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

EVOLUTION

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.

This open access article is distributed under Creative Commons Attribution-NonCommercial-NoDerivatives License 4.0 (CC BY-NC-ND).

[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

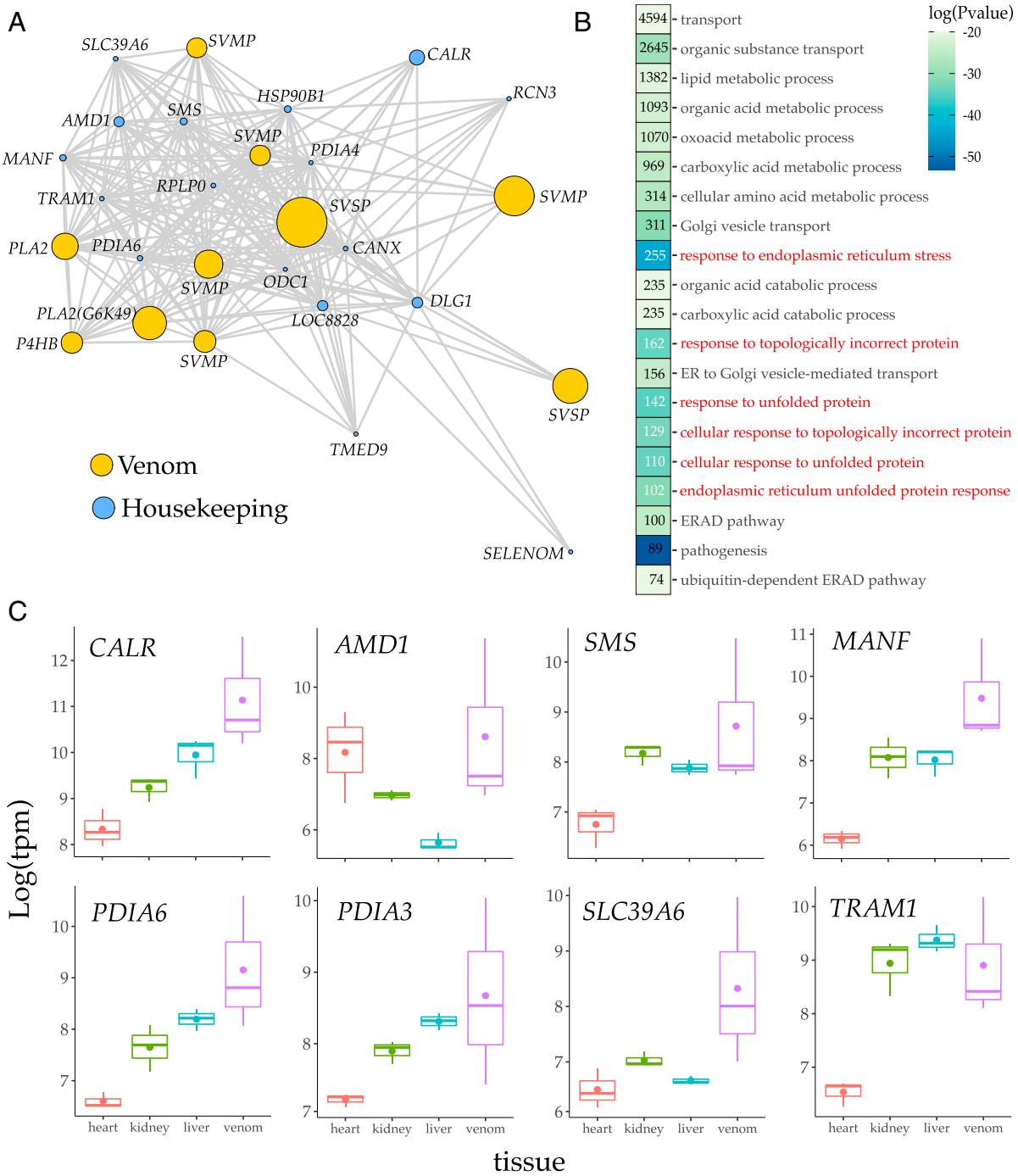

Fig. 1. The metavenom network module represents a group of coexpressed genes that are associated with production of toxin in the venom gland of the Taiwan habu. More than one-third of genes in the metavenom network are up-regulated in the venom gland and are involved in protein folding and protein modification. (*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

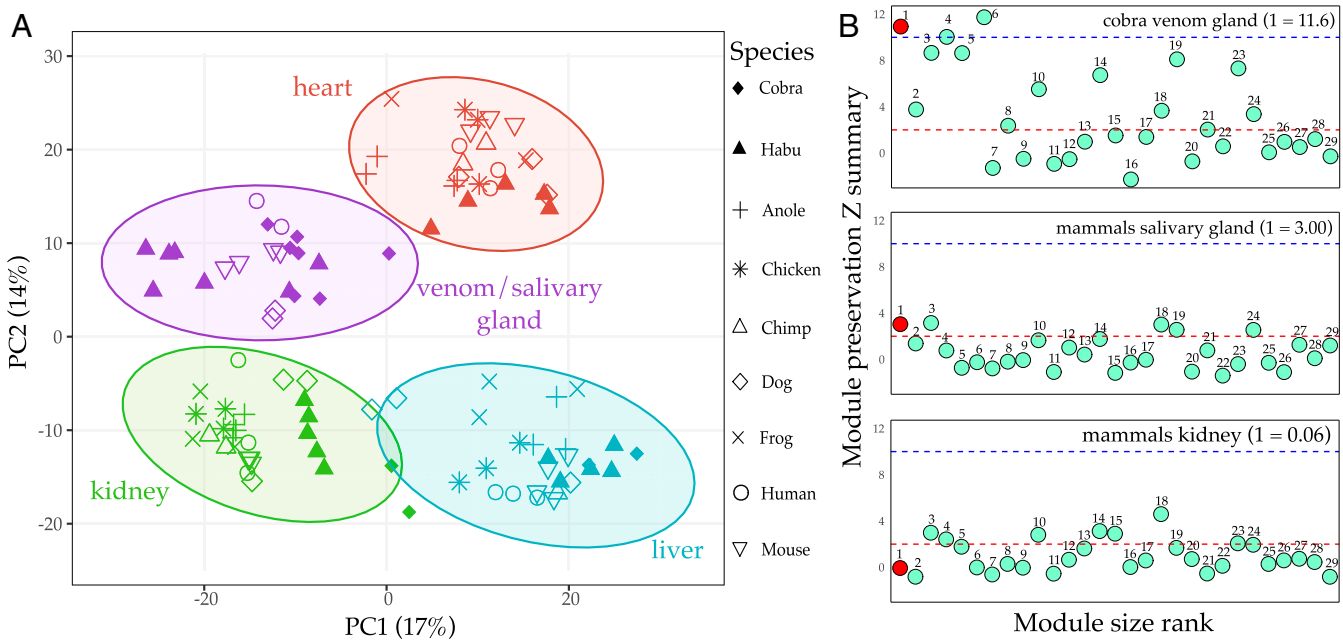

Fig. 2. Expression pattern of orthologs between venom gland in snakes and salivary gland in mammals was surprisingly well conserved. This conservation was also reflected in the preservation of the metavenom network module in the salivary gland of mammals. (*A*) When selecting the 546 one-to-one metavenom 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

EVOLUTION

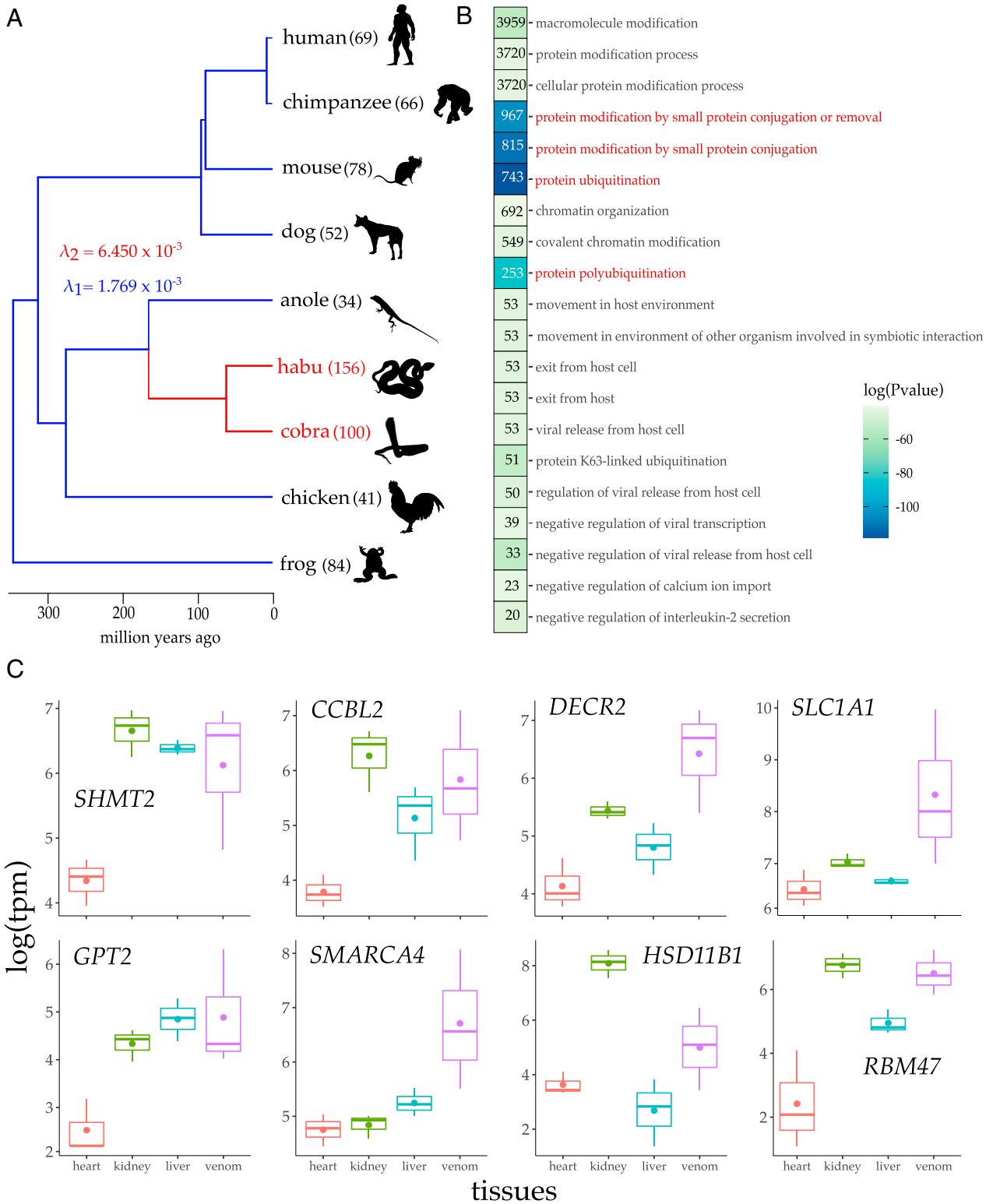

Fig. 3. Gene families in the metavenom network have not only evolved more rapidly in the lineage leading to snakes, but have also undergone more expansions in snakes than in other taxa. (*A*) Gene family evolution modeled as a "birth and death" process revealed higher rates of evolution in the branch 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.