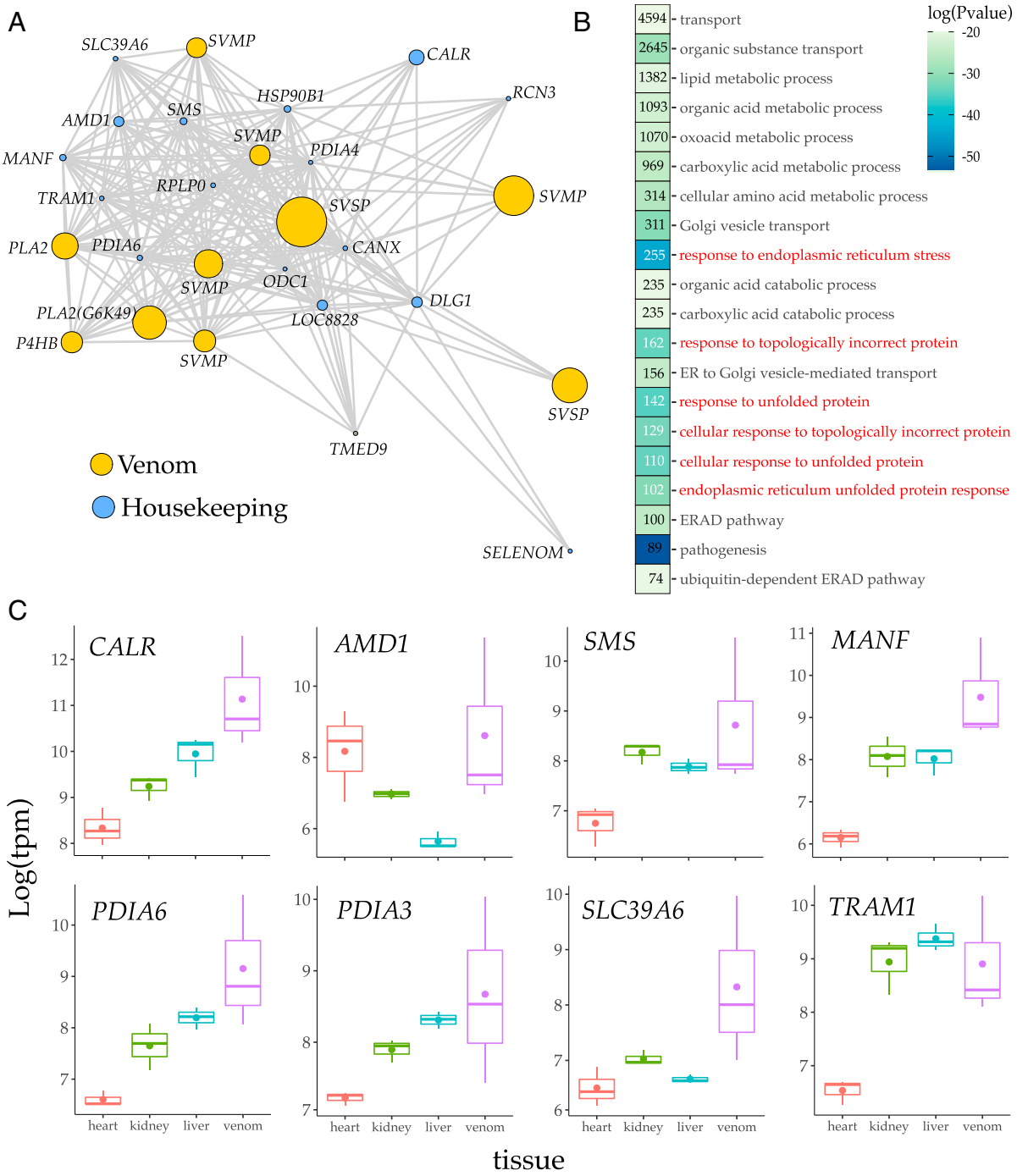


Fig. 1. The metavenom network module represents a group of coexpressed genes that are associated with production of toxin in the venom gland of the Taiwan habu. More than one-third of genes in the metavenom network are up-regulated in the venom gland and are involved in protein folding and protein modification. (*A*) The metavenom network module comprised a total of 3,380 genes. Out of them, 10 of the most expressed toxin genes and 20 of the most expressed nonvenom were plotted to visualize connections and overall module topography. An interactive version of the network graph is available at https://github.com/agneeshbarua/Metavenom (*SI Appendix*, Fig. S1). Most toxin genes and nontoxin genes are well interconnected. (LOC8828 represents a gene without a reliable annotation, but we believe it is a truncated SVMP as it is flanked very closely by a secreted SVMP.) (*B*) The 20 most significant GO terms enriched in the metavenom network module comprised processes related to molecular transport and metabolism. We focused on the most significantly enriched GO terms (in red) as they represent more specific biological processes and are less ambiguous as compared to more broadly defined terms like "transport" and "organic acid metabolic process." These specific terms refer to processes involved in protein folding and modification, in particular, the UPR and ERAD. The GO term "pathogenesis" has the highest significance and is attributed to the toxin genes present in the metavenom network. GO terms are arranged by descending order of size (given within panels). (*C*) Most of the genes with high module membership were on average up-regulated (with significance at $P < 0.05$) in the venom gland, with some up-regulated in nonvenom tissue. Dot within box plot indicates mean. *CALR*: calreticulin; *AMD1*: adenosylmethionine decarboxylase 1; *SMS*: spermine synthase; *MANF*: mesencephalic astrocyte derived neurotrophic factor; *PDIA6*: protein disulfide isomerase family A member 6; *PDIA3*: protein disulfide isomerase family A member 3; *SLC39A6*: solute carrier family 39 member 6; and *TRAM1*: translocation associated membrane protein 1. Therefore, the UPR and ERAD pathway seem particularly important for venom expression and likely helps maintain tissue homeostasis under the load of high protein secretion.

the taxa, differences among tissues explain more than 30% of variation present in the data (Fig. 2*A*). Performing a PCA using all 2,682 expressed orthologs between nine taxa, including those outside the metavenom network, homologous tissues clustered more tightly (*SI Appendix*, Fig. S4). As a sanity check we chose orthologs at random to check whether the transcriptomes would still be clustered by tissue; however, a random set of genes produced no clustering (*SI Appendix*, Fig. S5). This indicated that tissues cluster together based on some underlying structure in the expression patterns of specific sets of genes analyzed, and that this clustering cannot be reproduced by using any arbitrary set of genes (30, 31).

It is important to note that we are comparing expression patterns of orthologs that are expressed in all our sampled tissues in all our sampled taxa. Simply due to the different evolutionary histories of each sampled taxa, not all orthologs will be expressed equally across all tissues in all taxa. In other words, the more species we add to our dataset, the lower the number of genes we will get to compare because all the genes might not be equally expressed across tissues, and the number of one-to-one orthologs decrease, especially when comparing across animal classes (i.e., mammals, reptiles, birds, etc.). Despite this, we expanded the above analysis to include more taxa as well as diverse morphologies of salivary glands to determine the extent of conversation of expression patterns. We performed comparative transcriptomics with salivary glands of nonvenomous reptiles like the royal python (*Python regius*), corn snake (*Pantherophis guttatus*), and leopard gecko (*Eublepharis macularius*), as well as different morphologies of the mouse salivary gland. Even in this reduced dataset (2,291 one-to-one ortholog as opposed to 2,682) we still observed similar clustering patterns as with our original dataset (*SI Appendix*, Fig. S6). However, the overall resolution and variation (<30%) explained by this expanded dataset was low, due to reduction in the number of genes without a subsequent increase in the number of replicates. Although adding diverse morphologies of salivary glands did not change our results, understanding how changes in distinct salivary tissue morphologies gave rise to venom tissue would provide important clues to the origin of evolutionary innovation in venom glands.

Our comparative transcriptomic analysis using our original and expanded dataset showed that expression patterns between homologous tissues were well conserved, especially between venom glands in snakes and salivary glands in mammals. This suggests that the gene regulatory architecture of the metavenom network evolved in the common ancestor of amniotes and has for the most part remained conserved in extant taxa, while giving rise to the venom gland in snakes.

Network Characteristics of the Metavenom Network Are Conserved between the Salivary Glands of Mammals and Venom Glands of Snakes. The clustering of transcriptomes of venom gland in snakes and salivary gland in mammals was interesting because it suggests that both these tissues have a degree of molecular conservatism that likely originated in their common ancestor. Therefore to test whether the modular characteristics of the metavenom network are preserved in the salivary tissue of mammals we carried out module preservation analysis.

We estimated module preservation of the metavenom network in the venom gland of cobra and the salivary tissue of several mammals where sufficient transcriptomic data were available (mouse, human, and dog). The metavenom network was preserved in both venom glands of cobra as well as salivary tissue of mammals (Fig. 2*B*). In cobra the metavenom network had a $Z_{summary} > 10$ implying very high preservation, while in salivary

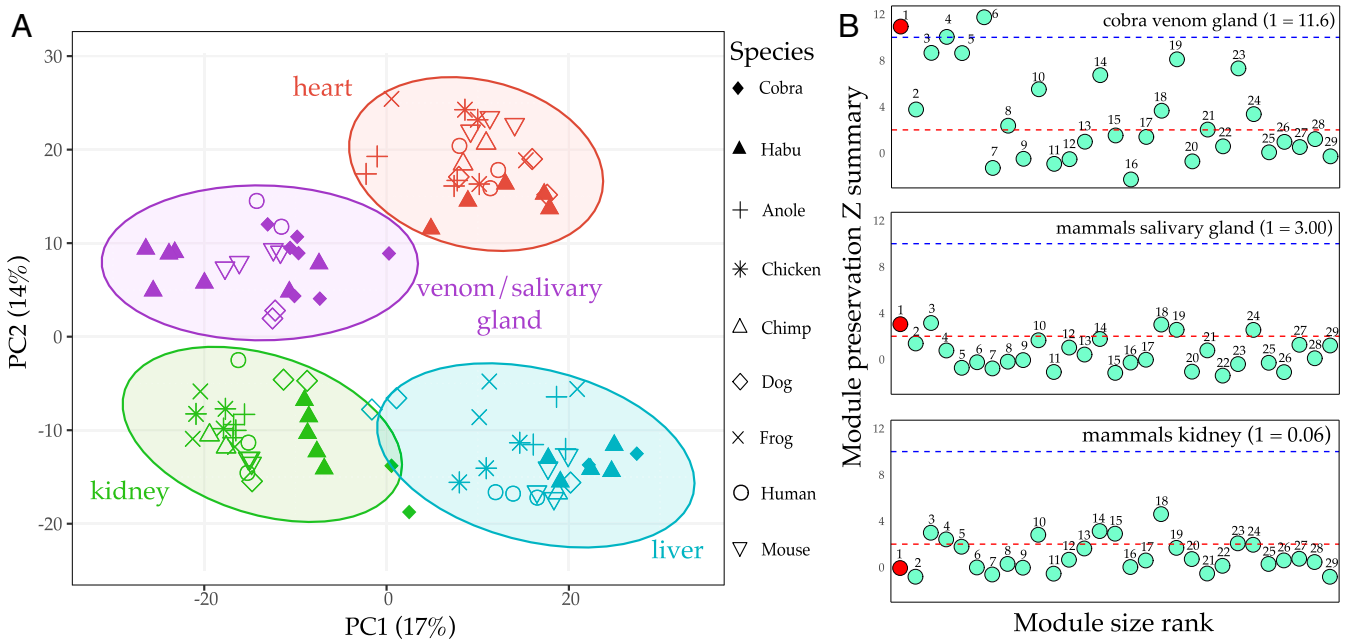


Fig. 2. Expression pattern of orthologs between venom gland in snakes and salivary gland in mammals was surprisingly well conserved. This conservation was also reflected in the preservation of the metavenom network module in the salivary gland of mammals. (*A*) When selecting the 546 one-to-one metavenom network orthologs expressed in all nine species, transcriptomes clustered based on tissue. Ellipses represent 95% confidence intervals. GO term enrichment of these 546 genes revealed that genes from the UPR and ERAD pathway are still significantly enriched, suggesting that even in a reduced dataset, the functional core of the metavenom network is still conserved (Dataset S2*B*). Despite the large evolutionary distance between species (most recent common ancestor ~300 million years ago), partitioning by tissue explains >30% of the variation in the data. (*B*) The metavenom network was highly preserved in the venom gland of cobra ($Z_{summary} > 10$) while it was weakly preserved in salivary gland of mammals ($Z_{summary} > 2$). The metavenom network, however, was not preserved in the kidneys of mammals ($Z_{summary} < 2$). These lines of evidence indicate that common regulatory architecture inherited from a common amniote ancestor gave rise to the snake venom gland. Despite the subsequent evolutionary elaboration of the venom gland, it has maintained this conserved regulatory core.

tissue of mammals the $Z_{summary}$ was 3, implying weak to moderate preservation.

To further test the extent of conservation of the metavenom network we carried out module preservation using an expanded dataset that comprised expression levels of orthologs in venom gland of prairie rattlesnake (*Crotalus viridis*) (32), and salivary glands of nonvenomous reptiles mentioned in the section above. We also included the data for different morphologies of the mouse salivary gland (33). In all these comparisons, the metavenom module was still highly preserved (Dataset S1 *E–G*). The high module preservation of the metavenom network in venomous snakes, nonvenomous reptiles, and across different morphologies of venom glands in mouse provides strong evidence of a degree of molecular conservatism that has persisted since the origin of oral secretory tissues in amniotes.

Gene Families in the Metavenom Network Evolve Rapidly and Have Undergone Greater Expansion in Venomous Snakes. Increasing the number of gene copies, especially in venom systems, are crucial to bringing about evolutionary novelty (2, 34, 35). The metavenom network in habu comprises genes that have many copies, which could have played a role in evolution of the venom system in snakes (Dataset S4). To determine whether gene families in the metavenom network evolved rapidly in venomous snakes, either by expansions or contractions, we examined gene family evolution using CAFE (36).

We used different rate parameters (λ) along the lineage leading up to venomous snakes to test the hypothesis that metavenom network gene families evolved faster in snakes as compared to other species. The rate parameter λ describes the probability that any gene will be either gained or lost, where a higher λ denotes rapid gene family evolution (37). Gene families in the branches leading up to snakes have a higher degree of family expansion, as well as higher evolution rates ($\lambda = 6.450 \times 10^{-3}$) as compared to the rest of the tree ($\lambda = 1.769 \times 10^{-3}$) (Fig. 3*A*). Among the orthogroups identified by CAFE, 23 groups were statistically rapid (see *Materials and Methods*). Ancestral estimations of gene family sizes showed that in the venomous snake lineage, most families (16 out of 23) underwent significant expansions, while a few families contracted (2 out of 23) or remained the same (5 out of 23) (Dataset S5 *A* and *B*). GO term enrichment of the 23 statistically rapid orthogroups revealed genes involved in protein modifications, protein ubiquitination, viral release from cells (genes from snakes, not of viral origin), and chromatin organization, among others (Fig. 3). We focused on genes having the most significant GO terms (Fig. 3*B*), namely, protein ubiquitination (GO:0016567), protein modification by small protein conjugation (GO:0032446), protein modification by small protein conjugation or removal (GO:0070647), and protein polyubiquitination (GO:0000209). Of the genes in the metavenom network that were enriched for these terms, almost half were significantly differentially expressed between venom gland, heart, liver, and kidney (Dataset S3*E*). While on average most of these genes were up-regulated in the venom gland, many were up-regulated in other tissues (Fig. 3*C*, only 8 shown, full list in Dataset S3*E*). Our results show that although genes involved in protein ubiquitination underwent significant expansion in venomous snakes, their overall activity is not strictly restricted to the venom gland but functions in other tissues as well.

Discussion

No biological system acts in isolation, even highly specific processes. Coexpression of genes regulates both cellular processes and maintains cellular homoeostasis (20, 38, 39). Toxin genes in the snake venom system are coexpressed with a large number of nontoxin genes. Together they form a GRN that we term the metavenom network. The metavenom network comprises genes that are involved in various processes, the most significant being the UPR and ERAD pathways. While toxin genes are evolutionarily labile (40), the conserved genes they interact with reveal the origins and repeated evolution of venom systems in vertebrates.

Genes with evolutionarily conserved expression represent functionally important groups in which coregulation is advantageous (20). Therefore, the conserved expression of metavenom network orthologs between venom glands in snakes and salivary glands in mammals was particularly important (Fig. 2*A*). While many snakes employ an oral venom system for securing prey, there are also mammals, such as shrews, and solenodons, that have evolved oral venom systems (based on salivary glands) for prey capture or defense (41). Therefore, the overall conservation of metavenom network expression, as well as preservation of the metavenom network module (Fig. 2*B*), suggests that salivary glands in mammals and venom glands in snakes share a functional core that was present in their common ancestor. Using this common molecular foundation as a starting point, snakes diversified their venom systems by recruiting a diverse array of toxins while mammals developed less complex venom systems with high similarity to saliva (42). Developing similar traits using common molecular building blocks is the hallmark of parallelism (43).

Despite the shared molecular foundation, however, the alternate path taken by snakes and the majority of mammals in developing an oral secretory system has led to the accumulation of large-scale phenotypic and functional differences between the two lineages. For instance, salivary tissue of most mammals produce large volumes of very dilute mixtures, while snake venom glands produce highly concentrated mixtures of diverse toxins (44). At the genetic level these differences are apparent when comparing evolutionary rates of gene families that comprise the metavenom network. In venomous snakes, gene families have undergone greater expansions, and have evolved at a significantly higher rate than in other lineages like mammals (Fig. 3*A*). The most enriched process among the groups of significantly expanded gene families is protein modification via ubiquitination (Fig. 3*B*). Along with tagging proteins for degradation, the ubiquitin system influences various aspects of protein functioning in the cell (45). The significant expansion of these genes in venomous snakes suggests a possible link between establishment of a complex venom system and the need for a molecular machinery which shapes a multitude of cellular processes.

The UPR and ERAD System Promoted the Evolution of an Oral Venom System. While it is difficult to attribute individual genes to a specific process without functional assays, knowing how the components of the metavenom network function in other species, we can hypothesize their roles in the venom gland of snakes and their ancestors. Even for the rapidly expanding gene families in the metavenom network, linking their direct role in the evolution of venom can only be confirmed by functional assays in both venomous and nonvenomous animals. We can nonetheless provide possible ways these genes could have functioned, painting a picture as to how incorporating these genes would enable the establishment of an oral venom system.

The UPR and ERAD act as "quality control" machinery ensuring that proteins undergo proper folding and maturation (46). Several hub genes in the metavenom network that are up-regulated in the venom gland can contribute to this quality control process (Fig. 1*C*). For example, Calreticulin (CALR) is a lectin-like chaperone that increases both the rate and yield of correctly folded proteins as well as preventing aggregation of partially folded proteins (47). Mesencephalic astrocyte derived neurotrophic factor (MANF) is induced during the UPR as a response to overexpression of misfolding-prone proteins to alleviate ER stress, and has an evolutionarily conserved cytoprotective function (48, 49). Disulfide bonds maintain structural stability and functional integrity

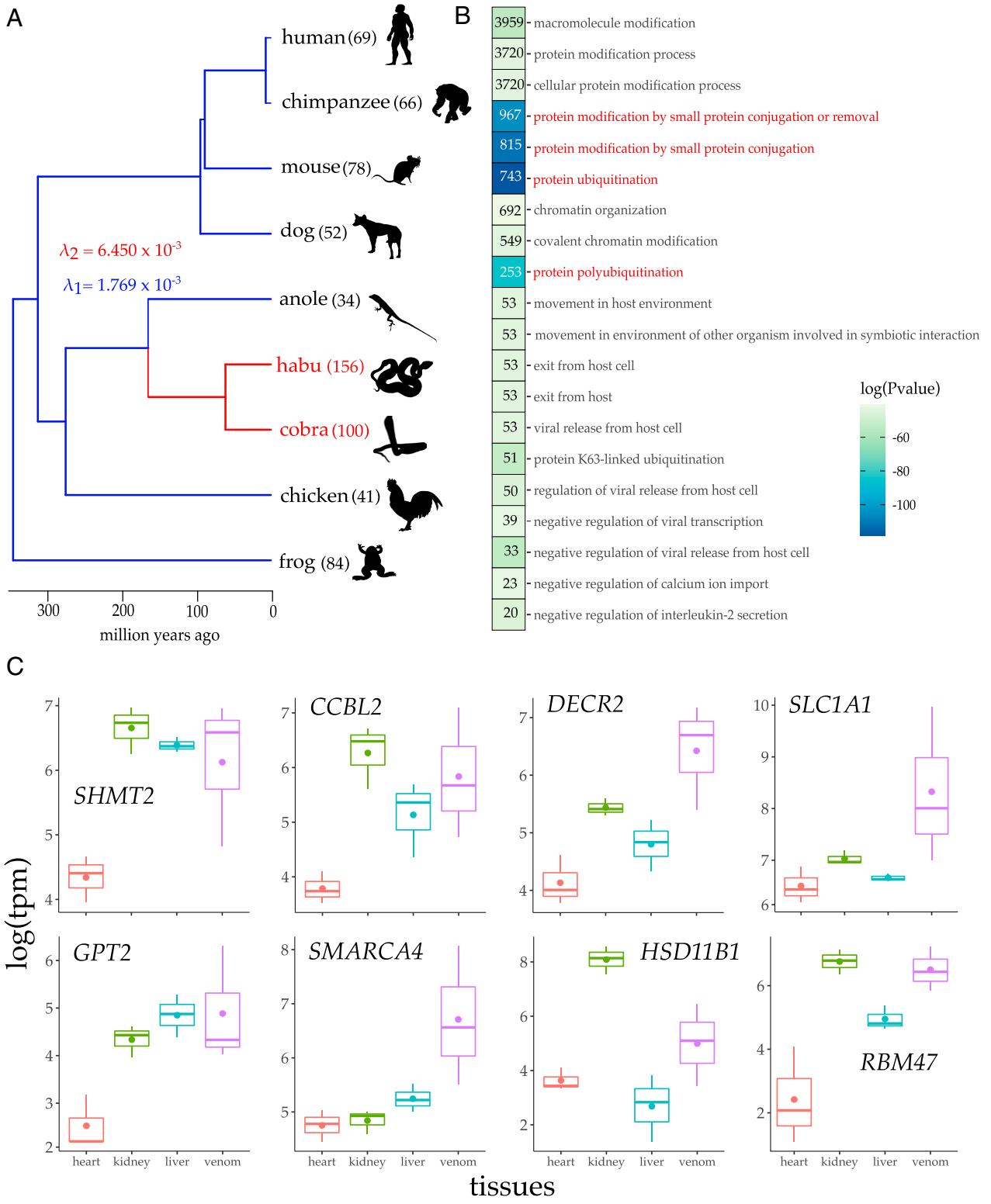


Fig. 3. Gene families in the metavenom network have not only evolved more rapidly in the lineage leading to snakes, but have also undergone more expansions in snakes than in other taxa. (*A*) Gene family evolution modeled as a "birth and death" process revealed higher rates of evolution in the branch leading up to venomous snakes (red; $\lambda = 6.4 \times 10^{-3}$) as compared to other taxa (blue; $\lambda = 1.7 \times 10^{-3}$). A model with dual rates (λ_1, λ_2) at different branches was a better fit than a uniform rate (single λ across the whole tree) model as estimated by a likelihood ratio test (*SI Appendix*, Fig. S5). (*B*) Orthogroups undergoing significant expansion were highly enriched for GO terms protein ubiquitination (*GO:0016567*), protein modification by small protein conjugation (GO:0032446), protein modification by small protein conjugation or removal (GO:0070647), and protein polyubiquitination (GO:0000209), among others. (*C*) On average, most of the genes that were associated with the above GO terms, were up-regulated (with significance at $P < 0.05$) in the venom gland, although a substantial portion was up-regulated in other tissues as well (only eight shown, full list in Dataset S3*D*). Dot within box plot indicates mean. *SHMT2*: serine hydroxymethyltransferase 2 (mitochondrial); *CCBL2*: cysteine conjugate-beta lyase 2; *DECR2*: 24-dienoyl-CoA reductase 2 peroxisomal; *SLC1A1*: solute carrier family 1 member 1; *GPT2*: glutamic pyruvate transaminase (alanine aminotransferase) 2; *SMARCA4*: SWI/SNF related matrix associated actin dependent regulator of chromatin subfamily a member 4; *HSD11B1*: hydroxysteroid (11-beta) dehydrogenase 1; and *RBM47*: RNA binding motif protein 47.

of many secreted proteins including venom toxins (50). Our results confirmed this as protein disulfide isomerase families (PDIA6 and PDIA3) were up-regulated in the venom gland and also occupied hub positions in the metavenom network (Fig. 1 *A* and *C*). PDI families catalyze disulfide bond formation, and are also vital in rearranging incorrect bonds to restore correct protein conformation (51). This restorative ability of PDI makes it an integral part of the metavenom network. Individual components of the UPR and ERAD also do not work in isolation. Feedback loops allow several components to communicate and coordinate their individual processes to relieve ER stress. For example, CALR and PDIA work in close association to equilibrate the removal of misfolded proteins and restore correct protein conformation (52). This is reflected in the metavenom network where they not only share connections, but also occupy hub positions.

Although UPR and ERAD are considered to be stress responses, they function in a stress-independent manner as well. The UPR system is activated by developmental, cell surface signaling, circadian, and various other physiological cues, implying that the system (or at least elements of it) are frequently and even continuously fine tuning cellular functions (53). In fact, consistent detection of key regulators of the UPR (ATF4, ATF6, and PERK) in nonstressed mouse tissues suggest their role in basal regulation of gene expression in vivo (54–56). Having UPR regulators contribute to the regulation of various cellular processes provides greater flexibility: a wide range of signals can be transmitted to multiple overlapping or branching pathways to fine tune their activity, a form of regulation that would be evolutionarily advantageous in organisms with diverse tissue types (53). This fine tuning is further enhanced by ubiquitin ligases that spatially and temporally modify the magnitude and duration of the UPR, impacting overall physiology (57). Therefore, the expansion of metavenom network genes associated with protein ubiquitination (Fig. 3*C*) would enable a high degree of fine tuning of cellular secretory processes in lineages leading up to venomous snakes.

The UPR anticipates, detects, and correctly folds misfolded proteins. The ERAD ensures that misfolded proteins are degraded so as to prevent cellular toxicity, and ubiquitin ligases add an overall level of regulation to fine tune these processes. These pathways support protein secretory functions, which are characterized by high demand for protein synthesis and quality control, mediating endoplasmic reticulum stress that takes place in many secretory glands, including salivary glands (58). Correspondingly, UPR and ERAD pathways are up-regulated in venom glands during venom biosynthesis in rattlesnakes (59). Having such a robust regulatory network in place would improve the tenacity of the ancestral secretory system, enabling it to tolerate an increase in tissue complexity through changes in composition and concentration of secreted proteins. Therefore, having these molecular systems already in place likely primed the ancestors of venomous animals to undergo a series of steps to attain a weaponized oral venom system. Diversification of the UPR and ERAD systems may accompany transitions from simple to complex secretory systems (60). As a result, understanding how these pathways have changed to handle additional stress of producing high venom loads, may be a productive area of future research.

Evolution of Oral Venoms from an Ancestral Salivary GRN. Given the existence of a conserved salivary GRN, venom can evolve in two ways: exaptation of existing components or through the addition of novel genes. Both mechanisms played a role in the evolution of snake venom. Furthermore, the architecture of the ancestral salivary GRN and comparisons to other venoms, such as those of solenodon and shrews, suggests a general model by which venoms have evolved across a range of taxa.

Stage 1: Exaptation of salivary enzymes, particularly kallikrein-like serine proteases. Kallikrein-like serine proteases are expressed in multiple tissues and are especially abundant in saliva of many amniotes (61, 62). Kallikrein proteolytic activity releases bradykinin and promotes inflammation. Interestingly, when injected, salivary kallikreins from nonvenomous animals, such as mice and rats, induce a hypotensive crisis leading to death (63, 64). In fact, Hiramatsu et al. (63) effectively blurred the lines between venomous and nonvenomous mammals by proposing that male mice secrete "toxic proteins (kallikrein-like enzymes) into saliva, as an effective weapon." Lethality of saliva differs between mouse strains, suggesting that heritable variability in this trait exists within species, a necessary prerequisite for adaptation (65). Thus, under ecological conditions where venom lethality promotes reproductive success, natural selection should favor the evolution of an envenomation system from this starting point. In other words, while mice probably don't use their saliva as a weapon, evolution may easily weaponize it under the right ecological conditions.

Serine protease-based toxins are nearly universal in amniote oral venoms. Mammalian oral venoms (e.g., solenodon and *Blarina* shrews), as well as those of reptiles (e.g., *Heloderma* lizards and possibly in varanids) all employ kallikrein-like serine protease overexpression (42, 66–68). Similarly, Fry noted that snake venom kallikreins arose by direct modification of salivary counterparts, based on their phylogenetic proximity to salivary proteins in lizards (14). This suggests a commonality of biochemical mechanisms inherited from the ancestral salivary GRN. Furthermore, kallikreins found in the ancestral salivary GRN's predispose the evolution of envenomation strategies based on hypotensive shock, one of two main strategies for prey immobilization by modern venomous snakes (69).

While kallikrein-like serine proteases represent the most striking and taxonomically diverse example of exaptation, other ancestral salivary components have been recruited by a range of taxa. For instance, cysteine-rich secretory proteins (CRISPs), which are expressed in many tissues including salivary glands, are commonly found in the venom of snakes and of lizards (*Heloderma*) (14, 70). CRISPs play a wide variety of roles in nonvenomous tissues, and their function appears likewise diverse in venoms (71). This illustrates that the ancestral expression of a gene need not be limited to saliva since many of them are also expressed in other tissues as well, as are many, if not all, elements of the metavenom network (Figs. 1*C* and 3*C*). Rather, these genes are united by pharmacology that could be easily repurposed and overexpressed in the novel venomous context. It further suggests that the salivary GRN is flexible, in that it can evolve to secrete high levels of a wide range of proteins.

Stage 2: Gene recruitment. Snake venoms arose from the same ancestral GRN and followed the same first evolutionary step relying on initial exaptation of existing components. Yet, today they contain numerous novel toxins and bear little resemblance to the composition of ancestral saliva. Incorporation of novel toxins has occurred relatively infrequently, and the process remains poorly understood at the transcriptional level. For example, recent insights into the evolution of snake venom metalloproteinases found that they are related to the mammalian *adam28* gene (35, 72). This gene is expressed in many tissues, but only weakly in the salivary glands of some species (73), and, furthermore, it is a transmembrane rather than a secreted protein. While the series of sequential deletions necessary for the protein sequence to acquire toxicity have been revealed (35), the corresponding changes in gene expression accompanying them remain a mystery. Similarly, while the origin of phospholipases A_2 has been traced to a common amniote ancestor, the steps required for its neofunctionalization remain obscure (74). One attribute common to these toxins is that prototoxin genes are expressed in a variety of tissues. As a result, metavenom network genes likely already interact with "future" toxin genes in other tissues, facilitating their eventual recruitment into the venom.

Conclusion

When comparing between organisms, it is important to remember that all lineages have experienced different degrees of trait loss and gain (75). Therefore, most organisms typically show combinations of both ancestral and derived characters (76). Despite being derived phenotypes experiencing strong selection, snake venoms rely on a conserved secretory GRN that is expressed in ancestral saliva and other tissues. Key components of the GRN appear to have been exapted for the evolution of snake and other vertebrate oral venoms. Rather than being nonhomologous products of convergent evolution, as previously believed (41, 42, 77), gene coexpression analysis revealed that these venom systems share a deep homology at the level of regulatory architectures. As a result, the evolution of toxicity in vertebrate saliva may be more common than currently recognized, and the line between vertebrates with and without oral venoms much less clear.

Materials and Methods

RNA Extraction and Sequencing. RNA was extracted from 30 specimens of *P. mucrosquamatus* which were collected from various localities throughout Okinawa, Japan. Venom glands were harvested from all 30 specimens while nonvenom tissues were harvested from 5 specimens. Specimens had almost equal distribution of male and female (m: 21, f: 26) (Dataset SM1). Venom was extracted from all specimens at day 0 and glands were harvested at several time points (days 1, 2, 4, and 8). RNA-seq libraries were prepared as described in refs. 78 and 10. Reads were mapped using Bowtie 2 within the RSEM package, which was also used to quantify transcript abundance (79). Raw RNA-seq reads are available under NCBI accession PRJDB4386. Further details like specific locations of sampling and generation of RNA data can be found in ref. 12.

Network Construction. Weighted gene coexpression analysis was conducted using the WGCNA package in R (23). The input data consisted of a regularized log transformed matrix of 18,313 genes (as columns) and 29 libraries (as rows) of the venom gland which was filtered for low expressed transcripts (transcripts per million [tpm] < 0.05). One of the venom gland libraries was excluded in all further analysis due to low spike (*SI Appendix, Supplementary Materials*). A characteristic organizational feature of biological networks is a "scale-free" topology, where connections follow a power-law distribution, such that there are very few nodes with very many connections and vice versa (80, 81). To attain scale-free topology, a soft threshold of 13 was selected based on results from the "pickSoftThreshold" function in the WGCNA package. After a soft threshold was estimated, a hierarchical clustering algorithm was used to identify modules of highly connected genes. A threshold of 0.2 and minimum module size = 30 was used to merge very similar expression profiles to obtain a total of 29 modules. We used the "modulePreservation" function to calculate preservation of module characteristics of the metavenom network module, between a reference and test dataset. In all cases, the reference dataset was the metavenom network module, while the test was a topological overlap matrix (TOM) from either nonvenom tissues or venom tissue in cobra. The Z_{summay} is a composite statistic that combines statistical summaries of network density and connectivity to get a reliable estimate of whether network characteristics are preserved between reference and test (24). Simulations revealed that a threshold of $2 > Z_{summary} < 10$ indicates weak to moderate evidence of preservation, while $Z_{summary} > 10$ implies strong preservation and $Z_{summary} < 2$ implies no preservation (24).

Differential gene expression analysis was carried out in edgeR (82). Transcripts with missing or very low read counts were filtered out before performing the tests. Libraries were normalized (using suggested TMM [trimmed mean of M] values) to account for compositional bias as well as account for any size variations between libraries. We performed an ANOVA-like test to identify differentially expressed genes between four tissue groups; venom gland, liver, kidney, and heart. A quasi-likelihood F test was then applied to identify differentially expressed genes between the four groups (at $P < 0.05$ significance). Gene expression plots were made using the same libraries that we used to estimate differential gene expression (at day = 1).

External validation of module preservation. We conducted an external validation of our data and WGCNA algorithm parameters using an external study of human salivary gland gene expression (25). This dataset uses specimens with salivary gland pathology and was carried out on microarrays. We expected that despite these differences, if the metavenom network is conserved, it will show overlap with one or more modules inferred in the human data. We tested for overlap using Fisher's exact tests correcting for multiple comparisons using the Benjamini–Hochberg procedure with the false discovery rate set at 0.05.

Functional Annotation of Gene Sets. GO terms of habu genes were annotated using Blast2GO software (using a BLAST e-value cutoff $\leq 10^{-3}$) (83). We used both BLAST and InterProt results of the *P. mucrosquamatus* genome (PRJDB4386) as input for Blast2GO. Using both nucleotide and protein sequences allowed more accurate annotation of GO terms. GO terms enrichment analysis was carried out using the GOstats package in R (84). Depending on the analysis (e.g., GO enrichment of metavenom network genes or enrichment of expanded gene families) different gene sets were used as the test data and GO annotations (of the set of all genes) from Blast2GO was used as the "universe."

Orthology Estimate and Comparative Transcriptomics. Orthologs for habu (*P. mucrosquamatus* ncbi tax id: 103944), human (*Homo sapiens* 9606), chimp (*Pan troglodytes*: 9598), mouse (*Mus musculus*: 10090), dog (*Canis familiaris*: 9615), anole (*Anolis carolinensis*: 28377), chicken (*Gallus gallus*: 9031), and frog (*Xenopus tropicalis*: 8364) were obtained from the "Gene" database of NCBI (ftp://ftp.ncbi.nlm.nih.gov/gene/DATA/gene_orthologs.gz). These orthologs were calculated by NCBI's Eukaryotic Genome Annotation pipeline that combines both protein sequence similarity as well as local synteny information. Furthermore, orthologous relations were additionally assigned after manual curation. A combination of command line and R scripts was used to extract a list of one-to-one orthologs shared between all eight taxa (*SI Appendix, Supplementary Materials*). In addition to using the orthologs defined by NCBI, we carried out phylogenetic ortholog estimation using OrthoFinder (OF) (28). OF uses protein sequences to infer orthogroups and then combines information from gene trees and species trees to distinguish between gene copies arising from speciation or duplication events within lineages. OF also has the added advantage of removing any errors that tend to occur during similarity-based assignment of orthologs (85). Protein sequences for eight taxa were obtained from Ensembl (86). Cobra (*Naja naja*: 35670) protein and transcript sequences were obtained by request from the authors (29). Using both these approaches we obtained a combined list of 2,682 one-to-one expressed orthologs (see next section) between nine taxa. From these we filtered metavenom network orthologs based on the habu genes present in the metavenom network. This results in a list of 546 expressed metavenom network orthologs found in all nine taxa. For the expanded dataset used in the supplementary analysis protein and transcript sequences were obtained from NCBI. Sequence data for the Leopard gecko was obtained from ref. 87. *C. viridis* sequence data were obtained from ref. 33. One-to-one orthologs were obtained using OrthoFinder, which resulted in a total of 2,291 orthologs, and 460 metavenom network orthologs, expressed across all tissues in 12 taxa (*SI Appendix, Supplementary Materials*). RNA data for each species and tissue were obtained from the Sequence Read Archive (SRA) database (Dataset SM2). Datasets were from a variety of sources including published studies (29, 31, 33, 88–91) and large-scale sequencing projects like the Broad Institute's canine genomic resources and the ENCODE project (92). Where possible, at least three libraries for each tissue from each taxa were used to compile our comparative dataset, and only data generated from healthy, adult tissues were used. All the sources did not distinguish between salivary gland subtypes and used whole tissue due to the high genetic similarity of subtypes (33, 93). We used the "fasterq-dump" function in SRA toolkit 2.9.1 (https://github.com/ncbi/sra-tools/wiki) to download fastq files, which were quantified using kallisto (94). Kallisto indices for human, mouse, chimp, dog, anole, frog, and chicken were created using GTF and cDNA files from the Ensembl database (86). Index for cobra was made using annotation and transcript files from Suryamohan et al. (29). Indices for all other studies were constructed from transcript data from NCBI (python, corn snake) or obtained from their respective studies (leopard gecko and *C. viridis*). For single end reads we set length parameter to 350 and SD of length fragment to 150. A custom R script was used to aggregate transcript-level read counts to gene-level read counts. Once total tpm was obtained for each tissue from each taxa, the data were filtered to obtain a final dataset of one-to-one orthologs expressed across all tissues across all nine taxa. To allow for comparisons across samples, expression levels were normalized. Normalization was carried out by adding a pseudo count of 1×10^{-5} (to prevent log[0] scores), followed by $\log_2$ transformation. The transformed data were then quantile normalized among samples. Quantile normalization ensured equal across sample distribution of gene expression levels so as to minimize the effects of technical artifacts (95, 96).

Our aim was to identify any conserved pattern of expression present between homologous tissues from multiple taxa; however, identifying patterns in expression data from multiple species as well as multiple studies

requires the removal of their respective batch effects (97). The batch effect imparted by species is due to the level of shared functionality of genetic processes, where evolutionary changes (during speciation) in shared molecular machinery will simultaneously alter the expression of genes in all tissues, thereby masking any historical signals of homology (98, 99). To remove these batch effects and identify patterns (if any) of homology in expression between tissues we used an empirical Bayes method (implemented via the ComBat function in the sva R package) (100). We used the plotPCA function in the DESeq2 package (101) to carry out principal component analysis. Using both species and study as batch effects produced similar results, although species explained more variation and provided better resolution of underlying tissue-specific trends (*SI Appendix, Supplementary Materials*).

Gene Family Evolution. Gene family evolution across amniotes was investigated using CAFE v5.0 (37, 102). CAFE models gene family evolution across a species tree using a stochastic birth and death process. An ultrametric species tree was drawn in Mesquite (103) and divergence times were estimated using http://www.timetree.org/. Protein sequences for seven taxa were obtained from Ensembl and the rest (habu and cobra) from NCBI. Gene families were inferred with BLAST and MCL (implemented in CAFE), using proteins present in the metavenom network as query sequences. This resulted in 250 estimated gene families. Although most of our taxa are model organisms with well-assembled genomes, for increased statistical robustness, we estimated an error model due to genome assembly error which was later used for λ analysis (36) (Dataset SM3). The rate parameter λ describes the probability that any gene will either be gained or lost, where a higher λ denotes rapid gene family evolution (37). We used a global λ (λ_1) as our null model and a different rate parameter (λ_2) for the lineage leading up to venomous snakes to test the hypothesis that gene families evolved faster in the lineage leading up to venomous snakes compared to other lineages. Simulations of gene families from observed data and a subsequent likelihood ratio test using the global λ (λ_1) estimate and lineage specific λ (λ_2) was used to determine significance. Once the log likelihoods were obtained, lhtest.R script (provided by CAFE) was used to create a histogram with a null distribution obtained from simulations. Significance is determined by how far left the observed likelihood ratio ($2 \times$ lnLglobal − lnLmulti) would fall on the tail of the distribution. In our case the likelihood ratio count would fall on the far left of the distribution indicating a very low *P* value (*SI Appendix*, Fig. S5). Along with inferring rates of gene family evolution, CAFE also determines expansions or contractions in gene size by calculating ancestral states at nodes along the tree. For each gene family CAFE computes a *P* value associated with the gene family size in extant species given the model of gene family evolution (102). This was used to determine which gene families underwent significant expansion, contraction, or stayed the same in venomous snakes (Dataset S5).

Data Availability. All code, data, figures, and tables can be found at https://github.com/agneeshbarua/Metavenom (104). All study data are included in the article and/or supporting information.

ACKNOWLEDGMENTS. We would like to acknowledge the DNA Sequencing Section (SQC) and Scientific Computing and Data Analysis Section (SCDA) of the Okinawa Institute of Science and Technology Graduate University for their assistance in sequencing of libraries and high-performance computing, respectively.

1. T. F. Duda Jr, S. R. Palumbi, Evolutionary diversification of multigene families: Allelic selection of toxins in predatory cone snails. *Mol. Biol. Evol.* **17**, 1286–1293 (2000).
2. Y. Moran *et al*., Concerted evolution of sea anemone neurotoxin genes is revealed through analysis of the Nematostella vectensis genome. *Mol. Biol. Evol.* **25**, 737–747 (2008).
3. D. R. Rokyta, A. R. Lemmon, M. J. Margres, K. Aronow, The venom-gland transcriptome of the eastern diamondback rattlesnake (Crotalus adamanteus). *BMC Genomics* **13**, 312 (2012).
4. D. R. Rokyta, K. P. Wray, M. J. Margres, The genesis of an exceptionally lethal venom in the timber rattlesnake (Crotalus horridus) revealed through comparative venom-gland transcriptomics. *BMC Genomics* **14**, 394 (2013).
5. J. M. Surm *et al*., A process of convergent amplification and tissue-specific expression dominates the evolution of toxin and toxin-like genes in sea anemones. *Mol. Ecol.* **28**, 2272–2289 (2019).
6. M. J. Margres *et al*., Quantity, not quality: Rapid adaptation in a polygenic trait proceeded exclusively through expression differentiation. *Mol. Biol. Evol.* **34**, 3099–3110 (2017).
7. A. Barua, A. S. Mikheyev, Many options, few solutions: Over 60 My snakes converged on a few optimal venom formulations. *Mol. Biol. Evol.* **36**, 1964–1974 (2019).
8. V. Schendel, L. D. Rash, R. A. Jenner, E. A. B. Undheim, The diversity of venom: The importance of behavior and venom system morphology in understanding its ecology and evolution. *Toxins (Basel)* **11**, 666 (2019).
9. G. Zancolli, N. R. Casewell, Venom systems as models for studying the origin and regulation of evolutionary novelties. *Mol. Biol. Evol.* **37**, 2777–2790 (2020).
10. S. D. Aird *et al*., Snake venoms are integrated systems, but abundant venom proteins evolve more rapidly. *BMC Genomics* **16**, 647 (2015).
11. A. Barua, A. S. Mikheyev, Toxin expression in snake venom evolves rapidly with constant shifts in evolutionary rates. *Proc. Biol. Sci.* **287**, 20200613 (2020).
12. S. D. Aird *et al*., Population genomic analysis of a pitviper reveals microevolutionary forces underlying venom chemistry. *Genome Biol. Evol.* **9**, 2640–2649 (2017).
13. A. Malhotra, S. Creer, J. B. Harris, R. S. Thorpe, The importance of being genomic: Non-coding and coding sequences suggest different models of toxin multi-gene family evolution. *Toxicon* **107**, 344–358 (2015).
14. B. G. Fry, 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.* **15**, 403–420 (2005).
15. B. G. Fry *et al*., Evolution of an arsenal: Structural and functional diversification of the venom system in the advanced snakes (Caenophidia). *Mol. Cell. Proteomics* **7**, 215–246 (2008).
16. A.-L. Barabási, Z. N. Oltvai, Network biology: Understanding the cell's functional organization. *Nat. Rev. Genet.* **5**, 101–113 (2004).
17. J. A. Miller, S. Horvath, D. H. Geschwind, Divergence of human and mouse brain transcriptome highlights Alzheimer disease pathways. *Proc. Natl. Acad. Sci. U.S.A.* **107**, 12698–12703 (2010).
18. M. C. Oldham, S. Horvath, D. H. Geschwind, Conservation and evolution of gene coexpression networks in human and chimpanzee brains. *Proc. Natl. Acad. Sci. U.S.A.* **103**, 17973–17978 (2006).
19. M. Filteau, S. A. Pavey, J. St-Cyr, L. Bernatchez, Gene coexpression networks reveal key drivers of phenotypic divergence in lake whitefish. *Mol. Biol. Evol.* **30**, 1384–1396 (2013).
20. J. M. Stuart, E. Segal, D. Koller, S. K. Kim, A gene-coexpression network for global discovery of conserved genetic modules. *Science* **302**, 249–255 (2003).
21. J. D. Allen, Y. Xie, M. Chen, L. Girard, G. Xiao, Comparing statistical methods for constructing large scale gene networks. *PLoS One* **7**, e29348 (2012).
22. V. A. H.-T. Guido Sanguinetti, Ed., *Gene Regulatory Networks.pdf* (Springer Science and Business Media, 2019).
23. P. Langfelder, S. Horvath, WGCNA: An R package for weighted correlation network analysis. *BMC Bioinformatics* **9**, 559 (2008).
24. P. Langfelder, R. Luo, M. C. Oldham, S. Horvath, Is my network module preserved and reproducible? *PLOS Comput. Biol.* **7**, e1001057 (2011).
25. S. Horvath *et al*., Systems analysis of primary Sjögren's syndrome pathogenesis in salivary glands identifies shared pathways in human and a mouse model. *Arthritis Res. Ther.* **14**, R238 (2012).
26. D. Ron, P. Walter, Signal integration in the endoplasmic reticulum unfolded protein response. *Nat. Rev. Mol. Cell Biol.* **8**, 519–529 (2007).
27. NCBI Resource Coordinators, Database resources of the National center for biotechnology information. *Nucleic Acids Res.* **44**, D7–D19 (2016).
28. D. M. Emms, S. Kelly, OrthoFinder: Phylogenetic orthology inference for comparative genomics. *Genome Biol.* **20**, 238 (2019).
29. K. Suryamohan *et al*., The Indian cobra reference genome and transcriptome enables comprehensive identification of venom toxins. *Nat. Genet.* **52**, 106–117 (2020).
30. A. Breschi *et al*., Gene-specific patterns of expression variation across organs and species. *Genome Biol.* **17**, 151 (2016).
31. A. D. Hargreaves, M. T. Swain, M. J. Hegarty, D. W. Logan, J. F. Mulley, Restriction and recruitment-gene duplication and the origin and evolution of snake venom toxins. *Genome Biol. Evol.* **6**, 2088–2095 (2014).
32. D. R. Schield *et al*., The origins and evolution of chromosomes, dosage compensation, and mechanisms underlying venom regulation in snakes. *Genome Res.* **29**, 590–601 (2019).
33. X. Gao, M. S. Oei, C. E. Ovitt, M. Sincan, J. E. Melvin, Transcriptional profiling reveals gland-specific differential expression in the three major salivary glands of the adult mouse. *Physiol. Genomics* **50**, 263–271 (2018).
34. N. L. Dowell *et al*., The deep origin and recent loss of venom toxin genes in rattlesnakes. *Curr. Biol.* **26**, 2434–2445 (2016).
35. M. W. Giorgianni *et al*., The origin and diversification of a novel protein family in venomous snakes. *Proc. Natl. Acad. Sci. U.S.A.* **117**, 10911–10920 (2020).
36. M. V. Han, G. W. C. Thomas, J. Lugo-Martinez, M. W. Hahn, Estimating gene gain and loss rates in the presence of error in genome assembly and annotation using CAFE 3. *Mol. Biol. Evol.* **30**, 1987–1997 (2013).
37. M. W. Hahn, T. De Bie, J. E. Stajich, C. Nguyen, N. Cristianini, Estimating the tempo and mode of gene family evolution from comparative genomic data. *Genome Res.* **15**, 1153–1160 (2005).
38. M. R. J. Carlson *et al*., Gene connectivity, function, and sequence conservation: Predictions from modular yeast co-expression networks. *BMC Genomics* **7**, 40 (2006).
39. M. Rotival, E. Petretto, Leveraging gene co-expression networks to pinpoint the regulation of complex traits and disease, with a focus on cardiovascular traits. *Brief. Funct. Genomics* **13**, 66–78 (2014).
40. A. J. Mason *et al*., Trait differentiation and modular toxin expression in palm-pitvipers. *BMC Genomics* **21**, 147 (2020).
41. R. Ligabue-Braun, H. Verli, C. R. Carlini, Venomous mammals: A review. *Toxicon* **59**, 680–695 (2012).

42. N. R. Casewell *et al*., Solenodon genome reveals convergent evolution of venom in eulipotyphlan mammals. *Proc. Natl. Acad. Sci. U.S.A.* **116**, 25745–25755 (2019).
43. E. B. Rosenblum, C. E. Parent, E. E. Brandt, The molecular basis of phenotypic convergence. *Annu. Rev. Ecol. Evol. Syst.* **45**, 203–226 (2014).
44. S. P. Mackessy, A. J. Saviola, Understanding biological roles of venoms among the Caenophidia: The importance of rear-fanged snakes. *Integr. Comp. Biol.* **56**, 1004–1021 (2016).
45. G. Xu, S. R. Jaffrey, The new landscape of protein ubiquitination. *Nat. Biotechnol.* **29**, 1098–1100 (2011).
46. J. Hwang, L. Qi, Quality control in the endoplasmic reticulum: Crosstalk between ERAD and UPR pathways. *Trends Biochem. Sci.* **43**, 593–605 (2018).
47. M. Michalak, J. Groenendyk, E. Szabo, L. I. Gold, M. Opas, Calreticulin, a multi-process calcium-buffering chaperone of the endoplasmic reticulum. *Biochem. J.* **417**, 651–666 (2009).
48. T. J. Bergmann *et al*., Chemical stresses fail to mimic the unfolded protein response resulting from luminal load with unfolded polypeptides. *J. Biol. Chem.* **293**, 5600–5612 (2018).
49. M. Bai *et al*., Conserved roles of C. elegans and human MANFs in sulfatide binding and cytoprotection. *Nat. Commun.* **9**, 897 (2018).
50. H. Safavi-Hemami *et al*., Rapid expansion of the protein disulfide isomerase gene family facilitates the folding of venom peptides. *Proc. Natl. Acad. Sci. U.S.A.* **113**, 3227–3232 (2016).
51. L. Ellgaard, L. W. Ruddock, The human protein disulphide isomerase family: Substrate interactions and functional properties. *EMBO Rep.* **6**, 28–32 (2005).
52. M. Michalak, E. F. Corbett, N. Mesaeli, K. Nakamura, M. Opas, Calreticulin: One protein, one gene, many functions. *Biochem. J.* **344**, 281–292 (1999).
53. D. T. Rutkowski, R. S. Hegde, Regulation of basal cellular physiology by the homeostatic unfolded protein response. *J. Cell Biol.* **189**, 783–794 (2010).
54. X. Yang *et al*., ATF4 is a substrate of RSK2 and an essential regulator of osteoblast biology; implication for Coffin-Lowry Syndrome. *Cell* **117**, 387–398 (2004).
55. A.-H. Lee, E. F. Scapa, D. E. Cohen, L. H. Glimcher, Regulation of hepatic lipogenesis by the transcription factor XBP1. *Science* **320**, 1492–1496 (2008).
56. Y. Wang, L. Vera, W. H. Fischer, M. Montminy, The CREB coactivator CRTC2 links hepatic ER stress and fasting gluconeogenesis. *Nature* **460**, 534–537 (2009).
57. D. Senft, Z. A. Ronai, UPR, autophagy, and mitochondria crosstalk underlies the ER stress response. *Trends Biochem. Sci.* **40**, 141–148 (2015).
58. M.-J. Barrera *et al*., Endoplasmic reticulum stress in autoimmune diseases: Can altered protein quality control and/or unfolded protein response contribute to autoimmunity? A critical review on sjögren's syndrome. *Autoimmun. Rev.* **17**, 796–808 (2018).
59. B. W. Perry, D. R. Schield, A. K. Westfall, S. P. Mackessy, T. A. Castoe, Physiological demands and signaling associated with snake venom production and storage illustrated by transcriptional analyses of venom glands. *Sci. Rep.* **10**, 18083 (2020).
60. A. Brückner, J. Parker, Molecular evolution of gland cell types and chemical interactions in animals. *J. Exp. Biol.* **223**(suppl. 1), jeb211938 (2020).
61. A. Pavlopoulou, G. Pampalakis, I. Michalopoulos, G. Sotiropoulou, Evolutionary history of tissue kallikreins. *PLoS One* **5**, e13781 (2010).
62. V. L. Koumandou, A. Scorilas, Evolution of the plasma and tissue kallikreins, and their alternative splicing isoforms. *PLoS One* **8**, e68074 (2013).
63. M. Hiramatsu, K. Hatakeyama, N. Minami, Male mouse submaxillary gland secretes highly toxic proteins. *Experientia* **36**, 940–942 (1980).
64. D. H. Dean, R. N. Hiramoto, Lethal effect of male rat submandibular gland homogenate for rat neonates. *J. Oral Pathol.* **14**, 666–669 (1985).
65. J. C. Huang, K. Hoshino, Y. T. Kim, F. S. Chebib, Species and strain differences in the lethal factor of the mouse submandibular gland. *Can. J. Physiol. Pharmacol.* **55**, 1107–1111 (1977).
66. M. Kita *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.* **101**, 7542–7547 (2004).
67. G. Datta, A. T. Tu, Structure and other chemical characterizations of gila toxin, a lethal toxin from lizard venom. *J. Pept. Res.* **50**, 443–450 (1997).
68. I. Koludarov *et al*., Enter the dragon: The dynamic and multifunctional evolution of anguimorpha lizard venoms. *Toxins (Basel)* **9**, 242 (2017).
69. S. D. Aird, Ophidian envenomation strategies and the role of purines. *Toxicon* **40**, 335–393 (2002).
70. K. Sunagar, W. E. Johnson, S. J. O'Brien, V. Vasconcelos, A. Antunes, Evolution of CRISPs associated with toxicoferan-reptilian venom and mammalian reproduction. *Mol. Biol. Evol.* **29**, 1807–1822 (2012).
71. W. H. Heyborne, S. P. Mackessy, "Cysteine-rich secretory proteins in reptile venoms" in *Handbook of Venoms and Toxins of Reptiles*, S. P. Mackessy, Ed. (CRC Press, 2016), pp. 325–335.
72. F. J. Vonk *et al*., The king cobra genome reveals dynamic gene evolution and adaptation in the snake venom system. *Proc. Natl. Acad. Sci. U.S.A.* **110**, 20651–20656 (2013).
73. J. A. Jury, A. C. Perry, L. Hall, Identification, sequence analysis and expression of transcripts encoding a putative metalloproteinase, eMDC II, in human and macaque epididymis. *Mol. Hum. Reprod.* **5**, 1127–1134 (1999).
74. I. Koludarov, T. N. W. Jackson, A. Pozzi, A. S. Mikheyev, Family saga: Reconstructing the evolutionary history of a functionally diverse gene family reveals complexity at the genetic origins of novelty. *bioRxiv* (2019).
75. R. R. Strathmann, D. J. Eernisse, What molecular phylogenies tell us about the evolution of larval forms. *Integr. Comp. Biol.* **34**, 502–512 (1994).
76. R. A. Jenner, Unburdening evo-devo: Ancestral attractions, model organisms, and basal baloney. *Dev. Genes Evol.* **216**, 385–394 (2006).
77. K. E. Folinsbee, Evolution of venom across extant and extinct eulipotyphlans. *C. R. Palevol* **12**, 531–542 (2013).
78. S. D. Aird *et al*., Quantitative high-throughput profiling of snake venom gland transcriptomes and proteomes (Ovophis okinavensis and Protobothrops flavoviridis). *BMC Genomics* **14**, 790 (2013).
79. B. Langmead, S. L. Salzberg, Fast gapped-read alignment with Bowtie 2. *Nat. Methods* **9**, 357–359 (2012).
80. B. Zhang, S. Horvath, A general framework for weighted gene co-expression network analysis. *Stat. Appl. Genet. Mol. Biol.* **4**, e17 (2005).
81. M. L. Siegal, D. E. L. Promislow, A. Bergman, Functional and evolutionary inference in gene networks: Does topology matter? *Genetica* **129**, 83–103 (2007).
82. M. D. Robinson, D. J. McCarthy, G. K. Smyth, edgeR: A Bioconductor package for differential expression analysis of digital gene expression data. *Bioinformatics* **26**, 139–140 (2010).
83. S. Götz *et al*., High-throughput functional annotation and data mining with the Blast2GO suite. *Nucleic Acids Res.* **36**, 3420–3435 (2008).
84. S. Falcon, R. Gentleman, Using GOstats to test gene lists for GO term association. *Bioinformatics* **23**, 257–258 (2007).
85. D. M. Emms, S. Kelly, OrthoFinder: Solving fundamental biases in whole genome comparisons dramatically improves orthogroup inference accuracy. *Genome Biol.* **16**, 157 (2015).
86. A. D. Yates *et al*., Ensembl 2020. *Nucleic Acids Res.* **48**, D682–D688 (2020).
87. Z. Xiong *et al*., Draft genome of the leopard gecko, Eublepharis macularius. *Gigascience* **5**, 47 (2016).
88. D. Brawand *et al*., The evolution of gene expression levels in mammalian organs. *Nature* **478**, 343–348 (2011).
89. N. L. Barbosa-Morais *et al*., The evolutionary landscape of alternative splicing in vertebrate species. *Science* **338**, 1587–1593 (2012).
90. R. Marin *et al*., Convergent origination of a *Drosophila*-like dosage compensation mechanism in a reptile lineage. *Genome Res.* **27**, 1974–1987 (2017).
91. M. Alame *et al*., The molecular landscape and microenvironment of salivary duct carcinoma reveal new therapeutic opportunities. *Theranostics* **10**, 4383–4394 (2020).
92. ENCODE Project Consortium, A user's guide to the encyclopedia of DNA elements (ENCODE). *PLoS Biol.* **9**, e1001046 (2011).
93. F. de Paula *et al*., Overview of human salivary glands: Highlights of morphology and developing processes. *Anat. Rec. (Hoboken)* **300**, 1180–1188 (2017).
94. N. L. Bray, H. Pimentel, P. Melsted, L. Pachter, Near-optimal probabilistic RNA-seq quantification. *Nat. Biotechnol.* **34**, 525–527 (2016).
95. B. M. Bolstad, R. A. Irizarry, M. Astrand, T. P. Speed, A comparison of normalization methods for high density oligonucleotide array data based on variance and bias. *Bioinformatics* **19**, 185–193 (2003).
96. B. Qiu *et al*., Towards reconstructing the ancestral brain gene-network regulating caste differentiation in ants. *Nat. Ecol. Evol.* **2**, 1782–1791 (2018).
97. Y. Gilad, O. Mizrahi-Man, A reanalysis of mouse ENCODE comparative gene expression data. *F1000 Res.* **4**, 121 (2015).
98. J. M. Musser, G. P. Wagner, Character trees from transcriptome data: Origin and individuation of morphological characters and the so-called "species signal". *J. Exp. Zoolog. B Mol. Dev. Evol.* **324**, 588–604 (2015).
99. C. Liang, J. M. Musser, A. Cloutier, R. O. Prum, G. P. Wagner, Pervasive correlated evolution in gene expression shapes cell and tissue type transcriptomes. *Genome Biol. Evol.* **10**, 538–552 (2018).
100. J. T. Leek, W. E. Johnson, H. S. Parker, A. E. Jaffe, J. D. Storey, The sva package for removing batch effects and other unwanted variation in high-throughput experiments. *Bioinformatics* **28**, 882–883 (2012).
101. M. I. Love, W. Huber, S. Anders, Moderated estimation of fold change and dispersion for RNA-seq data with DESeq2. *Genome Biol.* **15**, 550 (2014).
102. T. De Bie, N. Cristianini, J. P. Demuth, M. W. Hahn, CAFE: A computational tool for the study of gene family evolution. *Bioinformatics* **22**, 1269–1271 (2006).
103. W. P. A. D. R. M. Maddison, Mesquite: A modular system for evolutionary analysis. Version 3.61 (2019).
104. A. Barua, Metavenom. GitHub. https://github.com/agneeshbarua/Metavenom. Accessed 12 October 2020.

Appendix II:
Barua A, Koludarov I, Mikheyev AS. Co-option of the same ancestral gene family gave rise to mammalian and reptilian toxins**.** *BMC Biology*. 2021;19: 268.

Barua et al. BMC Biology (2021) 19:268
https://doi.org/10.1186/s12915-021-01191-1

BMC Biology

RESEARCH ARTICLE Open Access

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

Agneesh Barua[1*], Ivan Koludarov[2] 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.
[1]Ecology and Evolution Unit, Okinawa Institute of Science and Technology Graduate University, Okinawa, Japan
[3]Research School of Biology, Australian National University, Canberra, ACT, Australia
Full list of author information is available at the end of the article

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 Toxicoferan 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 monophy 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-toxicoferans 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-toxicoferans. 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,

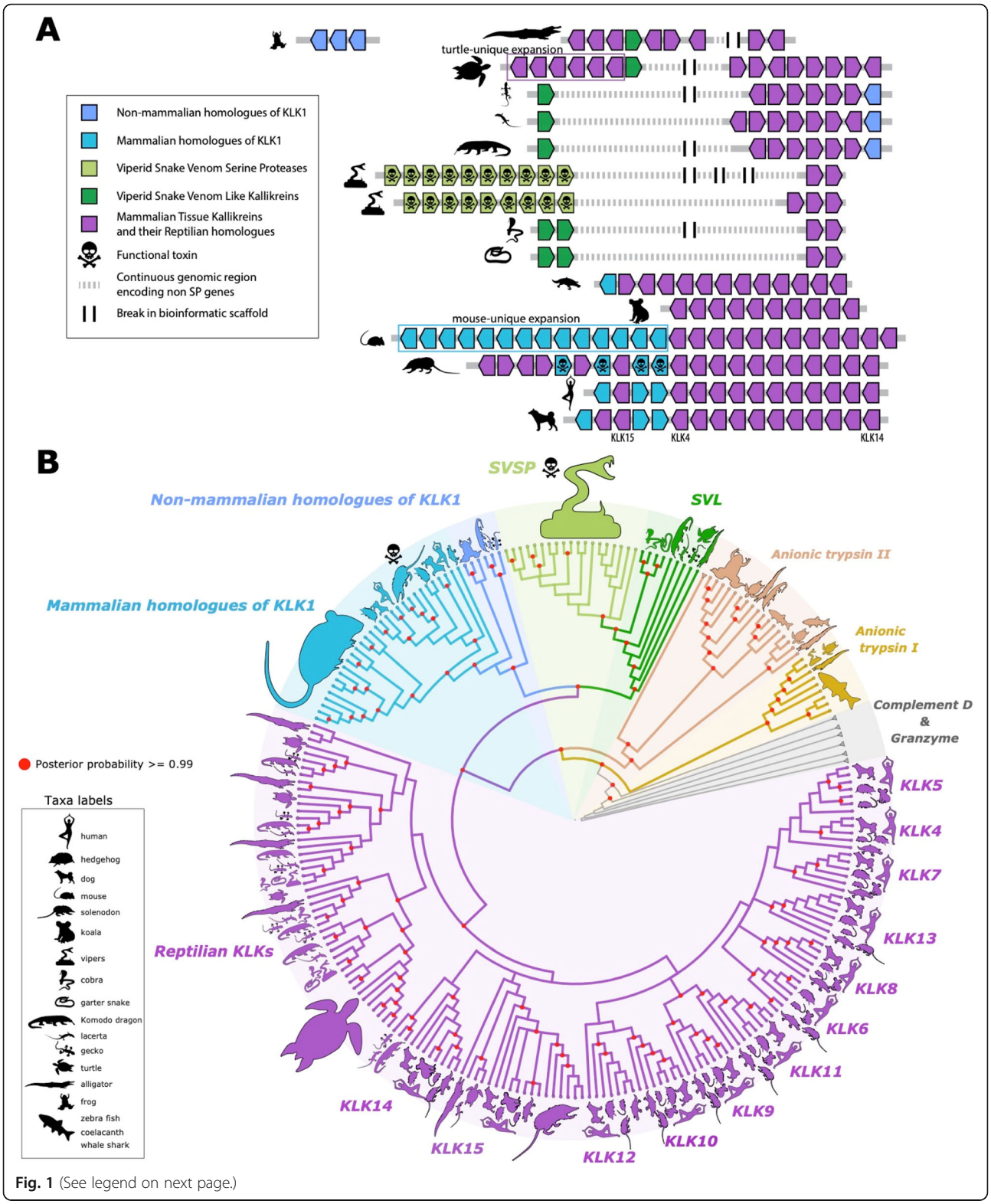


Fig. 1 (See legend on next page.)

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 trypsins, 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

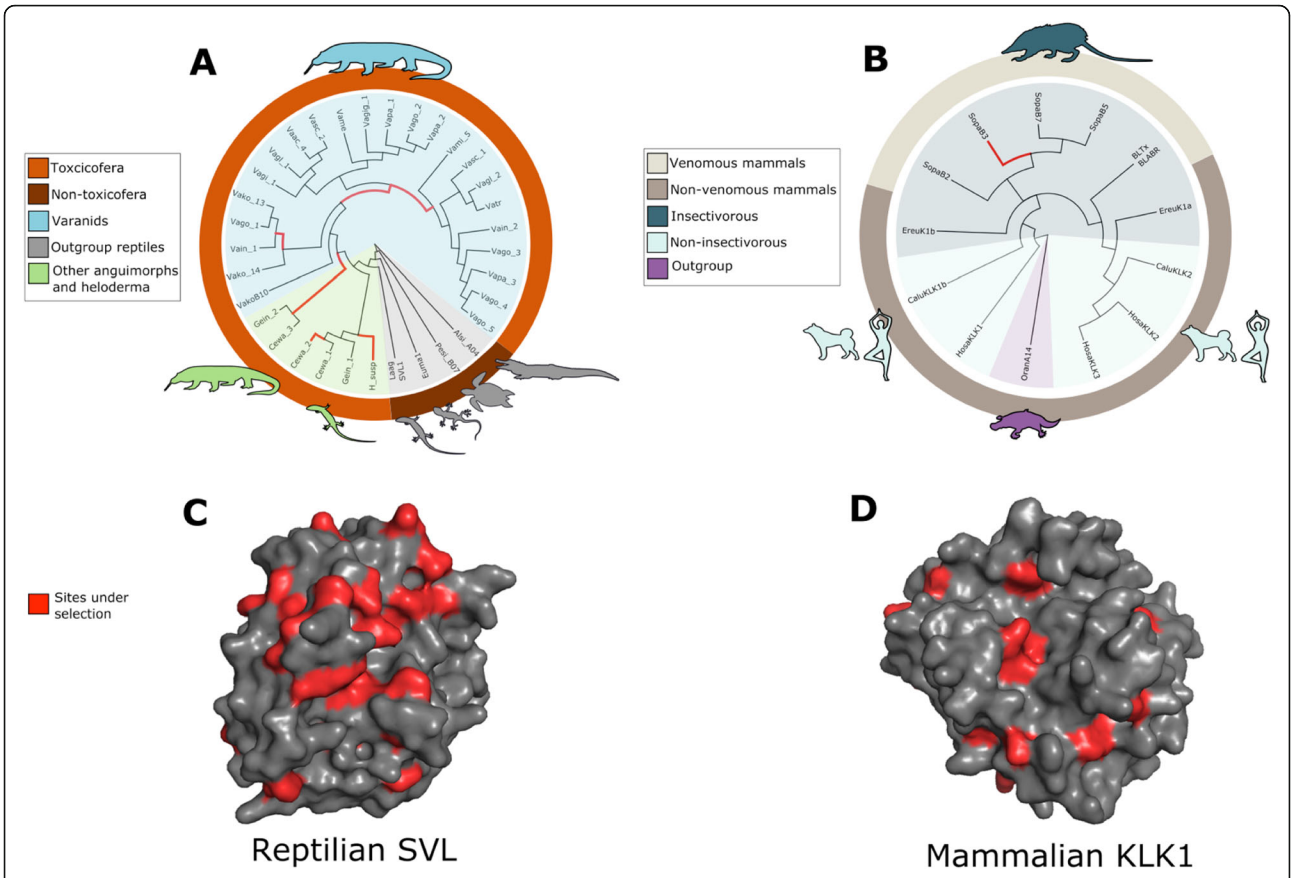


Fig. 2 Venomous lineages experienced different selective forces as compared to non-venomous ones. **A** Toxicofera experienced different selection as compared to non-toxicoferan reptiles. aBSREL found evidence for diversifying selection (red branches) in 6 branches within toxicofera. Alsi, *Alligator sinensis*; Cewa, *Celestus warreni*; Euma, *Eublepharis macularius*; Gein, *Gerrhonotus infernalis*; H_susp, *Heloderma suspectum*; Laag, *Lacerta agilis*; Pesi, *Pelodiscus sinensis*; Vaac, *Varanus acanthurus*; Vagi, *Varanus gilleni*; Vagl, *Varanus glauerti*; Vagig, *Varanus giganteus*; Vain, *Varanus indicus*; Vako, *Varanus komodoensis*; Vame, *Varanus mertens*; Vami, *Varanus mitchelli*; Vapa, *Varanus panoptes*; Vasc, *Varanus scalaris*. **B** Like in reptiles, venomous mammals experienced different selective pressures as compared to non-venomous mammals. aBSREL found evidence of diversifying selection one branch (red) leading up to a Solenodon copy (but see Supplementary Figure 11). Ereu, *Erinaceus europaeus*; Sopa, *Solenodon paradoxus*; BLTx, Blarina toxin; Calu, *Canis lupus*; Oran, *Ornithorhynchus anatinus*. **C** MEME identified 24 sites (in red) in the reptilian SVL that have experienced positive selection. Most of these sites are on the surface (raw output in supplementary dataset 9 and 10). These observations are consistent with previous estimates of high selection on surface residues of toxin serine protease [30]. **D** Unlike reptiles, however, only 10 sites on mammalian KLK1s showed evidence of positive selection, with a few on the surface.

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

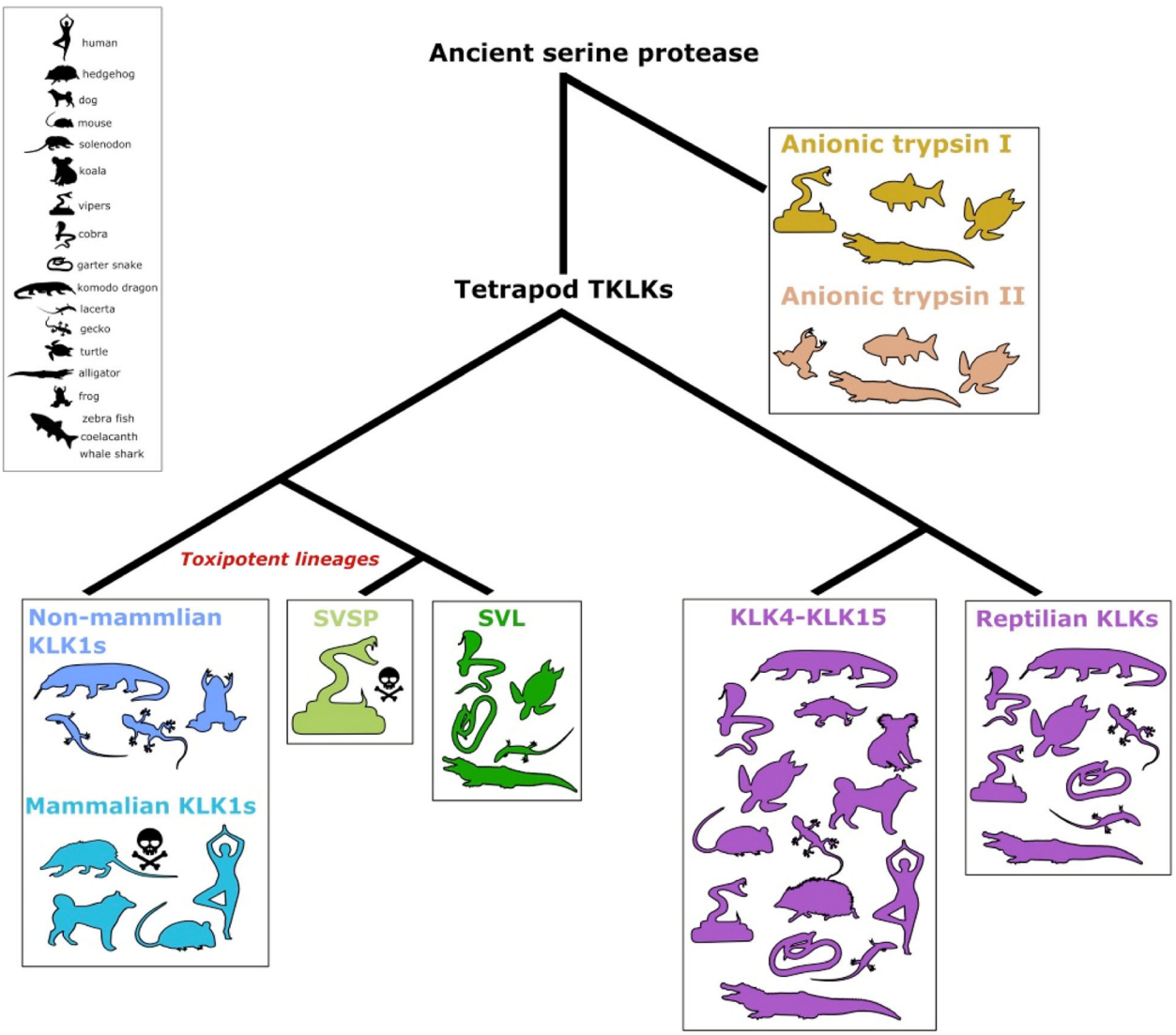


Fig. 3 Evolution of tetrapod toxin kallikreins. The tetrapod lineage of TKLKs evolved from an ancestral serine protease that also gave rise to vertebrate anionic trypsins. From here, the tetrapod TKLKs diverged into the KLK4-KLK15 group and the toxipotent KLK1-SVL-SVSP lineage. Proteins in KLK1-SVL-SVSP lineage are preadapted to become toxins, given their ability to regulate blood pressure when injected into the bloodstream

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, Schr dinger, 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 phylogeney.

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

Received: 26 August 2021 Accepted: 11 November 2021
Published online: 23 December 2021

References
1. 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.
2. 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.
3. 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.
4. Holland PWH. Evolution of homeobox genes. Wiley Interdiscip Rev Dev Biol. 2013;2(1):31–45. https://doi.org/10.1002/wdev.78.
5. 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.
6. 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.
7. 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.
8. 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.
9. 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.
10. 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.
11. 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.
12. 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.
13. 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.
14. 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.
15. 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.
16. 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.
17. Schachter M. Kallikreins and kinins. Physiol Rev. 1969;49(3):509–47. https://doi.org/10.1152/physrev.1969.49.3.509.
18. 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.
19. 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 MJP, 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 biodiscovery: 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, Hiramoto 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.
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 MJP, 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

Appendix III:

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;652;287:20200613.

PROCEEDINGS B

royalsocietypublishing.org/journal/rspb

Research

Cite this article: Barua A, Mikheyev AS. 2020 Toxin expression in snake venom evolves rapidly with constant shifts in evolutionary rates. *Proc. R. Soc. B* **287**: 20200613. http://dx.doi.org/10.1098/rspb.2020.0613

Received: 17 March 2020
Accepted: 30 March 2020

Subject Category:
Evolution

Subject Areas:
evolution, genetics, ecology

Keywords:
gene expression evolution, complex traits, early burst, key innovation, snake venom

Author for correspondence:
Agneesh Barua
e-mail: agneesh.barua@oist.jp

Electronic supplementary material is available online at https://doi.org/10.6084/m9.figshare.c.4938105.

THE ROYAL SOCIETY PUBLISHING

Toxin expression in snake venom evolves rapidly with constant shifts in evolutionary rates

Agneesh Barua[1] and Alexander S. Mikheyev[1,2]

[1]Ecology and Evolution Unit, Okinawa Institute of Science and Technology Graduate University, 1919-1 Tancha, Onna-son, Kunigami-gun, Okinawa-ken 904-0495, Japan
[2]Evolutionary genomics group, Australian National University, Canberra ACT 0200, Australia

AB, 0000-0002-8347-2171

Key innovations provide ecological opportunity by enabling access to new resources, colonization of new environments, and are associated with adaptive radiation. The most well-known pattern associated with adaptive radiation is an early burst of phenotypic diversification. Venoms facilitate prey capture and are widely believed to be key innovations leading to adaptive radiation. However, few studies have estimated their evolutionary rate dynamics. Here, we test for patterns of adaptive evolution in venom gene expression data from 52 venomous snake species. By identifying shifts in tempo and mode of evolution along with models of phenotypic evolution, we show that snake venom exhibits the macroevolutionary dynamics expected of key innovations. Namely, all toxin families undergo shifts in their rates of evolution, likely in response to changes in adaptive optima. Furthermore, we show that rapid-pulsed evolution modelled as a Lévy process better fits snake venom evolution than conventional early burst or Ornstein–Uhlenbeck models. While our results support the idea of snake venom being a key innovation, the innovation of venom chemistry lacks clear mechanisms that would lead to reproductive isolation and thus adaptive radiation. Therefore, the extent to which venom directly influences the diversification process is still a matter of contention.

1. Introduction

Key innovations are adaptations that provide an ecological opportunity by enabling the utilization of previously unexplored niches [1–3]. This enables animals to colonize new environments and in turn facilitates ecological speciation [4,5]. While this concept has considerable intuitive appeal, the idea of key innovations is not exempt from ambiguity and controversy. Throughout history, key innovations have been defined in numerous ways [1]. The many definitions often lead to confusion regarding what key innovations are, and the expected patterns they should exhibit. The long-standing belief is that key innovations lead to the spread of ecological adaptive zones whose eventual outcome is species diversification or adaptive radiation [3,4,6–8]. However, as reviewed by Rabosky [9], key innovations should not be considered the sole reason for differential rates of species diversification. Rather, the role of key innovations should be focused on providing entry into novel ecological niches or adaptive zones, and studies should aim to identify specific shifts in tempo and mode of phenotypic evolution of the assumed key innovation [9,10]. Ecological speciation, i.e. speciation driven by differences in ecology, is considered the primary mode by which adaptive radiation can take place, and as various traits produce specific differences in ecology, certain traits are more strongly associated with the radiation process than others [3,11]. Therefore, it is vital

to explain the effect of trait differences and how they contribute to species diversification, overall phenotypic disparity and ecological divergence.

Evolutionary models are extensively used to study trait evolution and have been used to model everything from body shape evolution to gene expression level evolution [12,13]. Therefore, it is not surprising that evolutionary models are also widely used to study key innovation. However, rarely does one model consistently explain the evolution of key innovations (or traits believed to be key innovations). Some traits are better explained by Brownian motion (BM), others by Ornstein–Uhlenbeck (OU) models; some traits fit a single-peak OU model better, while others a multi-peak OU model; other traits fit neither BM nor OU processes well [14–18]. Along with BM and OU models, it is also possible to model early burst (EB). An EB in speciation rate and trait evolution is believed to be the predominant pattern in adaptive radiation [11,19,20]. While it is possible to model EB, evidence for it is rarely observed in comparative data [21]. The often-conflicting results between these models warrant cautious interpretation of features like evolutionary rates [14]. Perhaps one limitation of these models is using a Gaussian process to model continuous trait evolution. Evolutionary processes can result in changes that are too abrupt to be accounted for by a Gaussian process [22]. Pulsed models, however, can account for abrupt shifts in the continuous character evolution that conventional evolutionary models cannot easily explain [22]. For example, using this approach, Landis & Schraiber found that body size evolution is better represented by rare stochastic pulses of diversification than by conventional EB or multi-optima OU models [11,20]. Therefore, examining traits using a pulsed model of evolution might reveal previously unresolved evolutionary trends.

The complex nature of traits makes it difficult to ascertain how individual components of the complex phenotype contribute towards evolutionary innovation. It also makes it difficult to discern the specific evolutionary trajectories experienced by individual genes. Since gene expression represents the contribution of an individual gene, especially in highly specialized tissues, it is ideal for identifying gene-specific trends in evolutionary rates. This modular nature makes gene expression in certain tissues highly autonomous, such that the activities of genes within that system depend very little on elements outside of it, facilitating the production of specific heritable variations and evolutionary innovations [23,24]. Highly tissue-specific genes would also likely reduce significant pleiotropic constraints and cross-phenotype associations, helping to discern the unique trajectories experienced by individual genes [25]. Despite the usefulness of modelling gene expression, relatively few phenotypes can be meaningfully reduced purely to gene expression levels, making the study of gene expression variation in phenotypic evolution difficult.

Exceptionally, snake venom, which is a complex phenotype composed of secreted proteinaceous mixtures, can essentially be reduced to expression levels of each of its constituent components. This enables us to understand the contribution of each of the constituent genes towards phenotypic variation. Venom toxins can have both agonistic and antagonistic interactions with other toxin components, but how they influence other traits outside the venom system is unclear. On one hand, venoms are integrated systems with different toxins acting in concert to immobilize prey [26]. On the other hand, whether this mode of action introduces an evolutionary constraint is less clear, since there is little phylogenetic covariance between components, and gene–environment constraints appear to act on individual loci, independent of co-expression patterns between toxin genes [27,28].

Each component of the snake venom cocktail is a toxin that can be quantified and traced to a distinct genomic locus [29–32]. Changes in expression levels of individual toxins alter their abundance in the venom, thereby influencing venom efficacy [32–34]. This alteration in venom efficacy impacts the feeding ecology of snakes, which in turn determines how snakes adapt and colonize new niches [35,36]. The strong ecological and evolutionary consequence of toxin expression variation allows us to characterize the gene expression levels of toxins as polygenic phenotypes and trace venom evolution over macroevolutionary timescales.

The idea that venom is a key innovation and that it underlies the extensive radiation of snakes is pervasive in the literature [26,37–41]. Yet, few studies have examined long-term changes in evolutionary rates of venom gene expression in snakes. There are numerous studies that have examined the role of venom in lineage diversification in other taxa [42–45]. In blenny fish, the presence of a venom system in the form of a buccal gland and fang is associated with higher rates of diversification [44]. In tetrapods, the evolution of venoms and poisons is typically associated with an increase in diversification rates (except in amphibians) [42]. There is also a substantial amount of literature suggesting the role of diet in lineage diversification [45–48]. Since snakes use venom primarily for prey procurement, alterations in venom and diet could have an effect on diversification in venomous snakes.

Key innovations, however, have more features than just causing lineage diversification. Key innovations contribute to the expansion of ecological ranges, represent optimal adaptations, and usually undergo changes in evolutionary rates to fill morphospace [49]. Restricting the role of key innovations to just diversification ignores these features and removes focus from evolution of the key innovation itself [9,49]. In this study, we specifically focus on the evolution of snake venom. We use a comparative dataset of snake venom gene expression to identify shifts in phenotypic macroevolutionary rates, which are characteristic of key innovations [9]. To further characterize the patterns of venom evolution, we estimated long-term changes in evolutionary rates of venom gene expression and also fitted the data to several trait evolution models. Our results revealed that toxin expression in snake venom evolves very rapidly and has experienced numerous shifts in evolutionary rates over the past 60 million years.

2. Results

(a) Phylogeny and expression data

We collected venom gene expression data for snakes from published literature that reported relative levels of toxin expression via transcriptome sequencing of cDNA libraries. From a list of 39 publications, we obtained data for a total of 52 different snake species from the three venomous

families (Colubridae, Elapidae and Viperidae). We included only species for which phylogenetic data were available, irrespective of transcriptome availability (see Table S1 in additional information (GitHub)), i.e. even if there were transcriptomic data available for a snake species, if the species was not present in our phylogeny, we excluded it. Our dataset included components that are found in at least 50% of the transcriptomes analysed here, this was done to focus on generally more widely abundant toxins (greater than 90% variation across 52 species) and because the sample sizes for the other components would be too low for accurate and phylogenetically unbiased inference, an approach similar to [27,50] (see Figs S1 and S2 in additional information (GitHub)). While changes in these toxins may well contribute significantly to the overall efficacy of venom, our goal was to trace the evolution of relatively ubiquitous components over time. As a result, our analyses are conducted on one component at a time, and the minor components do not greatly affect the percentage of the major components and thus do not affect the overall result [27, Supplementary Material]. Overall, 10 out of 27 toxins were retained (electronic supplementary material, figure S1). Viperid and elapid PLA2 are encoded by different loci and have evolved independently of one another [51]. Therefore, to make the interpretation of our data more intuitive for the reader, we classified elapid PLA2 (type I) as 'ePLA2' and viperid PLA2 (type II) as 'vPLA2'. The published time-calibrated phylogeny of squamates used in our study estimates the most recent common ancestor (root) of the three snake families at about 60 million years ago [37,52].

(b) What are the evolutionary rate dynamics of venom toxins?

Key innovations are predicted to experience shifts in tempo and mode of evolution in response to changes in optima [9]. Along with this, we would expect transitions in evolutionary rates, with a key innovation experiencing subsequent reduction in evolutionary rates since the time of the first occurrence along a branch [4]. We used the Bayesian analysis of macroevolutionary mixtures (BAMM) [53], to determine shifts in rates of toxin expression evolution that took place at different points throughout the history of snake venom evolution, as well as changes in evolutionary rates over time.

For all the 10 toxin families, BAMM revealed several rate shifts along the phylogeny, indicating that evolutionary rates for toxins do not remain constant (phylorate plot, figure 1). Bayes factor estimates support the occurrence of at least one rate shift in all toxin families, indicating that toxin families have experienced changes in their evolutionary rates since becoming a part of the venom arsenal (see Fig. S4 in additional information (GitHub)). Changes in evolutionary rates since the common ancestor of venomous snakes denote different evolutionary trajectories of toxin families (figure 1). CRISP, SVMP and TFTx had a larger distribution of high evolutionary rates (warmer colours) in the ancestors of all venomous snake families and experienced subsequent slowdown in evolution rates (cooler colours) as modern species emerged. The remaining toxins start with slower rates of evolution, which eventually increased in extant species. The phylorate plots also provide configurations of rate change that explain the occurrence and distribution of toxin families in venoms of modern snakes (figure 1). For example, SVMP shows a stark reduction in rates from the common ancestor of venomous snakes to elapid lineages, while it experiences increase in rates in viper lineages. TFTx shows the exact opposite trend, with an increase in elapids and reduction in vipers. BPP, vPLA2 and SVSP show rate trends consistent with their greater distribution in vipers.

Under the adaptive radiation hypothesis, ecomorphological rates should transition from rapid rates to slow, equilibrium rates as ecological niches get filled [54]. To identify these patterns, we estimated the rates of toxin expression evolution of each toxin family after the split of the three families. Our estimates of '*rate through time*' revealed that toxin families show unique evolutionary rates and rate dynamics in each venomous snake family (figure 2).

PLA2s, SVMP, SVSP and TFTx, which make up the largest portion of the venom, have higher evolutionary rates than the other minor components. In colubrids, there was evidence of a delayed rate increase in CTL, BPP and ePLA2. KSPI and TFTx showed an increase in evolutionary rate, while SVMP showed a steady decline. Among colubrids, TFTx was the only toxin to experience an increase in evolutionary rate since divergence of the family.

In vipers, toxin families generally showed an increase in evolutionary rates, with a majority of them occurring at around the 20 Ma mark, just after the diversification of the major viperid lineages (figure 2), which is potentially consistent with venom evolution being linked to ecological opportunity. Two most abundant toxins in the vipers, i.e. SVSP and vPLA2, showed an increase in evolutionary rates since the divergence of viperid lineages, while SVMP showed a decrease.

In elapids, there were very few instances of increase in evolutionary rate. BPP, SVSP and SVMP showed rate increases likely due to their high expression in *Ophiophagus hannah*. The most widespread toxin family in elapid venom TFTx showed a decrease in rate. ePLA2 showed an almost steady rate at the origin of colubrids but experienced a jump around 35 million ago.

SVMP and TFTx show an interesting pattern where they seem to represent alternate venom types. SVMP has high rates and is dominant in vipers (and to an extent in colubrids), while TFTx has higher rates and is predominant in elapids. The alternate lineage of these toxins could be evidence of trade-offs, a pattern that we previously observed [27].

(c) Which model of trait evolution best describes venom evolution?

We fitted a number of trait evolution models to our data to understand which evolutionary process best describes snake venom evolution. We tested BM, OU, EB and jump models (pulsed models) implemented in the pulsR package [20]. The Lévy process can be used to model jumps in trait evolution, which may be appropriate for traits like gene expression, which cannot be explained by simple stochastic models [22,55]. Furthermore, the REML estimation in pulsR can account for intraspecific variation in trait measurements, allowing a more robust model comparison (see Methods).

BM was used to model incremental phenotypic change based on stochastic changes in optima, while OU was used to model incremental evolution around a single optimum. The EB model aims to capture the slowdown in tempo over

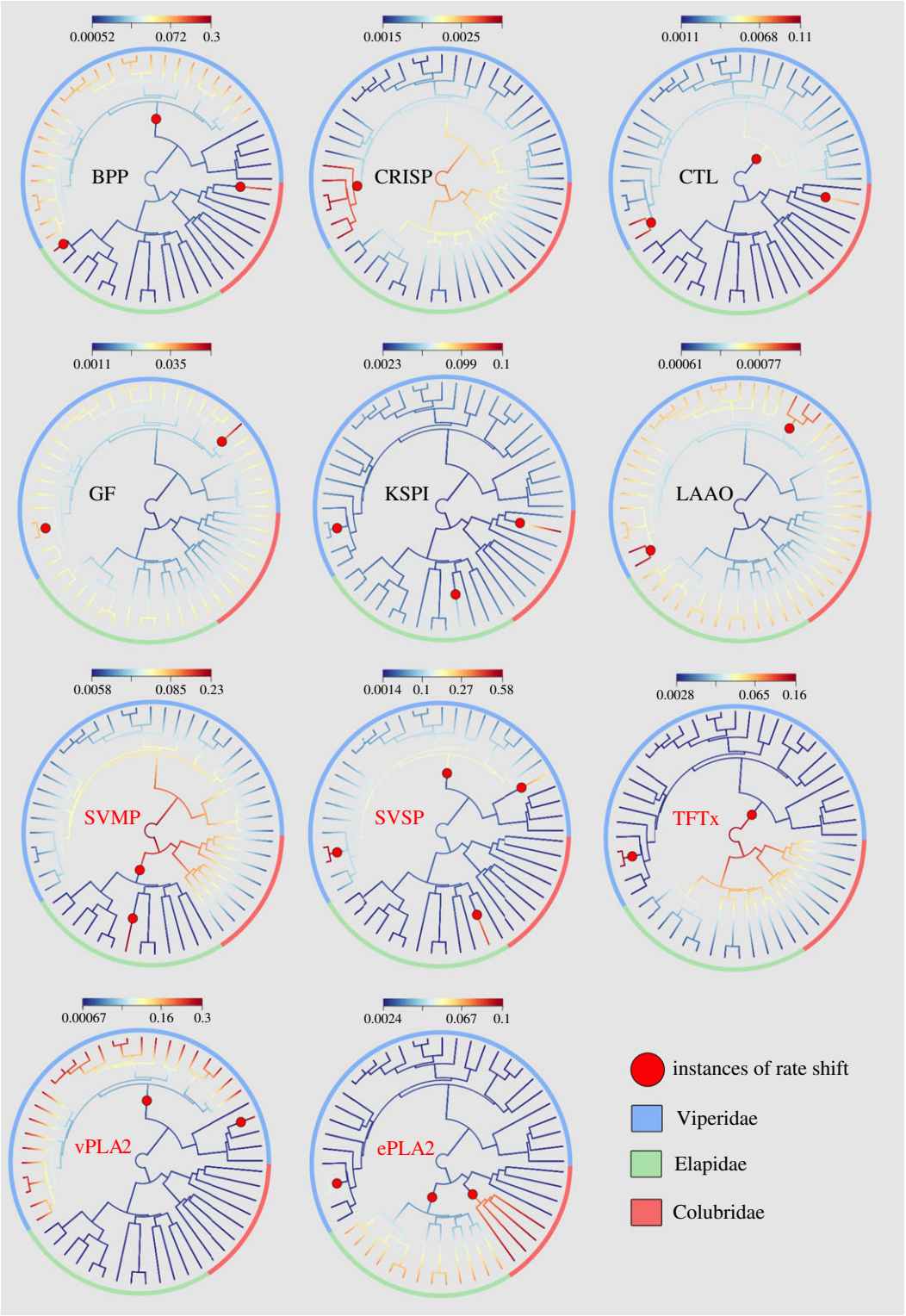


Figure 1. Snake venom phenotypes originated via multiple evolutionary rate shifts. BAMM phylorate plots show locations of the best rate shift configuration (red-filled circle) from among a large posterior distribution of shifts. Rate shift configurations are unique for each toxin family but all families experienced at least one rate shift, indicating a departure from the original evolutionary trajectory since the time of its first occurrence. The branches of the phylogeny are coloured based on distributions of evolutionary rates along the branch. Warmer colours denote a distribution of high evolutionary rates while cooler colours denote a distribution of low rates. With the exception of CRISP, TFTx and SVMP, all the other toxin families show slower rates near the root with a subsequent increase in modern snake lineages. (Online version in colour.)

time expected during adaptive radiation. Two variants of the jump process were modelled; jump normal (JN) and normal inverse Gaussian (NIG). The JN process represents infrequent jumps where stasis is followed by large-scale shifts in adaptive zones, while NIG represents more frequent jump processes, which captures the dynamics of constant phenotypic change that occurs by shifts within an adaptive zone [20]. Both these models represent a process of rapid-pulsed evolution. Jump models have the highest weighted AIC scores and are a better fit to snake venom gene expression data as compared with conventional incremental models of evolution (see Table S1 in additional information (GitHub)). The best model was one whose Akaike information criterion (AIC) weight was at least twice as high as other competing models. The

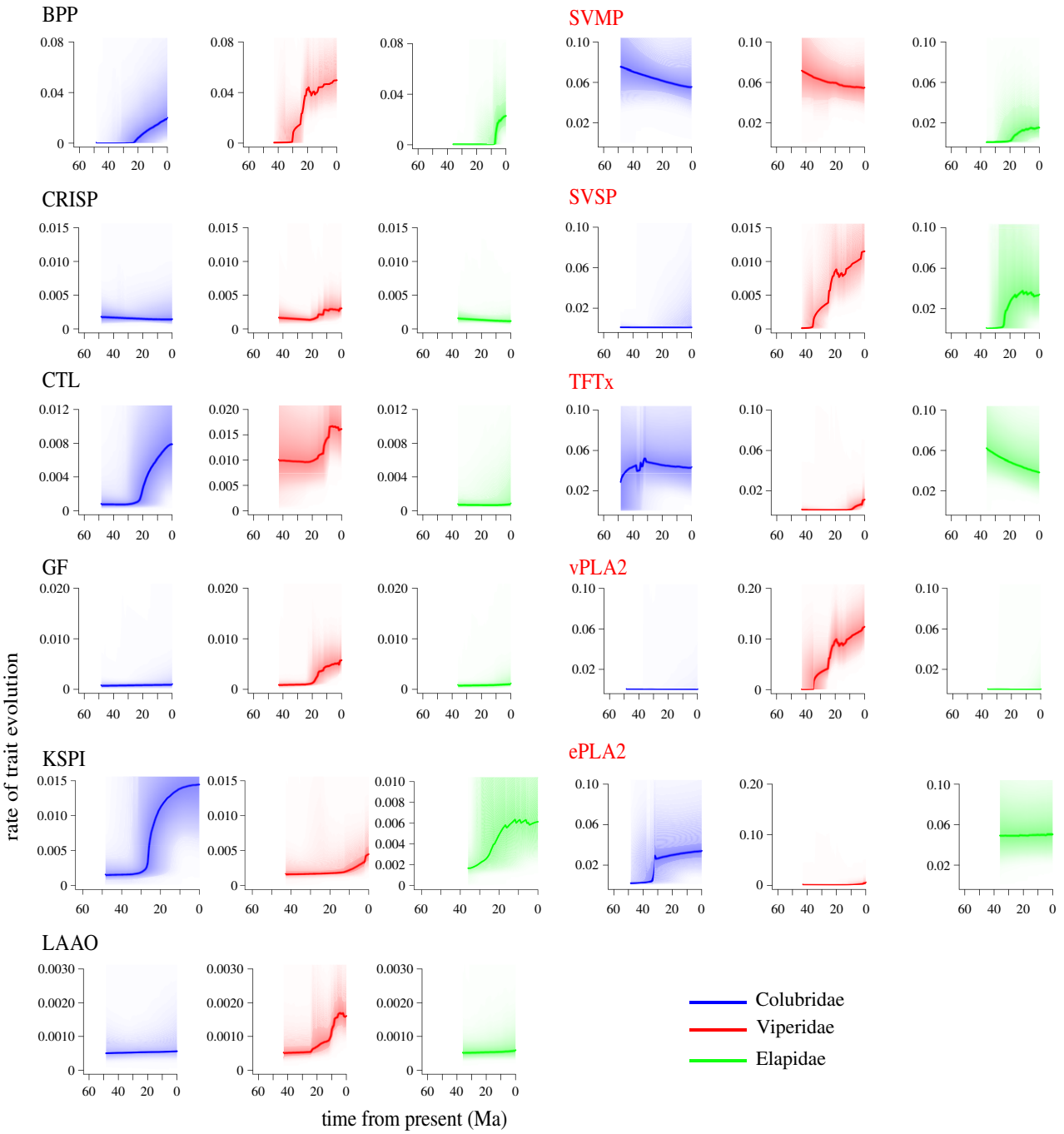


Figure 2. Family-specific trends in evolutionary rates of toxin expression can help explain variation in venom composition and toxin abundance observed between snake families. Blue, red and green (or leftmost, middle and right) represent evolutionary rates for Colubridae, Viperidae and Elapidae, respectively. The gradients represent confidence intervals for each of the estimated rates. The common trend in evolutionary rates between families is that they change through the course of venom evolution and different families have varying rates for individual toxins. (Online version in colour.)

jump models always have better fit than the incremental models; however, individual jump models (JN, NIG or any of their variants) do not differ much in their relative fit. To account for this, the Lévy process with the highest AIC weight was used to represent the entire class of jump processes.

Fitting our data to different models of phenotypic evolution showed that rapid-pulsed evolution explains the evolution of toxin gene expression better than incremental BM and OU process, or explosive EB process.

3. Discussion

One of the most intuitively appealing theories of how phenotypes influence long-term evolution of organisms is that of key innovations. Traits that provide ecological opportunity by allowing exploration of new potential niches, leading to adaptive radiation, seemed to be the perfect explanation of how trait evolution influences species evolution. But as mentioned before, this is far from the case. Many studies have shown that phenotypic disparity is only one of the many axes by which rapid radiation can take place [11]. Therefore, rather than looking at how specific traits correlate with species diversification rates, shifting focus towards identifying shifts in ecological and macroevolutionary space is a better representation of what key innovations achieve [9]. Our study found that gene expression in snake venom has experienced several shifts in evolutionary rates, and that rapid-pulsed evolution better explains snake venom

evolution (table 1). Both these results showcase the highly dynamic and rigorous process of venom evolution and remind us of its strong impact on both the ecology and evolution of venomous snakes.

(a) Shifts in adaptive optima and rapid-pulsed evolution

The idea that snake venom evolution would be characterized by constant shifts in the evolutionary rates of toxin expression is not entirely unexpected. The high variability in the snake venom phenotype is likely due to the presence of various optima. A previous study showed that distribution of toxin families on the macroevolutionary scale can be explained by the presence of convergent phylogenetic optima [27]. Furthermore, the effect of various environmental factors like temperature and longitudinal climatic gradient influences venom variation, hinting at the occurrence of optima that maintain disparate, locally adaptive venom complexes [28]. Therefore, the shifts in phenotypic macroevolutionary rates are likely due to shifts between these optima. A combination of changes in prey diversity and changes in environmental conditions would lead to changes in adaptive optima, which would require snakes to constantly shift expression of different toxin families to chase the optima [28,56]. The extent of this combinatorial action towards diversification of traits in general is not known and might actually be restricted only to venom systems.

Rapid-pulsed evolution, on the other hand, is common in a variety of ecological, palaeontological and comparative data [20,57–59]. One of the proposed explanations for this kind of evolution is Wright's shifting balance theory [60]. The theory states that stochastic forces like genetic drift have a non-trivial effect on adaptation, and that populations occupying local adaptive peaks would compete with each other till the single fittest peak spreads to the entire species [60,61]. Considering the importance of population variation, genetic drift and adaptive peaks in snake venom divergence, the shifting balance theory could well be one of the explanations of how venom evolves [27,31,62].

Beyond the complicated dynamics of shifts in adaptive optima and punctuated evolution, looking at our results under a common perspective of toxin abundance, toxin age and evolutionary rate dynamics bring about an interesting evolutionary pattern. Abundant toxin families along with ePLA2 and SVSP showed a higher net rate of evolution of gene expression (figure 2), a trend also observed in sequence data [31,63]. This suggests that the most abundant (and often toxicologically dominant) toxins have the strongest links to the ecology and evolution of snakes and their venom systems—they are also probably more exposed to selection as a result.

Rate dynamics and age also tend to be related, with older toxins experiencing higher rates in the past followed by reduction in modern lineages (figure 1). It should be noted that the probable origin of most of the toxin families in our study pre-dates the root of our tree (electronic supplementary material, figure S1) [38]. Typically, if a trait is responsible for lineage diversification, its origin should be at a point within the tree around the time of major branching events. However, we can only examine what happens to venom evolution after the most recent common ancestor of extant snakes. Some of the oldest toxins to be included in the venom: SVMP, TFTx and CRISP [39], all showed a larger distribution of high evolutionary rates near the root than the tips (figure 2). This could be because the major toxin families were likely present in the ancestral venom and experienced a uniform reduction in evolutionary rates as lineages diversified. These toxins that pre-date the root likely allowed ancestral snake lineages to realize their ecological potential, which led to niche specification, which in turn led to a slowdown in evolutionary rates in their descendants. While it might be tempting to declare the above results evidence for trait-dependent diversification, one has to carefully look at all the possible lines of evidence, or rather lack thereof.

Table 1. Rapid-pulsed evolution modelled as a Lévy process explain toxin expression evolution in snake venom better than conventional BM, OU and EB models. Model fits (weighted AIC) for BM, OU, EB and pulsed model of phenotypic evolution computed in pulsR [13]. Italic type indicates best fit. We use the AIC weight to determine which model best suits our data. The values in our table represent AIC weights for each of the nine models we tested (BM, OU, EB and six pulsed models). In all cases, the pulsed models were favoured as compared with the non-pulsed models. However, each pulsed model had very similar weights, which make it difficult to determine which pulsed model is better. For that reason, we club them together and report the highest AIC weight.

toxin family	BM	OU	EB	pulsed
BPP	0.039	0.014	0.051	*0.377*
CRISP	0.031	0.011	0.011	*0.455*
CTL	0.002	0	0	*0.979*
GF	0.171	0.062	0.062	*0.465*
KSPI	0	0	0	*0.711*
LAAO	0.123	0.021	0.122	*0.335*
SVMP	0.002	0	0	*0.678*
SVSP	0.127	0	0.046	*0.349*
TFTx	0	0	0	*0.976*
vPLA2	0.089	0.033	0.032	*0.363*
ePLA2	0	0	0	*0.900*

(b) Could venom be responsible for adaptive radiation in venomous snakes?

The increase in abundance of different toxin families in snake venom likely provided a diverse range of phenotypic effects. While these diverse phenotypes are potential key innovations and can contribute to ecological opportunity by opening up previously unexplored feeding niches, they might not necessarily lead to adaptive radiation or show patterns of trait-dependent diversification [10,64]. For example, the Cocos finch (*Pinaroloxias inornata*), the only geospizine finch found outside the Galapagos, has colonized various feeding niches on Cocos Island, but has not speciated into different lineages [65]. This has been attributed to the fact that feeding differences alone did not lead to morphological or behavioural changes, and thus populations that have different feeding habits can still interbreed [66]. Trophic morphology (morphological characters related to food intake) in Lake Tanganyika cichlids provide only part of the impetus needed for rapid speciation, as body shape and microhabitat

traits are undergoing higher degrees of specialization to impart differences between species [67,68]. Even in the poster species for adaptive radiation, Darwin's finches, purifying selection to maintain optimal bill morphology influenced other behavioural, ecological and population dynamics, which prevented homogenization of breeding populations, aiding in speciation [69]. Therefore, when traits influencing feeding niche specification can lead to broader morphological, behavioural and ecological changes, speciation might occur.

Self-contained modular traits that evolve independently of each other can actually reduce the potential of a species to attain large-scale diverse forms. For example, species with highly modular traits can individually evolve different aspects of those traits, without having a large influence on the overall biology of the animal [10,70]. The venom system comprises venom toxins, venom glands, fangs and muscle architecture responsible for delivering the venom into the prey. Numerous examples exist of toxin recruitments coinciding with the development of various morphological features like high-pressure venom delivery and certain hunting strategies like ambush feeding [39]. However, any modifications to enhance prey procurement would be restricted to the venom system and unlikely to affect changes in other parts of the animal [25]. In Darwin's finches, modification of bill morphology influences mating behaviours, where females do not choose males whose bill morphology starkly differs from theirs [69,71]. It is not known if snakes exhibit mating preference based on venom composition or related adaptations. For example, would a female prefer a male with more similar or dissimilar venom composition for mating? Venom might lead to indirect ecological consequences in terms of foraging style, habitat choice and temporal differences in activity. But speciation requires a level of reproductive isolation; how this is achieved either directly or indirectly by changes to the venom is not obvious.

4. Conclusion

Studies of adaptive radiation and character evolution are complex and often come with several caveats. Nearly all studies of adaptation focus on traits and processes in extant species, and this is a major disadvantage since there is no way of representing extinct taxa and thus no way of determining whether a clade with specific innovations was more species-rich in the past [72,73]. While most studies provide a microevolutionary perspective, extrapolating from processes that operate in the present day to what happened early in a clade's history is difficult; because conditions were different in the past, different processes may have been at work or may have produced different outcomes [73]. Perhaps in the past there were venomous snakes with venom compositions specific to the past environment. In response to any changes in this environment, snakes could have evolved venom compositions starkly different from the ones we see today. There might also have been venomous snake lineages in the past that became extinct, leaving a whole history of venom composition unexplored.

The selective and adaptive advantages of snake venom are in no doubt, and based on our results, venom in snakes can be rightly classified as a key innovation. Snakes usually need to produce large amounts of venom, and determining if venom is costlier compared with other offensive (or defensive strategies) is difficult, as it requires prey-handling experiments, taxon-specific toxicity testing, etc., which are both complicated, and difficult to implement [41]. However, considering the several ways snakes can modulate venom output (e.g. venom metering, secretions with reduced protein content etc.), venom might actually be an effective way of procuring energy-rich meals (by subduing large prey), making it a particularly cost-effective innovation [41,74]. Despite this, we believe venom is not the sole reason for the radiation of venomous snakes.

Key innovations are not the only sources of ecological opportunity. The effect of new habitat, antagonistic extinction and key innovations act in concert to promote ecological release which leads to adaptive radiation [3]. Therefore, key innovation plays only one part in the triumvirate of ecological opportunity; if the relative impacts of new habitat, antagonistic extinction and early-stage allopatry are sufficiently strong, release from natural selection and subsequent adaptive radiation might still take place [64,75,76]. Adaptive radiation is also subject to certain initial conditions. Some clades tend to radiate more than others, suggesting that evolvability and the propensity to speciate are vital for adaptive radiation to take place [10]. Looking at the evolution of snake venom in terms of its impacts on speciation would provide greater insight into the role of snake venom in adaptive radiation of venomous snakes. As venom transcriptomes of more snakes become available, revisiting our workflow would tell us to what extent our results represent a general trend in evolution of gene expression in snake venom.

5. Material and methods

(a) Data collection and phylogenetic tree

We used a dataset comprising 10 toxins, which account for greater than 90% of the total venom composition across 52 snake species. The data were collected from a list of 39 publications (see Table S1 in additional information (GitHub)). We included only species for which phylogenetic data were available, irrespective of transcriptome availability (see Table S1 in additional information (GitHub)), i.e. even if there were transcriptomic data available for a snake species, if the species was not present in our phylogeny, we excluded it. We scaled gene expression levels by the average within-species variance, allowing us to standardize the measurements and carry out comparisons across species. This scaled dataset was used in all subsequent analysis. We used a previously described time-calibrated phylogeny of squamates based on two large datasets comprising 44 nuclear genes for 161 squamates, and a dataset of 12 genes from 4161 squamate species; both these datasets represented families and subfamilies [27,37,52,77]. While we manage to sample the three main families of venomous snakes, how under-sampling some species affects our analysis has been discussed in the 'Analytical considerations' section in the additional information (GitHub).

(b) Evolutionary rate dynamics

We used BAMM [53] to estimate evolutionary rate dynamics for each toxin. We ran BAMM on normalized toxin values for each toxin family. We modified the BAMM control file to carry out analysis for phenotypic evolution. *Modeltype* was set to 'trait' and was run for 10^9 generations with MCMC write frequency of 10^5. Priors were obtained using the *setBAMMpriors* function in BAMMTools [78]. For our analysis, we used a conservative prior with *expectedNumberOfShifts* = 1; this model assumes zero rate shifts will have a higher prior probability. Using the

Bayesfactor calculations implemented in BAMMTools, we can determine if the rate shifts we obtain are significantly different from a model with zero rate shifts (or the lowest possible rate configuration where zero rate shifts cannot be computed). The convergence of MCMC chains was determined by visual inspection, by plotting *effective sample size* of log-likelihood and number of shifts in each sample, both of which well exceeded the recommended value of 200 (see Fig. S3 in additional information (GitHub)). We used the *credibleShiftSet* function to identify 95% credible set of distinct shift configurations (see additional information (GitHub)). The rate configuration reported in figure 1 was obtained using the *getBestShiftConfiguration* command in BAMMTools. Explanations behind two rate configurations potentially misrepresented in the phylorate plots are provided in the 'Analytical considerations' section in the additional information (GitHub).

(c) Trait evolution models

We used the pulsR package to fit classes of evolutionary models [20]. Standard variants of incremental evolution BM, OU and EB were modelled as a BM process with branch lengths rescaled as a function of the model parameters [79]. Pulsed evolution, on the other hand, was modelled as a Lévy process. The Lévy process is a stochastic process characterized by three components: (i) a constant directional drift μ, (ii) a Brownian motion with rate σ, and (iii) a jump measure $v(\mathrm{d}x)$. The Lévy process is represented mathematically using the Lévy–Khinchine representation, where one can compute the variance of trait change along a branch of length t. We model stasis followed by rapid adaptation using a compound Poisson process. This is the JN process, which assumes jump sizes are drawn from a normal distribution. The other pulsed evolution model is NIG, which uses an infinitely active Lévy process to model constant rapid adaptation. Since a trait measurement is usually a statistic of a population (trait mean), its value cannot be exactly known. For this reason, the REML estimation assumes that observed traits are drawn from a normal distribution around their true values. This is modelled as a '*tip noise*' parameter ($\boldsymbol{\sigma}_{\mathrm{tip}}$). This parameter is estimated as a combination of both sampling error due to intraspecific variation as well as measurement error. The parameter $\boldsymbol{\sigma}_{\mathrm{tip}}$ can be used as a proxy for $\boldsymbol{\sigma}_{\mathrm{intra}}$, and Landis & Schraiber [20] have shown that $\boldsymbol{\sigma}_{\mathrm{tip}}$ predicts $\boldsymbol{\sigma}_{\mathrm{intra}}$ moderately well. Weighted AIC was used as a measure of model fit. We decided to favour a particular model only if its Akaike weight is at least twice as high as its competing model, similar to [20]. While this arbitrary criterion indeed lacks elegance, it makes model comparison easier and removes ambiguity, especially considering that we are comparing many models. As the AIC between various jump models does not differ greatly, we clubbed them (JN + NIG + BMJN + BMNIG + EBJN + EBNIG) together, and the highest AIC weight was used to represent the entire class of jump processes.

Data accessibility. Additional information, comprising figures, tables, datasets, original plots, code and a section about analytical caveats, can be found at https://agneeshbarua.github.io/venom-phenotype-evolution/. AIC and code for Lévy models can be found at: https://agneeshbarua.github.io/LevyModels/.

Authors' contributions. A.B. and A.S.M. conceptualized the study. A.B. collected the data and carried out the analysis. Both A.B. and A.S.M. wrote the paper.

Competing interests. We declare we have no competing interests.

Funding. We received no funding for this study.

References

1. Hunter JP. 1998 Key innovations and the ecology of macroevolution. *Trends Ecol. Evol.* **13**, 31–36. (doi:10.1016/S0169-5347(97)01273-1)
2. Galis F. 2001 Key innovations and radiations. In *The character concept in evolutionary biology* (ed. GP Wagner), pp. 581–605. San Diego, CA: Academic Press.
3. Yoder JB *et al.* 2010 Ecological opportunity and the origin of adaptive radiations. *J. Evol. Biol.* **23**, 1581–1596. (doi:10.1111/j.1420-9101.2010.02029.x)
4. Simpson GG. 1953 *The major features of evolution*. New York: NY: Columbia University Press.
5. Seehausen O. 2006 African cichlid fish: a model system in adaptive radiation research. *Proc. Biol. Sci.* **273**, 1987–1998. (doi:10.1098/rspb.2006.3539)
6. Van Valen L. 1971 Adaptive zones and the orders of mammals. *Evolution* **25**, 420–428. (doi:10.1111/j.1558-5646.1971.tb01898.x)
7. Bond JE, Opell BD. 1998 Testing adaptive radiation and key innovation hypothesis in spiders. *Evolution* **52**, 403–414. (doi:10.1111/j.1558-5646.1998.tb01641.x)
8. Wheat CW, Vogel H, Wittstock U, Braby MF, Underwood D, Mitchell-Olds T. 2007 The genetic basis of a plant–insect coevolutionary key innovation. *Proc. Natl Acad. Sci. USA* **104**, 20 427–20 431. (doi:10.1073/pnas.0706229104)
9. Rabosky DL. 2017 Phylogenetic tests for evolutionary innovation: the problematic link between key innovations and exceptional diversification. *Phil. Trans. R Soc. B* **372**, 20160417. (doi:10.1098/rstb.2016.0417)
10. Stroud JT, Losos JB. 2016 Ecological opportunity and adaptive radiation. *Annu. Rev. Ecol. Evol. Syst.* **47**, 507–532. (doi:10.1146/annurev-ecolsys-121415-032254)
11. Martin CH, Richards EJ. 2019 The paradox behind the pattern of rapid adaptive radiation: how can the speciation process sustain itself through an early burst? *Annu. Rev. Ecol. Evol. Syst.* **50**, 569–593. (doi:10.1146/annurev-ecolsys-110617-062443)
12. Freckleton RP, Harvey PH, Pagel M. 2003 Bergmann's rule and body size in mammals. *Am. Nat.* **161**, 821–825. (doi:10.1086/374346)
13. Brawand D *et al.* 2011 The evolution of gene expression levels in mammalian organs. *Nature* **478**, 343–348. (doi:10.1038/nature10532)
14. Price SA, Wainwright PC, Bellwood DR, Kazancioglu E, Collar DC, Near TJ. 2010 Functional innovations and morphological diversification in parrotfish. *Evolution* **64**, 3057–3068. (doi:10.1111/j.1558-5646.2010.01036.x)
15. Colombo M, Damerau M, Hanel R, Salzburger W, Matschiner M. 2015 Diversity and disparity through time in the adaptive radiation of Antarctic notothenioid fishes. *J. Evol. Biol.* **28**, 376–394. (doi:10.1111/jeb.12570)
16. Poe S *et al.* 2018 Comparative evolution of an archetypal adaptive radiation: innovation and opportunity in *Anolis* lizards. *Am. Nat.* **191**, 185–194. (doi:10.1086/697223)
17. Burress ED, Tan M, Wainwright PC. 2019 Head shape modulates diversification of a classic cichlid pharyngeal jaw innovation. *Am. Nat.* **194**, 693–706. (doi:10.1086/705392)
18. Arbour JH, Curtis AA, Santana SE. 2019 Signatures of echolocation and dietary ecology in the adaptive evolution of skull shape in bats. *Nat. Commun.* **10**, 2036. (doi:10.1038/s41467-019-09951-y)
19. Glor RE. 2010 Phylogenetic insights on adaptive radiation. *Annu. Rev. Ecol. Evol. Syst.* **41**, 251–270. (doi:10.1146/annurev.ecolsys.39.110707.173447)
20. Landis MJ, Schraiber JG. 2017 Pulsed evolution shaped modern vertebrate body sizes. *Proc. Natl Acad. Sci. USA* **114**, 13 224–13 229. (doi:10.1073/pnas.1710920114)
21. Harmon LJ *et al.* 2010 Early bursts of body size and shape evolution are rare in comparative data. *Evolution.* **64**, 2385–2396. (doi:10.1111/j.1558-5646.2010.01025.x)
22. Landis MJ, Schraiber JG, Liang M. 2013 Phylogenetic analysis using Lévy processes: finding jumps in the evolution of continuous traits. *Syst. Biol.* **62**, 193–204. (doi:10.1093/sysbio/sys086)

23. Raff RA, Sly BJ. 2000 Modularity and dissociation in the evolution of gene expression territories in development. *Evol. Dev.* **2**, 102–113. (doi:10.1046/j.1525-142x.2000.00035.x)
24. Espinosa-Soto C, Wagner A. 2010 Specialization can drive the evolution of modularity. *PLoS Comput. Biol.* **6**, 1000719. (doi:10.1371/journal.pcbi.1000719)
25. Margres MJ, Wray KP, Hassinger ATB, Ward MJ, McGivern JJ, Lemmon EM, Lemmon AR, Rokyta DR. 2017 Quantity, not quality: rapid adaptation in a polygenic trait proceeded exclusively through expression differentiation. *Mol. Biol. Evol.* **34**, 3099–3110. (doi:10.1093/molbev/msx231)
26. Reeks TA, Fry BG, Alewood PF. 2015 Privileged frameworks from snake venom. *Cell. Mol. Life Sci.* **72**, 1939–1958. (doi:10.1007/s00018-015-1844-z)
27. Barua A, Mikheyev AS. 2019 Many options, few solutions: over 60 million snakes converged on a few optimal venom formulations. *Mol. Biol. Evol.* **36**, 1964–1974. (doi:10.1093/molbev/msz125)
28. Zancolli G *et al.* 2019 When one phenotype is not enough: divergent evolutionary trajectories govern venom variation in a widespread rattlesnake species. *Proc. R. Soc. B* **286**, 20182735. (doi:10.1098/rspb.2018.2735)
29. Rokyta DR, Lemmon AR, Margres MJ, Aronow K. 2012 The venom-gland transcriptome of the eastern diamondback rattlesnake (*Crotalus adamanteus*). *BMC Genomics* **13**, 312. (doi:10.1186/1471-2164-13-312)
30. Rokyta DR, Wray KP, Margres MJ. 2013 The genesis of an exceptionally lethal venom in the timber rattlesnake (*Crotalus horridus*) revealed through comparative venom-gland transcriptomics. *BMC Genomics* **14**, 394. (doi:10.1186/1471-2164-14-394)
31. Aird SD, Arora J, Barua A, Qiu L, Terada K, Mikheyev AS. 2017 Population genomic analysis of a pitviper reveals microevolutionary forces underlying venom chemistry. *Genome Biol. Evol.* **9**, 2640–2649. (doi:10.1093/gbe/evx199)
32. Casewell NR, Wagstaff SC, Harrison RA, Renjifo C, Wüster W. 2011 Domain loss facilitates accelerated evolution and neofunctionalization of duplicate snake venom metalloproteinase toxin genes. *Mol. Biol. Evol.* **28**, 2637–2649. (doi:10.1093/molbev/msr091)
33. Daltry JC, Wüster W, Thorpe RS. 1996 Diet and snake venom evolution. *Nature* **379**, 537–540. (doi:10.1038/379537a0)
34. Gibbs HL, Mackessy SP. 2009 Functional basis of a molecular adaptation: prey-specific toxic effects of venom from *Sistrurus* rattlesnakes. *Toxicon* **53**, 672–679. (doi:10.1016/j.toxicon.2009.01.034)
35. Sunagar K, Moran Y. 2015 The rise and fall of an evolutionary innovation: contrasting strategies of venom evolution in ancient and young animals. *PLoS Genet* **11**, 1005596. (doi:10.1371/journal.pgen.1005596)
36. Calvete JJ. 2017 Venomics: integrative venom proteomics and beyond. *Biochem. J.* **474**, 611–634. (doi:10.1042/BCJ20160577)
37. Pyron RA, Burbrink FT, Wiens JJ. 2013 A phylogeny and revised classification of Squamata, including 4161 species of lizards and snakes. *BMC Evol. Biol.* **13**, 93. (doi:10.1186/1471-2148-13-93)
38. Fry BG *et al.* 2006 Early evolution of the venom system in lizards and snakes. *Nature* **439**, 584–588. (doi:10.1038/nature04328)
39. Fry BG, Scheib H, van der Weerd L, Young B, McNaughtan J, Ramjan SFR, Vidal N, Poelmann RE, Norman JA. 2008 Evolution of an arsenal: structural and functional diversification of the venom system in the advanced snakes (Caenophidia). *Mol. Cell. Proteomics* **7**, 215–246. (doi:10.1074/mcp.M700094-MCP200)
40. Durban J *et al.* 2011 Profiling the venom gland transcriptomes of Costa Rican snakes by 454 pyrosequencing. *BMC Genomics* **12**, 259. (doi:10.1186/1471-2164-12-259)
41. Jackson TNW, Jouanne H, Vidal N. 2019 Snake venom in context: neglected clades and concepts. *Front. Ecol. Evol.* **7**, 332. (doi:10.3389/fevo.2019.00332)
42. Harris RJ, Arbuckle K. 2016 Tempo and mode of the evolution of venom and poison in tetrapods. *Toxins* **8**, 193. (doi:10.3390/toxins8070193)
43. Arbuckle K, Speed MP. 2015 Antipredator defenses predict diversification rates. *Proc. Natl Acad. Sci. USA* **112**, 13 597–13 602. (doi:10.1073/pnas.1509811112)
44. Liu S-YV, Frédérich B, Lavoué S, Chang J, Erdmann MV, Mahardika GN, Barber PH. 2018 Buccal venom gland associates with increased of diversification rate in the fang blenny fish *Meiacanthus* (Blenniidae; Teleostei). *Mol. Phylogenet. Evol.* **125**, 138–146. (doi:10.1016/j.ympev.2018.03.027)
45. Blanchard BD, Moreau CS. 2017 Defensive traits exhibit an evolutionary trade-off and drive diversification in ants. *Evolution* **71**, 315–328. (doi:10.1111/evo.13117)
46. Rainford JL, Mayhew PJ. 2015 Diet evolution and clade richness in Hexapoda: a phylogenetic study of higher taxa. *Am. Nat.* **186**, 777–791. (doi:10.1086/683461)
47. Burin G, Kissling WD, Guimarães Jr PR, Şekercioğlu ÇH, Quental TB. 2016 Omnivory in birds is a macroevolutionary sink. *Nat. Commun.* **7**, 11250. (doi:10.1038/ncomms11250)
48. Rojas D, Ramos Pereira MJ, Fonseca C, Dávalos LM. 2018 Eating down the food chain: generalism is not an evolutionary dead end for herbivores. *Ecol. Lett.* **21**, 402–410. (doi:10.1111/ele.12911)
49. Givnish TJ. 2015 Adaptive radiation versus 'radiation' and 'explosive diversification': why conceptual distinctions are fundamental to understanding evolution. *New Phytol.* **207**, 297–303. (doi:10.1111/nph.13482)
50. Junqueira-de-Azevedo ILM, Campos PF, Ching ATC, Mackessy SP. 2016 Colubrid venom composition: an -omics perspective. *Toxins* **8**, 230. (doi:10.3390/toxins8080230)
51. Dowell NL, Giorgianni MW, Kassner VA, Selegue JE, Sanchez EE, Carroll SB. 2016 The deep origin and recent loss of venom toxin genes in rattlesnakes. *Curr. Biol.* **26**, 2434–2445. (doi:10.1016/j.cub.2016.07.038)
52. Zheng Y, Wiens JJ. 2016 Combining phylogenomic and supermatrix approaches, and a time-calibrated phylogeny for squamate reptiles (lizards and snakes) based on 52 genes and 4162 species. *Mol. Phylogenet. Evol. B* **94**, 537–547. (doi:10.1016/j.ympev.2015.10.009)
53. Rabosky DL. 2014 Automatic detection of key innovations, rate shifts, and diversity-dependence on phylogenetic trees. *PLoS ONE* **9**, 89543. (doi:10.1371/journal.pone.0089543)
54. Slater GJ, Price SA, Santini F, Alfaro ME. 2010 Diversity versus disparity and the radiation of modern cetaceans. *Proc. R. Soc. B* **277**, 3097–3104. (doi:10.1098/rspb.2010.0408)
55. Chaix R, Somel M, Kreil DP, Khaitovich P, Lunter GA. 2008 Evolution of primate gene expression: drift and corrective sweeps? *Genetics* **180**, 1379–1389. (doi:10.1534/genetics.108.089623)
56. Davies E-L, Arbuckle K. 2019 Coevolution of snake venom toxic activities and diet: evidence that ecological generalism favours toxicological diversity. *Toxins* **11**, 711. (doi:10.3390/toxins11120711)
57. Thompson JN. 1998 Rapid evolution as an ecological process. *Trends Ecol. Evol.* **13**, 329–332. (doi:10.1016/S0169-5347(98)01378-0)
58. Carroll SP, Hendry AP, Reznick DN, Fox CW. 2007 Evolution on ecological time-scales. *Funct. Ecol.* **21**, 387–393. (doi:10.1111/j.1365-2435.2007.01289.x)
59. Hunt G, Bell MA, Travis MP. 2008 Evolution toward a new adaptive optimum: phenotypic evolution in a fossil stickleback lineage. *Evolution* **62**, 700–710. (doi:10.1111/j.1558-5646.2007.00310.x)
60. Wright S. 1932 The roles of mutation, inbreeding, crossbreeding, and selection in evolution. In *Proc. 6th Int. Congress Genet.* (ed. DF Jones), vol. 1, pp. 356–366. Oxford, UK: Blackwell Sciences.
61. Coyne JA, Barton NH, Turelli M. 1997 Perspective: a critique of Sewall Wright's shifting balance theory of evolution. *Evolution* **51**, 643–671. (doi:10.1111/j.1558-5646.1997.tb03650.x)
62. Margres MJ, Walls R, Suntravat M, Lucena S, Sánchez EE, Rokyta DR. 2016 Functional characterizations of venom phenotypes in the eastern diamondback rattlesnake (*Crotalus adamanteus*) and evidence for expression-driven divergence in toxic activities among populations. *Toxicon* **119**, 28–38. (doi:10.1016/j.toxicon.2016.05.005)
63. Aird SD, Aggarwal S, Villar-Briones A, Tin MM-Y, Terada K, Mikheyev AS. 2015 Snake venoms are integrated systems, but abundant venom proteins evolve more rapidly. *BMC Genomics* **16**, 647. (doi:10.1186/s12864-015-1832-6)
64. Losos JB. 2010 Adaptive radiation, ecological opportunity, and evolutionary determinism. American Society of Naturalists E. O. Wilson award address. *Am. Nat.* **175**, 623–639. (doi:10.1086/652433)
65. Smith JNM, Sweatman HPA. 1976 Feeding habits and morphological variation in Cocos

finches. *Condor* **78**, 244–248. (doi:10.2307/1366860)

66. Werner TK, Thomas S. 1987 Behavioral feeding specialization in *Pinaroloxias inornata*, the 'Darwin's finch' of Cocos Island, Costa Rica. *Proc. Natl Acad. Sci. USA* **84**, 5506–5510. (doi:10.1073/pnas.84.15.5506)
67. Kocher TD. 2004 Adaptive evolution and explosive speciation: the cichlid fish model. *Nat. Rev. Genet.* **5**, 288–298. (doi:10.1038/nrg1316)
68. Muschick M, Nosil P, Roesti M, Dittmann MT, Harmon L, Salzburger W. 2014 Testing the stages model in the adaptive radiation of cichlid fishes in East African Lake Tanganyika. *Proc. R. Soc. B* **281**, 20140605. (doi:10.1098/rspb.2014.0605)
69. Edelaar P. 2018 Ecological speciation: when and how variation among environments can drive population divergence. In *Bird species: how they arise, modify and vanish* (ed. DT Tietze), pp. 195–215. Cham, Switzerland: Springer International Publishing.
70. Clune J, Mouret J-B, Lipson H. 2013 The evolutionary origins of modularity. *Proc. R. Soc. B* **280**, 20122863. (doi:10.1098/rspb.2012.2863)
71. Grant PR, Grant BR. 2008 Pedigrees, assortative mating and speciation in Darwin's finches. *Proc. R. Soc. B* **275**, 661–668. (doi:10.1098/rspb.2007.0898)
72. Schluter D, Price T, Mooers AØ, Ludwig D. 1997 Likelihood of ancestor states in adaptive radiation. *Evolution* **51**, 1699–1711. (doi:10.1111/j.1558-5646.1997.tb05095.x)
73. Gavrilets S, Losos JB. 2009 Adaptive radiation: contrasting theory with data. *Science* **323**, 732–737. (doi:10.1126/science.1157966)
74. Schendel V, Rash LD, Jenner RA, Undheim EAB. 2019 The diversity of venom: the importance of behavior and venom system morphology in understanding its ecology and evolution. *Toxin* **11**, 666. (doi:10.3390/toxins11110666)
75. Losos JB, Ricklefs RE. 2009 Adaptation and diversification on islands. *Nature* **457**, 830–836. (doi:10.1038/nature07893)
76. Cotoras DD, Bi K, Brewer MS, Lindberg DR, Prost S, Gillespie RG. 2018 Co-occurrence of ecologically similar species of Hawaiian spiders reveals critical early phase of adaptive radiation. *BMC Evol. Biol.* **18**, 100. (doi:10.1186/s12862-018-1209-y)
77. Wiens JJ, Hutter CR, Mulcahy DG, Noonan BP, Townsend Jr TM, Sites JW, Reeder TW. 2012 Resolving the phylogeny of lizards and snakes (Squamata) with extensive sampling of genes and species. *Biol. Lett.* **8**, 1043–1046. (doi:10.1098/rsbl.2012.0703)
78. Rabosky DL, Grundler M, Anderson C, Title P, Shi JJ, Brown JW, Huang H, Larson JG. 2014 BAMMtools: an R package for the analysis of evolutionary dynamics on phylogenetic trees. *Methods Ecol. Evol.* **5**, 701–707. (doi:10.1111/2041-210X.12199)
79. Hansen TF. 1997 Stabilizing selection and the comparative analysis of adaptation. *Evolution* **51**, 1341–1351. (doi:10.1111/j.1558-5646.1997.tb01457.x)

Appendix IV:

Barua A, Mikheyev AS. Many Options, Few Solutions: Over 60 Million Snakes Converged on a Few Optimal Venom Formulations. *Molecular Biology and Evolution*. 2019;36:1964–74.

Many Options, Few Solutions: Over 60 My Snakes Converged on a Few Optimal Venom Formulations

Agneesh Barua*[,1] and Alexander S. Mikheyev[1,2]

[1]Okinawa Institute of Science and Technology Graduate University, Onna, Japan

[2]Evolutionary Genomics Research Group, Ecology and Evolution Unit, Australian National University, Canberra, Australia

***Corresponding author:** E-mail: agneesh.barua@oist.jp.

Associate editor: Claudia Russo

Abstract

Gene expression changes contribute to complex trait variations in both individuals and populations. However, the evolution of gene expression underlying complex traits over macroevolutionary timescales remains poorly understood. Snake venoms are proteinaceous cocktails where the expression of each toxin can be quantified and mapped to a distinct genomic locus and traced for millions of years. Using a phylogenetic generalized linear mixed model, we analyzed expression data of toxin genes from 52 snake species spanning the 3 venomous snake families and estimated phylogenetic covariance, which acts as a measure of evolutionary constraint. We find that evolution of toxin combinations is not constrained. However, although all combinations are in principle possible, the actual dimensionality of phylomorphic space is low, with envenomation strategies focused around only four major toxin families: metalloproteases, three-finger toxins, serine proteases, and phospholipases A2. Although most extant snakes prioritize either a single or a combination of major toxin families, they are repeatedly recruited and lost. We find that over macroevolutionary timescales, the venom phenotypes were not shaped by phylogenetic constraints, which include important microevolutionary constraints such as epistasis and pleiotropy, but more likely by ecological filtering that permits a small number of optimal solutions. As a result, phenotypic optima were repeatedly attained by distantly related species. These results indicate that venoms evolve by selection on biochemistry of prey envenomation, which permit diversity through parallelism, and impose strong limits, since only a few of the theoretically possible strategies seem to work well and are observed in extant snakes.

Key words: **gene expression, generalized linear mixed model, macroevolution, parallel evolution, venom.**

Article

Introduction

Single genes underlying major traits are the exception rather than the rule, and the dissection of polygenic trait variation has been at the forefront of biological research (Lander and Kruglyak 1995; Nadeau 2001; Morley et al. 2004). Much of the complexity resulting from interactions between genes is mediated through their expression, which plays a central role in determining phenotypic variation between individuals and populations (Deutsch et al. 2005; Cardoen et al. 2011; de Montaigu et al. 2015; Ghalambor et al. 2015; Catalán et al. 2016). In particular, levels of gene expression account for substantial sources of variation in natural populations, acting as potential targets of natural selection (Oleksiak et al. 2002; Deutsch et al. 2005; Harrison et al. 2012). Although population-level differences in expression may contribute to the onset of local adaptation and perhaps even eventual adaptive divergence (Nolte et al. 2009; Jeukens et al. 2010; Ghalambor et al. 2015), how changes in gene expression levels lead to evolution of complex traits over the course of millions of years remains largely unknown.

Interactions between genes and their effect in channeling of adaptive responses have been the focus of the field of quantitative genetics. How evolution results from the combined effects of the adaptive landscape, and the pattern of genetic variances and covariance among genes (the *G* matrix), is one of the key questions in this field (Lande 1979; Arnold et al. 2008). The covariance between genes plays a vital role in shaping complex traits by determining the evolutionary trajectory through natural selection (Arnold et al. 2001), and the occurrence of parallelism (Rosenblum et al. 2014). Although most quantitative genetics studies deal with populations, their conclusions can translate to macroevolutionary processes as well. For example, estimates of divergence between populations show that the direction of greatest phenotypic divergence can be predicted by the multivariate direction of greatest additive genetic variance within populations (Schluter 1996). Unfortunately, the *G* matrix cannot be extrapolated across macroevolutionary timescales, as it itself evolves (Steppan et al. 2002). Fortunately, it is possible to compute a phylogenetic covariance (PCOV) matrix for multivariate traits, which can serve as a useful analogy to the *G* matrix, but over much larger timescales, and incorporating a broader range of constraints (Lynch 1991; Adams and Felice 2014). We can then examine whether the structure of the PCOV matrix corresponds to evolutionary trajectories of complex traits.

Here, we use the analogy between the *G* matrix and the PCOV matrix to understand how gene expression evolves in a complex trait, namely snake venom. Being composed of

proteinaceous cocktails, snake venoms are unique in that the expression of each toxin type can be quantified and traced to a distinct genomic locus (Rokyta et al. 2012, 2013; Aird et al. 2017; Margres, Bigelow, et al. 2017; Shibata et al. 2018). Variations in gene expression alter the abundance of proteins in the venom, thereby influencing venom efficacy (Daltry et al. 1996; Gibbs and Mackessy 2009; Casewell et al. 2011; Holding et al. 2016; Margres, Wray, et al. 2017). Thus, toxin expression levels constitute the polygenic phenotype that is the venom, allowing us to examine how selection affects gene expression over tens of millions of years. To examine the features of complex trait evolution at the level of gene expression, we estimated phylogenetic covariance of 10 toxins families using data from 52 snake species covering the 3 venomous snake families (Elapidae, Viperidae, and Colubridae) and asked the extent to which our observed patterns corroborate already known instances of evolutionary change across taxa.

Although we find that extant snake venoms occupy a limited area of phenotypic space, largely centered on four major toxin families, we find no evidence of phylogenetic constraints on the number of possible venom combinations. These data show that the relatively small number of molecular strategies used by the snakes result from consistent and often convergent selection on the biochemistry of envenomation, rather than from intrinsic constraints on gene interactions. Thus, over tens of millions of years selection likely plays a greater role in shaping the venom phenotype than intrinsic constraints.

Results

Expression Data and Phylogeny

Expression data for snakes were collected from published studies that reported relative levels of toxin expression via next-generation transcriptome sequencing of cDNA libraries. We obtained data for a total of 52 different snake species from the 3 major venomous families (Colubridae, Elapidae, and Viperidae), from a list of 39 publications (supplementary table 1, Supplementary Material online). For inclusion, each study had to provide quantitative data on toxin component abundance and had species for which phylogenetic data were available. We restricted our data set to include components that are found in at least 50% of snakes (supplementary fig. 35, Supplementary Material online). We focused on generally important toxin families, because sample sizes for the other components would be too low for accurate and phylogenetically unbiased inference, an approach similar to that of Junqueira-de-Azevedo et al. (2016). Incidentally, this cut-off also eliminated many low-abundance toxin families (on average $<1\%$ of the venom, supplementary fig. 34, Supplementary Material online). The abundance of these toxins would be more difficult to estimate, as they are closer to the signal to noise to threshold of gene expression experiments. Overall 10 out of 25 toxin families we retained. For comparative analyses, we used a published time-calibrated phylogeny of squamates, which estimated the most recent common ancestor (root) of the three snake families to about 60 Ma (Zheng and Wiens 2016).

Evolutionary Covariance between Venom Components

By limiting the range of responses to natural selection, the covariances between genes reflect constraints that shape a phenotype. The PCOV matrix accounts for the effect of phylogeny on the interrelationships between genes coding for the snake venom phenotype, providing an approximation of the presence or absence of constraint behind the evolution of gene expression levels. To estimate the PCOV, we used a phylogenetic generalized linear mixed model (PGLMM) under a Bayesian framework. The concept of PGLMM was devised in the early 90s as a method to infer evolutionary constraints of characters using only phylogeny and measures of phenotypes and is based on the animal model in quantitative genetics (see Materials and Methods) (Lynch 1991; Wilson et al. 2010). As an extension of maximum likelihood–based techniques like phylogenetic least squares, PGLMM was notable for its versatility as a comparative method (Miles and Dunham 1993; de Villemereuil and Shinichi Nakagawa 2014). We use a modern rendition of the PGLMM devised by Hadfield and Nakagawa, which was optimized for faster and better performance (Hadfield and Nakagawa 2010; de Villemereuil and Shinichi Nakagawa 2014). The mean effective sample size for all parameters was greater than 11,000 (supplementary fig. 4, Supplementary Material online). The diagnostics revealed suitable convergence of the chains with negligible autocorrelation in the Markov chain Monte Carlo (MCMC; supplementary figs. 1–3, Supplementary Material online). Significant values in the PCOV matrix denote the presence of phylogenetic constraint, whereas nonsignificant values denote its absence. We observed a lack of significant values in the PCOV (fig. 1) for all the venom components that we modeled. In addition to estimating a PCOV, the model was used to compute λ values which denote the phylogenetic signal (fig. 1). Phylogenetic heritability of a trait is defined as the proportion of variance explained by the relationship among species given by the phylogeny, and in our case it is equivalent to Pagel's lambda model of phylogenetic signal which is similar to that of Lynch's original phylogenetic heritability (Freckleton et al. 2002; Housworth et al. 2004; de Villemereuil and Shinichi Nakagawa 2014). The λ values are a measure of statistical dependence of trait values and phylogeny. They indicate whether certain components in modern snakes were likely similar as in their ancestors. In our case, most venom components show strong phylogenetic signals of greater than 0.5, albeit with large confidence intervals. However, all venom components have λ significantly greater than 0. A few, in particular cysteine-rich secretory proteins (CRISPs), snake venom metalloproteinase (SVMP), three-finger toxins (TFTx), and Kunitz-type serine protease inhibitor (KSPI), show very strong phylogenetic signals (>0.8) and narrow confidence intervals, indicating the presence of strong phylogenetic inertia.

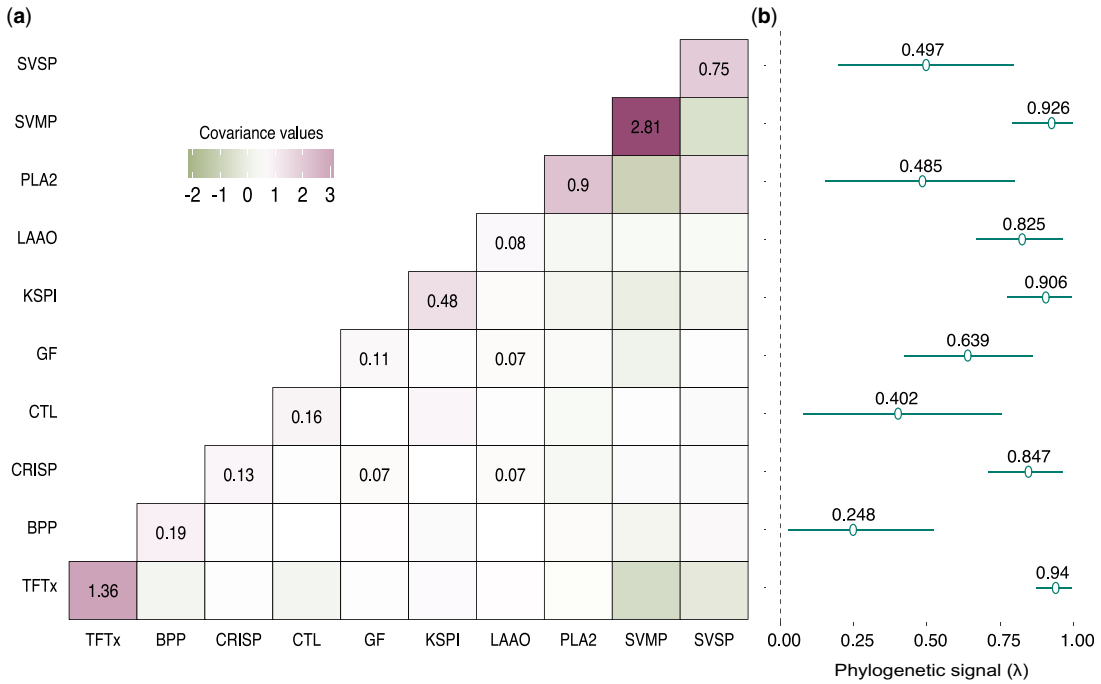


Fig. 1. Phylogenetic constraints on individual toxins and their combinations. (*a*) A lack of significant values (only significant values labeled) in the PCOV matrix denotes a lack of phylogenetic constraint between toxin families. (*b*) Components show a significant presence of a phylogenetic signal, indicating that closely related species are likely to evolve the same way. Lambda, represents phylogenetic signal, which is a measure of dependency of trait evolution with phylogeny. Lambda values, are estimated as toxin variance on the diagonal, divided by the sum of diagonal variance and residuals. TFTx, SVMP, KSPI, LAAO, and CRISP showed the highest signal, with greatest significance, whereas the rest showed comparatively weaker signals. Phylogenetic constraints determine convergence and parallel evolution, where high constraint reduces the likelihood of genes contributing to different convergent regimes (Rosenblum et al. 2014). Yet, for snake venom genes, we see no such constraints in gene expression despite the high phylogenetic signal, suggesting that all toxin combinations, in principle, are possible.

Compositional Data Considerations

It should be noted that the main analyses were performed on compositional (sum-constrained) data, which has the potential of introducing spurious correlations. A range of common solutions to this problem involve log-transformations of the data (Aitchison and Egozcue 2005), which allows for the comparison of relative quantities of the components. However, structural zeros cannot undergo log-transformations, but also cannot be excluded from a comparative analysis because they represent biologically valid characters. Nonetheless, we validated the robustness of the main results using the centered-log-ratio (clr) transform using the *compositions* R package (van den Boogaart and Tolosana-Delgado 2008) and imputed zero values in our compositional data using the "cmultRepl" function in the *zCompositons* R package (Palarea-Albaladejo and Martín-Fernández 2015), to confirm that the overall structure of the covariance matrix is unchanged. Indeed, although the PGLMM using transformed data had significantly worse fit, we did not detect more off-diagonal correlations, and the on-diagonal values were still high (supplementary fig. 37, Supplementary Material online).

Four Toxin Families Drive the Evolution of the Snake Venom Arsenal

The PCOV is a measure of additive phylogenetic covariance, that can be used to estimate the direction of greatest adaptive phenotypic variation (Schluter 1996; Wilson et al. 2010). We identified axes of maximum variations in the toxin components using Principal component analysis (PCA) on the phylogenetic covariances, using it to visualize the dimensionality of the venom phenotype (Uyeda et al. 2015). The venom phylomorphospace has very low dimensionality as the first two components jointly explained 74.3% of the variation. The largest loadings were from four families of toxins: TFTx, SVMP, phospholipase A2 (PLA2), and snake venom serine protease (SVSP) (fig. 2). We therefore classified them as "major" toxins, representing three largely distinct envenomation strategies focussed around SVMP, TFTx, and a combination of PLA2 and SVSP.

The clustering of snakes on this phylomorphic venom space shows a clear association between family and the major component in the venom. For example, most elapids venoms form a cluster dominated by TFTx, which is the principal family found in their venom. On the other hand, vipers occupy a larger region of phylomorphospace because some have venoms dominated by SVMP, whereas others use different combinations of SVMP, SVSP, and PLA2. Finally, colubrid venoms are the most diverse in composition, employing all of the different strategies. A key observation in the PCA is that some distantly related species cluster together around the same envenomation strategy, suggesting parallel evolution.

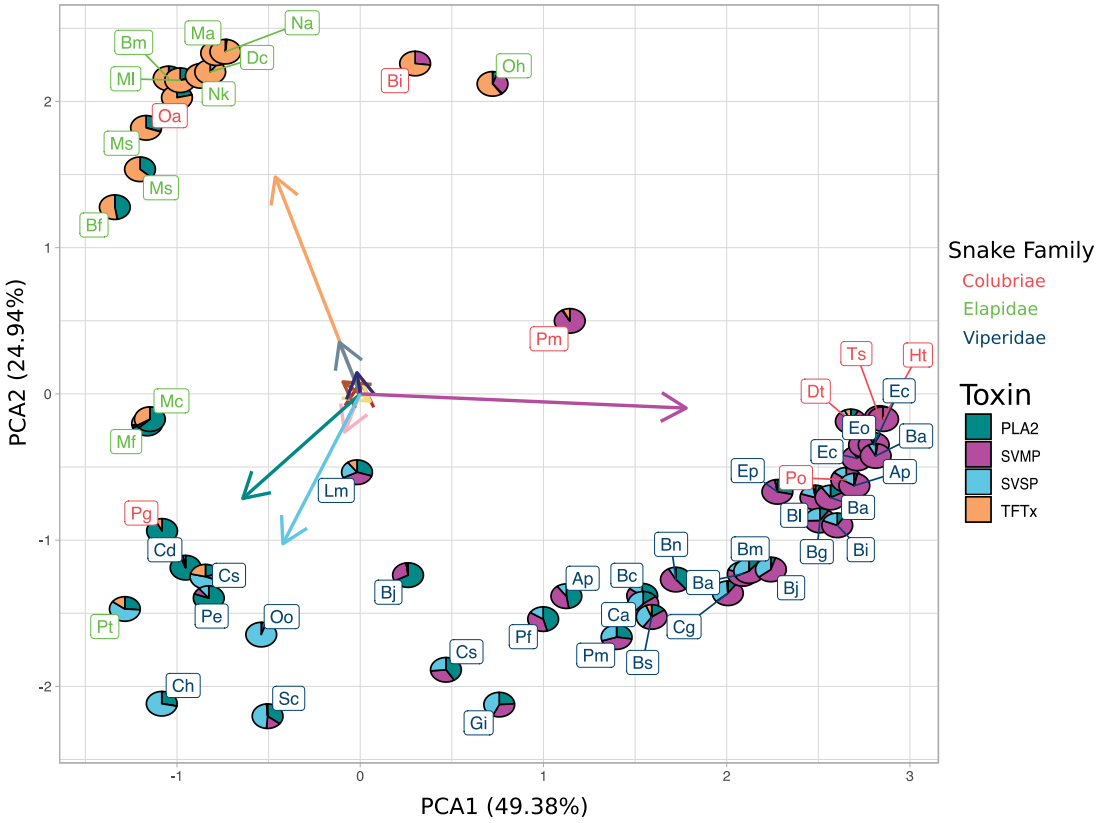


FIG. 2. Snakes (species codes provided in Supplementary Material online) cluster on phylomorphospace along the axes of four toxins: PLA2, SVSPs, SVMPs, and TFTx. These axes represent three distinct envenomation strategies employed by the snakes. Vipers in our data employ a wide spectrum of strategies, from being focussed primarily on SVMP, to employing a mixture of PLA2 and SVSP. Most elapids in our data employ a strategy primarily based on TFTx, whereas two Micrurus species (Mc, Mf) have a combination of PLA2 and TFTx. Colubrids show a unique trend of being scattered throughout the phylomorphospace, having at least one species adopting each of the three strategies. Despite the lack of constraint in gene expression, the snake venom phenotype has very low dimensionality with the four major components accounting for 74.3% of the variation. Clustering of distantly related snakes to around a similar strategy hints at the likely parallelism of these major toxins.

It is important to note that PLA2s in elapids (group I) and vipers (group II) are produced by different loci and have apparently evolved independently (Lynch 2007; Gibbs and Rossiter 2008; Vonk et al. 2013; Dowell et al. 2016; Shibata et al. 2018). In order to account for any underlying family-specific evolutionary trend, we conducted a parallel analysis by splitting PLA2 into elapid PLA2 and viperid PLA2 (supplementary fig. 5, Supplementary Material online). This analysis produced qualitatively the same results as the combined analysis, though the first two components of the PCA explained less variance (62.3% as opposed to 74.3%). In particular, loadings for both elapid and viperid PLA2 were oriented in the same direction (supplementary fig. 10, Supplementary Material online), indicating that the direction of variation in the phylomorphospace is the same for both groups. Thus, we carried out all subsequent analysis by combining them into a joint functional category.

Parallelism of Envenomation Strategies

The clustering of distantly related species in the PCA despite the generally high phylogenetic inertia hinted at the likely parallelism of envenomation strategies across snakes. We use parallelism rather than convergence because parallelism describes a shared molecular basis, where phenotypes arise by using the same molecular mechanisms (Rosenblum et al. 2014). Thus, for our study, parallelism is a type of convergence brought about by a shared molecular basis, and we use both the terms interchangeably. Also, since our data consist of gene families shared by all the snakes, describing convergence in terms of a shared genetic structure seemed appropriate. To test for parallelism across the phylogeny, we used SURFACE (Ingram and Mahler 2013), which fits a series of stabilizing selection models to identify instances where multiple lineages adopt the same selective regime (Ingram and Mahler 2013). Our goal was to test whether macroevolutionary models involving convergent shifts to optima on a Simpsonian landscape can explain the clustering, and similarities in the venom phenotype. We do not test whether the presence or absence of a toxin family is due to convergence. SURFACE uses AIC as criterion to determine goodness of fit and keeps adding models until the AIC does not improve further (Ingram and Mahler 2013). SURFACE provides two measures of convergence: Δk and c, which represent refinement of the adaptive landscape due to convergence, and shifts toward convergent regimes occupied by multiple lineages respectively. Although both these measures represent convergence, we use shifts to convergent regime (c) to classify convergence, but report both. The final model included nine regime shifts and three

distinct regimes ($\Delta k = 3$) and a $c = 6$ convergent shifts. The AIC improved from 298.4 to 229.5 in the forward phase, to a final AIC of 211.38 in the backward phase (S11) which indicated that the final model was a better fit than the initial ones. The SURFACE model revealed widespread convergent shifts as a result of optima (the software considers parallelism and convergence to be one in the same) in elapids, vipers, and colubrids (fig. 3). Vipers showed evidence of two distinct optima, one focussed on SVMP and another on a combination of SVMP, SVSP, and PLA2 (fig. 3 and supplementary fig. 12, Supplementary Material online). One of these regimes evolved in parallel due to multiple shifts toward an optima (highlighted species names in fig. 3 and supplementary fig. 12, Supplementary Material online). The other regime focused on SVMP represents an optima in both viperids and colubrids (fig. 3) that has been achieved not due to multiple shifts but likely due to consistent use of SVMP throughout their evolutionary history. In elapids, there was greater evidence for a single convergent regime focused around TFTx that was reached by multiple shifts (fig. 3). *Pseudonaja textilis* and *Pantherophis guttatus* were the only species to have converged toward an optima focused around PLA2 via multiple shifts (fig. 3 and supplementary fig. 12, Supplementary Material online).

We used the inbuilt simulation function in SURFACE to obtain a null distribution on a simulated data set using a Hansen model that lacked true convergence (Ingram and Mahler 2013). Comparison to the null model simulations (supplementary table 2, Supplementary Material online) revealed significantly more convergent regimes (c) obtained from our analysis than would be obtained by chance ($p_c = 0.038$). This allowed us to reject the null hypothesis and conclude that species cluster together due to convergence toward some optima in the phenotypic adaptive landscape.

Strategies Based on Major Components Evolved at Different Times

Understanding the ancestral state of a trait can paint a picture of the journey taken by the trait through evolution. We used ancestral state reconstruction (ASR) analysis to estimate

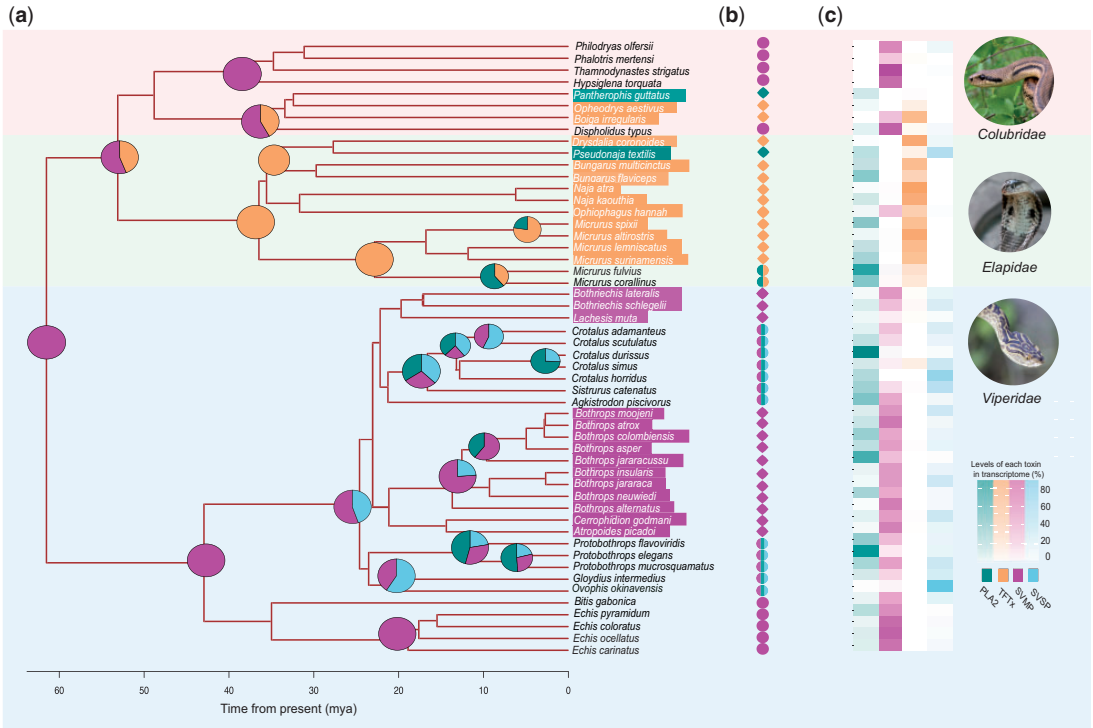


FIG. 3. Evolution of the four major venom toxins and convergent phenotypic regimes in snakes. (*a*) Pie charts at selected nodes represent ASRs of the four major toxins (PLA2, TFTx, SVMP, and SVSP). For clarity, only the nodes where substantial changes in toxin levels took place are shown. Because snake venom composition has evolved dynamically, the ancestral venom (at the root 60 Ma) is difficult to estimate. Although only SVMP was reconstructed as present with a high degree of likelihood, albeit at low levels in our analysis, we would not rule out the presence of other venom components, particularly at low levels. For instance, SVSP does occur in all three families, though not detected at the root. Also, ancestral recruitment of a number of toxin compounds has been argued previously (Fry and Wüster 2004). Lineage-specific specialization occurred relatively recently, in the past 20–40 My. (*b*) Common selective regimes estimated by SURFACE are indicated by symbols. The analysis was conducted using the first two PCA axes of the ten-toxin covariance matrix, but most of the convergent strategies are centered on the four major toxins. Highlighted species names and diamonds represent optima attained by many species via multiple convergent shifts. Circles represent convergent optima due to single shifts. Symbols are colored based on the toxin axes the estimated optima lie on (fig. 2 and supplementary fig. S12, Supplementary Material online). (*c*) Tiles represent the relative abundance of venom toxin in extant snakes. The overall trend is that starting from a relatively undifferentiated ancestor, snakes have increasingly focused on specific toxin families, occasionally investing in new toxin categories for their arsenals (e.g., PLA2s and SVSPs).

recruitment times of the major venom components into the venom arsenal, and how venoms have changed throughout the course of evolution. Because of the diversity and plasticity of the venom phenotype, confidence intervals at the root were very large, and the inference of the venom in the most recent common ancestor should be interpreted with caution, particularly concerning absence of individual toxin families. Of the four major toxins that are responsible for venom diversification, the ASR detected only SVMP in the most common ancestor of the snakes (~60 Ma, henceforth referred to as "the ancestral venom") (fig. 3). The ASR reveals SVMP to be a major and widespread component for most of the evolutionary history of snakes. However, at the base of elapid radiation, SVMP was largely replaced by TFTx as the major component in elapid venoms. TFTx was likely present prior to the split of colubrids and elapids, but while elapids have focused primarily on TFTx, colubrids employed a combination of TFTx and SVMP throughout their evolution (McGivern et al. 2014). In vipers, SVMP has taken various paths, from being the predominant component in Viperinae (*Echis* and *Bitis*), to diversifying substantially in the Crotaline clade (*Protobothrops*, *Bothrops*, *Crotalus*, etc). The ASR suggests that high levels of PLA2 and SVSP (which is mostly restricted to vipers) are more recent additions to the venom. Although not shown in our analysis, PLA2 (both group I and group II) was most likely present at the common ancestors of both Elapids and Crotalids (Dowell et al. 2016), but became substantial parts of the venom from around 20 Ma in both these taxa as observed from their increased occurrence. Although we had estimated ancestral states for the other six components as well (supplementary figs. 23–33, Supplementary Material online), we limited our discussion to only the major toxin families since they dominate adaptive optima in the venom phylomorphospace.

Discussion

We set out to understand how changes in gene expression underlie the evolution of a complex trait, the snake venom. First, we examined the dimensionality of this trait by estimating phylogenetic covariances between expression levels of individual toxin families. The covariances between toxin expression levels can be viewed as constraints that limit the evolution of a trait, analogously to the *G* matrix in quantitative genetics. Unlike the *G* matrix, which arises largely from pleiotropic interactions between genes, phylogenetic constraints may additionally include ecological, developmental, physiological, and other factors. Significant covariance between individual components would reflect constraints on evolutionary change and the total phenotypic space attainable by selection (Pavličev and Cheverud 2015). Thus, traits that are constituted by genes under high constraint would not be able to diversify as freely as traits with no constraint. Genetic constraints also determine convergence and parallel evolution, where high constraint reduces the likelihood of genes contributing to different convergent regimes (Rosenblum et al. 2014). Yet, for snake venom genes, we see no evidence for such constraints in gene expression, suggesting that all toxin combinations, in principle, are possible (fig. 1).

Although the lack of constraint between components implies that venom has the potential to diversify freely and fully fill the possible phenotypic space, this is far from what we observe. Rather, the total phenotypic space has surprisingly low dimensionality, with two principal components explaining 74% of the variance. Venoms form three distinct clusters around the major toxin components in the phylomorphospace, indicating the possible presence of distinct adaptive optima focussed around these toxins (figs. 2 and 3). Although snakes cluster around the major toxin components, this does not diminish the utility of the other minor components which likely impart a more nuanced and refined mode of action to the venom. However, since most species have not yet evolved the lineage specific minor components, their role in the long-term evolution of the snake venom phenotype is limited. Similar toxin-specific strategies have been observed between populations of snakes, but we show that the trend extends phylogenetically to different species as well as different families (Calvete 2017; Strickland et al. 2018). Although individual venom components do exhibit significant phylogenetic inertia (fig. 1), the phylomorphospace clusters often include unrelated taxa, suggesting shifts in envenomation strategies between adaptive optima. These shifts likely result from parallelism, which may be facilitated by lack of constraints between components (fig. 3).

Is a lack of constraint surprising for a trait like snake venom? To answer this, we need to understand one of the key processes by which novel functions and variations in gene families arise—gene duplication (Ohno 1970; Lynch and Conery 2000; Fuchs et al. 2007; Xiao et al. 2008). One of the ways gene duplication can cause functional redundancy is by producing gene copies where one of the copies carries out its designated function, whereas the other copy has no active role in the biological process, thus freeing it from selective constraints (Ohno 1970; Kondrashov et al. 2002; Innan and Kondrashov 2010). This relaxed selective constraint could allow the duplicated genes to diversify freely, as long as one of the copies performs the essential function, and the presence or absence of another copy does not affect fitness. Therefore, a system that comprises many duplicated gene families would also likely have the ability to diversify freely. Snake venom fits this characteristic since it consists of gene families that have undergone varying degrees of duplications throughout their history (Oguiura et al. 2009; Margres, Bigelow, et al. 2017). We hypothesize that the lack of constraint observed between expression levels of genes encoding for snake venom could be due to the fact that snake venom comprises duplicated genes.

One of the most prevalent theories about the origins of venom composition suggests that they originated after ancestral physiological genes underwent duplication and neofunctionalization (Casewell et al. 2013). Since venom phenotypes need to be flexible and to adapt quickly, duplicated genes make ideal toxin candidates as they are under lower selective constraints (Wong and Belov 2012; McCabe and Mackessy 2015; Sunagar et al. 2016). In addition to

sequence-level changes, changes in gene expression also contribute to microevolution in snake venom (Margres, Wray, et al. 2017). To get a complete picture of the evolution of the snake venom phenotype, we need to understand how microevolution (changes in gene expression over short time scales) relates to macroevolution (selection over large time scales). From our observations, we propose a model for snake venom evolution that could potentially link the two, and explain why in spite of having the potential to freely evolve, snake venom has such low dimensionality. We propose that gene duplication facilitated recruitment of physiological genes into the venom system, following which expression levels were free to respond to natural selection due to their low constraint and to potentially occupy a wide phenotypic space. The venom compositions that provided the greatest adaptive advantage due to their favorable biochemistry of envenomation is what we see in present-day species. These observed adaptive optima are dominated by the four main toxin families leading to a high degree of parallelism. This model could likely explain why snake venom, like other systems composed of duplicated genes, experience both positive and relaxed purifying selection (Persi et al. 2016; Aird et al. 2017).

Temporal Patterns in Venom Evolution

Ancestral snake venom composition has received considerable attention, but until now the analyses have been qualitative in nature (Calvete 2017). Although the confidence intervals for ASR are large (supplementary figs. 14–33, Supplementary Material online) owing to the remarkable evolutionary lability of venom, we can nonetheless make a number of observations about the course of evolution of major components. Among the major components, the ancestral venom most likely contained only appreciable amounts of SVMP (fig. 3). This finding is consistent with previous estimates of a likely recruitment of SVMP into the venom prior to the split of vipers from their common ancestor (~62 Ma) (Wüster et al. 2008; Casewell et al. 2011). While we could not detect PLA2, TFTx, and SVSP with confidence in the most recent common ancestor, they could have been present at lower levels in the ancestral venom, or as ancestral precursor molecules (Jin et al. 2007; Lynch 2007; Dowell et al. 2016). This is especially likely for SVSP and PLA2 given that all three families have it in their venom at some level (fig. 3).

Being present in the ancestral venom, SVMP continued to be used as a major toxin by viperids and is still the dominant toxin family in some genera (*Echis* and *Bitis*), as well as some colubrids. However, other toxin families were recruited (or increased in quantity) later in venomous snake evolution. For example, consistent with previous work that placed recruitment of TFTx before the divergence of modern elapids (Fry et al. 2003), we also show that TFTx was likely present at the node prior to the split between elapids and colubrids. At that time TFTx may have co-occurred with SVMP prior to the split of Elapids and Colubrids, perhaps as a specific strategy, one that is quite rare in present-day snakes, being found only in the colubrid brown tree snake (*Boiga irregularis*), and to an extent in the king cobra (*Ophiophagus hannah*). With the proliferation of the TFTx family, elapids have largely lost their reliance on SVMPs.

Viperid and elapid subfamilies have convergently evolved greater reliance on PLA2 toxins (group I in elapids and group II in viperids) but have diverged in venom phenospace due to the previous co-option of different major components (TFTx for elapids and SVSP for vipers). The likely presence of PLA2 (group II) gene copies at the common ancestors of Crotalids raises questions about when the complex expanded in the course of snake evolution (Dowell et al. 2016). From our analysis, we believe that the expansion started somewhere around 20–25 Ma in vipers and was already established as a substantial part of the venom before the split of *Crotalus*, and *Protobothrops* genera. In elapids, ASR does not detect the use of PLA2 before its recruitment as a major component of coral snakes (Micrurus) about 20 Ma, but it was likely present at the common ancestor of elapids and maybe even colubrids given the convergent regime experienced by *Pseudonaja textilis* and *Pantherophis guttatus*, and its presence in many extant species. Interestingly, the recruitment of the two PLA2 families by elapids and viperids occurred at roughly the same time, perhaps as a result of convergent selection driven by radiations in prey lineages, such as mammals.

The overall trend is that recruitment of major toxin families took place at different times, and has progressed along different trajectories in different lineages, with instances of both loss and heightened expression. Snakes have then shifted focus on specific toxin families, occasionally investing into new toxin categories for their arsenals (e.g., PLA2s and SVSPs). The increased concentration of specific venom components, relative to the ancestors, has most likely happened by increases in copy number of the specific gene families (Oguiura et al. 2009; Junqueira-de-Azevedo et al. 2015; Margres, Wray, et al. 2017). Interestingly, shifts in selective regimes produced parallel specialization on the same toxin family by different snakes (fig. 3), suggesting that at the level of toxin family selection generally favors specialization as opposed to diversity.

Conclusion

The extent to which traits are constrained by their history, versus reaching their fitness optima has been a major debate in evolutionary biology. Numerous studies have relied on phylogenetic regression to estimate morphological covariation between traits while accounting for phylogenetic non-independence (Arnqvist and Rowe 2002; Nogueira et al. 2009; Monteiro and Nogueira 2010; Meloro et al. 2011; Adams and Felice 2014). In our approach, we analyze more than one response variable simultaneously and incorporate effects on trait relationships that arise through shared ancestry using the principles behind the animal model (Hadfield 2010; Wilson et al. 2010). We show that the structure of the gene expression PCOV can give insights into how traits evolve, by providing a conceptual bridge between micro- and macroevolutionary forces. By showing that the phenotypic space is inherently unconstrained, we are able to highlight the existence of fitness optima and explain the existence of

widespread parallelism seen in snake venoms. These findings show that in the long-term snakes are able to overcome the inherent trade-off between fitness and phylogenetic constraints. Once genes underlying more traits are known in other systems, subsequent studies will show to what extent snake venoms are typical of a general evolutionary pattern.

Materials and Methods

Data Collection

Toxin expression data were collected from 39 publications (Supplementary Material online). Out of the 25 reported toxin families, we selected only 10 as they were the most ubiquitous toxins amongst all snakes. We restricted our data set to include components that are found in at least 50% of snakes and eliminated low-abundance toxin families (supplementary figs. 34 and 35, Supplementary Material online). Toxins levels were recorded as per publication. Toxin values reported as absolute Fragments Per Kilobase of transcript per Million (FPKM) values were converted to a percentages of total toxin transcript expression. The phylogenetic modeling and ASR were carried out using this curated data set. The toxin values were normalized for calculating the PCOV.

Phylogenetic Tree

We used a time-calibrated tree of squamate reptiles (snakes and lizards) based on two large data sets comprising of 44 nuclear genes for 161 squamates, and a data set of 12 genes from 4,161 squamate species, both these data sets represented families and subfamilies (Wiens et al. 2012; Pyron et al. 2013; Zheng and Wiens 2016). The result was an extensive phylogeny of squamates both in terms of sampling of genes and species. Fossil-based age constraints were used in time-calibrating the tree making it ideal for studies of biogeography, diversification, and trait evolution (Zheng and Wiens 2016). All analyses were carried out using a pruned version of this tree (supplementary fig. 13, Supplementary Material online) that contained the 52 snake species for which we collected gene expression data. This pruned tree had a time at root estimated to be ~60 Ma.

Estimating PCOV Matrix

To familiarize the reader with the rationale behind how the model was constructed, we provide a brief introduction to the animal model and refer the reader to Kruuk et al. (2000), Garant et al. (2004), and Wilson et al. (2010) and chapter 11 of de Villemereuil and Shinichi Nakagawa (2014) for more details. The animal model in quantitative genetics is based on the concept that provided adequate knowledge about the relationships between individuals, and measures of their phenotypic traits, we can make inferences about the patterns of inheritance and evolutionary potential of traits. At its heart is the assumption that if closely related individuals, who share most of their genes, are phenotypically more similar than unrelated individuals, who share fewer genes, we can infer that genes make a significant contribution to phenotypic variance (Wilson et al. 2010). The most basic interpretation would be that phenotypic variation (V_P) is a result of additive genetic variation (V_A) and a residual variance from environmental effects (V_R), where the additive genetic variance (V_A) is the independent effect of inherited alleles on the phenotype.

$$V_P = V_A + V_R. \quad (1)$$

The partitioning of variance can also be done for multiple, covarying traits where the phenotypic covariance (COV_P) would be the sum of additive genetic covariance (COV_A) and covariance of residuals (COV_R). In the animal model, "breeding value" is used as an explanatory variable for a phenotypic trait such that

$$y_i = \mu + a_i + e_i, \quad (2)$$

where y_i is our phenotypic trait of interest, μ is the population mean, a_i is the breeding value, and e_i is the residual error. Although a_i is used as an explanatory variable, its actual value is unknown and thus cannot be used to fit the model. To overcome this, we can specify the above model as a mixed effects model, with a_i being modeled as random effect (Galwey 2014). By incorporating a random effect based on the pedigree of individuals, we can get an estimate of among-individual variance for the phenotypic trait (y) in the population. This allows us to obtain an estimate of among-individual variance in breeding values, which is defined as the additive genetic variance (V_A) (Wilson et al. 2010). Therefore, the key concept behind the statistical interpretation of the animal model is that: population pedigree structure provides insights into how breeding values should covary among individuals, allowing us to solve genetic parameters like V_A, and in multivariate models, COV_A. For n individuals in a pedigree, the matrix of additive genetic covariance of a trait is given as AV_A where A is an $n \times n$ additive genetic relationship matrix containing pairwise values of relatedness. The phylogenetic linear mixed model is exactly the same as the animal model, except that instead of using a pedigree we use a phylogeny to infer additive phylogenetic covariances.

For a simple univariate trait thinking in terms of variance is sufficient, however, for multivariate models it is useful to think in terms of variance–covariance matrices. Thus, for a bivariate model of say trait 1 and trait 2, the phenotypic matrix P would comprise of variances for both trait 1 and trait 2 along the diagonal (V_{P1}, V_{P2}) and covariance between the traits (COV_{P12}) such that $P = G + R$, where G is the additive genetic covariance matrix (or in our case the phylogenetic covariance [PCOV]), and R is the residual matrix. Our model was similar to model (2) and written based on the description given in section 3 on the MCMCglmm vignette for modeling multiresponse traits (Hadfield 2010). The only difference is that our model is a multivariate model with the ten toxins as response variable (y).

Although the genetic (or phylogenetic) effect has the potential to explain a substantial amount of phenotypic similarity, in actuality, a number of intrinsic and extrinsic variables may also be responsible. If there is speculation that such variables are important, they may be added to the model

as fixed effects. This would allow us to interpret the resultant variance as having been conditioned on the specific fixed effect. However, if the additional explanatory variables are not associated with the pedigree (phylogeny in our case) then their inclusion would not alter the estimate of genetic (or phylogenetic) effect (Wilson et al. 2010). In our study, we obtained data from various studies that employ different sequencing technologies and protocols. But, since sequencing technology does not influence the phylogeny of the species, we believe that there would be no substantial change to the PCOV. For the sake of statistical fidelity however, we included sequencing technology, as reported by each study, as a fixed effect and found that there was no change to the overall PCOV structure which still largely consisted of insignificant values (Supplementary Material online).

Phylogenetic generalized linear mixed models allow for testing slightly complicated models, provide more than a simple qualitative estimate of the existence of phylogenetic structure, and have greater statistical power than typically used metric randomization approaches (Ives and Helmus 2011). The MCMC was run for a total of 20 million iterations, with burnin and thinning values of 1 million and 1,500, respectively. Diagnostics for the MCMC run were done by obtaining the plot for the MCMC and autocorrelation. The phylogenetic signal was obtained by dividing the covariance for each toxin by the total covariance of the toxin and the residuals, as mentioned in de Villemereuil and Shinichi Nakagawa (2014). More details regarding passing of fixed and random effect can be found in the Supplementary Material online. We performed principal components analysis using the phylogenetic covariances obtained from the MCMCglmm analysis. Species codes are provided in supplementary note 1, Supplementary Material online.

Analysis of Parallelism

We used the default Ornstein–Uhlenbeck process, a convenient representation of evolution toward adaptive peaks for modeling parallelism in the SURFACE analysis (Ingram and Mahler 2013). The SURFACE method considers parallelism and convergence to be one in the same, and uses Hansen's approach (Hansen model) of modeling evolution toward different adaptive optima by painting multiple adaptive hypothesis onto branches of a phylogenetic tree (Hansen 1997; Ingram and Mahler 2013). SURFACE is unique because unlike previous methods that utilize Hansen models, the placements of regime shifts is guided by trait data as opposed to some a priori hypothesis regarding the location of convergence (Ingram and Mahler 2013). The SURFACE method is divided into two phases. The forward phase adds successive regimes to a basic Hansen model using input from continuous trait measurements, which in our study were the first two principal components estimated from the PCOV. Using principal components from the PCOV allows us to incorporate phylogenetic effect in estimation of an adaptive landscape comprising all ten toxins in our analysis, and because the principal component axes are orthogonal, it nicely deals with the compositional nature of the data. The performance of each successive model was measured using AIC by balancing improvements in log-likelihood against increase in model complexity (Ingram and Mahler 2013). Since AIC for the models are calculated after adding log-likelihoods, the AIC for successive models may improve. The regime shift representing the best model is painted onto the tree. The backward phase is the second phase in the analysis. During this phase of SURFACE all subsets of regimes are collapsed to yield distinct regimes. The collapse is continued till the AIC of the models does not increase further. The final model has k regime shifts, and k' distinct regimes, in addition to the extent of convergence which is defined as the difference of these terms (Δk), c is used to represent shifts toward different convergent regimes in multiple lineages (Ingram and Mahler 2013). We used all standard parameters as mentioned in the SURFACE vignette. To obtain a null distribution, we ran 500 iterations of the inbuilt *surfaceSimulate* function using a Hansen-fit model and concatenated the output from each iteration.

Ancestral State Reconstruction

The default parameters for the *fastAnc* function implemented in the Phytools package was used to perform the ASR (Revell 2012). *fastAnc* performs a maximum likelihood–based reconstruction by computing the root value using Felsenstein 1985 contrasts algorithm (Revell 2012). A phenogram, which shows relative positions of species in evolutionary phenospace, was plotted for each toxin using a spread cost of 0.1 (supplementary figs. 14–32, Supplementary Material online). We used the *contMap* function in Phytools to obtain a tree for changing trait values on a continuous scale represented by a color spectrum. Confidence intervals were plotted on the nodes as bars. Only traits whose confidence intervals did not overlap zero (only positive values) were considered to be present at the root. Pie charts in the main figure were drawn by calculating the relative levels of each of the major toxins estimated by the ASR at the specific node. The ancestral states for each toxin was estimated separately, and thus could not capture any (unlikely) constraint between toxin families that might have been present in the past. The ancestral states were clubbed together to only give a representative picture of what venom configuration at a particular node might have looked like. Two images in the main were obtained from Wikimedia under the creative commons license (Elapidiae: Thomas Jaehnel and Colubridae: Carlo Catoni) image for Viperidae provided by Alexander S. Mikheyev.

Data Availability

Supplementary information including code, data, original figures, and additional analysis with 25 toxin classes are available at: https://agneeshbarua.github.io/Many-options-supplementary/

Supplementary Material

Supplementary data are available at *Molecular Biology and Evolution* online.

Acknowledgments

We would like to thank all the members of the Ecology and Evolution unit at OIST for their input and feedback. Ivan Koludarov and Steven D. Aird for useful discussions about snake venom biochemistry and evolution. We are especially grateful to Steven Aird for locating several additional data sets. Nick Friedman from the Biodiversity and Biocomplexity unit at OIST for discussions regarding comparative methods. We also thank Okinawa Institute of Science and Technology Graduate University for funding the study.

Author Contributions

Data set was collected by A.B. Both A.B. and A.S.M. analyzed the data. A.B. and A.S.M. wrote the article.

References

Adams DC, Felice RN. 2014. Assessing trait covariation and morphological integration on phylogenies using evolutionary covariance matrices. *PLoS One* 9(4):e94335.

Aird SD, Arora J, Barua A, Qiu L, Terada K, Mikheyev AS. 2017. Population genomic analysis of a pitviper reveals microevolutionary forces underlying venom chemistry. *Genome Biol Evol*. 9(10):2640–2649.

Aitchison J, Egozcue JJ. 2005. Compositional data analysis: where are we and where should we be heading? *Math Geol*. 37(7):829–850.

Arnold SJ, Bürger R, Hohenlohe PA, Ajie BC, Jones AG. 2008. Understanding the evolution and stability of the G-matrix. *Evolution* 62(10):2451–2461.

Arnold SJ, Pfrender ME, Jones AG. 2001. The adaptive landscape as a conceptual bridge between micro- and macroevolution. *Genetica* 112-113:9–32.

Arnqvist G, Rowe L. 2002. Correlated evolution of male and female morphologies in water striders. *Evolution* 56(5):936–947.

Calvete JJ. 2017. Venomics: integrative venom proteomics and beyond. *Biochem J*. 474(5):611–634.

Cardoen D, Wenseleers T, Ernst UR, Danneels EL, Laget D, DE Graaf DC, Schoofs L, Verleyen P. 2011. Genome-wide analysis of alternative reproductive phenotypes in honeybee workers. *Mol Ecol*. 20(19):4070–4084.

Casewell NR, Wagstaff SC, Harrison RA, Renjifo C, Wüster W. 2011. Domain loss facilitates accelerated evolution and neofunctionalization of duplicate snake venom metalloproteinase toxin genes. *Mol Biol Evol*. 28(9):2637–2649.

Casewell NR, Wüster W, Vonk FJ, Harrison RA, Fry BG. 2013. Complex cocktails: the evolutionary novelty of venoms. *Trends Ecol Evol (Amst)*. 28(4):219–229.

Catalán A, Glaser-Schmitt A, Argyridou E, Duchen P, Parsch J. 2016. An indel polymorphism in the MtnA 3′ untranslated region is associated with gene expression variation and local adaptation in *Drosophila melanogaster*. *PLoS Genet*. 12:1–34.

Daltry JC, Wüster W, Thorpe RS. 1996. Diet and snake venom evolution. *Nature* 379(6565):537–540.

de Montaigu A, Giakountis A, Rubin M, Tóth R, Cremer F, Sokolova V, Porri A, Reymond M, Weinig C, Coupland G. 2015. Natural diversity in daily rhythms of gene expression contributes to phenotypic variation. *Proc Natl Acad Sci U S A*. 112(3):905–910.

de Villemereuil D, Shinichi Nakagawa P. 2014. General quantitative genetics methods for comparative biology. In: Zsolt GL, editor. Modern phylogenetic comparative methods and their application in evolutionary biology: concepts and practice. Berlin, Germany:Springer. p. 287–304.

Deutsch S, Lyle R, Dermitzakis ET, Attar H, Subrahmanyan L, Gehrig C, Parand L, Gagnebin M, Rougemont J, Jongeneel CV, et al. 2005. Gene expression variation and expression quantitative trait mapping of human chromosome 21 genes. *Hum Mol Genet*. 14(23):3741–3749.

Dowell NL, Giorgianni MW, Kassner VA, Selegue JE, Sanchez EE, Carroll SB. 2016. The deep origin and recent loss of venom toxin genes in rattlesnakes. *Curr Biol*. 26:2434–2445.

Felsenstein J. 1985. Phylogenies and the Comparative Method. *Am. Nat*. 125:1–15.

Freckleton RP, Harvey PH, Pagel M. 2002. Phylogenetic analysis and comparative data: a test and review of evidence. *Am Nat*. 160(6):712–726.

Fry BG, Wüster W. 2004. Assembling an arsenal: origin and evolution of the snake venom proteome inferred from phylogenetic analysis of toxin sequences. *Mol Biol Evol*. 21(5):870–883.

Fry BG, Wüster W, Kini RM, Brusic V, Khan A, Venkataraman D, Rooney AP. 2003. Molecular evolution and phylogeny of elapid snake venom three-finger toxins. *J Mol Evol*. 57(1):110–129.

Fuchs J, Nilsson C, Kachergus J, Munz M, Larsson E-M, Schüle B, Langston JW, Middleton FA, Ross OA, Hulihan M, et al. 2007. Phenotypic variation in a large Swedish pedigree due to SNCA duplication and triplication. *Neurology* 68(12):916–922.

Galwey NW. 2014. Introduction to mixed modelling: beyond regression and analysis of variance. Hoboken, New Jersey, USA:John Wiley & Sons.

Garant D, Kruuk LEB, McCleery RH, Sheldon BC. 2004. Evolution in a changing environment: a case study with great tit fledging mass. *Am Nat*. 164:115–129.

Ghalambor CK, Hoke KL, Ruell EW, Fischer EK, Reznick DN, Hughes KA. 2015. Non-adaptive plasticity potentiates rapid adaptive evolution of gene expression in nature. *Nature* 525(7569):372–375.

Gibbs HL, Mackessy SP. 2009. Functional basis of a molecular adaptation: prey-specific toxic effects of venom from Sistrurus rattlesnakes. *Toxicon* 53(6):672–679.

Gibbs HL, Rossiter W. 2008. Rapid evolution by positive selection and gene gain and loss: pLA(2) venom genes in closely related Sistrurus rattlesnakes with divergent diets. *J Mol Evol*. 66(2):151–166.

Hadfield JD. 2010. MCMC methods for multi-response generalized linear mixed models: the MCMCglmm R package. *J Stat Software* 33(2).

Hadfield JD, Nakagawa S. 2010. General quantitative genetic methods for comparative biology: phylogenies, taxonomies and multi-trait models for continuous and categorical characters. *J Evol Biol*. 23(3):494–508.

Hansen TF. 1997. Stabilizing selection and the comparative analysis of adaptation. *Evolution* 51(5):1341–1351.

Harrison PW, Wright AE, Mank JE. 2012. The evolution of gene expression and the transcriptome–phenotype relationship. *Semin Cell Dev Biol*. 23(2):222–229.

Holding ML, Biardi JE, Gibbs HL. 2016. Coevolution of venom function and venom resistance in a rattlesnake predator and its squirrel prey. *Proc R Soc B* 283(1829):20152841.

Housworth EA, Martins EP, Lynch M. 2004. The phylogenetic mixed model. *Am Nat*. 163(1):84–96.

Ingram T, Mahler DL. 2013. SURFACE: detecting convergent evolution from comparative data by fitting Ornstein–Uhlenbeck models with stepwise Akaike Information Criterion. *Methods Ecol Evol*. 4(5):416–425.

Innan H, Kondrashov F. 2010. The evolution of gene duplications: classifying and distinguishing between models. *Nat Rev Genet*. 11(2):97–108.

Ives AR, Helmus MR. 2011. Generalized linear mixed models for phylogenetic analyses of community structure. *Ecol Monogr*. 81(3):511–525.

Jeukens J, Renaut S, St-cyr J, Nolte AW, Bernatchez L. 2010. The transcriptomics of sympatric dwarf and normal lake whitefish (*Coregonus clupeaformis* spp., Salmonidae) divergence as revealed by next-generation sequencing. *Mol Ecol*. 19(24):5389–5403.

Jin Y, Lee W-H, Zhang Y. 2007. Molecular cloning of serine proteases from elapid snake venoms. *Toxicon* 49(8):1200–1207.

Junqueira-de-Azevedo ILM, Bastos CMV, Ho PL, Luna MS, Yamanouye N, Casewell NR. 2015. Venom-related transcripts from *Bothrops*

jararaca tissues provide novel molecular insights into the production and evolution of snake venom. *Mol Biol Evol*. 32(3):754–766.
Junqueira-de-Azevedo ILM, Campos PF, Ching ATC, Mackessy SP. 2016. Colubrid venom composition: an -omics perspective. *Toxins* 8(8):230.
Kondrashov FA, Rogozin IB, Wolf YI, Koonin EV. 2002. Selection in the evolution of gene duplications. *Genome Biol*. 3:2.
Kruuk LE, Clutton-Brock TH, Slate J, Pemberton JM, Brotherstone S, Guinness FE. 2000. Heritability of fitness in a wild mammal population. *Proc Natl Acad Sci U S A*. 97(2):698–703.
Lande R. 1979. Quantitative genetic analysis of multivariate evolution applied to brain: body size allometry. *Evolution* 33(1 Part 2):402–416.
Lander E, Kruglyak L. 1995. Genetic dissection of complex traits: guidelines for interpreting and reporting linkage results. *Nat Genet*. 11(3):241–247.
Lynch M. 1991. Methods for the analysis of comparative data in evolutionary biology. *Evolution* 45(5):1065–1080.
Lynch M, Conery JS. 2000. The evolutionary fate and consequences of duplicate genes. *Science* 290(5494):1151–1155.
Lynch VJ. 2007. Inventing an arsenal: adaptive evolution and neofunctionalization of snake venom phospholipase A2 genes. *BMC Evol Biol*. 7:2.
Margres MJ, Bigelow AT, Lemmon EM, Lemmon AR, Rokyta DR. 2017. Selection to increase expression, not sequence diversity, precedes gene family origin and expansion in rattlesnake venom. *Genetics* 206(3):1569–1580.
Margres MJ, Wray KP, Hassinger ATB, Ward MJ, McGivern JJ, Lemmon EM, Lemmon AR, Rokyta DR. 2017. Quantity, not quality: rapid adaptation in a polygenic trait proceeded exclusively through expression differentiation. *Mol Biol Evol*. 34(12):3099–3110.
McCabe TM, Mackessy SP. 2015. Evolution of resistance to toxins in prey. In: Gopalakrishnakone P, Malhotra A, editors. Evolution of venomous animals and their toxins. Dordrecht (the Netherlands): Springer. p. 1–19.
McGivern JJ, Wray KP, Margres MJ, Couch ME, Mackessy SP, Rokyta DR. 2014. RNA-seq and high-definition mass spectrometry reveal the complex and divergent venoms of two rear-fanged colubrid snakes. *BMC Genomics*. 15:1061.
Meloro C, Raia P, Carotenuto F, Cobb SN. 2011. Phylogenetic signal, function and integration in the subunits of the carnivoran mandible. *Evol Biol*. 38(4):465–475.
Miles DB, Dunham AE. 1993. Historical perspectives in ecology and evolutionary biology: the use of phylogenetic comparative analyses. *Annu Rev Ecol Syst*. 24(1):587–619.
Monteiro LR, Nogueira MR. 2010. Adaptive radiations, ecological specialization, and the evolutionary integration of complex morphological structures. *Evolution* 64(3):724–744.
Morley M, Molony CM, Weber TM, Devlin JL, Ewens KG, Spielman RS, Cheung VG. 2004. Genetic analysis of genome-wide variation in human gene expression. *Nature* 430(7001):743–747.
Nadeau JH. 2001. Modifier genes in mice and humans. *Nat Rev Genet*. 2(3):165–174.
Nogueira MR, Peracchi AL, Monteiro LR. 2009. Morphological correlates of bite force and diet in the skull and mandible of phyllostomid bats. *Funct Ecol*. 23(4):715–723.
Nolte AW, Renaut S, Bernatchez L. 2009. Divergence in gene regulation at young life history stages of whitefish (*Coregonus* sp.) and the emergence of genomic isolation. *BMC Evol Biol*. 9:59.
Oguiura N, Collares MA, Furtado MFD, Ferrarezzi H, Suzuki H. 2009. Intraspecific variation of the crotamine and crotasin genes in *Crotalus durissus* rattlesnakes. *Gene* 446(1):35–40.
Ohno S. 1970. Evolution by gene duplication. Berlin, Germany:Springer Science.
Oleksiak MF, Churchill GA, Crawford DL. 2002. Variation in gene expression within and among natural populations. *Nat Genet*. 32(2):261–266.
Palarea-Albaladejo J, Martín-Fernández JA. 2015. zCompositions—R package for multivariate imputation of left-censored data under a compositional approach. *Chemometrics Intellig Lab Syst*. 143:85–96.
Pavličev M, Cheverud JM. 2015. Constraints evolve: context dependency of gene effects allows evolution of pleiotropy. *Annu Rev Ecol Evol Syst*. 46(1):413–434.
Persi E, Wolf YI, Koonin EV. 2016. Positive and strongly relaxed purifying selection drive the evolution of repeats in proteins. *Nat Commun*. 7:13570.
Pyron RA, Burbrink FT, Wiens JJ. 2013. A phylogeny and revised classification of Squamata, including 4161 species of lizards and snakes. *BMC Evol Biol*. 13:93.
Revell LJ. 2012. phytools: an R package for phylogenetic comparative biology (and other things). *Methods Ecol Evol*. 3(2):217–223.
Rokyta DR, Lemmon AR, Margres MJ, Aronow K. 2012. The venom-gland transcriptome of the eastern diamondback rattlesnake (*Crotalus adamanteus*). *BMC Genomics*. 13:312.
Rokyta DR, Wray KP, Margres MJ. 2013. The genesis of an exceptionally lethal venom in the timber rattlesnake (*Crotalus horridus*) revealed through comparative venom-gland transcriptomics. *BMC Genomics*. 14:394.
Rosenblum EB, Parent CE, Brandt EE. 2014. The molecular basis of phenotypic convergence. *Annu Rev Ecol Evol Syst*. 45(1):203–226.
Schluter D. 1996. Adaptive radiation along genetic lines of least resistance. *Evolution* 50(5):1766–1774.
Shibata H, Chijiwa T, Oda-Ueda N, Nakamura H, Yamaguchi K, Hattori S, Matsubara K, Matsuda Y, Yamashita A, Isomoto A, et al. 2018. The habu genome reveals accelerated evolution of venom protein genes. *Sci Rep*. 8(1):11300.
Steppan SJ, Phillips PC, Houle D. 2002. Comparative quantitative genetics: evolution of the *G* matrix. *Trends Ecol Evol*. 17(7):320–327.
Strickland JL, Mason AJ, Rokyta DR, Parkinson CL. 2018. Phenotypic variation in Mojave rattlesnake (*Crotalus scutulatus*) venom is driven by four toxin families. *Toxins* 10(4):135.
Sunagar K, Casewell NR, Varma S, Kolla R, Antunes A, Moran Y. 2016. Deadly innovations: unraveling the molecular evolution of animal venoms. In: Venom genomics and proteomics. Toxinology. Dordrecht (the Netherlands): Springer. p. 1–27.
Uyeda JC, Caetano DS, Pennell MW. 2015. Comparative analysis of principal components can be misleading. *Syst Biol*. 64(4):677–689.
van den Boogaart KG, Tolosana-Delgado R. 2008. "compositions": a unified R package to analyze compositional data. *Comput Geosci*. 34(4):320–338.
Vonk FJ, Casewell NR, Henkel CV, Heimberg AM, Jansen HJ, McCleary RJR, Kerkkamp HME, Vos RA, Guerreiro I, Calvete JJ, et al. 2013. The king cobra genome reveals dynamic gene evolution and adaptation in the snake venom system. *Proc Natl Acad Sci U S A*. 110(51):20651–20656.
Wiens JJ, Hutter CR, Mulcahy DG, Noonan BP, Townsend TM, Sites JW Jr, Reeder TW. 2012. Resolving the phylogeny of lizards and snakes (Squamata) with extensive sampling of genes and species. *Biol Lett*. 8(6):1043–1046.
Wilson AJ, Réale D, Clements MN, Morrissey MM, Postma E, Walling CA, Kruuk LEB, Nussey DH. 2010. An ecologist's guide to the animal model. *J Anim Ecol*. 79(1):13–26.
Wong ESW, Belov K. 2012. Venom evolution through gene duplications. *Gene* 496(1):1–7.
Wüster W, Peppin L, Pook CE, Walker DE. 2008. A nesting of vipers: phylogeny and historical biogeography of the Viperidae (Squamata: Serpentes). *Mol Phylogenet Evol*. 49(2):445–459.
Xiao H, Jiang N, Schaffner E, Stockinger EJ, van der Knaap E. 2008. A retrotransposon-mediated gene duplication underlies morphological variation of tomato fruit. *Science* 319(5869):1527–1530.
Zheng Y, Wiens JJ. 2016. Combining phylogenomic and supermatrix approaches, and a time-calibrated phylogeny for squamate reptiles (lizards and snakes) based on 52 genes and 4162 species. *Mol Phylogenet Evol*. 94:537–547.